Reliability and Statistics in Geotechnical Engineering

Gregory B. Baecher

Department of Civil & Environmental Engineering, University of Maryland, USA

John T. Christian

Consulting Engineer, Waban, Massachusetts, USA

Reliability and Statistics in Geotechnical Engineering

Reliability and Statistics in Geotechnical Engineering

Gregory B. Baecher

Department of Civil & Environmental Engineering, University of Maryland, USA

John T. Christian

Consulting Engineer, Waban, Massachusetts, USA

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Baecher, Gregory B.
Reliability and statistics in geotechnical engineering / Gregory B. Baecher, John T. Christian.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-49833-5
1. Engineering geology – Statistical methods. I. Christian, John T. II. Title.

TA705.B25 2003
624.1′51 – dc21

2003053830

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-49833-5

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by TJ International, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

Applications of probabilistic methods in geotechnical engineering have increased remarkably in recent years. As we have worked on this book, we have encountered uses ranging from practical design and construction problems to advanced research publications. We have also found that many engineers are concerned about what these developments mean and how they can be applied with confidence. Our main goal in preparing this book has been to try to bring to the geotechnical and geological engineering profession a description of probabilistic methods that emphasizes both the underpinnings of the methodologies and their practical applications. There are many books on structural reliability theory and practice, and, of course, there are entire libraries of books on statistics and probability. However, the geotechnical world is different in many ways from the structural and mechanical worlds, and a book is needed that deals with reliability methods from a geotechnical point of view. Geotechnical engineers and geologists deal with materials whose properties and spatial distribution are poorly known and with problems in which loads and resistances are often coupled. Thus, a somewhat different philosophical approach is necessary. Historically, the geotechnical profession has dealt with uncertainty on important projects by using the 'observational' approach; this is quite compatible with reliability-based methods.

We have tried to steer a middle course between an abstract, mathematical treatment so typical of many graduate texts on statistics and a purely practical presentation that reduces all methods to a set of recipes. We have also had to choose among the wide range of probabilistic material in geotechnical engineering and geology. Inevitably, this means that we go more deeply into some subjects than others. In making such choices, we have been guided by our view of what is important to geotechnical practice.

The book is organized in four parts. Part I introduces concepts of uncertainty, probability, reliability, statistics, and risk. It discusses both practical considerations and philosophical issues that have existed as long as probability theory itself. Part II deals with uncertainty in a geologic or geotechnical context. It deals with issues of uncertainty in engineering properties and the spatial variation of soils and rocks. Part III describes how reliability analyses are performed. It surveys important methods and provides detail on various models. Part IV presents applications of reliability and probabilistic methods to practical problems. It also addresses how probabilistic information is obtained and managed. These major parts of the book could be viewed in an alternate way as well: they deal, in turn, with (1) the nature of uncertainty in geotechnical engineering and how uncertainty is modeled; (2) inference and estimation from data involving the use of statistical techniques; (3) predictions of the uncertainty in engineering performance involving

the use of probabilistic techniques, and (4) integrating uncertainties and consequences by means of risk analysis.

Many people have contributed to advances in probabilistic approaches to geotechnical engineering and have assisted us in preparing this book. At the risk of leaving out many significant contributors, we would like to mention the pioneering work of (alphabetically) C. Allin Cornell, Herbert H. Einstein, Milton E. Harr, Wilson H. Tang, and Tien H. Wu. We would also like to acknowledge the intellectual contributions of David Bowles, Karl Dise, Desmond Hartford, and Andrew Zielinski to Parts 1 and 4. Both of us spent several years as students and faculty members in the geotechnical group of the Department of Civil Engineering at the Massachusetts Institute of Technology in an environment of active debate on uncertainty and how to deal with its consequences. We learned a lot from the experience. In addition to those already mentioned, we acknowledge our debt to Charles C. Ladd, T. William Lambe, W. Allen Marr, Richard de Neufville, Erik Vanmarcke, Daniele Veneziano, Robert V. Whitman, and the many graduate students of the period who worked with us. Finally, we are thankful for the help and patience of the staff of John Wiley & Sons, especially Jan de Landtsheer, Wendy Hunter, and Céline Durand.

Part I

1 Introduction – Uncertainty and Risk in Geotechnical Engineering

Human beings built tunnels, dams, canals, fortifications, roads, and other geotechnical structures long before there was a formal discipline of geotechnical engineering. Although many impressive projects were built, the ability of engineers to deal with geotechnical problems in analytically rigorous ways that allowed them to learn from experience and to catalogue the behavior of soils and rocks was limited. During the first two-thirds of the 20th century, a group of engineers and researchers, led by Karl Terzaghi (1883–1963), changed all that by applying the methods of physics and engineering mechanics to the study of geological materials (Terzaghi 1925). They developed methods of theoretical analysis, procedures in laboratory testing, and techniques for field measurements. This allowed a rational approach to design and, in the process, provided that most important prerequisite for technological advancement: a parsimonious system of reference within which to catalogue observation and experience.

The developments of the early 1900s came almost a century after the introduction of rational methods in structural engineering and machine design. The lag is not surprising, for the enterprise of dealing with materials as nature laid them down is profoundly different from that of dealing with man-made materials, even such complicated materials as reinforced concrete. Structural and mechanical engineers deal with a world almost entirely of their making. Geometries and material properties are specified; systems of components and their points of connection are planned in advance. The principal uncertainties have to do with the tolerances to which a structure or mechanical device can be built and with the loads and environmental conditions to which the structure or device will be exposed.

The enterprise of geotechnical engineering is different. The geotechnical engineer or engineering geologist deals mostly with geometries and materials that nature provides.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

These natural conditions are unknown to the designer and must be inferred from limited and costly observations. The principal uncertainties have to do with the accuracy and completeness with which subsurface conditions are known and with the resistances that the materials will be able to mobilize. The uncertainties in structural and mechanical engineering are largely deductive: starting from reasonably well known conditions, models are employed to deduce the behavior of a reasonably well-specified universe. The uncertainties in geotechnical engineering are largely inductive: starting from limited observations, judgment, knowledge of geology, and statistical reasoning are employed to infer the behavior of a poorly-defined universe. At this point, some examples are called for.

1.1 Offshore Platforms

The search for oil in ever deeper ocean waters and ever more hostile environments has led engineers to develop sophisticated artificial islands for exploration and production. These offshore structures can be as tall as the tallest building, or can float tethered to a sea bottom 1000 meters below. They are designed to withstand storm waves 30 meters or more in height, as well as collisions with ships, scour at their mud line, earthquake ground shaking, and other environmental hazards. These challenges have led to innovative new technologies for characterizing site conditions on the deep sea bed and for designing foundation systems. They have also forced the engineer to confront uncertainties directly and to bring modern statistical and probabilistic tools to bear in dealing with these uncertainties.

Among the earliest attempts at engineered offshore structures were those built for military purposes in the 1960s along the northeast coast of the United States. A series of five 'Texas Towers' (Figure 1.1) was planned and three were put in place by the US

Figure 1.1 Texas Tower 2, off New England (photo courtesy US Coast Guard).

Government as sites for early-warning radar off New York and on the Georges Bank. The experience with the towers was not good. On the night of January 5, 1961 the first of the towers broke apart in heavy seas, with a loss of 28 lives. Within the year, all three of the original towers had failed or were abandoned due to large dynamic movements under wave loading.

Later attempts at building offshore structures for the oil industry have been more successful, and through the 1970s and 1980s structures were placed in ever deeper waters. Whereas early developments in the Gulf of Mexico were in 30 or 60 meters of water, North Sea structures of the 1980s are placed in 200 meters. The 1990s have seen bottom-founded structures in 350 meters, and tethered structures in 1000 meters. Yet, hurricanes, high seas, and accidents still take their toll. Figure 1.2 (from Bea 1998) shows a frequency-severity ('F-N') chart of risks faced by offshore structures. Charts like this are a common way to portray probabilities of failure and potential consequences, and are often used to relate risks faced in one situation to those faced in others.

What are the uncertainties an engineer faces in designing offshore structures? These can be divided into two groups: uncertainties about what loads to design for (loading conditions), and uncertainties about how much load a structure can sustain (resistances). As a first approximation, uncertainties about loading conditions have to do with operational loads (dead and live loads on the structure itself), environmental loads (principally waves, currents, and winds), and accidental loads (for example, from vessels colliding with the structure). Uncertainties about resistances have to do with site conditions, static and dynamic soil properties, and how the structure behaves when subject to load.

How large are the largest waves that the structure will be exposed to over its operating life? This depends on how often large storms occur, how large the largest waves can be that are generated by those storms, and how long the structure is intended to be in place.

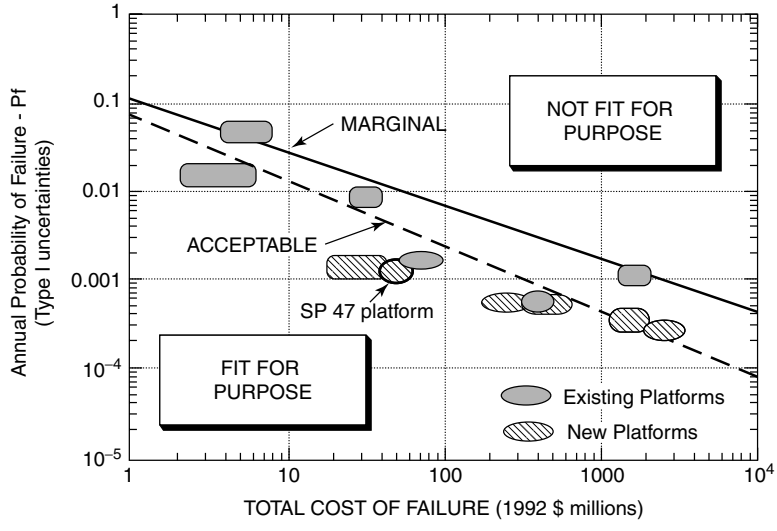


Figure 1.2 Probability-consequence chart showing risks faced at various offshore projects (Bea, R. G. 1998, 'Oceanographic and reliability characteristics of a platform in the Mississippi River Delta,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 124, No. 8, pp. 779–786, reproduced by permission of the American Society of Civil Engineers).

Like the weather patterns that generate large waves, storms and their associated wave heights are usually described by the average period between their reoccurrence. Thus, we speak of the '100-year wave' as we might of the 100-year flood in planning river works. This is the wave height that, on average, occurs once every 100 years, or more precisely, it is the wave height that has probability 0.01 of occurring in a given year. Figure 1.3 shows an exceedance probability curve for wave height at a particular site in

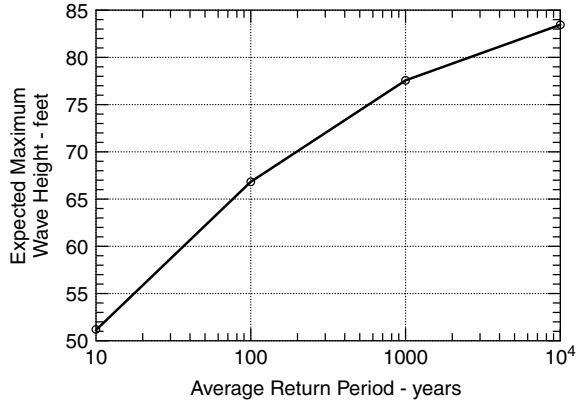


Figure 1.3 Exceedance probability curve for wave height (Bea, R. G. 1998, 'Oceanographic and reliability characteristics of a platform in the Mississippi River Delta,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 124, No. 8, pp. 779–786, reproduced by permission of the American Society of Civil Engineers).

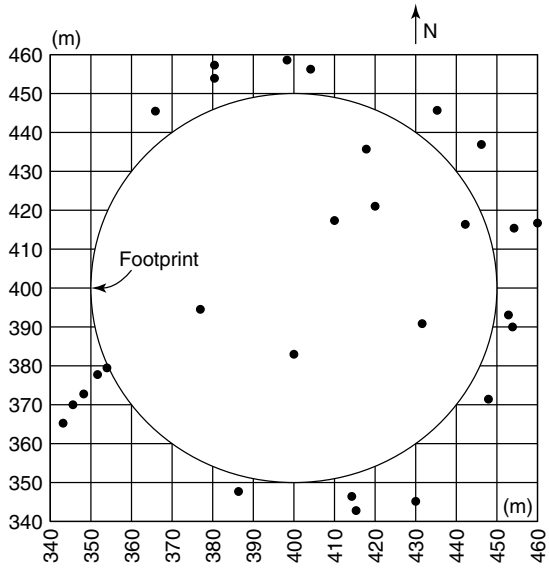


Figure 1.4 Pattern of soundings and cone penetration tests at a site in the North Sea (Lacasse, S., and Nadim, F. 1996, 'Uncertainties in characterising soil properties,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, pp. 49–75, reproduced by permission of the American Society of Civil Engineers).

the Gulf of Mexico. Such relations always depend on the particular site, because weather, currents, and ocean conditions are site specific. Based on such information, the engineer can calculate probabilities that, during its design life, the largest wave a structure will face will be 60 feet, or 100 feet, or perhaps even higher.

To investigate foundation conditions, the engineer measures penetrations of bottom sediments using tools such as the cone penetrometer, takes specimens in borings, and performs laboratory tests. Figure 1.4 (from Lacasse and Nadim 1996) shows a typical pattern of soundings and cone penetration tests at a site in the North Sea. The soil profile consists of 7–10 m of sand over a weaker, partly laminated clay of variable shear strength, which controls the foundation design. The location and undrained shear strength of this weak clay dictates the feasibility of the foundation, and using conservative lower bound estimates results in prohibitive costs. Therefore, the design needs to be based on best-estimates of the clay strength and extent, adjusted by a sensible assessment of spatial variation and statistically calculated uncertainties in estimates of engineering properties. Figures 1.5 and 1.6 show Lacasse and Nadim's interpolated map of clay strength and estimate of uncertainty in the undrained clay strength profile.

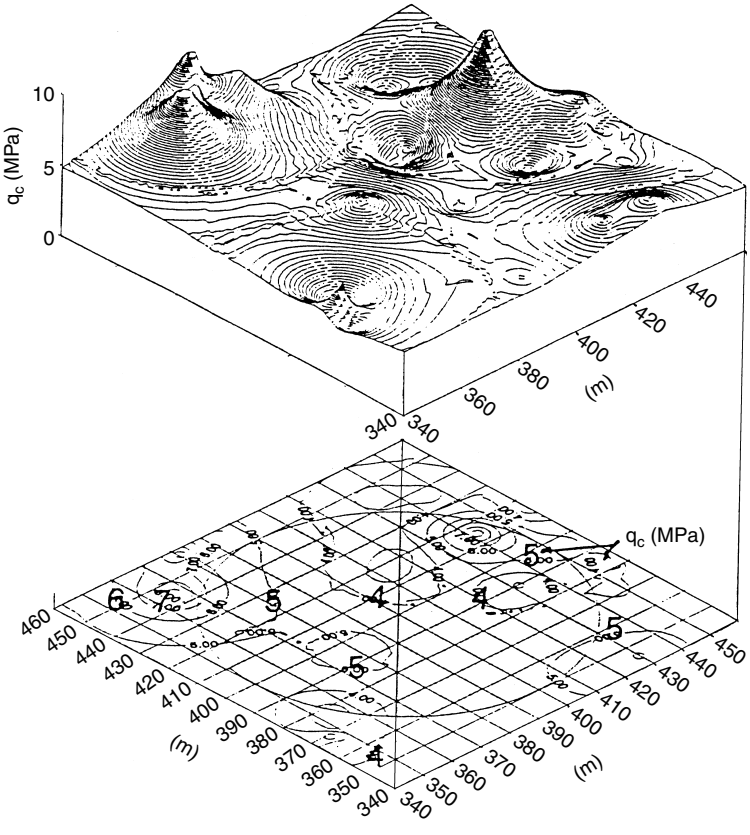


Figure 1.5 Autocorrelation function and interpolated map of clay penetration resistance (Lacasse, S., and Nadim, F. 1996, 'Uncertainties in characterising soil properties,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, pp. 49–75, reproduced by permission of the American Society of Civil Engineers).

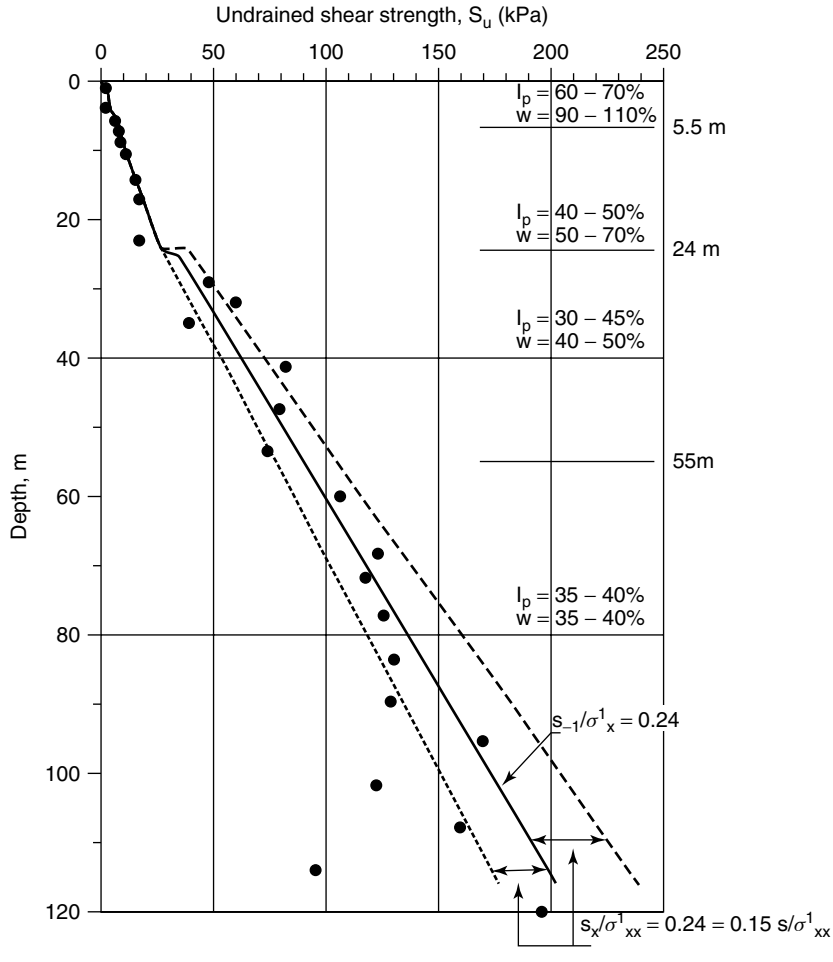


Figure 1.6 Uncertainty in undrained strength profile of clay (Lacasse, S., and Nadim, F. 1996, 'Uncertainties in characterising soil properties,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, pp. 49–75, reproduced by permission of the American Society of Civil Engineers).

Gravity structures in the North Sea are floated into position by being towed out to sea from deep fjord construction sites along the coast of Norway. Once in position, ballast tanks are filled with water, and the structure is allowed to sink slowly to rest on the sea floor. The dead weight of the structure holds the structure in place upon the sea floor, and the great width and strength of the structure resists lateral forces from wave action and currents. In calculating the stability of the gravity foundation, Lacasse and Nadim identify many of factors influencing uncertainty in model predictions of performance (Tables 1.1 and 1.2).

Reliability analyses were performed for the design geometry of Figure 1.7, involving a number of potential slip surfaces. Spatial averaging of the uncertainties in soil properties along the slip surfaces reduced the uncertainty in overall predictions. The coefficient of variation (standard deviation divided by the mean or best estimate) of wave loading was approximated as 15%, and variations in horizontal load and moment caused by

Table 1.1 Typical load and resistance uncertainties associated with offshore structures

Load uncertainties	Examples
Operational loads	Dead loads of structural weight
	Live loads of operations or equipment
Environmental loads	Hurricane winds
	Storm wave heights
	Seismic ground shaking
Accidental loads	Ship impact effects
	Fire or explosions
Resistance uncertainties	**Examples**
Site conditions	Location and character of stratification
	Thickness of important strata
	Existence of soft inclusions or gas pockets
Soil properties	Strength-deformation in static loading
	Consolidation parameters
	Cyclic strength parameters
Structural performance	Strength of critical members
	Fatigue behavior of structural materials
Interaction effects	Structural behavior of pile foundations

Table 1.2 Uncertainties in model predictions

Uncertainties
Position of the critical slip surface
Modeling of static and cyclic load history
Strain-softening
Progressive failure
Testing procedures in reference tests
Scale effect
Rate of shear
Stress conditions
Redistribution of stresses
Anisotropy
Structure stiffness
Model of soil profile
Drainage assumptions
Plane strain versus 3D analysis

environmental loads were approximated as perfectly correlated. The analysis showed, as is often the case, that the slip surface with the highest probability of failure (lowest reliability) is not that with the lowest conventional factor of safety (FS), but rather is that with the least favorable combination of mean factor of safety and uncertainty. Figure 1.7 summarizes this combination using a 'reliability index,' β, defined as the number of standard deviations of uncertainty separating the best estimate of FS from the nominal failure condition at FS = 1.

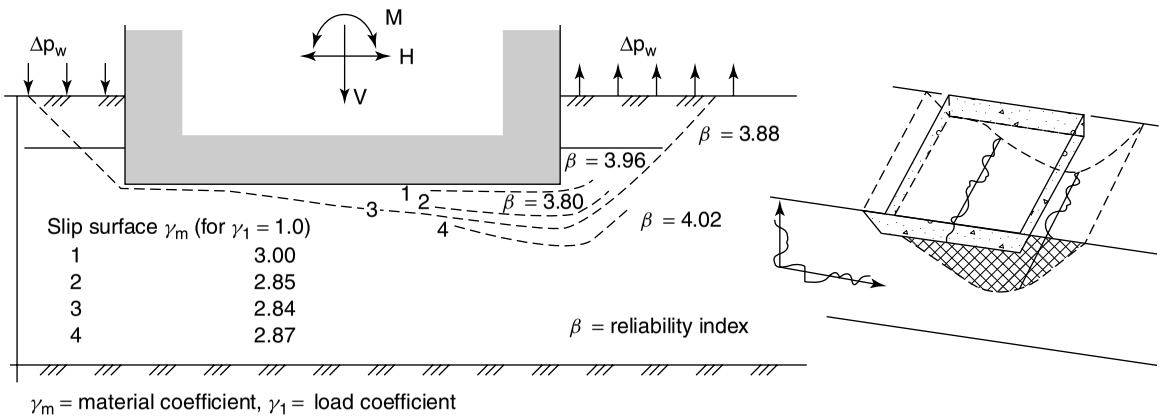


Figure 1.7 Results of probabilistic analysis of bearing capacity of shallow offshore foundation for a gravity platform (Lacasse, S., and Nadim, F. 1996, 'Uncertainties in characterising soil properties,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, pp. 49–75, reproduced by permission of the American Society of Civil Engineers).

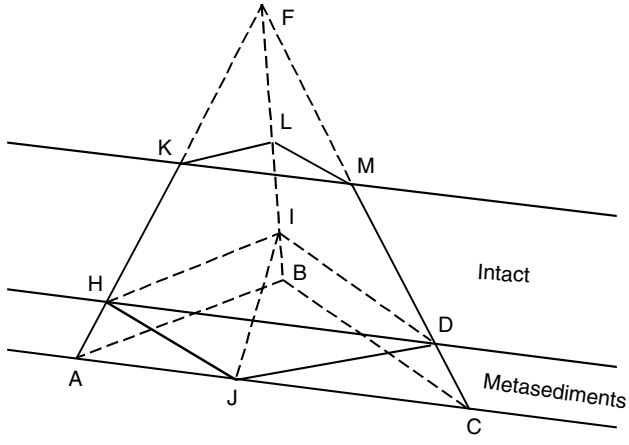


Figure 1.8 Wedge failure Mechanism for Chuquicamata mine slope.

1.2 Pit Mine Slopes

Pit mines are among the largest geotechnical structures in the world; the Chuquicamata copper mine in northern Chile is 750 m deep. They are necessarily located where ore, coal, or other material to be extracted is located, and such locations are often inconvenient for both the geotechnical engineer and the owner of the mine. In many cases seepage processes concentrate the ore along a fault, which then passes through the pit. The strata on opposite sides of the fault may have different geotechnical properties; it is not uncommon to find them dipping in different senses – one side toward the pit and the other at an oblique angle. Excavation usually takes place in a series of benches, each of which may be many meters high, so the engineer must evaluate the stability of individual benches as well as the overall slope, under conditions that vary from place to place.

Excavating and disposing of material is a major cost of operating the mine. The simplest way to hold down the cost of excavation and disposal is to reduce the volume that needs to be excavated to reach ore-bearing strata, and this implies that the slopes of the pit should be as steep as possible. On the other hand, the steeper the slopes, the greater the danger of slope failures, with attendant costs in life and money. Balancing the costs and benefits involves estimating the properties of the materials in the slopes, calculating the stability of various slope geometries, and monitoring the performance of the slopes as the pit is developed. Figure 1.8 shows a typical failure mechanism.

A related problem is the stability of tailings embankments (Vick 1983). These are piles of waste material from the mining and extraction processes. Although the operators of the mine have some control over the design and placement of the embankments, the materials themselves often have very poor geotechnical properties. Steeper slopes allow more material to be disposed of in a particular area, but they increase the danger of failure. There have been spectacular failures of tailings dams, and, since they tend to be located in areas that are more accessible than the open pits themselves, loss of life often involves the public.

Both open pit mine slopes and tailings embankments have traditionally been designed and operated on a deterministic basis. In many mining operations, this was based on

the empirical, and often undocumented, experience of the operators. In recent years, the scientific methods of geotechnical engineering, including soil and rock mechanics, have been employed, and there is now a much better understanding of the principles involved in their behavior. However, much of the design is still based on manuals developed over the years by mining organizations.

Because each mine or tailings slope tends to be unique, there is little empirical information on the overall risk of slope failure. However, mine operators recognize that instability of slopes poses a serious risk both to the safety of the operators and to the economic viability of the mine. In recent years, there has been growing interest in quantifying these risks.

The most important loads affecting the stability of a slope in a mine or tailings embankment are those due to gravity. Since the unit weight of soils and rocks can usually be determined accurately, there is little uncertainty in gravitational loads. In addition the patterns of percolation of pore fluids have major effects on the stability, and these are much more uncertain. Finally, some external factors can be important; probably the most significant of these are earthquakes (Dobry and Alvarez 1967).

The major resistance to failure is the shear strength of the soil or rock. This is a major concern of geotechnical engineers, and much of the application of reliability methods to geotechnical practice involves the probabilistic and statistical description of strength and its distribution.

The geology of a mine site is usually well known. Geotechnical properties will vary from place to place around the mine site, but the investigations that led to the establishment of the mine in the first place usually provide a much better description of the site than is common in geotechnical engineering – and certainly much better than the descriptions of underwater conditions with which the designers of offshore platforms must deal. In contrast, engineering properties of soils and rocks in the slopes are often less well known. Since the slopes are part of the cost structure of the mine rather than the revenue stream, there is sometimes not as much money available for site characterization as would be desirable.

Mapping the geologic characteristics of faces exposed during excavation is part of mine operation, so the operators have a good idea of the faults, joints, and interfaces in the slopes. Observation of mine operations also provides information on incipient problems. Thus, potential failure modes can be identified and analyzed, but surprises do occur. One typical mode of failure involves a block of material sliding on two or three planes of discontinuity. Riela *et al*. (1999) describe the use of reliability analysis to develop the results of the analysis of one of the potential sliding blocks of the Chuquicamata mine in probabilistic terms. For a particular mode of failure they found that the probability of failure was between 13% and 17%, depending on the method of analysis. They also found that the largest contributors to the probability were uncertainty in the friction angle and in depth to the water table.

Decisions regarding mine operation, and even which mines to open and close, are based on costs of operation, amount of ore to be extracted, and the value of the product on world markets. These are all inherently uncertain quantities, and modern management methods incorporate such uncertainties formally in the decision-making process. The probability of failures, whether in existing mines or as a function of the steepness of proposed excavations, is part of the cost structure for the mine. Reliability analysis provides the input to such an analysis.

1.3 Balancing Risk and Reliability in a Geotechnical Design

Most of the early pioneers of geotechnical engineering were educated in structural and mechanical engineering while those disciplines were developing rational and scientific bases. For example, Terzaghi's first degree was in mechanical engineering, Peck started out as a structural engineer, and Casagrande's education was in hydraulics. However, as these and other pioneers went about putting a rational foundation under what eventually came to be known as geotechnical engineering, they were keenly aware of the limitations of purely rational, deductive approaches to the uncertain conditions that prevail in the geological world. Their later writings are full of warnings not to take the results of laboratory tests and analytical calculations too literally. Indeed, one of the factors that attracted many students into the field was this very uncertainty. Even at the undergraduate level, things did not seem to be cut and dried. Each project presented a new challenge. There was scope for research on each new project, and the exercise of judgment.

The most widely accepted and successful way to deal with the uncertainties inherent in dealing with geological materials came to be known as the *observational method*, described succinctly by Casagrande (1965) and Peck (1969). It is also an essential part of the New Austrian Tunneling Method (Rabcewitz 1964a; 1964b; 1965; Einstein *et al.* 1996). The observational method grew out of the fact that it is not feasible in many geotechnical applications to assume very conservative values of the loads and material properties and design for those conditions. The resulting design is often physically or financially impossible to build. Instead the engineer makes reasonable estimates of the parameters and of the amounts by which they could deviate from the expected values. Then the design is based on expected values – or on some conservative but feasible extension of the expected values – but provision is made for action to deal with the occurrence of loads or resistances that fall outside the design range. During construction and operation of the facility, observations of its performance are made so that appropriate corrective action can be made. This is not simply a matter of designing for an expected set of conditions and doing something to fix any troubles that arise. It involves considering the effects of the possible range of values of the parameters and having in place a plan to deal with occurrences that fall outside the expected range. It requires the ongoing involvement of the designers during the construction and operation of the facility.

Recent years have shown a trend to place the treatment of uncertainty on a more formal basis, in particular by applying the results of *reliability theory* to geotechnical engineering. Reliability theory itself evolved from the structural, aerospace, and manufacturing industries; it has required special adaptation to deal with the geological environment. This book deals with current methods of reliability theory that are most useful in geotechnical applications and with the difficulties that arise in trying to make those applications. It must be emphasized at the outset that reliability approaches do not remove uncertainty and do not alleviate the need for judgment in dealing with the world. They do provide a way of quantifying those uncertainties and handling them consistently. In essence, they are an accounting scheme. The experienced geotechnical engineer has already made the first step in applying reliability methods – recognizing that the world is imperfectly knowable. The rest of the process is to discover how to deal with that imperfection.

Today the geotechnical engineer must increasingly be able to deal with reliability. There are several reasons for this. First, regulatory and legal pressures force geotechnical engineers to provide answers about the reliability of their designs. This is most notable

in heavily regulated areas of practice such as nuclear power, offshore technology, and waste disposal. It is a trend that will affect other areas of practice in the future. Secondly, management decisions on whether to proceed with a projected course of action, how to finance it, and when to schedule it are increasingly based on statistical decision analysis. A brief review of almost any textbook on modern financial management demonstrates that today's managers are trained to assess the value of a course of action from probabilistic estimates of the value of money, future profits, costs of production, and so on. The performance of major civil engineering facilities enters into such evaluations, and it, too, must be stated probabilistically. Thirdly, modern building codes are based on Load and Resistance Factor Design (LRFD) approaches, which are in turn based on reliability methods. These techniques are now being introduced into such areas as pile design for highway structures. Fourthly, reliability theory provides a rational way to deal with some historically vexed questions. For example, how much confidence should the engineer place on a calculated factor of safety? How should the engineer quantify the well-founded belief that the value of the friction angle is more dependable than that of the cohesion? How can the engineer demonstrate that a design based on more data and more consistent data is more robust than one based on partial information – and therefore worth the extra cost of obtaining those data? How can the engineer distinguish between different consequences of failure or separate cases in which progressive failure occurs from those in which an average behavior is to be expected? Reliability approaches provide insights in these areas and, in some cases, numerical procedures for analyzing them.

1.4 Historical Development of Reliability Methods in Civil Engineering

To find the antecedents of today's risk and reliability methods in civil engineering, one must look back to the allied field of structural reliability and to such pioneers as Alfred Freudenthal (1890–1975). In the 1950s and 1960s, Freudenthal published a series of fundamental papers in which many of the precepts of modern risk and reliability theory first appeared (Freudenthal 1951, Freudenthal *et al*. 1966; Freudenthal and Gumbel 1956). Among these were the concept of statistical description of material properties, state-space representation of failure conditions, and non-parametric reliability indices. Freudenthal's work was followed by a generation of researchers in structural engineering, including A. H.-S. Ang, C. A. Cornell, O. Ditlevsen, A. M. Hasofer, N. Lind, and R. Rackwitz. In the early 1970s the emerging field of structural reliability began to spill over into geotechnical engineering research, and this book is based on the results of that work.

As has already been stated, major government programs and economic trends of the 1970s and 1980s exerted significant influence on the direction of the field. The most important were the regulatory environment surrounding nuclear power generation, nuclear and solid waste disposal, and the energy crisis of the 1970s, which led to the development of offshore oil and gas production facilities in water of unprecedented depth. Each of these trends placed increased attention on uncertainties attending site characterization and quantified assessments of geotechnical performance. Other industrial interest in geotechnical risk and reliability came from surface mining, where high rock slopes are designed

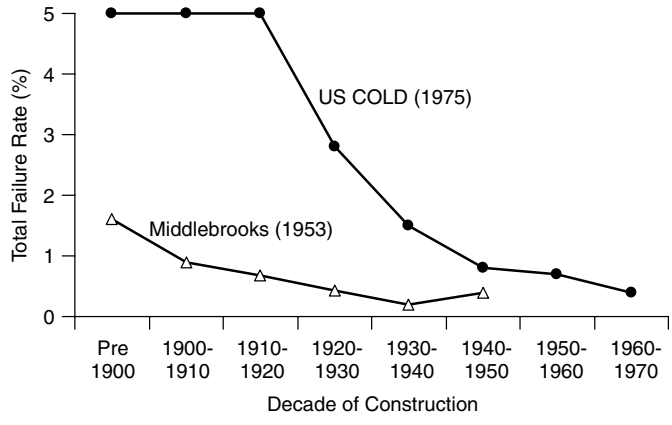


Figure 1.9 Historical numbers of modern dam failures (Baecher *et al.* 1980).

to small factors of safety, and seismic safety, where lifelines and other critical facilities could be corrupted by violent but infrequent ground motions.

With the failure of the Teton Dam in 1976, the dam-building agencies in the United States became strongly involved in risk assessment. Failures of dams and near-failures without loss of containment, while fortunately not common, are far from unknown. Figure 1.9 shows the frequency of modern day dam failures, which has led some workers to conclude that the annual risk of failure of a modern dam, absent other information, is on the order of 10^{-5} to 10^{-4} per dam-year (Baecher *et al.* 1980). Today, the U. S. Bureau of Reclamation has become a leading exponent of risk assessment for dams, and the U. S. Army Corps of Engineers has produced numerous manuals and workshops to guide their staff and contractors in applying reliability theory to their projects. They have both developed broad technical expertise in risk methodology.

1.5 Some Terminological and Philosophical Issues

The evaluation of risk and reliability is closely connected to the fields of probability and statistics. This creates certain problems for presenting reliability concepts to an engineering audience. First, the various disciplines that developed modern probability and statistics have erected elaborate, precise, and formidable systems of notation and nomenclature that often seem designed less to assist the reader in grasping concepts than to prevent the uninitiated from participating in sacerdotal rites. Some of this difficulty is the confusion attendant on any unfamiliar discipline. Some of the distinctions are in fact important and subtle. Therefore, the notation and nomenclature are largely necessary, but in this book we have tried to explain what the notation means as we go along. We hope this will make the material accessible to the intelligent and interested geotechnical engineer, even at the risk of explaining concepts well known to the probabilistic community.

Secondly, probability and statistics, though often lumped together in the public's mind, in textbooks, and in introductory college courses, are actually different subjects with different sets of procedures and definitions. Probability can be thought of as an algebra – a set of results derived by rigorous reasoning from a set of postulates. Statistics deals with

description of the observed world. That they often arrive at similar statements and conclusions does not change the fact that the basic reasoning processes of the two disciplines are different. Some of the confusion in the literature on reliability is due to this difference.

Thirdly, there is a related distinction between the *frequentist* approach and the *degree-of-belief* approach to uncertainty. The frequentist concept is that the uncertainties described by probabilistic and statistical work have to do with long series of similar events in the world. The degree-of-belief concept is that the uncertainties have to do with the confidence one places in knowing the state of the world. The dispute between the two schools can become at times a precious philosophical argument, but there is a meaningful difference that has practical implications. For example, an insurance company sells life insurance from a frequentist point of view based on actuarial tables. The consumer buys it on the basis of a degree of belief in his or her longevity. The modeling of geotechnical problems involves both types of reasoning, but most geotechnical engineers seem more at home with the degree-of-belief view of uncertainty than with the frequentist view.

This distinction between frequency and belief can be elaborated by considering the difference between uncertainties that are inherent to a process and those that reflect lack of knowledge. A good example is the difference between the uncertainties associated with a pair of dice and those of a shuffled deck of cards. A fair (unloaded) die is a randomizing device. The number that will turn up is random, and no practically obtainable knowledge about how the die is thrown, or when, or with what force affects one's ability to predict the outcome. This type of underlying uncertainty governs some natural processes. Radioactive decay is one example; quantum mechanics demonstrates that it is impossible to know precisely which atom will decay or when it will decay. The deck of cards typifies the other type of uncertainty. The deck has a definite arrangement, and anyone dealing honestly from the deck will find the same order of cards. The uncertainty lies entirely in our ignorance of the deck's arrangement, and sophisticated play in such games as poker and bridge is largely concerned with trying to garner information about the arrangement of the cards from the play of the hand. A great deal of geological uncertainty is of this type, for there is at any site a definite arrangement of geological materials and their properties, but we do not know what it is.

Hacking (1975) popularized the terms *aleatory* (after the Latin word *aleator*, meaning 'gambler' or 'die caster') for the first type of uncertainty, which reflects underlying physical randomness, and *epistemic* (after the Greek, $\varepsilon\pi\iota\sigma\tau\eta\mu\eta$ meaning 'knowledge') for the second, which reflects lack of knowledge. A National Research Council (2000) study suggested the terms *natural variability* for aleatory uncertainty, and *knowledge uncertainty* for epistemic uncertainty. These are also useful terms, which enjoy the benefit of familiarity. Workers in the field of seismic hazard evaluation have used the terms *random* for aleatory uncertainty and *uncertain* for epistemic uncertainty. Because 'random' and 'uncertain' already mean so many other things, we have chosen to use Hacking's nomenclature in this book. The distinction is important. Much of the confusion in geotechnical reliability analysis arises from a failure to recognize that there are these two different sources of uncertainty and that they have different consequences. For example, one can propose a program of investigations to reduce epistemic uncertainty, but such a program will not do much for the aleatory uncertainty.

1.6 The Organization of this Book

This book is divided into four sections, the last three of which correspond approximately to the stages of a geotechnical design project. The first part describes uncertainty in geotechnical engineering and sets out some basic principles of probability, statistics, and decision theory. The second describes procedures for data analysis and site characterization. Every project starts with the site, and the site must be characterized. The engineers must understand the soil, rock, and groundwater conditions and how these will interact with the proposed facilities. Thus, the first phase is to gather background information on the site and to conduct a program of exploration and testing to gather detailed information. From this, the engineer expects detailed maps, description of surface features, testing results from which to estimate engineering parameters, and identification of anomalous conditions that might affect the project. All these are subject to statistical analysis and interpretation.

The third part deals with reliability analysis itself. This corresponds to the project's design and analysis phase, in which analytical tools are employed to predict the behavior of various components of the project and to refine the design. This may involve detailed analytical or numerical modeling, evaluation of design alternatives, and sensitivity studies. Reliability analysis is part of a modern design and analysis effort and enhances traditional approaches.

The last part describes procedures for making decisions and evaluating risk. In a project, this involves comparing different alternatives with respect to feasibility, cost, safety, and other concerns. It also involves the monitoring during and after construction in keeping with the observational method. In effect, this is a summarizing and integrating phase for both the design and the reliability evaluation.

1.7 A Comment on Notation and Nomenclature

This book deals with material in several well-developed disciplines. Inevitably the notation and nomenclature of different fields clash. For example, a 'sample' of soil means one thing to a geotechnical engineer logging a boring, but a 'sample' means something quite different to a statistician estimating a trend. The Greek letter σ denotes a standard deviation in probability and statistics, but it denotes a normal stress in soil mechanics. There are two ways to deal with this problem. One is to develop a new, unambiguous notation for the purposes of this book. We have chosen not to do this because such a notation would inevitably not agree with that used in any of the other literature and would be difficult for workers in any field to understand. The other approach is to stick with the traditional notation and provide explanations when there is a chance of ambiguity. That is the approach adopted here. We have followed the common probabilistic convention of using an upper case letter to identify a variable and a lower case letter to indicate a particular value of that variable. Bold face letters denotes vectors and matrices.

2 Uncertainty[1]

We use terms like *uncertainty*, *randomness*, and *chance* in the course of daily life but seldom devote thought to what those terms mean precisely. Similarly, we use those terms in dealing with engineering problems, but overlook philosophical questions that use of the terms implies. Implicitly, we think we know what those terms have to do with, but sometimes the more we think about the sorts of uncertainties that are part of an engineering analysis, the less clear they seem to become. This chapter takes a look at the different sorts of uncertainties encountered in geotechnical risk and reliability studies and at the corresponding meanings of the probability models used to analyze them.

Most engineers, at least those who deal with the macroscopic world of civil infrastructure and geological processes, think of nature as deterministic. That is, for any effect there is a cause, and any cause and its effects are directly linked. There are grand laws of nature that control the behavior of the physical world, and if we knew initial conditions with precision, we could forecast the future state of the world with precision. But if it is true that the world is deterministic, what does it mean for something to be random? If the world is deterministic, what does it mean to speak of the world in probabilities? Such questions have to be answered if the results of risk and reliability analysis are to be correctly interpreted in practice.

In this chapter we delve into what it means for things to be uncertain, what it means for things to be random, and what it means to use probability measures to represent uncertainties. For example, is an "0.1 probability of rain" the same thing as an "0.1 probability of a factor of safety being less than 1"? Rain is a recurring natural phenomenon; factor of safety is a calculated parameter. Is the meaning of uncertainty the same in each case, and do the two probabilities carry the same meaning?

2.1 Randomness, uncertainty, and the world

We start the discussion with the philosophical question of *necessity* and *chance*. That is, is the world deterministic as suggested above, or is it random, or both? Is this distinction

[1] The authors are grateful to David Bowles of Utah State University, Karl Dise of the U.S. Bureau of Reclamation, Desmond Hartford of BC Hydro, and Andrew Zielinsky of Ontario Power Generation for their involvement in developing the concepts discussed in this chapter. Expanded discussion of these concepts appears in the Canadian Electricity Association's *Guide to Risk Analysis for Dam Safety* (forthcoming).

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

important? Then we turn to the categories of uncertainty that arise in engineering analysis, and consider a taxonomy of uncertainty for geotechnical applications. Finally, we consider what it means to assign quantitative probabilities to uncertainty.

2.1.1 Necessity or chance?

In a simple way, there are two opposing schools of philosophy on the question of fortune and fate, and these can be traced to antiquity. One school holds that things in the world happen by *necessity*; the other that they happen by *chance*. The difference has implications even to the mundane enterprise of engineering analysis.

As early as the 5th century BCE, Leucippus and, shortly thereafter, Democritus held that there is a cause for everything and that nothing in the world happens by chance. To these Atomists, 'chance' meant only that the cause of a thing was unknown and not that the thing itself was unknowable in advance of its happening. Centuries later, Laplace (1814) adopted the same point of view. In the introduction to his *Philosophical Essay on Probabilities*, Laplace wrote:

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes.

Laplace held that both the future and retrospectively the past (i.e. all history) are knowable from natural laws, combined with what today we would call boundary and initial conditions. This theme was picked up by historians (e.g. Buckle 1858) and social philosophers (e.g. Quetelet 1827), as well as by scientists of the standing of Boltzmann, Bousinesq, and Maxwell, all of whom struggled over reconciling this *Doctrine of Necessity*, with contemporary concepts of free will (see, e.g. Hacking 1990).

Two centuries after Leucippus, Epicurus argued, contrary to the Atomist position, that since people's decisions are not constrained, that is, that people have *free will*, things in the world cannot happen by necessity. Were things to follow from necessity, so, too, must human actions, and this would negate free will and consequently negate that people should be responsible for their own actions. In the late decades of the 19th century this stochastic view began to gain adherents, driven by Maxwell's development of statistical mechanics and by Peirce's criticisms of the Doctrine of Necessity. Peirce (1998) refuted Laplace's quotation by taking its argument to an extreme:

> The proposition in question is that the state of things in any time, together with certain immutable laws, completely determine the state of things at every other time (for a limitation to future time is indefensible). Thus, given the state of the universe in the original nebula, and given the laws of mechanics, a sufficiently powerful mind could deduce from these data the precise form of every curlicue of every letter I am now writing.

This is seen as an early example of the stochastic view of the world, now broadly held within the high-energy physics community.

Probabilism is a reconciliation of the extreme views of necessity and chance. Probabilism holds that the world is indeed unpredictable, but principally because our knowledge is inadequate to the task. That is, the world may or may not behave on the basis of necessity, but in any event we are ignorant of its initial state and of the details of natural law, and thus the world can only be described as if it were random. This is a point of view that might be imputed to Laplace's *Essay*, which goes on to develop probabilistic concepts and methods, and it is a view expanded on by Jaynes (1996). The present volume adopts this view as proper to geotechnical analysis.

In recent years a new consideration, compatible with probabilism, has arisen: *chaos*. The origins of chaos theory lie in the observation that, for non-linear systems, a small change in initial conditions can lead to large changes in outcomes. Since variations in initial conditions are too small to be modeled or possibly even observed, outcomes seem for practical purposes to be random (Gleick 1988). The often cited example is meteorological modeling, wherein small disturbances in input on one side of the world may generate large differences in output for the other side of the world. Wolfram (2002) expands on this concept in relation to computing. One can speculate that it is chaos effects that explain the old question of our inability to predict deterministically the simple toss of a coin.

2.1.2 Randomness and uncertainty

We have used the terms *random* and *uncertain* loosely in the foregoing discussion; what is it that we mean by these terms more precisely?

Since ancient times, the notion of *randomness* has concerned natural processes that are inherently unpredictable. For example, a modern dictionary definition of the word *random* is:

> **Random** (adjective). Date: 1565. 1. a: lacking a definite plan, purpose, or pattern b: made, done, or chosen at random; 2. a: relating to, having, or being elements or events with definite probability of occurrence. b: being or relating to a set or to an element of a set each of whose elements has equal probability of occurrence. (Merriam-Webster Inc. 1997)

Random has the sense of completely unpredictable, except in the relative frequencies with which something occurs. Think of simple games of chance involving coins, dice, and roulette wheels. The distinctive feature of games of chance is that the outcome of a given trial cannot be predicted, although the collective results of a large number of trials display regularity. There are many examples of the same sort involving collections of people, molecules of a gas, sets of genes, and so on. In this sense, *randomness* is a property of nature, independent of anyone's knowledge of it. It is innate. There are "true" values of these relative frequencies, although we may not know what they are.[2]

In recent times, the notion of random processes has often been described as *aleatory*, as discussed in Chapter 1. This is especially the case in seismic hazard analysis, nuclear safety, severe storm modeling, and related natural hazards (Table 2.1 gives a list of

[2] In this sense one encounters risk analyses in which the probability associated with an aleatory process is, itself, treated as a random variable, that is, probability is assumed to be a parameter (U. S. Army Corps of Engineers 1996).

Table 2.1 Terms used in the literature to describe the duality of meaning for "uncertainty"

Terms pertaining to uncertainty due to naturally variable phenomena in time or space: "Uncertainties of nature."	Terms pertaining to uncertainty due to lack of knowledge or understanding:"Uncertainties of the mind."	Citation
Natural variability	Knowledge uncertainty	(National Research Council 2000)
Aleatory uncertainty	Epistemic uncertainty	(McCann 1999)
Random or stochastic variation	Functional uncertainty	(Stedinger *et al*. 1996)
Objective uncertainty	Subjective uncertainty	(Chow *et al*. 1988)
External uncertainty	Internal uncertainty	(Chow *et al*. 1988)
Statistical probability	Inductive probability	(Carnap 1936)
Chance	*Probabilité*	Poisson, Cournot (Hacking 1975)

synonyms for random uncertainties that are commonly found in the literature of engineering analysis).

In this book, we adopt the word *random* to refer to those things which, for the purposes of modeling, we assume to be caused by chance. We also call these *aleatory uncertainties*. Even though one might adhere to a Laplacian philosophy that the world is deterministic, it is often convenient in risk and reliability analysis to presume that some part of the uncertainty we have about the world is due to randomness. This allows us to bring powerful mathematical tools to bear on a problem that might otherwise be difficult to address. Usually, the notion of uncertainty when applied to such random events is taken to mean the frequency of occurrence in a long series of similar trials. Two observers, given the same evidence, and enough of it, should converge to the same numerical value for this frequency or uncertainty.

In contrast to randomness, since ancient times the notion of *uncertainty* has concerned propositions of unproven veracity. For example, a modern dictionary definition of the word *uncertain* is:

> **Uncertain** (adjective). Date: 14th century. 1: Indefinite, indeterminate 2 : not certain to occur : Problematical 3: not reliable: Untrustworthy 4 a : not known beyond doubt : Dubious b: not having certain knowledge: Doubtful c: not clearly identified or defined 5: not constant: Variable, Fitful. Synonyms: *doubt, dubiety, skepticism, suspicion, mistrust*. (Merriam-Webster Inc. 1997).

Uncertain has the sense of unknown or unverified, but not unpredictable. An individual can be uncertain about the truth of a scientific theory, a religious doctrine, or even about the occurrence of a specific historical event when inadequate or conflicting eyewitness accounts are involved. Such propositions cannot be defined as repeatable experiments. In this sense, uncertainty is a property of the mind; it can only be learned by self-interrogation, and is unique to the individual. Two observers, given the same evidence, can arrive at different values and both be right. There is no "true" value of such degrees of belief, and correspondingly no "true" value of the probability assigned to them.

In recent times, this notion of knowledge uncertainty has been described as *epistemic*, as discussed in Chapter 1. The notion of uncertainty when applied to such propositions is taken to mean the strength of belief in their truth or falsity.

In modern practice, risk and reliability analysis usually incorporates probabilities of both aleatory and epistemic varieties. The separation is not an immutable property of nature, but an artifact of how we model things. Some aspects of geotechnical engineering can be treated as if they were random and thus describable by relative frequencies (e.g. flood frequency, spatial variations of soil properties), but not all. Others may not have to do with real world processes that are repeatable; they may have to do with a unique event that we are unsure about. In this case, uncertainty has a meaning of strength of opinion or degree of belief. Such strength of opinion may not be one-to-one identifiable with observed responses in the past but may depend on qualitative experience, reasoning from first principles, and intuition.

2.2 Modeling uncertainties in risk and reliability analysis

To understand the implications of how we conceive the world – whether of necessity or of chance – on the practice of risk and reliability analysis, it is useful to consider the types of uncertainties that arise in engineering practice.

2.2.1 Categories of uncertainty

As a first approximation, the uncertainties that are dealt with in geotechnical engineering fall into three major categories (Figure 2.1):

Natural variability is associated with the "inherent" randomness of natural processes, manifesting as variability over time for phenomena that take place at a single location (temporal variability), or as variability over space for phenomena at that take place at different locations but at a single time (spatial variability), or as variability over both time and space. Such natural variability is approximated using mathematical simplifications, or models. These models many or may not provide a good fit to naturally phenomena. In the best case, they are close but only approximate fits.

Knowledge uncertainty is attributed to lack of data, lack of information about events and processes, or lack of understanding of physical laws that limits our ability to model the real world. Knowledge uncertainty is just a more common description of epistemic uncertainty. This has sometimes also been called subjective uncertainty or internal uncertainty. Knowledge uncertainty divides into three major sub-categories for geotechnical

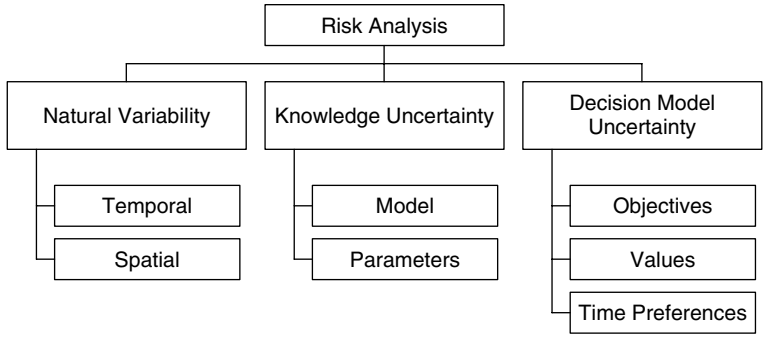


Figure 2.1 Categories of uncertainty entering risk analysis.

applications: site characterization uncertainty, model uncertainty, and parameter uncertainty.

Site characterization uncertainty has to do with the adequacy of interpretations we make about the subsurface geology. It results from data and exploration uncertainties, including (i) measurement errors, (ii) inconsistency or inhomogeneity of data, (iii) data handling and transcription errors, and (iv) inadequate representativeness of data samples due to time and space limitations.

Model uncertainty has to do with the degree to which a chosen mathematical model accurately mimics reality. Model uncertainty, reflecting the inability of a model or design technique to represent precisely a system's true physical behavior, or our inability to identify the best model, or a model that may be changing in time in poorly known ways. The models we fit to naturally varying phenomena need to be fit to natural process by observing how those processes work, by measuring important features, and by statistically estimating parameters of the models.

Parameter uncertainty has to do with the precision to which model parameters can be estimated. Parameter uncertainty results from our inaccuracy in assessing parameter values from test or calibration data and is exacerbated by limited numbers of observations and resulting statistical imprecision.

In addition to natural variability and knowledge uncertainties, two practical types of uncertainty also sometimes enter geotechnical risk and reliability analysis. These have to do with the implementation of designs in practice and with the economic issues attending benefit-cost calculations. These are *operational uncertainties*, including those associated with construction, manufacture, deterioration, maintenance, and human factors not accounted for in models of engineering performance; and *decision uncertainties*, which describe our inability to know social objectives or to prescribe social discount rates, the length of a planning horizon, desirable temporal consumption-investment trade-offs, or the social aversion to risk. For the most part, these latter uncertainties are not considered in the present volume.

2.2.2 Natural variation

How does aleatory uncertainty – randomness – arise in testing and performance monitoring? Consider in more detail how we estimate soil parameters. We observe scatter in test data and treat that scatter as if it derived from some random process. Presume for the moment that the test data are free from measurement error. We then use statistical theory to summarize the data and to draw inferences about some hypothetical population of soil samples or test sections. Most people would agree that the things being observed, that is the soil properties or engineering performances, are not random. One may not know the properties or performances at every point in the soil deposit, but the properties and performances are knowable. They could, in principle, be observed. The variation in the data is spatial.

To say that the variability being observed is spatial rather than random is to liken the soil deposit to a deck of playing cards, rather than, say, to a pair of dice. Once the deck of cards has been shuffled and placed on the table, the order of cards in the deck is fixed. As with a soil deposit, the variation in the deck is spatial, not random. A player simply does not know the order of the cards before the game begins. Indeed, in many card games, such as Bridge or Blackjack, the winning strategy is for players to try to infer the order

of cards remaining in the deck as play proceeds. In this way, card games are profoundly different from dice games, in which the sequential outcomes are assumed to be random. Thus, geotechnical engineering can be thought akin to card games, but not to dice games (for further discussion see Chapter 9).

To simplify modeling and inferences from data, the assumption is sometimes made to treat natural variations as if they were random – even though they are not. In this case, the variations are a function of time or space, and said to be *stochastic*, the term merely meaning a random process defined over some domain. This is a modeling – and presumably simplifying – assumption that transfers some of the uncertainty about natural variation from the epistemic column to the aleatory column, where it is easier to handle. Moving the boundary between the two types of uncertainty does not lessen the total. It does, however, create an irreducible level of uncertainty in the analysis and corresponding predictions, by presuming a fraction of the total uncertainty to be random, and thus unknowable. The trade off against this irreducible level of uncertainty is that one hopes the modeling assumption will allow more powerful methods of mathematics to be applied to the problem of inference and estimation and thus, in the end, a more precise outcome to be achieved.

2.2.3 Trade-off between aleatory and epistemic uncertainty

An implication of making the trade-off between aleatory and epistemic uncertainty is that what is meant by a predictive probability may change. Consider the "probability of excessive settlement" of a long levee, or by the same token, the "probability of excessive lateral deformation" in a long excavation. What does it mean that this probability equals, say, 0.1? Does it mean that 10% of the levee or excavation should be expected to fail? Does it mean that there is a 0.1 chance that the entire levee or excavation will fail? Does it mean something between these two? Confusion over this issue is frequent in the literature, where the temporal or spatial fraction of adverse performance of a large structure is often used to verify a probabilistic prediction.

The answer to the question is that it depends on how the modeling assumptions are made; specifically, on how the total uncertainty is divided between aleatory (temporal or spatial) and epistemic (parametric). To the extent that all the uncertainty is assumed aleatory, the probability refers to a temporal or spatial fraction. To the extent that all the uncertainty is assumed epistemic, the probability refers to a chance of complete failure. Almost always, the uncertainty is apportioned between aleatory and epistemic, so the probability itself is a mixture.

A second implication of the trade-off between aleatory and epistemic uncertainty is the variability of performance as a function of scale. To the extent that uncertainty is presumed to be aleatory, the uncertainty averages over space and perhaps time. The variability of measured performance among long test sections will be less than the variability among short test sections. This was alluded to above. The variability of soil properties among large specimens will be less than among small specimens. The variability among *in situ* tests that mobilize large soil volumes will be less than the variability among *in situ* tests that mobilize small soil volumes. The converse is true of behaviors that rest on extreme soil or formation properties. Seepage conditions and piping that depend on the most conductive element of a formation become both more variable with scale and also on average more extreme. Rock slope failures that depend on the least favorably inclined joint become more variable and also more probable as the volume of rock mass considered becomes larger.

2.2.4 Are frequencies "objective"?

There is much debate in the literature of probability over the issue of frequency vs. belief, over whether probability is a property of the world or a property of the mind. We argue that from a modeling point of view, probability is neither a property of the world nor (uniquely) a property of the mind, but a property of the model used to make sense of a physical world. Frequency and belief coexist in our modeling endeavors; they are not opposed to one another. Furthermore, one can have a degree of belief about a frequency, as well as a frequency of degrees of belief (as in the case of expert judgments).

The modeling activity circumscribes a particular phenomenon in space, time, and extent. In doing so, it defines a frequentist process for which parameters exist, and these parameters are frequencies. Are these frequency parameters properties of the world or of the mind? Insofar as they are associated with worldly phenomena but have no meaning outside the modeling enterprise, they are properties of the world as circumscribed by the model. Within the narrow confines of the definition of the frequentist process being modeled, these frequency parameters are properties of reality. But their meanings depend on the model, and thus changing the modeling configuration can profoundly change their values. The richer the modeling of nature, the narrower the purview of frequency-based probability and the broader the purview of belief-based probability, and conversely.

So, are frequencies objective? When a model is fixed, the processes that we assume to be stochastic are fixed as well. For example, the domain and mean-trend of a spatially or temporally varying process are thus determined. The frequency parameters that define the processes are therefore properties of the world, given the model that has been adopted. If one changes the modeling assumptions, of course, the definitions and values of the frequency parameters change. But, while the modeling assumptions obtain, the frequency parameters are properties of the world, as the world is defined in the model. Typically, we do not know exact values for these frequency parameters. We have to estimate them from sample data. Thus, it is permissible for one to have a degree of belief about such a frequency.

2.3 Probability

James Clerk Maxwell, the eminent 19th century physicist, in 1850 wrote (Tolstoy 1981),

> The [...] science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

Most engineers and scientists today use quantified probabilities as the standard way to talk about uncertainty; and, to an increasing extent, so, too, do people in health care, education, natural resources, and other areas of public policy (Morgan and Henrion 1990). Geotechnical engineering – almost unique among the professions – has until recently been reluctant to make this transition to modernity. The present section discusses the meaning of *probability* as a philosophical concept and the use of mathematical probability as a measure of uncertainty.

We take for granted today that uncertainties and risks should be expressed in the language of probability and that calculations of risk should be based on inferences made using statistics. Underlying this supposition is a history of debate over the human enterprises to which probability theory applies, and to the very meaning of the term, *probability*. The historical development of the theory of probability since the time of Pascal and Fermat has been associated with a rich and subtle body of thought, and has led to a duality in the way we interpret what is meant by the term *probability*. It should surprise no one that this duality in the concept of probability parallels the similar duality, discussed in preceding sections, in the concept of uncertainty, which is interpreted on the one hand as a natural frequency in the world, and on the other hand as a lack of knowledge in ourselves. This parallel between the duality of probability and the duality of uncertainty needs to be reflected in practical engineering applications.

The debate over the interpretation of *probability* is often simplified to contrasting two alternative views of the meaning of the term. One is that *probability* reflects frequencies of occurrences in the natural world which operate outside human agency, or at least outside human manipulation. The other is that *probability* reflects degrees of rational belief in the outcome of an event or truth of a proposition that an individual holds in the face of evidence. This often is referred to as the debate between frequency and belief, but the philosophical convolutions of this debate are more complex and subtle than the simple distinction suggests (Hacking 1975).

For engineering purposes, on the other hand, the dual concepts of probability-as-frequency and probability-as-belief complement one another. They are not competing notions of uncertainty; rather they are measures of different things. Just as differential fields can be used to represent such different things as groundwater and magnetism, so, too, probability can be used to represent such different things as frequency and belief. Thus, probability is the over-arching framework for grappling with the dual nature of uncertainty: probability-as-frequency is used to grapple with natural variations in the world, while probability-as-belief is used to grapple with limited knowledge. Probability is also the *lingua franca* for integrating the two notions of uncertainty within a uniform analytical reference. Frequency and belief each has a rôle to play, and there is no contest between them.

This is a practical view. There are those who argue that only objective frequencies of the world can be logically treated as probabilities (von Mises *et al*. 1939) and those who argue that even the frequencies we observe in the world can only be interpreted as the subjective belief of the observer (De Finetti 1974). These extreme positions are philosophically engaging but difficult to implement in practical risk and reliability modeling. As Gigerenzer (1991) points out, many people perceive uncertain events in the world through the lens of frequency, no matter the philosopher's view. Thus, a modeling approach that combines frequencies of natural variations with uncertain degrees of belief seems practical.

2.3.1 What is probability and why use it?

Ideas about human behavior and reasoning have long shaped the meaning of *probability* and remain central to understanding the interpretations in use today (Bernstein 1996). The thing most practitioners overlook, however, is that, while all other terms in the probability calculus have well-defined meanings, the term *probability* itself does not; it is a primitive term which is not part of the formal system of probability theory (Salmon 1998).

Probability theory is a logical, mathematical construct based on a small number of axioms. The mathematical aspects of these axioms and their implications are discussed in Chapter 3. As long as one accepts the axioms, all the results of mathematical probability theory follow necessarily. But, within these axioms, only the mathematical properties of probability as a measure are defined. Nowhere do these axioms, and therefore mathematical probability theory as a whole, address the meaning of the concept of probability. That is, probability theory is simply a logical construct, the meaning we give the construct in risk analysis is a philosophical question independent of mathematics. The meaning we assign, however, does affect what can be done with the mathematical theory and the meaning of the numbers or functions that derive from our modeling. As a result, it is pertinent to review briefly how probability theory evolved and the philosophical interpretations imputed to probability over the course of history.

2.3.2 Historical development of probability

The history of probability theory is rich and fascinating, and the interested reader can find a sampling of its intellectual development in a number of places (David 1962; Hacking 1975; Pearson and Pearson 1978; Porter 1986; Stigler 1986; Daston 1988; Gigerenzer *et al*. 1989; Hacking 1990). The early modern references to the development of a probability theory, as related to games of chance, date to Cardano in the 16th century, but today the origin of probability theory is usually traced to a correspondence between Fermat and Pascal just prior to 1654. Out of this arose the combinatorial mathematics that underlie modern probability theory. The Chevalier de Méré became a footnote in history by having suggested to Pascal the famous "problem of points," which has to do with how to divide the stakes in a game of chance that is interrupted before one side wins. The correspondence with Fermat on the problem of points led to the arithmetic triangle of binomial coefficients (David 1962).

Shortly after the Fermat-Pascal correspondence, the Englishman John Graunt in 1662 compiled a bill of mortality for the city of London, charting births and deaths from 1604–1661, and published it with commentary on the meaning of the data. Thus, for the first time statistical data were collected on a social process and published with an intent to inform trade and government.[3] The popularity of the endeavor quickly expanded to continental Europe, where the French undertook a similar compilation for Paris in 1667 (Bernstein 1996). Graunt foreshadowed modern concerns with risk, in commenting that, while many people fear "notorious diseases, [the author] will set down quantitatively how many actually die of them, so that those persons may the better understand the hazard they are in."

The development of probability theory, beyond questions of games and population statistics to physics, principally astronomy and geodesy in the earlier years and a broad range of applications by the end, flowered in the 18th and 19th centuries. This was the work of many people, whom we now think of as among the great scientists, mathematicians, and philosophers of the modern age. Among these were James and Daniel Bernoulli, Laplace, Poisson, Legendre, Gauss, Maxwell, Boltzmann, and Gibbs.

[3] The quote from Graunt, J. *Natural and political observations made upon the bills of mortality*, is "[. . .] to know how many people there be of each Sex, state, age, religious, trade, rank, or degree &c. by the knowing whereof trade and government may be made more certain [. . .]. "(Newman 1966).

So, too, during this age of the Enlightenment, did the nature of uncertainty became part of the emerging inquiry into human knowledge among the adherents of logical deduction from first principles and the proponents of inductive reasoning from empiricism. The former were concerned with fundamental truths and the latter with observed frequencies (David 1962; Hacking 1975; Daston 1988). Between 1710 and 1750, Jacob Bernoulli, John Locke, David Hartley, and David Hume put forward views on knowledge and reason that accommodated both objective and subjective concepts of uncertainty (Gigerenzer, 1994).

A seminal work of this period was Jacob Bernoulli's *Ars Conjectandi* (published posthumously in 1713), which set the stage for the use of probability as logic, or, in his terms, the art of conjecturing. The probability of Bernoulli was to play the central role in moral philosophy, and its implementation was to use the model of jurisprudence rather than that of empirical science. This view followed closely upon that of Leibniz who had recommended that a serious effort be made toward developing a theory of probabilistic reasoning for public purposes. Within this practical discipline, mathematics was to provide a rigorous method of reasoning and a way of insuring consistency when the complexity of the situation might otherwise overwhelm normal reason. Maxwell, at the beginning of the Victorian age, echoes this view, in the quotation at the start of this section.

By about 1850, growing misgivings about the rationality of the common man gave way to the rise of statistical inference and its stature in the increasingly influential fields of physics and social science (Porter 1986; 1995). Throughout most of the 19th and early 20th centuries, probability became defined by frequencies of occurrence in long series of similar trials, largely to the exclusion of degree of belief. In engineering, this view was most associated with Richard von Mises (Wald 1957). Beginning with the pioneering work of Ramsey (Ramsey and Braithwaite 1931) and De Finetti (1937), followed by Savage (1954) after World War II, the degree-of-belief view gradually regained acceptance. The rapid advance of operations research and decision analysis during the war had been an important factor. This duality of frequency and belief remains today, with frequency dominating the experimental sciences, and belief prevailing in economics, management, and public policy. An exposition of the practical differences is given by Barnett (1999).

2.3.3 The meaning of probability

The theory of probability is an internally coherent mathematical construct based on axioms originally due to Kolmogorov (Carnap 1936). There exist parallel, equivalent sets of axioms upon which the theory of probability can also be founded. The point made earlier, but worth repeating, is that probability is a purely mathematical theory. If the axioms are accepted, then all the results of the theory hold necessarily. Thus, there is no disagreement about the properties that probability obeys, and these properties can be found in any textbook on probability theory.

The issue is not what properties *probability* has, but what the term means. Since the beginning of modern probability theory in the 1600's, there has been a dual meaning: (a) relative frequency in a long or infinite number of trials, and (b) objective or subjective degree of belief. Poison and Cournot used the French words *chance* and *probabilité* to refer to the two concepts of frequency and belief. The terms have more distinct meaning in French than their translations in English suggest. *Chance* denotes an objective propensity of an event, the facility with which it can occur. An event will have, by its very nature, a

larger or smaller *chance*, known or unknown. *Probabilité* denotes credibility or degree of reasonable belief. The probability of an event is the reason for thinking that the event did or will occur. The French terms parallel the distinction between aleatory and epistemic uncertainty.

How can a theory describe two quite different concepts? As Pólya (1954) notes, this is not unusual in science, as the example of groundwater flow and magnetism suggests. The meanings of probability are not in opposition, although strongly stated points of view in the literature might lead one to think otherwise. The meanings describe distinct concepts, both of which are part of most geotechnical risk and reliability studies. The concepts coexist in practical applications because the natural variability and knowledge uncertainty coexist in practical applications. The issue is not which meaning to adopt, but how to convolve the two types of uncertainty and thus the two types of probability in a useful way.

In recent years, a third concept of uncertainty and correspondingly a third sense of probability has appeared in the engineering literature of risk and reliability. This is the uncertainty caused by vagueness of language and definition. For example, it is suggested in this literature that concepts such as 'stiff' clay or 'poorly graded' filter material are imprecisely defined, therefore a new way of capturing the uncertainty attending this vagueness is needed. A variety of non-bivalent measures of uncertainty (e.g. 'fuzzy sets') have been proposed for the purpose. The present discussion takes the position that what is needed is not a replacement for probability theory but better definitions for the vague terms, as in the traditional scientific approach.

2.3.4 Probability as relative frequency

In much of engineering, probability is interpreted to be the frequency of occurrence of some event in a long series of similar trials. A principal early proponent of this view was John Venn (1834–1923), who summarized the notion as,

> The fundamental Conception is that of a series which combines individual irregularity with aggregate regularity. Probability has no meaning except in connection with such a series. Any probability must be referred to a series. The probability of an event is its relative frequency in the series.

A more modern proponent was Richard von Mises (1883–1953).

A trial is an individual occurrence producing an outcome of some sort. For example, each individual lift of soil placed in a compacted embankment might be considered a trial. The frequency of soils having low moisture content among these lifts (i.e., among the trials) would be the probability of soil with low moisture content. Similarly, each maximum annual discharge in a stream might be considered a trial. The frequency of peak annual flows above a certain magnitude would be the probability of floods of at least a certain size.

Frequency definitions of probability are the ones that non-statisticians usually think of first when asked to define probability. They are based on the relative number of times a certain outcome will occur in a long or infinite number of similar trials. There are three traditional definitions of frequency, which might be called, the classic, the Venn limit, and the infinite trials. The classic definition defines probability *a priori*. If there are n possible

alternatives, for m of which A is true, then the probability of A is defined as m/n. The Venn limit says that if an event occurs a large number of times, the probability of A is the limit of the ratio of the number of trials on which A is true to the whole number of trials, as the number of trials tends to infinity. The infinite trials definition assumes an infinite number of trials and defines the probability of A as the ratio of cases in which A is true to the whole number. The classic definition is associated with the work of de Moivre and Cramer, the Venn limit with that of Venn and von Mises, and the infinite trials with that of Gibbs and Fisher. The implicit assumption underlying all these definitions is that one cannot place probabilities on states of nature or on unique events. States of nature are constants (perhaps unknown), not random variables. To ask the probability of a flood stage higher than some value occurring this year is meaningless; the probability of a single occurrence is undefined.

The frequency theory gives answers in the form of the probability of different sets of observations occurring, given some hypothesis. The practical problem is just the reverse of this. That is, the practical question usually is, what is the probability of some hypothesis being true, given the observations recorded. This is illustrated in Figure 2.2 (Baecher 1972). To say, "if a fault existed at this site, the aerial photographs would show its trace 90% of the time," is not to say that if a trace does not appear, the probability of the fault existing at the site is only 10%. The probability of a trace showing in the air photos given that a fault exists is $B/(A + B)$, which is permitted by frequentist theory; whereas the probability of a fault existing given a trace is observed in the air photos is $B/(B + D)$, which is undefined in frequentist theory.

The frequency approach requires at least the potential for large numbers of similar trials, for example, many realizations of a process, or a sample of many measurements from a population. The frequency interpretation pertains to variability among collections of things. A frequency interpretation cannot be applied to the unique occurrence of an event, or to a condition where statistical sampling with representative trials is impossible, or to the direct inference of a state of nature. Proponents would say that the frequentist interpretation is objective, because it deals only with observable patterns of behavior within groups of similar things. Inference from these patterns about the state of nature does not follow necessarily from the mathematics, but this is acceptable within the frequentist

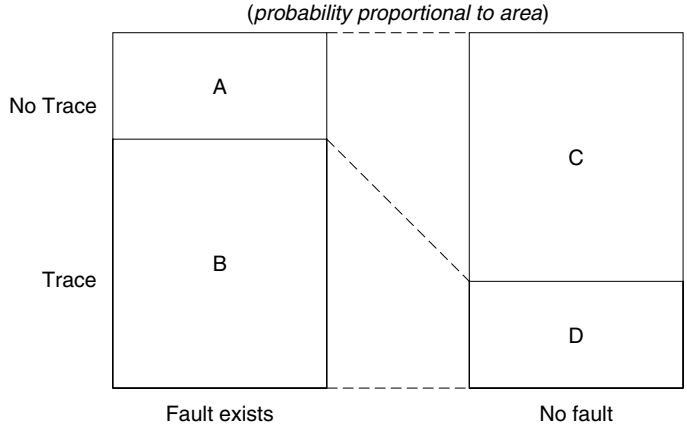


Figure 2.2 Conditional probabilities of observing evidence of a geological fault.

school. Inference within the frequentist school is based on refutation and is thus *ad hoc* (see, e.g., Barnett 1999).

Consider again a deck of cards. The order of the cards is not random, although we may not know what it is. A frequentist interpretation allows statements to be made about where the Ace of Clubs may appear in many re-shuffles of the deck. A belief interpretation allows statements about the location of the Ace of Clubs within one particular shuffle. Unlike a deck of cards that can be shuffled repeatedly forming a long series of trials, a geological formation is created once. Thus, uncertainties about soil and rock properties are not averages over ensembles but degrees of belief. For example, drilling to discover the presence of a geologic fault at some specified location is not replicable. The fault is found or it is not. A frequentist would consider any statement about the chance the fault's existence to be outside probability altogether, even though geologic knowledge may provide partial evidence.

2.3.5 Probability as degree of belief

The belief interpretation, quite common in civil engineering, holds that probability is a rational degree of belief. The probability that an important solution cavity exists in a limestone dam abutment is typical of a dam-related problem that cannot be easily approached using the frequency definition of probability, since the uncertainty is fundamentally epistemic. Such probabilities have to do with one-time events, past experience, and amounts of information. They are personal and subjective, and not easily related to frequencies, actual or conceptual, as summarized by Augustus de Morgan (1845),

> Probability is the feeling of the mind, not the inherent property of a set of circumstances. [...] the amount of our belief of one proposition varies with the amount of our belief of other propositions with which it is connected.

According to Ramsey and Braithwaite (1931), a degree of belief is the propensity to take action when faced with a situation of uncertainty. To say that someone has degree of belief p that a proposition is true means that for integers r and b with $r/(r+b) < p$, if that individual is offered an opportunity to bet an equal amount on the truth of a proposition or on "red in a single draw" from an urn containing r red and b black balls, he will prefer the first bet; if $r/(r+b) > p$, he will prefer the second bet. Degree of belief is manifest in action, not in introspection.

The notion of degrees of belief as subjective probability was much debated in the early half of the 20th century. While, today, the notion is widely accepted, in earlier times significant efforts were devoted to developing an objective degree-of-belief theory. This theory holds that two observers of the same data necessarily arrive at the same degree of belief in a proposition. Subjective degree-of-belief theory says that two observers of the same data may arrive at different degrees of belief in a proposition. Objective degree-of-belief theory is most notably associated with the work of Jeffreys, the great Cambridge geophysicist, but it has fallen in to disfavor in recent times.

Jeffreys postulates an objective degree of belief by presuming that every observer starts from a condition of complete ignorance about the quantity to be inferred and in this state should have the same prior probability distribution. Jeffreys' *principle of indifference* states that in a condition of ignorance, one should assign equal probabilities to all possible

outcomes of an experiment or values of a parameter. Thus, if a coin can land heads-up or heads-down, and if one has no information about the possible bias (unfairness) of the coin, then the rational assignment of probability *a priori* is to treat each possible outcome as equi-probable, that is, to assign each the value 0.5. This uniform assignment of prior probability is called "non-informative." In the same way, if one had no information about the value of, say, the mean strength of a soil, then the prior probability distribution over its possible values should be taken as uniform within the region of the possible sample outcomes. The principal difficulty of this approach is that the resulting prior probability distributions expressing total ignorance are not invariant to transformations of variable. As a result, objective degree-of-belief theory has fallen from favor.

The degree-of-belief theory does not require a number of trials. It provides that probabilities can be assigned to unknown outcomes of particular realizations of a process, or to the very properties (states) of nature itself. The belief interpretation pertains to the weight of evidence leading to a quantitative statement of uncertainty about nature. Thus, a belief interpretation can be applied to the unique occurrence of an event. On the other hand, the belief interpretation necessarily implies that probability is in the eye of the beholder, it is necessarily subjective. Two observers can view the same data, assess different probabilities, and each, working from his or her experience and knowledge, can be right.

A subjective approach considers information of all kinds, statistical and otherwise, as admissible in developing a probability estimate. The way people use various types of information in formulating subjective probabilities is not fixed; they adopt statistical and knowledge-based reasoning or a mixture of the two depending on the circumstances. A familiar case is a weather forecaster's prediction of rain. While climatological records for the day in question provide a base-rate frequency, information from satellite photographs, atmospheric pressures, and even a glance out the window are also incorporated in a probability estimate. In fact, weather forecasters have served as subjects in behavioral studies for just this reason, and they have been found to forecast probabilities that correspond closely to long-run frequencies (Murphy and Winkler, 1974).

Likewise in geotechnical applications, all the information at hand, together with personal experience and judgment are brought to bear in estimating the subjective probability of an event or condition. Because subjective probability inherently depends on one's state of knowledge at the time it is formulated, any such value can be expected to vary, both from one person to another and with time, as information and knowledge are gained, in the same way as (engineering) judgment itself. Subjective probability can be said to be a reflection of judgment.

3 Probability

Engineering data on soil or rock mass properties are usually scattered. Graphical and simple probabilistic methods are useful in summarizing this scatter so that a better understanding of the data – and of the corresponding uncertainties associated with engineering performance – can be developed. In this chapter, we introduce basic concepts of probability theory and the use of probability in describing geotechnical uncertainties. This chapter is intentionally brief. The fundamental concepts of probability theory are well described, and in greater detail than possible here, in many texts. For example, Benjamin and Cornell (1970) and Ang and Tang (1975, 1990) provide eminently readable introductions to basic probabilistic concepts in the context of civil and environmental engineering, while Feller (1967, 1971) provides a general introduction of substantial depth. A principal purpose of the present chapter, then, is to establish a vocabulary to follow through the rest of the book. Appendix A contains a more detailed discussion of probabilistic concepts and of the properties of various distributions.

3.1 Histograms and Frequency Diagrams

The most common way to represent scattered data graphically is a histogram. A histogram graphs the number of measurements falling within specific intervals of value as a vertical bar. Thus, a histogram is sometimes called a bar chart. The height of the bar above each interval shows the number of measured values within the interval, and the sum of the heights of the bars equals the total number of measurements (Figure 3.1). The histogram divides data into fixed intervals. The choice of intervals is arbitrary, but they should be of uniform width and have convenient end points. If too many intervals are chosen the general picture of relative frequencies will be jagged, and, conversely, if too few intervals are chosen, the general picture will be blurred. A frequency distribution is constructed from a histogram by dividing each vertical bar by the total number of measurements. This gives the relative frequency of observed values in each interval as a decimal fraction. The cumulative frequency distribution, shown as a solid curve in Figure 3.1, sums the number of data less than or equal to a particular value.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

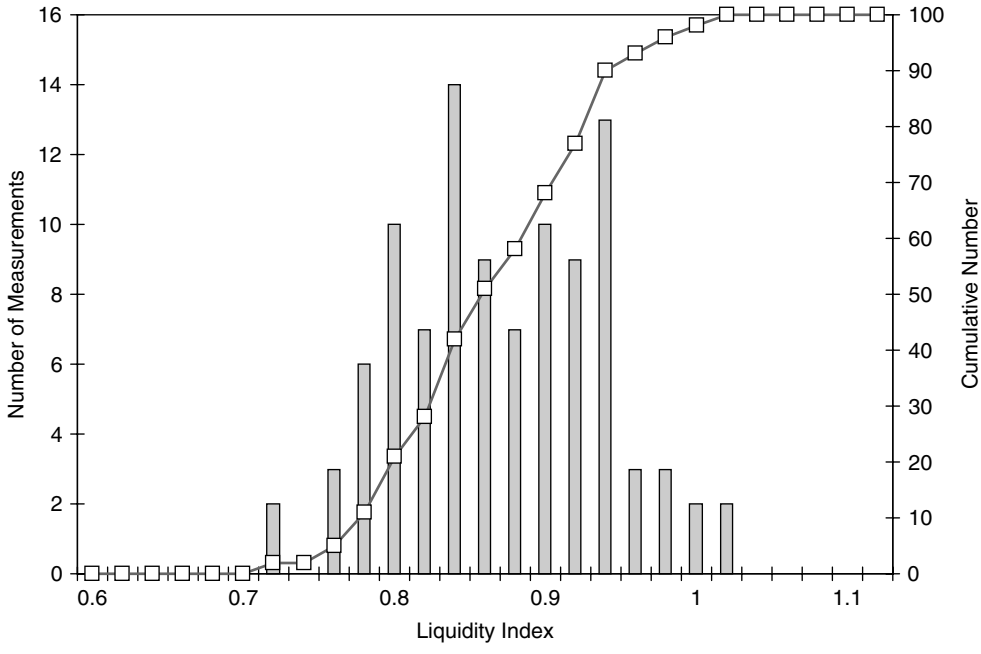


Figure 3.1 Example of a histogram and cumulative distribution of liquidity index data from a marine clay.

Many important features of the data are apparent by inspection of the frequency distribution. For example, the data in Figure 3.1 vary about a central peak between 0.84 and 0.86. The data are more or less symmetric about this peak, and data that vary substantially from the peak are infrequent. The bulk of the data lie within an interval approximately between 0.75 and 0.95, with extreme values ranging to 0.71 and 1.02. A symmetric distribution like this is sometimes described as 'bell-shaped.'

A histogram gives a summary view of variation. The shape suggests whether the data have a central tendency and, if so, where along the x-axis the bulk of the data is concentrated. The width suggests the dispersion or scale of variation. Some histograms have one point of concentration and are thus called unimodal. Others have more than one and are called multimodal. Usually, hydrological and geotechnical data have unimodal distributions. Multimodal distributions may indicate an inhomogeneous mixture of data from different soil types, hydraulic regimes, construction procedures, or sites. The histogram also shows whether the variation is symmetric or asymmetric, that is, whether high and low variations are evenly balanced, whether variations from the central tendency are more frequent on one side than on the other.

A histogram and cumulative distribution of a set of cone penetration data from a copper porphyry tailings dam are shown in Figure 3.2. These data are not symmetric about a peak frequency, but are skewed with a long upper tail. The highest frequency occurs near the lower end of the scale, at about 2 psi (13.8 kPa), and, while frequencies decline on both sides of the peak, they do so more slowly on the upper side. Such distributions are said to be skewed.

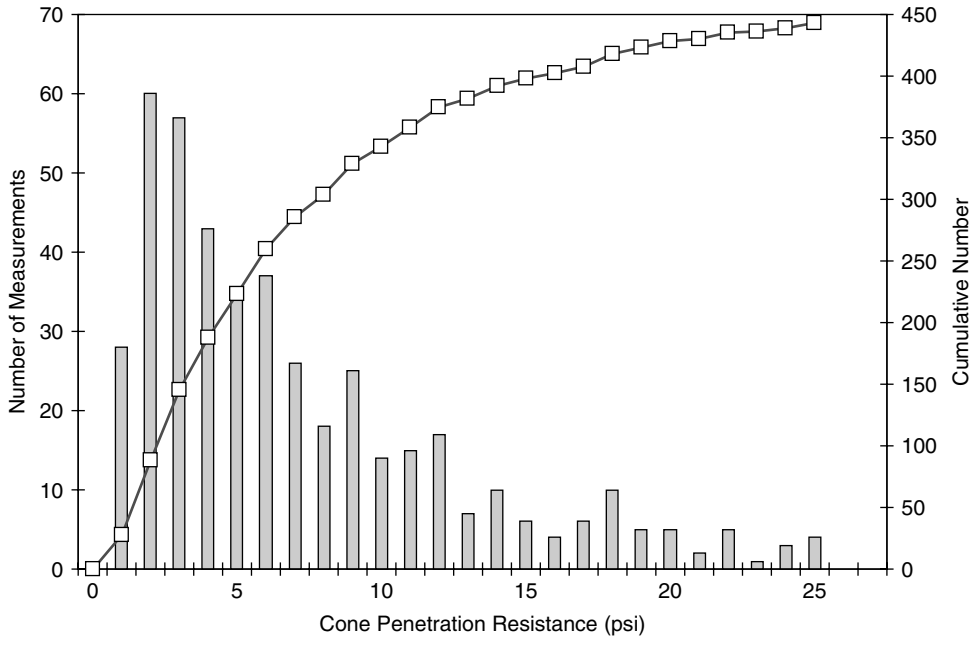


Figure 3.2 Histogram and cumulative distribution of a set of cone penetration data from a copper porphyry tailings dam (Baecher *et al*. 1980).

To construct a histogram, a simple procedure can be used: (1) divide the data range into about five to ten intervals of constant width; (2) count the number of data having values within each interval; and (3) plot this number as a vertical bar. Five to ten intervals are used because this number typically allows a sufficient number of data in each interval for the observed frequencies to vary smoothly, yet provides adequate definition to the shape of the distribution. If larger numbers of data are available, larger numbers of intervals might be used. For small numbers of data, a convenient rule-of-thumb for choosing the number of intervals is

$$k = 1 + 3.3 \log_{10} n \tag{3.1}$$

in which $k =$ the number of intervals, and $n =$ the number of data values (Sturges 1926). Usually, it is convenient to specify interval boundaries to one fewer decimal place that to which the data are measured, avoiding the problem of where to place values falling directly on an interval boundary. When this is not possible, a consistent procedure should be adopted for deciding how to count data that fall directly on a boundary. Since the choice of number of intervals can affect visual interpretation of data scatter, it is sometimes useful to construct more than one histogram, using a different number of intervals for each.

A cumulative distribution is constructed by summing relative frequencies starting at the lower end of the data and proceeding toward the upper end. The cumulative frequency distribution gives the fraction of measurements less than or equal to a particular value:

$$F_X(x_0) = \textit{fraction of measurements } \leq x_0 \tag{3.2}$$

An empirical cumulative frequency distribution is found by plotting the data of Figures 3.1 or 3.2 against fraction-less-than. The plotting positions in a cumulative frequency distribution are usually taken as

$$F(x_m) = \frac{m}{N+1} \tag{3.3}$$

in which $F(x_m)$ is the cumulative frequency of the mth point, with $m = 1$ being the lowest value, and N being the total number of points (McCuen and Snyder 1986). Sometimes the plotting points are taken as $F(x_m) = m/N$, for all but the largest value, which is then plotted as in Equation (3.3). An advantage of the cumulative distribution is that the data do not need to be grouped into an arbitrary number of intervals. A disadvantage is that the shape of the distribution is not as readily apparent as in the histogram.

3.2 Summary Statistics

Frequency distributions are convenient representations of data for visual inspection, but statistics of distribution characteristics are often more useful for calculations or for setting standards. The term *statistic* refers to any mathematical function of a set of measured data. For example, given measurements $\boldsymbol{x} = \{x_1, \ldots, x_m\}$, any function $T(\boldsymbol{x})$ is a statistic of the data. The arithmetical average is such a statistic, the largest value $x_{\max}$ or the smallest value $x_{\min}$ is a statistic, and so on. There are infinitely many statistics that could be calculated from a set of data, but the most useful usually have to do either with central tendency or dispersion (spread) of the data.

3.2.1 Central Tendency

The common measures of central tendency are the mean, median, and mode. The *mean* is the arithmetic average of a set of data. The *median* is the value for which half the observations are smaller and half larger. The *mode* is the most frequent value.

The mean of a set of n data $x = \{x_1, \ldots, x_n\}$, denoted $\overline{x}$, is the arithmetical average

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3.4}$$

The mean is the center of gravity of the frequency distribution along the x-axis. For example, the mean of the liquidity index data of Figure 3.1 is 0.87, while that of the cone penetration data in Figure 3.2 is 7.13.

The median of the set of data, denoted $x_{0.5}$, is that value of x_n for which half the data are less and half more. Correspondingly, the cumulative distribution evaluated at the median is 0.5:

$$F_X(x_{0.5}) = 0.5 \tag{3.5}$$

The median is the midpoint of the data when listed in increasing or decreasing order. Common practice in the case of an even number of data is to define the median as intermediate between the two middle data.

The mode of a set of data is most common data element or interval. For theoretical distributions the mode is usually obvious, but in sample data it can be erratic or unstable depending of the vagaries with which data appear. In Figure 3.1 the mode is 0.84, although it might nearly have been 0.94. In Figure 3.2 the mode is 2.

3.2.2 Dispersion

Common measures of dispersion are the *standard deviation, range*, and *inner quartiles* of the frequency distribution. The standard deviation is the root-mean-square (rms) value of the difference between the data and the mean, the range is the spread between maximum and minimum values, and the inner quartiles bracket the middle 50% of the data.

The standard deviation of a set of data $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, denoted s_x, is

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}} \tag{3.6}$$

in which $\overline{x}$ = the mean of the data. Sometimes the denominator $(n-1)$ rather than (n) is used to correct a statistical bias (Chapter 4), due to the fact that the mean, too, must be estimated from the same data. The *coefficient of variation* of a set of data is defined as the standard deviation divided by the mean

$$\Omega_X = \frac{s_x}{\overline{x}} \tag{3.7}$$

which expresses relative dispersion. The *variance* of a set of data, denoted *Var(x)*, is the square of the standard deviation

$$Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{3.8}$$

In many statistical calculations, the variance is a more convenient measure than the standard deviation. The variance is the moment of inertia of the frequency distribution about $\overline{x}$.

The range of a set of data, denoted r, is the difference between the largest and smallest values,

$$r_x = |x_{\max} - x_{\min}| \tag{3.9}$$

The range has poor statistical properties in that it is sensitive to extreme values in a data set; however, it is easily evaluated and therefore often useful.

The inner quartiles of a set of data, denoted $x_{0.25}$ and $x_{0.75}$, are the data values for which one-quarter of the data are smaller and one-quarter larger, respectively. The quartiles may be found from the cumulative distribution as

$$F(x_{0.25}) = 0.25 \tag{3.10}$$

$$F(x_{0.75}) = 0.75 \tag{3.11}$$

Table 3.1 Summary statistics for the data of Figures 3.1 and 3.2

Statistic	Liquidity index (Figure 3.1)	Cone penetration (Figure 3.2)
Mean	0.87	7.13
Median	0.86	5.00
Mode	0.84	2.00
Standard Deviation	0.067	8.64
Coefficient of Variation	0.077	1.21
Variance	0.0045	74.63
Lower Quartile	0.81	2.50
Upper Quartile	0.91	9.30
Interquartile Range	0.10	5.80

The *inter-quartile range*, denoted $r_{0.5}$,

$$r_{0.5} = (x_{0.75} - x_{0.25}) \tag{3.12}$$

is less influenced by extreme values than is the range itself, but it is correspondingly more troublesome to compute. Various summary statistics applied to the data of Figures 3.1 and 3.2 are shown in Table 3.1.

3.2.3 Quick estimates of central tendency and dispersion

Often one wants quick, approximate estimates of means, standard deviations, or other summary statistics from limited numbers of data. Shortcut techniques are available for this purpose. These provide economies of time and effort while sometimes causing only minor loss of precision.

A quick and often good estimate of central tendency can be obtained from the median. For data scatter that is symmetric about its central value and for small numbers of data, the sample median is a good estimate of the mean. On the other hand, if the data scatter is asymmetric – for example, if there are many small values and a few large values – the sample median is not such a good estimator. A second shortcut for estimating the mean is taking one-half the sum of the largest and smallest values, $(1/2)(x_{\max} + x_{\min})$. This estimator is sensitive to extreme values in a set of measurements, and thus fluctuates considerably. It should be used with caution.

A quick estimate of dispersion from small numbers of tests can be made from the sample range $r_x = (x_{\max} - x_{\min})$. Like the standard deviation, the range is a measure of dispersion in a set of data. However, the relationship between the standard deviation and the sample range, on average, depends on how many observations are made. If the data are Normally distributed, to obtain an estimate of s_x from the range of data r_x one uses a multiplier N_n (i.e. a correction factor) dependent on sample size (Table 3.2). The estimate is then

$$\hat{s}_x = N_n(x_{\max} - x_{\min}) \tag{3.13}$$

in which $\hat{s}_x$ indicates an estimate.

As in the case of the sample median, the range is a good estimator of the standard deviation for symmetric data scatter with small n. However, for asymmetric data scatter

Table 3.2 Multiplier for estimating standard deviation from sample range for normally distributed variable (Burington and May 1970; Snedecor and Cochran 1989)

n	N_n	n	N_n	n	N_n
		11	0.315	30	0.244
2	0.886	12	0.307	50	0.222
3	0.510	13	0.300	75	0.208
4	0.486	14	0.294	100	0.199
5	0.430	15	0.288	150	0.19
6	0.395	16	0.283	200	0.18
7	0.370	17	0.279		
8	0.351	18	0.275		
9	0.337	19	0.271		
10	0.325	20	0.268		

the range, which is strongly affected by extremes, is not a good estimator of standard deviation. Fortunately, with the notable exception of hydraulic parameters such as permeability, most geotechnical data display symmetric scatter. In the case of hydraulic data a logarithmic transformation usually makes data scatter symmetric (Kitanidis 1997), and again the median and range become reasonable estimators.

3.3 Probability Theory

Probability theory is a branch of mathematics. It is logically consistent in the sense that the mathematics of probability theory all can be derived from a set of axioms. These axioms specify properties that *probability* must have, but do not say what *probability* is. As a result many interpretations are possible.

One set of axioms that probability theory can be based on is the following, although there are others (Kolmogorov 1950). The technical terms used in the axioms are defined in the discussion to follow:

Axiom 1: the probability $P[A]$ of event A has a value between 0 and 1: $0 \leq P[A] \leq 1$.
Axiom 2: the sum of the respective probabilities of each of a set of mutually exclusive and collectively exhaustive events $\{A_i\}$ is 1.0: $\sum_i P[A_i] = 1$.
Axiom 3: the probability that two independent events A_i and A_j both occur equals the product of their individual probabilities: $P[A_i \text{ and } A_j] = P[A_i]P[A_j]$.

All the formal mathematical relationships of probability theory derive from these simple axioms.

3.3.1 Random events

The mathematical theory of probability deals with experiments and their outcomes. An *experiment* is a random process generating specific and *a priori* unknown results or *outcomes*. The set all possible outcomes of an experiment is called the *sample space*, S

(sometimes called, the *outcome space*). Individual outcomes (points) within the space are called *sample points*. The sample space can be discrete or continuous.

The example sample space of Figure 3.3 contains 77 discrete sample points. A simple event consists of a single sample point; a compound event consists of more than one sample point. The event, A, as shown, is a collection of 15 sample points within S. The *complement* of A, denoted $\overline{A}$, comprises the $62 = 77 - 15$ sample points in S not in A.

The *union* of two events A and B (denoted A-or-B, or symbolically $A \cup B$) is the collection of sample points contained in A or B or both. The *intersection* of two events A and B (denoted A-and-B, or symbolically $A \cap B$) is the collection of sample points contained in both A and B. The well known Venn diagram (Figure 3.4) conventionally illustrates these relations (Venn 1866). The term *mutually exclusive* means that two (or more) events cannot occur together, that is, the events share no sample points in common, while *collectively exhaustive* means that at least one event within the group of events must occur, that is, together the events comprise all the sample points in S.

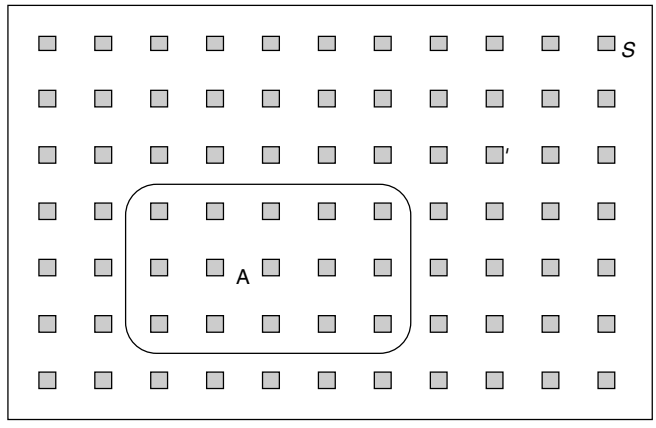


Figure 3.3 Sample space showing sample points and an event A.

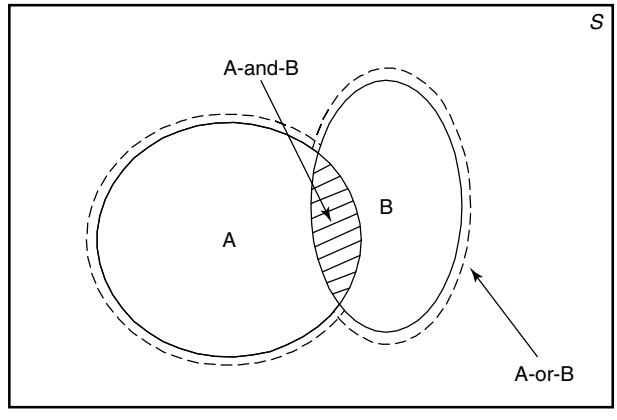


Figure 3.4 Venn diagram of events A and B in sample space S.

3.3.2 Conditional probability, dependence, and independence

One of the most important concepts in probability theory is *conditional probability*, with its corollaries, *dependence* and *independence*. Informally, conditional probability has to do with the notion that the probability of an event A may be changed if another event B is known to obtain. For example, the probability assigned to high values for the undrained strength of a clay is presumably changed by knowing the preconsolidation pressure of the clay. From empirical observation and theory, high values of undrained strength are associated with high preconsolidation pressures, and conversely. Thus, the probability of high undrained strength should rise if we learn that the preconsolidation pressure is high.

The conditional probability of event A, given event B, is written, $P[A|B]$. The probability $P[A]$, irrespective of B, is said to be the *marginal probability*, for reasons that will become clear later in the chapter. For the case in which the conditional probability is the same as the marginal

$$P[A|B] = P[A] \tag{3.14}$$

A is said to be *independent* of B. For the case in which the conditional probability differs from the marginal, $P[A|B] \neq P[A]$, A is said to be *dependent* on B. The intuitive meaning of independence is that the probability of A is not changed by knowing that B obtains. The definition is symmetric in A and B, and when applying to both, A and B are said to be *mutually independent*.

From an informal consideration of Figure 3.4

$$\begin{aligned} P[A \text{ and } B] &= P[A]P[B|A] \\ &= P[B]P[A|B] \end{aligned} \tag{3.15}$$

can be rearranged and combined with the second axiom to give the relationship among conditional probabilities known as *Bayes' Theorem*:

$$P[A|B] = \frac{P[A \text{ and } B]}{P[B]} = \frac{P[A]P[B|A]}{P[A]P[B|A] + P[\overline{A}]P[B|\overline{A}]} \tag{3.16}$$

The expression $\overline{A}$ means 'not A.'

Table 3.3 Elementary relationships among probabilities of events

Description	Equation
Event and its complement	$P[\overline{A}] = 1 - P[A]$
Intersection of two events	$P[A \text{ and } B] = P[A] + P[B] - P[A \text{ or } B]$ $P[A \text{ and } B] = P[A]P[B\|A]$
Intersection of mutually exclusive and collectively exhaustive events	$P[A_1 \text{ and } \ldots A_n] = \prod_{i=1}^{n} P[A_i]$
Union of two events	$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$
Union of mutually exclusive and collectively exhaustive events	$P[A_1 \text{ or } \ldots A_n] = \sum_{i=1}^{n} P[A_i]$
Total probability	$P[A] = \sum_{i=1}^{n} P[A \text{ and } B_i]$, s.t. B_i me & ce
Conditional probability	$P[B\|A] = \dfrac{P[A \text{ and } B]}{P[A]}$

In summary, a number of mathematical relationships derive immediately from the axioms of probability. These are presented without proof in Table 3.3. The derivations of these and other elementary relationships of probability theory can be found in Benjamin and Cornell (1970), or Ang and Tang (1975), or in any introductory text on probability.

3.4 Random Variables

In practice, we are often concerned with mathematical functions of natural variations or of uncertainties. Thus, the question arises, is there not a more convenient way of describing the distribution of probability over a space of outcomes than fully enumerating separate values of $P[A]$ for every possible outcome?

For many distributions of probability over a sample space, S, there exists a function $f_X(x)$ from which the probability $P[A]$ for any event A can be obtained by a summation of the form

$$P[A] = \sum\nolimits_A f_X[x] \quad x \in A \tag{3.17}$$

For discrete sample spaces, this function is called, a *probability mass function* (pmf), and for continuous sample spaces, a *probability density function* (pdf). In either case, we will use the generic notation, $f(x)$, to indicate the probability distribution. In applying Equation (3.17) to a continuous variable, the summation is replaced by integration.

3.4.1 Probability mass functions

The simplest case of a discrete sample space is an ordered series, for example, the frequency distribution corresponding to either Figures 3.1 or 3.2. This ordered series associates a probability, p_i, with each sample point, x_i

$$X : \begin{matrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{matrix} \tag{3.18}$$

The probability that the uncertain quantity X takes on a particular value, x, is

$$f_X(x) = P(X = x) \tag{3.19}$$

which is the probability mass function. Since the outcome takes on exactly one value, the sum of the pmf over S is

$$\sum_{i=1}^{n} f_X(x_i) = 1 \tag{3.20}$$

The Cumulative Mass Function (CMF) describes the probability that the outcome of X is less than or equal to a particular value:

$$F_X(x_j) = \sum_{i=1}^{j} f_X(x_i) \tag{3.21}$$

The CMF rises monotonically from zero and approaches one as ever more of the sample space is included in the summation.

3.4.2 Probability density functions

For a continuous sample space, S, it makes no sense to speak of masses of probability, since there are infinitely many outcomes within the space. Instead, we deal with probability as a density, as we do with the density of gravitational mass in a soil formation. Probability is then found by integrating the probability mass over a finite region of S.

The distribution of probability mass over the sample space is the pdf, above. This function has properties corresponding to those of the pmf. Specifically, for the case of a scalar real-valued variable

$$P(A) = \int_A f_X(x)dx \tag{3.22}$$

or correspondingly

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f_X(x)dx \tag{3.23}$$

Since the outcome takes on exactly one value, the integral of the pdf over the sample space is

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1 \tag{3.24}$$

The Cumulative Density Function (CDF) describes the probability that the outcome of X is less than or equal to a particular value

$$\begin{aligned} F_X(x_i) &= P(x \leq x_i) \\ &= \int_{-\infty}^{x_i} f_X(x)dx \end{aligned} \tag{3.25}$$

As with the CMF, the CDF rises monotonically from zero approaching one as ever more of S is included in the integral

$$\begin{aligned} F_X(-\infty) &= 0 \\ F_X(+\infty) &= 1 \end{aligned} \tag{3.26}$$

3.4.3 Moments of probability distributions

As in engineering mechanics, is it mathematically convenient sometimes to represent probability distributions by their moments. The nth moment of a probability distribution about the origin – for either pmf or pdf, but we will use the common notation $f_X(x)$ to indicate either – is

$$E(x^n) = \int_{-\infty}^{+\infty} x^n f_X(x)dx \tag{3.27}$$

in which the weighted integral over the probability distribution is said to be the *expectation.* For the case, $n = 1$, this is the arithmetic average or mean, also called, the *expected value of* x, and is often denoted $E(x) = \mu$.

The nth moment about the mean, said to be the nth *central moment*, is

$$E[x - E(x)]^n = \int_{-\infty}^{+\infty} [x - E(x)]^n f_X(x) dx \tag{3.28}$$

The most common is the second central moment, called the *variance*

$$E[x - E(x)]^2 = \int_{-\infty}^{+\infty} [x - E(x)]^2 f_X(x) dx \tag{3.29}$$

often denoted $E[x - E(x)]^2 = \sigma^2$.

The third and fourth central moments also capture qualities of the distribution of probability that are useful in practice. The third central moment, or *skew*, $E[x - E(x)]^3$, is used as a measure of asymmetry in the distribution of probability over X. The larger this absolute value, the more asymmetric the pmf or pdf. A symmetric distribution has zero skew, a skew toward higher values of x has positive value, and a skew toward lower values of x has negative value. The fourth central moment, or *kurtosis*, $E[x - E(x)]^4$, is used to measure the degree of peakedness of the distribution of probability in the central region around the mean. The greater the kurtosis, the more peaked the pmf or pdf, and correspondingly, the thinner the tails of the distribution for values of x distant from the mean.[1]

3.4.4 Mathematical models for probability distributions

For many problems it is convenient to approximate the probability distribution by a mathematical function. Surprisingly, a comparatively small set of mathematical functions can be used to fit a broad range of frequency distributions encountered in practice, and we return to these in the next section in discussing random process models. Johnson and Kotz (1969; 1970) and Evans *et al.* (1993) give mathematical equations, ranges, and moments for a variety of commonly used probability mass and density functions.

In the late nineteenth century, Karl Pearson attempted to systematize families of probability distributions. The system consists of seven solutions (of twelve originally enumerated by Pearson) to a differential equation, which also approximate a wide range of distributions of different shapes. Gruska, Mirkhani, and Lamberson (1973) describe in detail how the different Pearson curves can be fit to an empirical distribution. This system, using a simple graph to plot the relationship of families of distributions to functions of their moments, is widely used in hydrologic and geotechnical engineering. Lumb (1994) gives an example, and the technique is also used in Chapter 8 of this volume. Detailed explanation of Pearson's families of distributions is given by Ord (1972).

Johnson (1960) described another system of frequency curves representing transformations of the standard Normal curve (see Hahn and Shapiro 1967). By applying such

[1] For reference, the kurtosis of a Normal distribution is three (3), and this value is often used as a reference in comparing the peakedness of other probability distributions.

transformations, a wide variety of non-Normal distributions can be approximated, including distributions which are bounded on either one or both sides (e.g. U-shaped distributions). The advantage of this approach is that once a particular Johnson curve has been fit, the normal integral can be used to compute expected percentage points under the curve. Methods for fitting Johnson curves, so as to approximate the first four moments of an empirical distribution, are described by Hahn and Shapiro, (1967: 199–220).

3.4.5 Multiple variables

When more than one variable is of concern, the probability mass function is directly extended to a joint pmf. For the bivariate case

$$f_{x,y}(x, y) = P[(X = x) \text{ and } (Y = y)] \tag{3.30}$$

$$F_{x,y}(x, y) = P[(X \leq x) \text{ and } (Y \leq y)] \tag{3.31}$$

The joint probability density function is the continuous analog of the joint pmf

$$f_{x,y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{x,y}(x, y) \tag{3.32}$$

The marginal distribution of one variable irrespective of the other(s) is found by integrating (summing in the discrete case) over the distribution of probability in the other variable(s). For example,

$$f_X(x) = \int_Y f_{x,y}(x, y) dy \tag{3.33}$$

This is called 'marginal' because historically common practice was to plot it along the axis, i. e. in the margin, of the plot of a bivariate distribution (Pearson and Pearson 1978).

The conditional distribution of a variable x given another variable y is found from the joint distribution by renormalizing with respect to the second variable

$$f_{X|Y}(x, y_0) = \frac{f_{X,Y}(x, y)}{f_Y(y_0)} \tag{3.34}$$

For the special case in which X and Y are independent, the following relations hold:

$$f_{X|Y}(x|y) = f_X(x) \tag{3.35}$$

$$f_{Y|X}(x|y) = f_Y(y) \tag{3.36}$$

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \tag{3.37}$$

$$F_{X,Y}(x, y) = F_X(x) F_Y(y) \tag{3.38}$$

$$F_{X|Y}(x|y) = F_X(x) \tag{3.39}$$

$$F_{Y|X}(y|x) = F_Y(y) \tag{3.40}$$

When the variables are dependent, these simple multiplicative relations do not hold.

3.5 Random Process Models

In many cases, there are physical considerations that suggest appropriate forms for the probability distribution function of an uncertain quantity. In such cases there may be cogent reasons for favoring one distributional form over another, no matter the behavior of limited numbers of observed data. The term *uncertain quantity* is used to encompass natural variability and lack of knowledge. In the follow sections we consider a pair of pmf's resulting from random process models, and several pdf's.

3.5.1 Success and failures: The Binomial pmf

The number of North American dam failures in any year is the sum of all the failures of individual dams across the continent. For purposes of illustration, assume that these failures are independent from one dam to another; that is, the failure of one dam is assumed for the purposes of data modeling in no way to affect potential failures of other dams; and furthermore, that no common cause initiating event simultaneously affects failures of multiple dams.[2]

Presuming independence among the failures of large dams, let the probability of any one dam failing in a particular year be p. The probability that x dams fail in one year is then, p^x. The probability that the remaining $n - x$ dams do not fail is $(1 - p)^{n-x}$.

Thus, the probability of x dams out of a total on n dams failing is the product of the probability of x failing and the complement $n - x$ not failing, or $p^x(1 - p)^{n-x}$. Since there are many different combinations of how x dams can be chosen from a set of n total dams, the probability distribution (pmf) of the number of failures of x out of n dams in any given year is

$$F(x|n) = \binom{x}{n} p^x(1 - p)^{n-x} \tag{3.41}$$

in which

$$\binom{x}{n} = \frac{n!}{x!(n - x)!} \tag{3.42}$$

is the number of combinations of n things taken x at a time. The mean of x is $E[x] = np$, and the variance is $Var[x] = np(1 - p)$. Equation (3.41) is known as the *Binomial distribution,* and is originally due to Pascal and Fermat's correspondence alluded to in Chapter 2. Figure 3.5 shows the pmf for one scenario of dam failures.

3.5.2 Poisson pmf and Exponential pdf

In surveying rock joints in a dam abutment, sampling lines are laid out across outcrops, borings are cored into the abutment, and the number of fractures intersected by the

[2] Of course, this assumption of independence is not strictly true even for large dams, and may not be even approximately true for small dams. The failure of a dam upstream may directly cause the failure of a dam down stream by releasing a flood wave that over tops the lower dam. When the Austin, Pennsylvania Dam failed in 1911, it washed out seven dams downstream (PEMA 2000). Similarly, extreme weather events may cause multiple dam failures across a region. When Hurricane Floyd struck North Carolina in 1998 it led to the failure of 36 dams in 44 North Carolina counties.

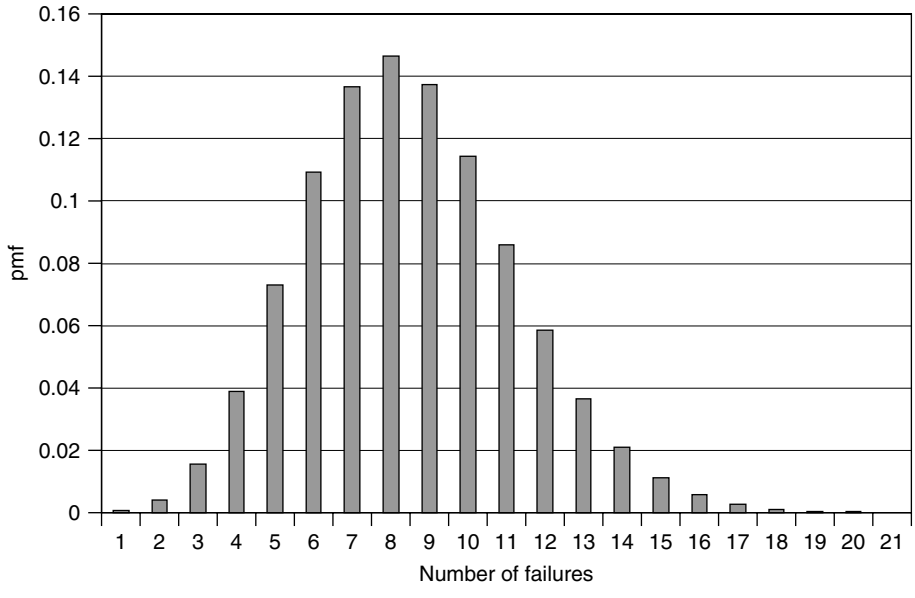


Figure 3.5 Binomial pmf of the annual number of dam failures in the US, for which $p = 10^{-4}$ and $n = 75{,}000$. (Baecher, G. B., Paté, E. M., and de Neufville, R. 1980, "Risk of dam failure in benefit/cost analysis," *Water Resources Research*, **16**(3), pp. 449–456. Copyright 1980 American Geophysical Union. Reproduced by permission of the American Geophysical Union.)

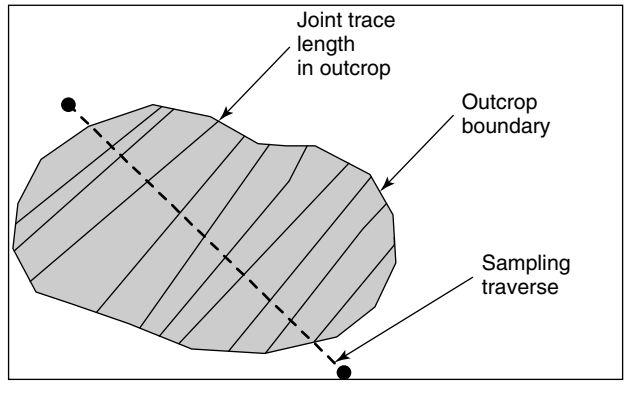


Figure 3.6 Sampling configuration for joint spacings in rock outcrop.

sampling line or boring is recorded. Jointing patterns in rock masses are known to display distinct orientation clusters, but it is often assumed that the locations of fractures within clusters are random (Einstein and Baecher 1983) (Figure 3.6).

Under the assumption that joints of a particular cluster are random and independent, the number intersecting a unit length of sampling line follows the Binomial pmf of Equation (3.41), in which x is the number occurring within a given interval, p is the probability of any one fracture occurring within the interval, and n is the total number of fractures. In the field, the total number of fractures n is large, and thus it is convenient

to replace n and p with their product $\lambda = np$, the density of fractures per unit length. Taking the limit of the Binomial pmf as $n \to \infty$ and $p \to 0$, while keeping their product λ constant, leads to

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{3.43}$$

is known as the *Poisson distribution*. The mean of x is $E[x] = \lambda$ and the variance $Var[x] = \lambda$. For sampling widths w other than unit length, the density λ is replaced by the density over w, which is λw, to obtain

$$f(x|\lambda, w) = \frac{(\lambda w)^x e^{-\lambda w}}{x!} \tag{3.44}$$

for which $E[x] = \lambda w$ and $Var(x) = \lambda w$ (Figure 3.7).

The distribution of spacings among adjacent rock fractures is obtained by noting that, if starting at a given intersection, the next closest fracture is at distance s, then there must be no occurrences of fractures within the interval $(0, s)$. Thus, setting the sampling width equal to s and making it the argument of the distribution, and setting $x = 0$, one obtains

$$f(s|\lambda) = \lambda e^{-\lambda s} \tag{3.45}$$

known as the *Exponential distribution*. For example, Priest and Hudson (1976) have used this model for rock fractures surveys. The mean of s is $E[s] = 1/\lambda$ and the variance $Var[s] = 1/\lambda^2$ (Figure 3.8).

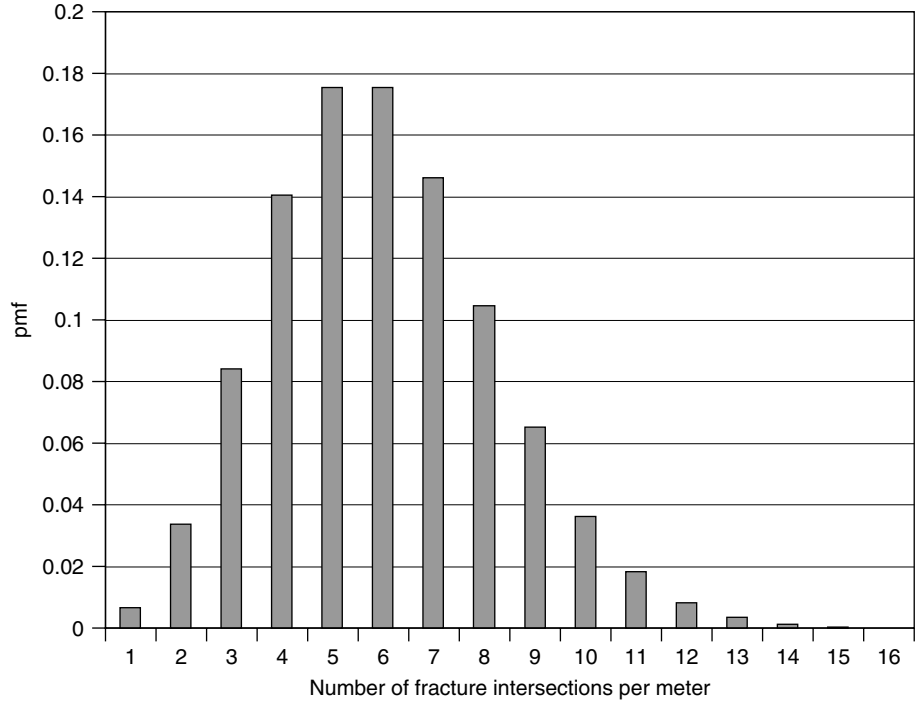


Figure 3.7 Poisson pmf for $\lambda = 5$.

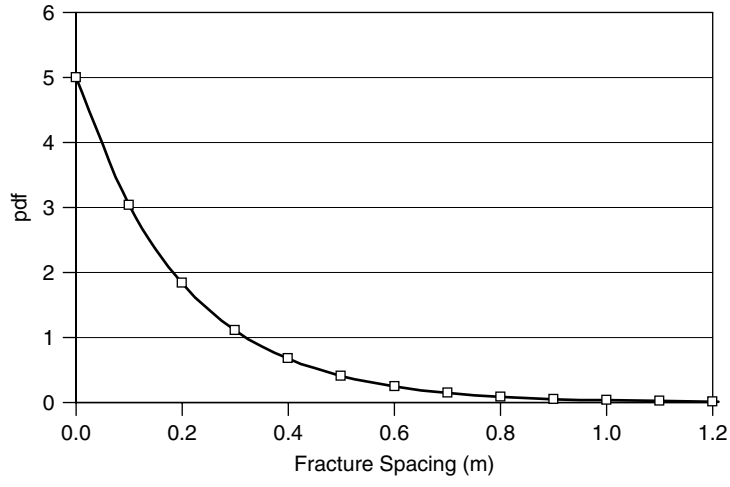


Figure 3.8 Exponential pdf for $\lambda = 5$.

3.5.3 Adding variables: the Normal pdf

In calculating the limiting equilibrium strength of a potential failure arc through an earthen embankment, the contributing strengths of a number of segments of the arc passing through different zones are added together to estimate total resistance. The contributing strengths of each of these segments are known only with some uncertainty, say up to a mean and a variance. What distributional form is appropriate for total resistance?

Figure 3.9 shows the normalized sums of uniformly distributed random variables in which one, two, five, and ten variables, respectively, are summed, and then divided by the number of variables in the sum. Thus, the plot labeled $N = 1$ shows 100 realizations of a random variable distributed as

$$f_x(x) \sim U(0, 1) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.46}$$

The plot labeled $N = 2$ shows 100 realizations of a random variable that is the sum of two uniformly distributed variables, divided by 2. That is, $x_2 = (x_i + x_j)/2$, in which the x_i and x_j are independently and identically (IID) distributed. Similarly, for the plots labeled $N = 5$ and $N = 10$; they are the sums of five and ten uniform variables, normalized by 5 and 10, respectively.

As N increases, the distributions of Figure 3.9 become more concentrated and more bell-shaped. In fact, as N becomes large, the distribution of the sum of independent random variables asymptotically approaches a Normal distribution, almost regardless of the distributions of the underlying variables. This is reflected in the Central Limit Theorem (Feller 1967), which states that the distribution of the sum of N random variables approaches Normality as N becomes large.[3]

[3] The Central Limit Theorem is not without constraints upon the underlying or component distributions. For example, Kaufman (1963) has shown in the context of oil reserve estimation that the sum of logNormal variables does not satisfy these conditions, and thus the distribution of reserve estimates involving logNormally distributed pool volumes is more complex than would they be if the Central Limit Theorem applied to them.

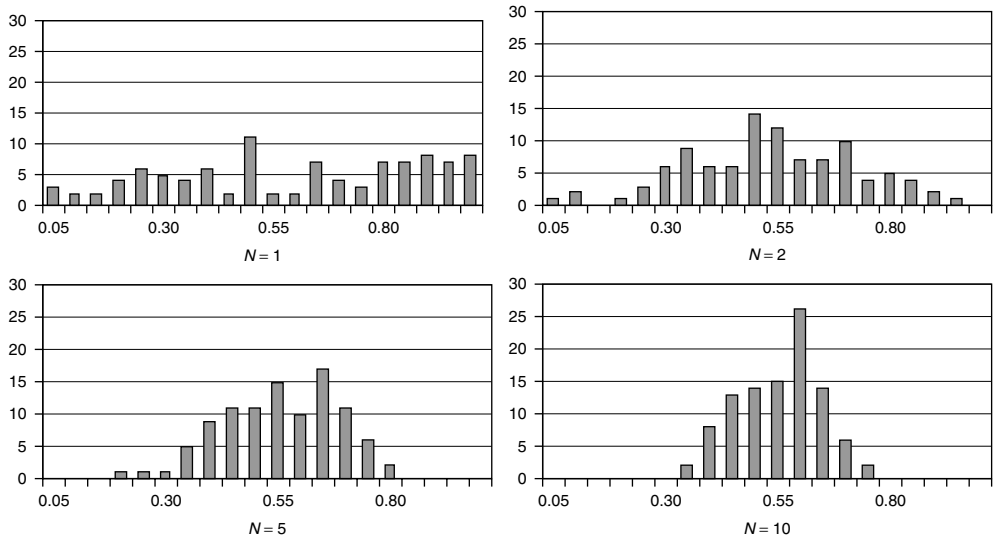


Figure 3.9 Distributions of the normalized sums of 1, 2, 5, and 10 uniformly distributed random variables.

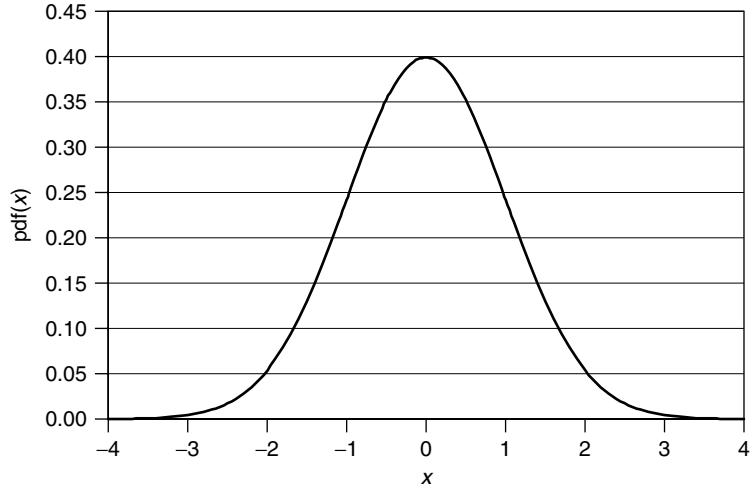


Figure 3.10 Normal pdf for $\mu = 0$ and $\sigma = 1.0$.

Thus, for uncertainties such as the average strength across a large failure arc which averages random variations, the Normal pdf is an appropriate model.

The Normal pdf is (Figure 3.10),

$$f_X(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \tag{3.47}$$

in which μ is the mean and σ^2 the variance. As can be seen from the exponent, the pdf is symmetric about a mode at μ, and falls off quickly as x deviates from the mean. The Normal pdf is usually tabulated or calculated numerically using its Standard Normal form, with mean zero and variance one, $N(0, 1)$.

3.5.4 Multiplying variables: the logNormal distribution

In the same way that calculating limiting equilibrium strengths involves the sum of random variables, other calculations involve the product of random variables. For example, in calculating certain material properties a series of modifying terms is sometimes applied to measurements. These modifying terms may have to do with known testing and measurement biases, model uncertainties, and the like, resulting in a term of the form

$$z = k_1 k_2 k_n x \tag{3.48}$$

Taking the logarithm of each side of the preceding equation leads to an equation involving the sum of logarithms, or more specifically,

$$\log z = \log k_1 + \log k_2 + \cdots + \log k_n + \log x \tag{3.49}$$

Thus, since $\log z$ is the sum of the logarithms of the k and x, with only loose assumptions on the distributions of the logarithms of the k and x, the distribution of $\log z$ should approach Normality (and hence z approach logNormality) as n becomes large. Figure 3.11 shows a normal distribution on the logarithm of X; Figure 3.12 shows the corresponding logNormal distribution on X.

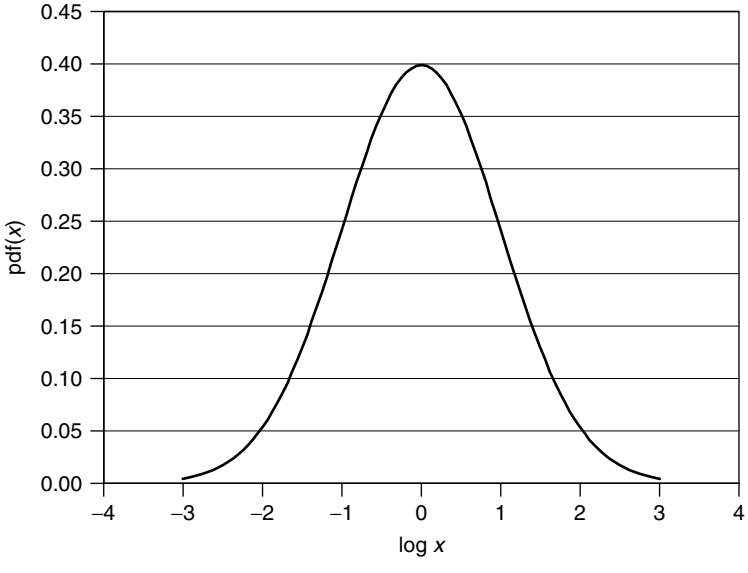


Figure 3.11 Normal distribution on Log Z.

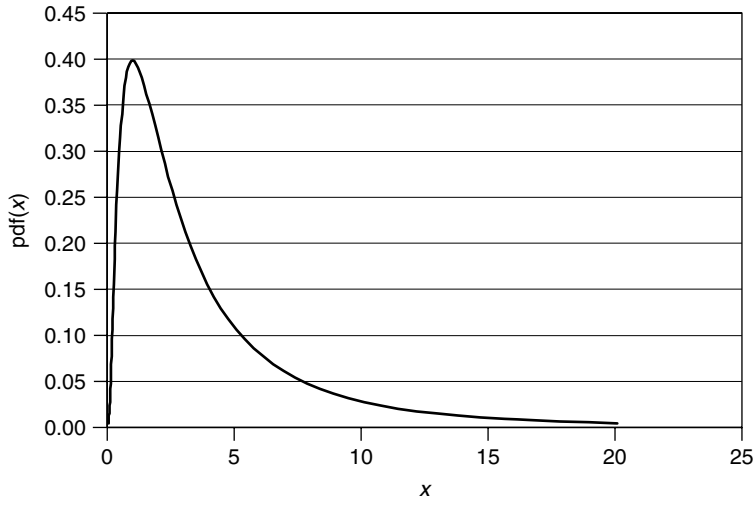


Figure 3.12 Distribution on Z corresponding to Figure 3.18.

If the mean and standard deviation of the data themselves are μ and σ, respectively, and the mean and standard deviation of the logarithms of the data are λ and ζ, respectively, the following relations apply and are useful in working with logNormal distributions:

$$\begin{aligned} \zeta^2 &= \ln\left(1 + \frac{\sigma^2}{\mu^2}\right) \\ \lambda &= \ln(\mu) - \frac{1}{2}\zeta^2 \\ \mu &= \exp\left(\lambda + \frac{1}{2}\zeta^2\right) \end{aligned} \tag{3.50}$$

3.5.5 Tails of probability distributions: Extreme Value pdf's

The peak annual flow in a stream is the largest discharge to occur in any of the weeks in the year. Presume for the purpose of discussion that peak weekly flows are independent of one another. (This is clearly not the case for daily flows, since storms typically last more than one day.) If the peak weekly discharge, say, can be modeled by a Normal pdf, what is the appropriate distributional form the peak annual discharge? In essence, the question reduces to the distribution of the maximum value within samples of size 52 drawn from a Normal population. In concept, this question is the same as that of the weakest link of a chain, which can be thought of as the minimum value within samples of a size reflecting the number of links in the chain. Such problems are said to involve the distribution of extreme values.

Let the population from which peak weekly discharges are drawn be $f_{Q_w}(q)$, in which Q_w is peak weekly discharge. For the peak annual discharge to be less than or equal to some value q_o, all the 52 weekly peak discharges must be less than q_o. Thus

$$F_{Q_{\max}}(q|n = 52) = F_Q^{52}(q|n = 1) \tag{3.51}$$

In the more general case, in which the largest of n uncertain quantities is sought, the cdf of the maximum becomes

$$F_{Q_{\max}}(q|n) = F_Q^n(q) \tag{3.52}$$

and, after taking the derivative of each side, the pdf of the maximum is

$$f_{Q_{\max}}(q|n) = \frac{d}{dq} F_{Q_{\max}}(q|n) = n F_{Q_{\max}}^{n-1}(q) f_Q(q) \tag{3.53}$$

Figure 3.13 shows the pdf's of the largest value within samples of 1, 10, 100, and 1000 drawn from a Normal parent distribution. As would be expected, as the sample size becomes larger, the mean value of the largest observation within the sample also becomes larger, but the variance becomes smaller. The classical reference on the statistics of extreme values is Gumbel (1954; 1958).

The shape of the extreme value is sensitive to the shape of the tails of the parent distribution. The pdf of the largest value within a sample is sensitive to the shape of the upper tail of the parent pdf, and likewise the pdf of the smallest value is sensitive to the lower tail. For a Normal distribution these are symmetric, but this is not the case for all parent distributions. Gumbel classified extreme value distributions into three categories, depending on their asymptotic behavior as $n \to \infty$. The Type I limiting distribution arises for the largest variable from a parent distribution with an exponentially decaying upper tail, that is, an upper tail that falls off as

$$F_X(x) = 1 - \exp(-g(x)) \tag{3.54}$$

in which $g(x)$ is an increasing function of x. For example, the Normal, Gamma, and Exponential distributions are all of this type. Gumbel showed that for large n this distribution approaches

$$F_X(x) = \exp(-e^{-\alpha(y-u)}) \tag{3.55}$$

$$f_X(x) = \alpha \exp(-\alpha(y-u) - e^{-\alpha(y-u)}) \tag{3.56}$$

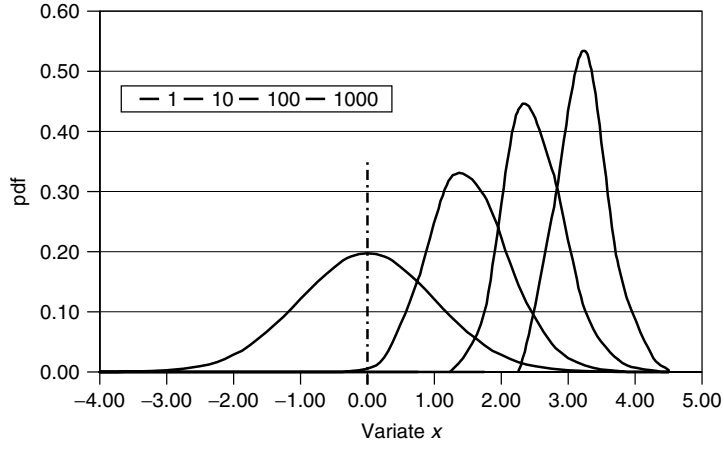


Figure 3.13 Distribution on Z corresponding to Figure 3.11.

The parameters of the distribution are typically estimated from observed data, the presumption being that the extreme variable arises from a large but perhaps unknown number n of, say, large stream discharges during the course of a year.

The Type II limiting distribution arises for the largest variable from a parent distribution with an upper tail that falls off as

$$F_X(x) = 1 - \beta \left(\frac{1}{x}\right)^k \tag{3.57}$$

Gumbel showed that for large n this Type II distribution approaches

$$F_X(x) = e^{(-u/y)^k} \tag{3.58}$$

$$f_X(x) = \frac{k}{u}\left(\frac{u}{y}\right)^{k-1} \tag{3.59}$$

The Type III limiting distribution arises for either the largest or smallest variable from a parent distribution with a limited tail, that is, a tail that falls off as

$$F_X(x) = 1 - c(w - x)^k \tag{3.60}$$

in which $x \leq w$ and $k > 0$. In practice, this distribution is most often used to model smallest values, having a lower tail of the form $F_X(x) = c(w - \varepsilon)^k$ for $x \geq \varepsilon$, where ε is the lower limit for x (Benjamin and Cornell 1970). For example, the Gamma distribution is of this type. Gumbel showed that for large n the Type III distribution approaches

$$F_X(x) = 1 - \exp\left[-\left(\frac{x - \varepsilon}{u - \varepsilon}\right)^k\right] x \geq \varepsilon \tag{3.61}$$

$$f_X(x) = \frac{k}{u - \varepsilon}\exp\left[-\left(\frac{x - \varepsilon}{u - \varepsilon}\right)^{k-1}\right]\exp\left[-\left(\frac{x - \varepsilon}{u - \varepsilon}\right)^k\right] x \geq \varepsilon \tag{3.62}$$

3.6 Fitting Mathematical Pdf Models to Data

Sometimes probability distributions are chosen simply because they appear to fit observed data. The rationale has nothing to do with physical law. This is often the case in fitting probability distributions to, say, soil data. One observes that when the frequencies of undrained strength measurements are plotted against strength value they might exhibit a bell-shaped form characteristic of the Normal pdf, and thus a Normal distribution model is used to approximate them.

Various statistical methods can be used to test whether the match between the empirical data frequencies and the theoretical pdf model is close enough that it is reasonable to assume the differences between observed and modeled could be simply due to sampling variability (Kendall and Stuart 1977). The most common of these tests are the Chi-square goodness-of-fit test and the Kolmogorov-Smirnov maximum deviation test.

The next chapter discusses some of these tests. In many cases, however, the easiest way to assess the quality of the fit of a theoretical distribution to an observed distribution is to plot the observed distribution against the theoretical distribution.

Probability grid is a specialized graph paper with horizontal and vertical scales designed such that the cumulative frequencies of a particular form of pdf (e.g. a Normal distribution) plot as straight lines. Thus, empirical data can be plotted on probability grid and a visual determination made whether they are well approximated. The effect of Normal probability grid can be achieved by using cumulative frequencies as arguments to an inverse standard Normal cumulative distribution function. The results are Figures 3.14 and 3.15 for the data of Figures 3.1 and 3.2, respectively. The liquidity index data fall nearly along a line in

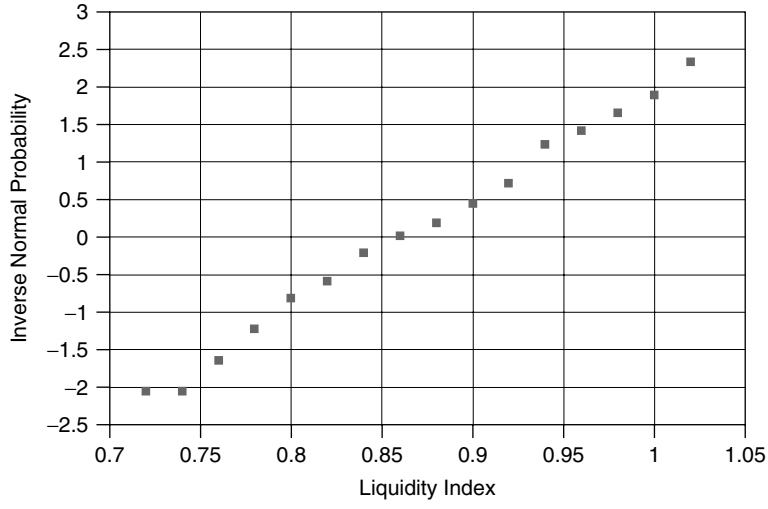


Figure 3.14 Liquidity index data from a marine clay plotted against probability axes.

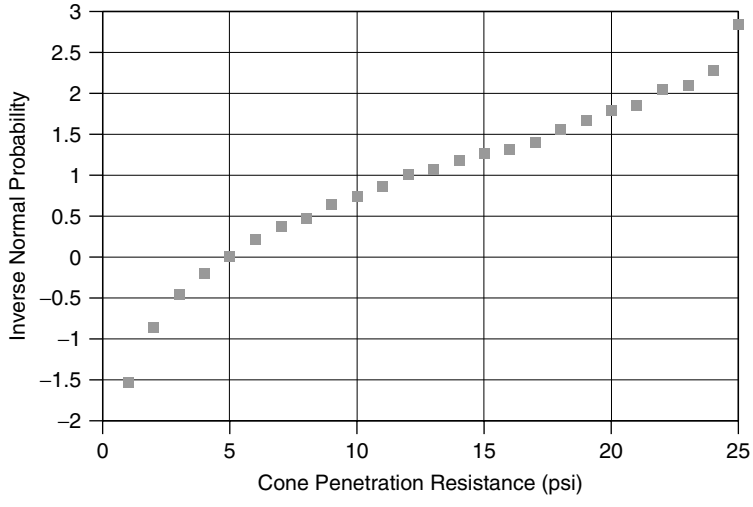


Figure 3.15 Cone penetration resistance data plotted against probability axes.

this plot, and one would infer that that they are well modeled as a Normal distribution. The cone penetration data do not fall along a line in this plot, and one would infer that they are poorly modeled as a Normal distribution. As a first improvement one could plot the logarithms of the data, as in Figure 3.16. This comes closer to a straight line, but at the high end of the data there is some deviation from a straight line fit.

Another graphical tool for comparing data with analytical distributions is the quantile-quantile (Q-Q) graph (Tukey 1977; Hoaglin, *et al.* 2000). A Q-Q graph shows observed quantiles of the data plotted against theoretical quantiles of the best fitting distribution. If the frequency distribution of the sample data is reasonably approximated, the empirical and analytical quantiles should be approximately the same, and the data pairs should lie on a 45-degree line (Figures 3.17 and 3.18).

In selecting a pdf empirically, the argument is occasionally made that domain constraints on the variable, X, should apply equally to the pdf. For example, the friction angle of a soil is an angular variable defined over $[0^\circ, 90^\circ]$. It cannot take on values outside this range; thus, the argument goes, the pdf fit to friction angle data should likewise be defined over a closed interval. For example, one such pdf is the Beta distribution,

$$f_X(x|\alpha, \beta) = \begin{cases} Kx^{(\alpha-1)}(1-x)^{(\beta-1)} & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.63}$$

in which $K = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ is a normalizing constant to force the integral over x to be 1. The gamma function $\Gamma(x)$ is a tabulated function that is available in modern spreadsheets, and for integer values of x it is related to the factorial by $\Gamma(n+1) = n!$

In practice, this is seldom an important restriction, since probability density falls rapidly outside a few standard deviations above or below the mean of the distribution, and thus the modeled probability of the variable falling outside the physically-defined range is

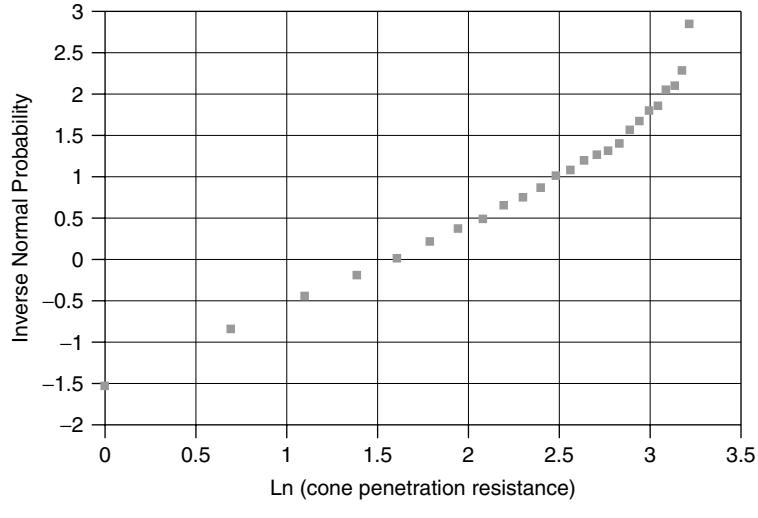


Figure 3.16 Natural logarithms of cone penetration resistance data plotted against probability axes.

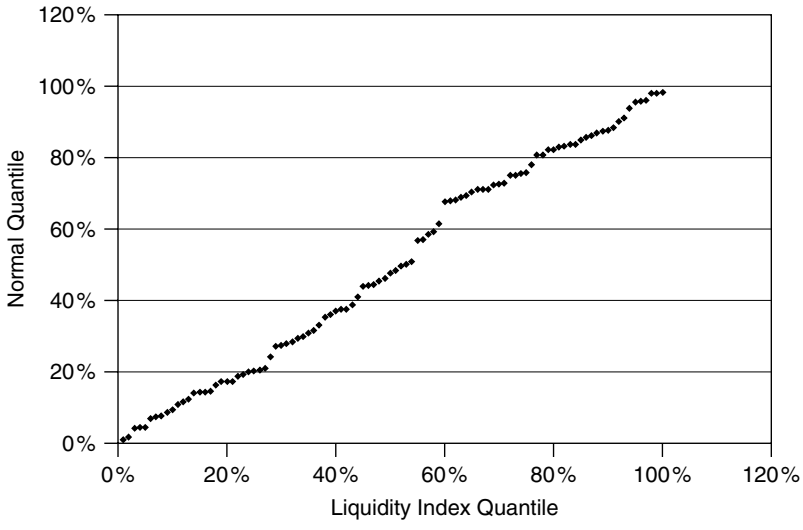


Figure 3.17 Q-Q plot for liquidity index data from a marine clay.

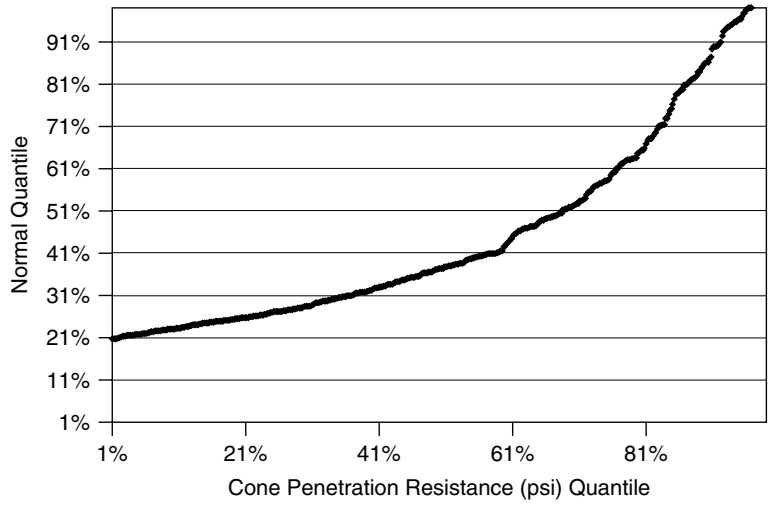


Figure 3.18 Q-Q plot for cone penetration data from a copper porphyry tailings dam.

negligible. For example, the American Army uses a Normal pdf to size uniforms for its troops, even though few American soldiers are shorter than four feet or taller than eight.

This same argument holds for derived variables such as the factor of safety, which cannot be less than zero. For commonly encountered means and standard deviations of factor of safety there is little justification in forcibly fitting a bounded distribution such as the logNormal just because that pdf has a theoretical lower bound at zero, when another distribution such as the Normal provides a more satisfying fit. For example, fitting a Normal pdf to a factor of safety prediction having mean 1.5 and standard deviation 0.2 leads to a probability of approximately 10^{-14} that $Fs < 0$.

In addition to fitting distributions by inspection, as mentioned earlier, one can also fit general 'families' of distributions. Two such families are common, the Johnson curves and the Pearson curves. We have already encountered the first moment of the distribution in the form of the mean and the second in the variance. The third and fourth moments are simply the sums or integrals of x^3 and x^4 over the distribution of x. The shapes of most continuous distributions can be sufficiently summarized in the first four moments. Put another way, if one fits to a histogram of observed data a distribution that has the same mean (first moment), variance (second moment), skewness (third moment) and kurtosis (fourth moment) as the observed data, then one can usually approximate the overall shape of the distribution very well. Once a distribution has been fitted, one can then calculate the expected percentile values under the (standardized) fitted curve, and estimate the proportion of items produced by the process that fall within the specification limits.

3.7 Covariance Among Variables

When dealing with more than one variable, uncertainties in one may be associated with uncertainties in the another. That is, the uncertainties may not be independent. Consider estimating the 'cohesion' and 'friction' parameters of a Mohr–Coulomb strength envelope. If the slope of the envelope, ϕ, is mistakenly estimated too high, then for the line to fit the data, the intercept, c, will have to be too low. The reverse is true if the slope is too high. Thus, uncertainties about the slope and intercept are associated with one another. Dependencies among events or among the uncertainties in estimates can be critical to obtaining proper numerical results in reliability analyses. To complicate matters, these dependencies can also be subtle to identify and difficult to estimate.

3.7.1 Causes of covariation and dependence

Probabilities associated with separate events or uncertainties can be made dependent through a number of circumstances. *Causal dependence* means that one event physically causes another; for example, liquefaction-induced settlement may directly lead to overtopping of an embankment, thus the liquefaction event and the overtopping event are not be independent of one another. If liquefaction settlement occurs, the probability of overtopping is enhanced. Table 3.4 describes such a case.

Table 3.4 Conditional probabilities of liquefaction and overtopping given the existence or non-existence of low-density, soft lenses in an embankment

	Pr.	Liquefaction	No liquefaction	Overtopping	No overtopping
Low-density lenses exist	0.1	0.9	0.1	0.67	0.33
No low-density lenses exist	0.9	0.0	1.0	0.01	0.99
Marginal (sum of Pr x column entries)		0.09	0.91	0.076	0.924

Probabilistic correlation means that two uncertainties may share a common dependence on a third uncertainty, as in the case of the low-density soil lenses in the example above. The realization that the low-density soil lenses exist (or do not exist) simultaneously changes the probabilities of both liquefaction cracking and of overtopping.

Spatial or temporal autocorrelation means that two uncertainties depend on the spatial or temporal realization of some third uncertainty which itself exhibits stochastic dependence in space or time. The performances of two sections of a long levee may depend on soil engineering properties in the naturally occurring valley bottom, which, when modeled as a stochastic (aleatory) process, exhibit a long wave length of correlation in space; thus, adjacent sections will exhibit similar settlements or factors of safety against strength instability.

Statistical correlation means that two uncertainties are simultaneously estimated from a given set of data and therefore are influenced by a common sampling variability error. In soil mechanics, a common – and mostly overlooked – statistical correlation is that between soil cohesion and friction angle just described, which being regression parameters, are negatively correlated.

3.7.2 Covariance and the correlation coefficient

The most common measure of dependence among uncertain quantities is the *correlation coefficient*. This measures the degree to which one uncertain quantity varies linearly with another uncertain quantity. Note, the notion of linearity here is important. Two uncertain quantities may be deterministically related to one another but have negligible correlation if the relationship is strongly non-linear.

The correlation coefficient for two uncertain quantities x and y is defined as the ratio of the covariance of x and y to the square root of the product of the variances of x and y (i.e. the product of the standard deviations of x and y):

$$\begin{aligned}\rho_{xy} &= \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} \\ &= \frac{E[(x-\mu_x)(y-\mu_y)]}{\sqrt{E[(x-\mu_x)^2]E[(y-\mu_y)^2]}}\end{aligned} \tag{3.64}$$

The correlation coefficient varies within $[-1,+1]$, with the higher bound implying a strict linear relation of positive slope and the lower bound a strict linear relation of negative slope. The higher the magnitude, the more closely the data fall on a straight line (Figure 3.19). Zero correlation coefficient implies no (linear) association between x and y. The correlation coefficient is equivalent to a normalized product moment of inertia in solid mechanics. It expresses the degree to which two parameters vary together. The correlation coefficient is non-dimensional because deviations of x and y are measured in the same units as their respective means.

3.7.3 Quick estimates of correlation

Calculation of correlation coefficients can be time consuming. A quick graphically approximation is obtained from the shape of the scatter plot. The method works well whenever

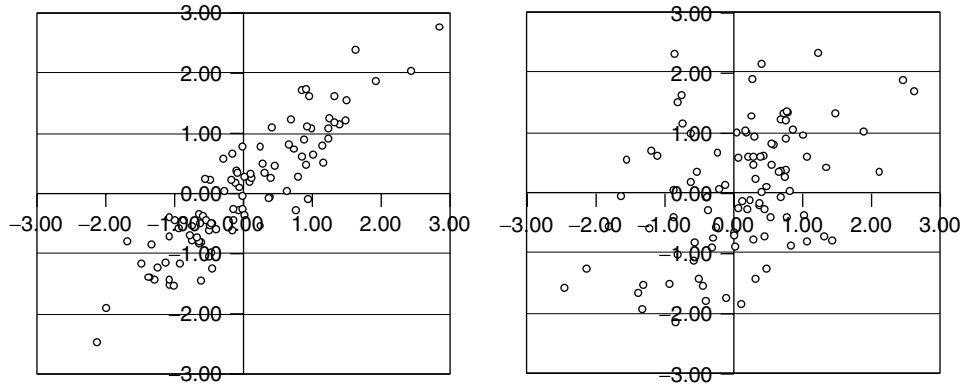


Figure 3.19 Pseudo-random Normal correlated variables: left plot $\rho = 0.9$, right plot $\rho = 0.4$.

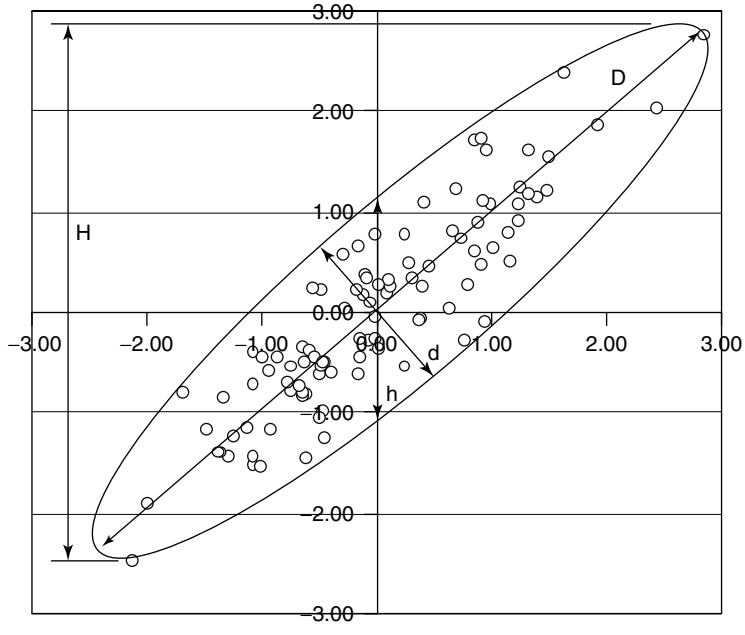


Figure 3.20 Example of 'balloon' shortcut method for estimating correlation coefficient (Chatillon 1984; Schilling 1984).

the outline of the scatter plot is approximately elliptical, and works even with small numbers of observations. Chatillon (1984) called this, the *balloon method*:

1. Plot a scatter diagram of y vs. x.
2. Draw an ellipse (balloon) surrounding all or most of the points on the data.
3. Measure the vertical height of the ellipse at its center, h, and the vertical height of the ellipse at its extremes, H.
4. Approximate the correlation coefficient as $\hat{\rho} = [1 - (h/H)^2]^{1/2}$.

An example of the method is shown in Figure 3.20 Chatillon's method gives a correlation coefficient of about 0.92, whereas the calculated value is 0.90. Empirically, the method works well for $\rho > 0.5$.

Schilling (1984) suggests a similar method:

1. Plot a scatter diagram of $(y - m_y)/s_x$ vs.$(m - m_x)/s_y$.
2. Draw an ellipse surrounding most of the data.
3. Measure the length of the principal axis of the ellipse having positive slope, D, and the length of the principal axis of the ellipse having negative slope, d.
4. Approximate the correlation coefficient as $\hat{\rho} = (D^2 - d^2)/(D^2 + d^2)$.

This method seems to work as well as Chatillon's. For the data of Figure 3.20, Schilling's method gives a correlation coefficient of about 0.91.

4 Inference

In this chapter we consider the use of *statistical methods* for drawing inferences from observations. Statistical methods are based on probability theory. As we saw in the last chapter, probability theory is a branch of mathematics; that is, it has to do with formal logic. Starting from axioms, all the relations of probability theory follow necessarily (Kolmogorov 1950). There is no room for different points of view. Statistical methods, in contrast, are more *ad hoc*. They do not derive from a set of axioms and, as a result, are open to interpretation. Thus, it is not surprising that schools of thought with strongly held points of view have grown up around the proper way of approaching statistical inference.

The chapter begins with a look at the two predominant schools of statistical thought: the *frequentists* and *Bayesians.* These grow out of the duality in the meaning of *probability* discussed in Chapter 2. The frequentist school builds on the notion of probability as a frequency with which things occur in a long series of trials; the Bayesian school builds on the notion of probability as a degree of belief. The resulting differences in the methods that have developed under the two schools of thought are profound. These differences directly and substantially affect how data are analyzed and the sorts of inferences that can be drawn from observations.

In this book we favor the Bayesian approach over the frequentist approach. We find it more consistent with geotechnical practice. This is not to say that frequentist approaches are never useful, but we do not devote a great deal of attention to them. The first section of the present chapter briefly summarizes concepts and vocabulary from frequentist statistics because these are necessary to understand at least some of the applications in the geotechnical literature. The rest of the chapter develops the Bayesian approach to statistical inference, regression theory and hypothesis testing. The chapter ends with a consideration of lessons learned from sampling theory for planning site characterization.

4.1 Frequentist Theory

The division between frequentist or 'traditional' school[1] and the Bayesian school follows from the meaning given by each to the term *probability*. The frequentist school takes

[1] What is here called the 'traditional' school has been called 'orthodox' by Jaynes (1996) and 'classical' by Raiffa and Schlaifer (1968), but neither of these names is wholly satisfactory. Each conflicts with other uses

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

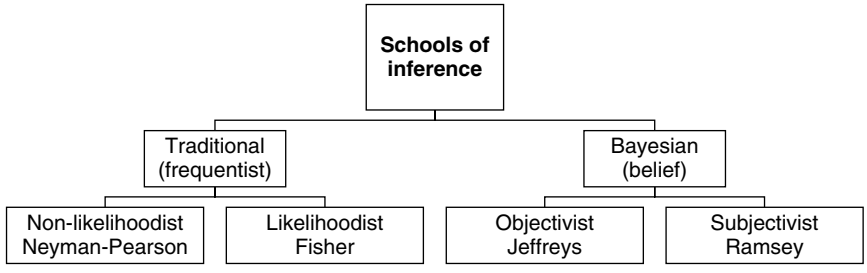


Figure 4.1 Taxonomy of statistical thought, adapted from Raiffa (1968) and Baecher (1972).

probability to mean frequency, while the Bayesian school takes probability to mean belief. A schematic taxonomy is show in Figure 4.1.

The frequentist school is based on a theory of refutation. Probabilities are not placed directly on states of nature, but confidences are calculated in a form of reverse reasoning by excluding states of nature under which the chance of observing the data that actually were observed would be small. The point of view, according to Fisher (1925), is that sufficient information should be collected until one cannot explain them, beyond a reasonable doubt, except by concluding that the conjecture under consideration is correct. The focus of traditional theory is on the probability of rejecting an hypothesis when in fact it is true (what epidemiologists call a false negative).

Consider a set of observations, $\mathbf{z} = \{z_1, \ldots, z_n\}$ arising from an experiment on an uncertain process. Let the process be described by a probability distribution $f_Z(z|\theta)$, with parameter θ (which could be a vector). One would like to say something about the uncertainty in θ that exists after having made the observations. The frequentist school rejects this question as undefined because, to it, θ is not a random variable. The data are random variables; the parameter of the process generating the data is a *state of nature*. This parameter has a particular value even if that value is unknown; that is, θ can be thought of as an unknown constant.

The frequentist-school approach to imputing a range of values to θ, given a set of observations, is based on the concept of *estimators*. Estimators are statistics of the observed data that can be used to make estimates of the value of an unknown but constant state of nature, θ. Any mathematical function of the sample observations

$$T = g(\mathbf{z}) \tag{4.1}$$

is said to be a *statistic* of the data, or simply a sample statistic. For example, commonly used sample statistics are the average, the standard deviation, the range, and so on. Appropriately chosen sample statistics are used to estimate a population parameter, as

$$\hat{\theta} = T = g(\mathbf{z}) \tag{4.2}$$

where 'appropriately chosen' means any of a set of desirable properties that estimators might have.

---

of the terms in the historical literature on probability. While 'traditional' itself is also not wholly satisfactory, it seems less objectionable then the other two. To avoid the issue, we have adopted the term, *frequentist*.

Note that, because an estimator is a mathematical function of the observed data and the observed data are taken to be random variables, the estimator is also a random variable. The *sampling distribution* of an estimator is its probability distribution in repeated sampling from the same uncertain process. The sampling distribution describes the frequencies of the estimator over all possible ways a set of observations might be made from the sampled process with given parameter, θ.

While a good estimator is one whose sampling distribution is concentrated in a narrow region near the true value of θ, certain mathematical properties of the sampling distribution are considered desirable. Some of the more common of these good-estimator properties are given in Table 4.1. These include *unbiasedness*, meaning that the expected value of T equals the true value of θ; *consistency*, meaning that T converges to θ as n becomes large; *efficiency*, meaning that the variance of the sampling distribution of T is as small as possible; *sufficiency*, meaning that the estimator T summarizes all the information contained in the sample observations relevant to making estimates of θ (i.e., knowing T, the actual data $\boldsymbol{z}$ carry no additional information with respect to θ); and *robustness*, meaning that the statistical properties of T in relation to θ are insensitive to deviations from the assumed form of $f_Z(z|\theta)$.

As an example, consider sampling from a Normal process. The sampling distribution of $\mathbf{z} = \{z_1, \ldots, z_n\}$ is

$$f_{\mathbf{Z}}(\mathbf{z}|\mu, \sigma) \propto \prod_{i=1,n} \exp\left(-\sum_{i=1}^{n} \frac{z_i - \mu}{2\sigma^2}\right) \tag{4.3}$$

in which μ is the mean and σ the standard deviation, and the individual observations are assumed independent and identically distributed (IID). Since the observations are assumed independent, their joint probability is simply the product of their marginal probabilities. Taking the sample average $\overline{z} = 1/n \sum z_i$ as an estimator, its sampling distribution is (Benjamin and Cornell 1970)

$$f_{\overline{z}}(\overline{z}|\mu, \sigma) \propto \exp\left(-\frac{n(\overline{z} - \mu)^2}{2\sigma^2}\right) \tag{4.4}$$

The expected value of $\overline{z}$ equals the mean of the process being sampled, $E[\overline{z}] = \mu$, and thus, $\overline{z}$ is an *unbiased estimator* of the population parameter μ.

Table 4.1 Desirable properties for statistical estimators

Estimator property	Definition
Unbiasedness	The expected value of T over all ways the sample might have been realized from the parent population equals the parameter to be estimated ($E[T] = \theta$).
Consistency	T converges to θ, as n becomes large ($T \underset{n\to\infty}{\rightarrow} \theta$).
Efficiency	The variance of the sampling distribution of T is minimum.
Sufficiency	The estimator T makes maximal use of the information contained in the sample observations.
Robustness	The statistical properties of T in relation to θ are insensitive to deviations from the assumed underlying pdf of z.

Estimators yield specific values for parameters; likely ranges of values are based on confidence intervals. The concept of a *confidence interval* arises from refutation. The confidence interval of an estimate of θ is that interval of the θ-axis for which the probability of observing the sample data, conditioned on θ, is larger than some arbitrary small value, say, $p = 0.05$. For any value of θ outside this interval, the probability of observing the z that we did observe would be small; thus, those values of θ are refuted. Those values of θ within the interval are not (yet) refuted. The gist of this estimate is not that there is a probability 0.05 that the true value of θ is within the confidence interval, but that, if θ is outside the interval, the probability of observing the actual sample outcomes could be no larger than 0.05. This distinction is often confused in the geotechnical literature.

A major subdivision within the frequentist school is that between statisticians who believe that the only relevant information from an experiment is contained in the *likelihood function*, which is the conditional probability of $\mathbf{z}$ given θ, and those who believe that the entire experimental frame of reference carries information. Consider, for example, someone trying to decide whether a clay core has adequately low permeability. While selectively taking only data that support the hypothesis causes a bias, one might run tests only until a favorable case is established, and then stop. This is legitimate to the 'likelihoodist,' to whom all the information resides in $L(\theta|\mathbf{z})$, but not to the 'non-likelihoodist.' The latter would say the experiment was biased in favor of the hypothesis, because the number of tests depended on the early outcomes.

The *likelihood principle*, due to Fisher (1921), says that all relevant information about θ arising out of an experiment with outcome $\mathbf{z} = \{z_1, z_2, \ldots, z_n\}$ is contained in the likelihood function

$$L(\theta|\mathbf{z}) = L(\theta|z_1, z_2, \ldots, z_n) = \Pr[z_1, z_2, \ldots, z_n|\theta] \tag{4.5}$$

For IID observations from $f_Z(z|\theta)$, the likelihood becomes

$$\begin{aligned} L(\theta|\mathbf{z}) &= f_{\mathbf{z}}[z_1, z_2, \ldots, z_n|\theta]dz_1dz_2, \ldots, dz_n \\ &= \prod_{i=1}^{n} f_z[z_i|\theta]dz_i \end{aligned} \tag{4.6}$$

Fisher recommends using the value of θ that maximizes the likelihood as an estimator, calling this the *maximum likelihood estimator* (MLE). The MLE has many desirable properties. It is asymptotically ($n \to \infty$) unbiased, consistent, sufficient, and asymptotically Normal.

The Neyman–Pearson school of inference, in contrast to the Fisher school, is not based on the likelihood principle and holds that statistical inference is about hypothesis testing, which is an action (or decision) problem of accepting or rejecting an hypothesis. Probability enters the problem of inference as a way of characterizing the experimental process itself: to express how reliably the testing process discriminates between alternative hypotheses, balancing the chance of rejecting a correct hypothesis with the chance of accepting a wrong hypothesis. It is from the Neyman-Pearson school that we inherit the concepts of *Type I error* (rejecting the null hypothesis when actually it is true) and *Type II error* (accepting the null hypothesis when actually it is false). The tool for balancing these error rates is called the *power function*, which, for a given frequency of Type I error, describes the frequency of Type II errors resulting from a particular testing strategy.

Wald (1971) expanded the solution to the decision problem by adding the concepts of cost and worth of consequences. As in the Neyman–Pearson and Fisher schools, Wald did not place probabilities on the state of nature, but in arguing the worth of consequences, he reduced the choices among which a decision must be made.

4.2 Bayesian Theory

The core difference between Bayesian and frequentist approaches is that the former treats uncertainty as degrees of belief and therefore admits probability statements on states of nature. To this way of thinking, states of nature are variables, not unknown constants.

4.2.1 Inferring probabilities

Consider a simple site characterization example in which standard penetration test (SPT) borings are used to explore for the possible liquefiable layers in the subsurface. The SPT tests are taken every five (5) feet (1.5 m) vertically, and presume that the target layers are two (2) feet (0.6 m) thick. Given the data that result, what is the probability that a liquefiable layer exists?

If a layer exists, then the conditional probability of finding it with a single boring is $p = 2/5 = 0.4$, but engineers sometimes mistake a low blow count as an outlier. Thus, assume the conditional probability is 0.3. Also, it is possible to get a false positive when no liquefiable material exists, so the assumptions are made that

$$\begin{aligned} P[\textit{find}|\textit{exists}] &= 0.3 \\ P[\textit{no find}|\textit{exists}] &= 0.7 \\ P[\textit{find}|\textit{does not exist}] &= 0.1 \\ P[\textit{no find}|\textit{does not exist}] &= 0.9 \end{aligned} \tag{4.7}$$

From Bayes' Theorem (Equation (3.16)),

$$P[E|F] = \frac{P[E]P[F|E]}{P[E]P[F|E] + P[\overline{E}]P[F|\overline{E}]} \tag{4.8}$$

in which $E =$ layer exists, $F =$ positive result, $\overline{E} =$ layer does not exist. Presuming no prior information on whether a layer exists, we take the prior probability to be 0.5. If one boring is drilled, and returns a positive result, then,

$$P[E|F] = \frac{0.5 \times 0.3}{0.5 \times 0.3 + 0.5 \times 0.1} = 0.75 \tag{4.9}$$

If the probabilities of hitting a layer are independent from one boring to another, the results of Figure 4.2 are obtained for three borings and zero to three hits. Parametric results for various prior probabilities are shown in Figure 4.3.

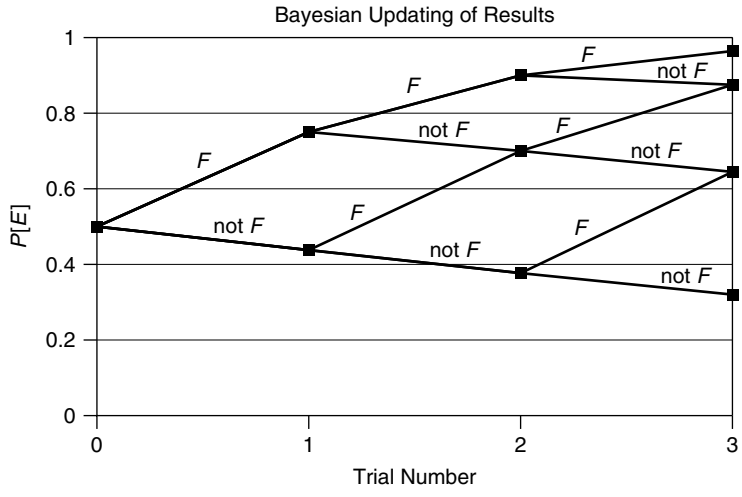


Figure 4.2 Posterior probabilities of a liquefiable lens given the results of three borings. The upper branch in each case indicates a 'find' and the lower branch 'no find.' Prior probability that the lens exists is assumed to be 0.5.

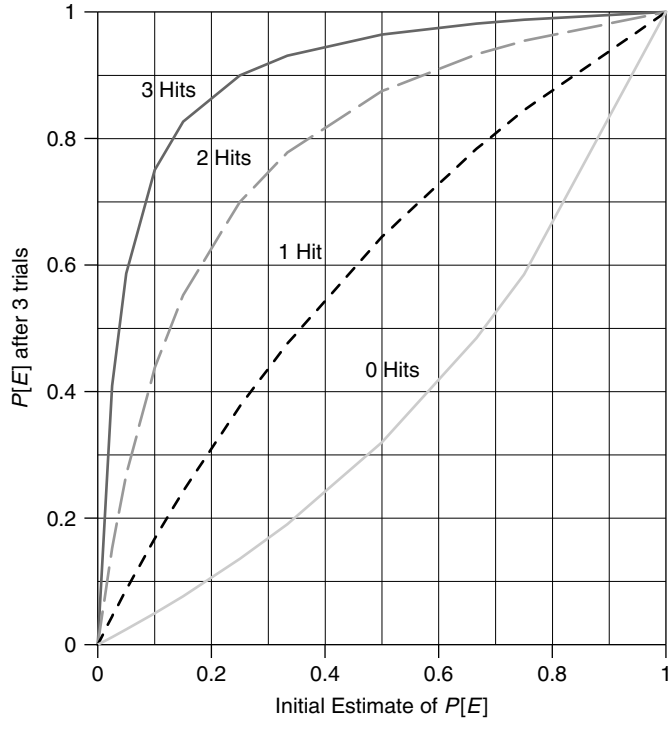


Figure 4.3 Parametric results for the posterior probability of the existence of a liquefiable lens for various exploration outcomes and prior probabilities.

An intuitive and easily verified result is that it does not matter whether the revision of the posterior probability is done for the three borings all at once, or incrementally as each boring is made. In the former case,

$$P[E|B_1, B_2, B_3] = P[E]P[B_1, B_2, B_3|E] \tag{4.10}$$

In the latter case,

$$\begin{aligned} P[E|B_1] &= P[E]P[B_1|E] \\ P[E|B_1, B_2] &= P[E|B_1]P[B_2|E, B_1] \\ P[E|B_1, B_2, B_3] &= P[E|B_1, B_2]P[B_3|E, B_1, B_2] \\ &= P[E]P[B_1, B_2, B_3|E] \end{aligned} \tag{4.11}$$

Sequential application of Bayes' Theorem yields the same result as incremental application, in which the intermediate posterior probabilities from the previous increment are used as the prior probabilities for the next round of sampling (Figure 4.4).

Bayesian methods use Bayes' Theorem and thus the Likelihood function as the basis for inference, so in a way they are philosophically similar to Fisher's later work on fiducial probabilities (Fisher and Bennett 1971; Hacking 1965). Common practice is to call the marginal probability before taking account of the observations, the *prior* or *a priori* probability; and to call the conditional probability accounting for the observations, the *posterior* or *a posteriori* probability. Sometimes, the prior and posterior are denoted with noughts and primes, respectively, as p^0 and p'.

Starting in the 1930s, a great deal of effort was invested in exploring the implications of Bayesian inference. Among those early authors were Fisher, Jeffreys, and de Finetti. This work flourished after World War II, in part growing upon the wartime investment in operations research and decision sciences. Among these later authors were Barnard, Savage, Lindley, Raiffa and Pratt, Box and Tiao, Zellner, and DeGroot. In addition to

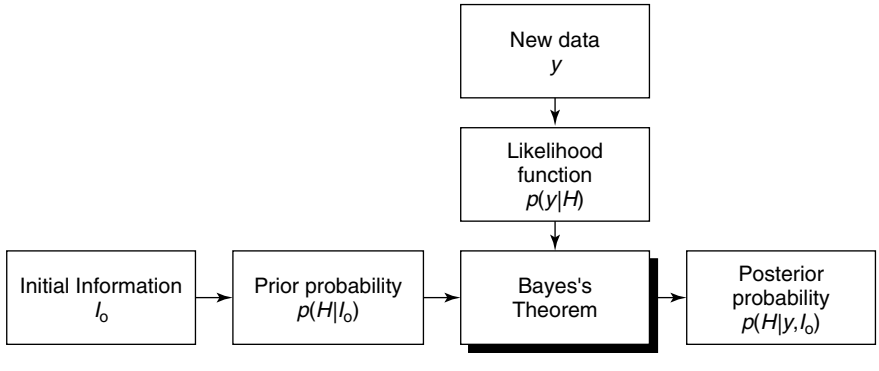


Figure 4.4 The process of revising probabilities of an hypothesis H, given new data.

other things, this work led to a realization that Bayesian methods enjoy favorable mathematical properties that frequentist methods do not. Important among these is mathematical coherency and simplicity.[2]

4.2.2 Inferring probability distributions

The case of inferring probability distributions over a discrete or continuous random variable is exactly analogous to the preceding case. Starting from the marginal probability distribution (pmf or pdf) over a state of nature, $f_\Theta(\theta)$, the likelihood function of the data is used to update the distribution via Bayes' Theorem to a posterior probability distribution conditioned on the observations,

$$f_\Theta(\theta|\mathbf{z}) = N\ f_\Theta(\theta)L(\theta|\mathbf{z}) \tag{4.12}$$

in which N is a normalizing constant ensuring that the posterior distribution is proper, that is, that it integrates to one, $N^{-1} = \int_{-\infty}^{+\infty} f_\Theta(\theta)L(\theta|\mathbf{z})\,d\theta$. Often, Equation (4.12) is simply written as a proportionality.

In the example of finding liquefiable layers in SPT borings, we assumed a value for the detection probability, that is, we assumed a probability that an existing layer would be detected in an SPT boring log if the boring in fact penetrates the layer. If empirical data were available from other sites or from other areas of the present site, we might instead have tried to use those data to infer the detection probability statistically. Consider that we have a set of twenty (20) SPT borings that we know *ex ante* to have intersected a liquefiable layer, and that in six (6) of those the layer was detected. What can we say about the value of the detection probability?

Assume that the observations are independent of one another, and that the probability of detection in any one boring is π. The individual outcomes are Bernoulli random variables (i. e., zeros or ones with some fixed probability), and thus the likelihood of observing six 'hits' out of twenty attempts can be modeled by a Binomial pmf (Equation 3.41),

$$\begin{aligned} f(x|n,\pi) &= \binom{x}{n} \pi^x (1-\pi)^{n-x} \\ &= L(\pi|x,n) \end{aligned} \tag{4.13}$$

in which $x = 6$, the number of 'hits,' and $n = 20$, the number of attempts. From Bayes' Theorem (Equation 4.12),

$$\begin{aligned} f_\pi(\pi|x,n) &\propto f_\pi(\pi)L(\pi|x,n) \\ &\propto (1.0)\pi^x(1-\pi)^{n-x} \end{aligned} \tag{4.14}$$

in which the prior pmf is taken to be uniform over [0,1], $p_\pi(\pi) = 1$, and terms not involving π are ignored (Figure 4.5). The maximum (mode) of the posterior pmf is at

[2] This can be seen, for example, in the complexity of Fisher's derivations of sampling distributions (Fisher and Bennett 1971), compared with the simple elegance of Jeffreys' (1983) results for corresponding cases.

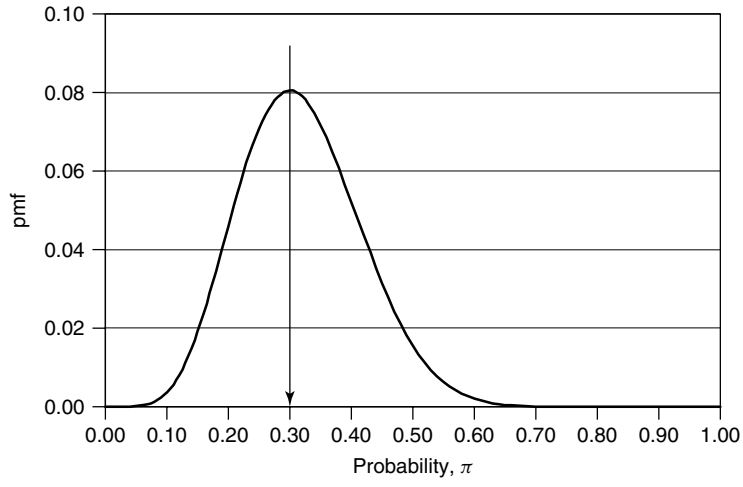


Figure 4.5 Posterior probability mass function of the Binomial parameter, π.

$\pi = 6/20 = 0.3$, corresponding to the naïve estimate, although the distribution is not perfectly symmetrical.

4.3 Prior Probabilities

The sticking point in Bayesian inference is specifying a prior distribution, especially in the case of no prior information. In his posthumous paper, Bayes (1763) introduces a postulate that says, in the case of ignorance every possible value of an uncertain quantity should be given equal probability. This suggests, for example, a uniform prior on a parameter such as π in the example above. Jeffreys (1983) builds on this postulate, calling it the *principle of indifference.*

A difficulty with this approach is that the resulting prior probability distributions expressing total ignorance are not invariant to transformations of variables. For example, if one is ignorant of the relative probabilities of values of π, one would be equally ignorant of the relative probabilities of values of π^2, so why not specify a uniform prior on the latter? Doing so results in a different posterior on π. This problem is discussed at length by Jeffreys and Zellner, and a variety of pragmatic justifications have been suggested by other authors (Box and Tiao 1992; Lindley 1971). Our view is that a great many assumptions enter any modeling activity – for example, the assumption that a random variable is Binomial, that the observations are independent, or even that the process is stationary – and the modest invariance introduced by the flat prior is inconsequential in this mix. In practical situations with little prior information, as the number of observations grows, the posterior pmf or pdf rapidly approaches the likelihood function.

4.3.1 Non-informative priors

The case of assigning prior probabilities to discrete variables when no information is available – such as 'presence' vs. 'non-presence' of a geological anomaly – is conceptually

straightforward. There are n categories, each is assigned the same prior probability, $1/n$, and the sum is 1.0. But, what do we do about variables with infinite domain?

Consider an event such as extreme flood or severe earthquake. This is usually modeled as a Poisson process with parameter, λ, describing the number of occurrences per unit time

$$f_n(n|\lambda, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \tag{4.15}$$

in which n is the number of occurrences, and t is duration. The parameter λ has domain $[0,\infty)$. What prior distribution should be used? A 'flat' pdf, $f_\lambda^0(\lambda) \propto k$, $k =$ constant, is improper because its area is infinite; and as before, we could just as reasonably make the pdf flat over λ^2 or even λ^3.

Several authors, notably Jeffreys (1983), have considered the best choice of *non-informative* prior distributions. Jeffreys argues that prior probabilities should be taken as uniform over the parameter for variables with domain $(-\infty, \infty)$, and uniform over the logarithm for variables with domain $[0,\infty]$

$$f_X^0(x) \propto \begin{cases} k & -\infty \le x \le \infty \\ x^{-1} & 0 \le x \le \infty \end{cases} \tag{4.16}$$

The latter becomes proportional to the inverse of the parameter in arithmetic space (Figure 4.6). Jeffreys presents a long and subtle discussion of the issues surrounding pdf's to represent ignorance. He notes that both Bayes and Laplace used uniform prior distributions in problems relating to sampling, but that this choice is not always satisfactory.

The so-called Jeffreys prior, $f_X(x) \propto 1/x$, is widely used in practice. Jeffreys argues for this prior on the basis of power transformation invariance. Sivia (1996) argues for it on the basis of scale transformation invariance. Zellner (1971) summarizes much of the

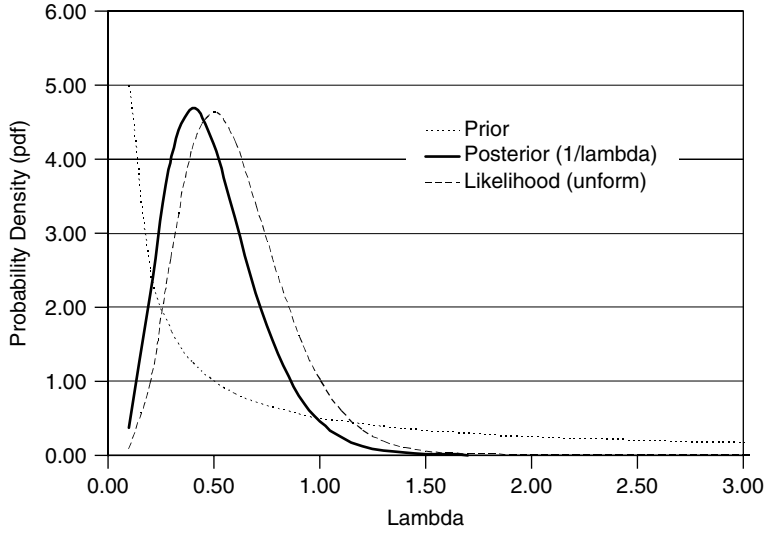


Figure 4.6 Prior and posterior pdf's of the Poisson pdf rate parameter λ, given an observation of five (5) events within ten (10) years, starting from a uniform prior on λ^{-1}.

Table 4.2 Prior distributions suggested to represent ignorance under various conditions

Constraint	Description	pdf	Colloquial name	Citation
any	Uniform	$f(x) \propto= constant$	Laplacian/Bayes prior	(Bayes 1763; Laplace 1814)
$a \le x \le b$	Uniform	$f(x\|a, b) = 1/(b-a)$	Uniform prior	(Jeffreys 1983)
$-\infty \le x \le +\infty$	Uniform	$f(x) \propto= constant$	Uniform prior	(Jeffreys 1983)
$0 \le x \le \infty$	Uniform in ln x	$f(x) \propto 1/x$	Jeffreys prior	(Jeffreys 1983)
$0 \le x \le 1$;	Uniform in x	$f(x) \propto= \dfrac{1}{x(1-x)}$	Beta prior	(Haldane 1932)

Jeffreys argument in favor of the prior. Suggestions of other non-informative priors are given in Table 4.2.

Another line of reasoning in selecting non-informative priors is based on the information content in the pdf. One measure of information common in modern communications theory is Shannon's Entropy (Shannon and Weaver 1949):

$$H = \int f_X(x) \log f_X(x)\, dx \tag{4.17}$$

in which $f_X(x)$ is the pdf of x, and H is said to be the informational entropy. This approach has been championed by Jaynes (Jaynes 1996; Sivia 1996) and by Harr (1987). Table 4.3 summarizes maximum entropy priors for various constraints and levels of information (Baecher 1972).

Table 4.3 Maximum entropy probability distributions for various constraints and levels of information (Baecher 1972)

Constraint	Maximum Entropy pdf	pdf	Range	Mean	Variance
$a \le x \le b$	Uniform	$f(x\|a, b) = 1/(b-a)$	$a \le x \le b$	(b − a)/2	$(b-a)^2/12$
$-\infty \le x \le +\infty$; mean, variance known	Normal	$f(x\|\mu, \sigma) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty \le x \le \infty$	μ	σ^2
$x \ge 0$; mean known	Exponential	$f(x\|\lambda) = \lambda e^{-\lambda x}$	$0 \le x \le \infty$	$1/\lambda$	$1/\lambda^2$
$x \ge 0$; mean, mean of ln x known	Gamma	$f(x\|\lambda, k) = \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$	$0 \le x \le \infty$	k/λ	k/λ^2
$0 \le x \le 1$; mean of ln x, and (1-ln x) known	Beta	$f(x\|c, d) = \dfrac{1}{B(c,d)} x^{c-1} \times (1-x)^{d-1}$	$0 \le x \le \infty$	$\dfrac{c}{c+d}$	$\dfrac{cd}{(c+d)^2(c+d+1)}$

Jeffreys made a broader observation respecting the non-invariance of prior distributions based on Fisher's information matrix. Fisher (1925) proposed a measure of the information contained in a set of observations $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ with respect to a parameter $\boldsymbol{\theta}$ to be

$$I_\theta = E_{\mathbf{y}|\theta}\left(-\frac{\partial^2 \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right) \tag{4.18}$$

Note, that $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood of $\boldsymbol{\theta}$. Jeffreys observed that taking the non-informative prior proportional to the square root of the determinant of the information matrix

$$f(\boldsymbol{\theta}) \propto |I_\theta|^{1/2} \tag{4.19}$$

provides invariance to any one-to-one differentiable transformation $\boldsymbol{\eta} = g(\boldsymbol{\theta})$. That is, similarly taking the prior $f(\boldsymbol{\eta}) \propto |I_\eta|^{1/2}$ leads to a consistent posterior probability for both cases (Zellner 1971).

Consider again sampling from a Binomial process with parameter π. The likelihood of an individual observation is

$$f_x(x|\pi) \propto \pi^x(1-\pi)^{1-x} \tag{4.20}$$

in which x is the observation (either zero or one), and π is the probability of success on the individual trial. Thus, since $E[x|\pi] = \pi$

$$\begin{aligned} I_\pi &= E_{x|\pi}\left(-\frac{\partial^2 \log f(x|\pi)}{\partial\pi^2}\right) \\ &= \pi^{-1}(1-\pi)^{-1} \end{aligned} \tag{4.21}$$

and the non-informative prior pdf on π should be taken proportional to

$$f_\pi(\pi) \propto \pi^{-1/2}(1-\pi)^{-1/2} \tag{4.22}$$

Note that this differs from the naïve prior, $f_\pi(\pi) \propto k$, a constant. Box and Tiao (1992) discuss the justification for this rule in more detail.

4.3.2 Informative and conjugate priors

It is sometimes the case that prior information exists about the value of a parameter so that the prior distribution is informative rather than non-informative. While prior information can be modeled by any probability distribution, in practice it is convenient to choose a functional form for the prior distribution that simplifies multiplication by the likelihood function in Bayes' Theorem. The most common choice is a distribution closed under multiplication by the likelihood function. This is called the *conjugate* (or natural conjugate) *distribution*.

Table 4.4 Conjugate distributions to commonly encountered likelihoods

Likelihood	pdf	Non-informative pdf	Conjugate	Informative pdf
Binomial	$p(n\|x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$	$f_\pi(\pi) \propto \theta^{-1/2}(1-\theta)^{-1/2}$	Beta	$f(x\|c,d) = \frac{1}{B(c,d)}x^{c-1}(1-x)^{d-1}$
Poisson	$f(n\|\lambda) = \lambda^n \exp(-\lambda)$	$f(\lambda) \propto= \lambda^{-1/2}$	Gamma	$f(\lambda\|a,b) = \frac{1}{\Gamma(a)}b^a\lambda^{a-1}\exp(-b\lambda)$
Multi-nomial	$p(n\|x) = \binom{n}{\mathbf{x}}\theta_1^{x_t}\ldots\theta_d^{x_d}\left(1-\sum\theta_i\right)^{n-\sum x_i}$	$f(x\|g,h) \propto \theta_1^{-1/2}\ldots\theta_d^{-1/2}(1-\Sigma\theta_i)^{-1/2}$	Dirichlet	$f(x\|g,h) = \frac{1}{D(g,h)}\theta_1^{g_1-1}\ldots\theta_d^{g_d-1}(1-\Sigma\theta_i)^{h-1}$
Uniform $(0,\theta)$	$f(x\|\theta) = \frac{1}{\theta}$	$f(\theta) = \theta^{-1/2}$	Reciprocal	$f(\theta\|a,r) = ra^r\theta^{-(r+1)}$
Exponential	$f(x\|\lambda) = \lambda e^{-\lambda x}$	$f(\lambda) \propto= \lambda^{-1/2}$	Gamma	$f(\lambda\|a,b) = \frac{1}{\Gamma(a)}b^a\lambda^{a-1}\exp(-b\lambda)$
Gamma (given k)	$f(x\|\lambda,k) = \frac{\lambda^k x^{k-1}e^{-\lambda x}}{\Gamma(k)}$	$f(\lambda\|k,a,b) \propto \lambda^{-1/2}$	Gamma	$f(\lambda,k\|a,b) = \frac{1}{\Gamma(a)}b^a\lambda^{a-1}\exp(-b\lambda)$
Normal	$f(x\|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$f(\mu,\sigma) \propto \sigma^{-1}$	Normal-Inverted Gamma	$f(\mu,\sigma\|a,d,m) \propto (\sigma^2)^{-(d+2)/2}\exp\{-[(\mu-m)^2+a]/2\sigma^2\}$

From the example of Binomial sampling, the likelihood of an individual observation x_i is Equation (4.20), thus the inference from a set of observations $\mathbf{x} = \{x_1, \ldots, x_n\}$ on the parameter π by Bayes' Theorem is

$$\begin{aligned} f_\pi(\pi|\mathbf{x}) &\propto f_\pi(\pi) L(\pi|\mathbf{x}) \\ &\propto f_\pi(\pi)[\pi^m(1-\pi)^{n-m}] \end{aligned} \tag{4.23}$$

in which $m = \Sigma x_i$. Taking a prior pdf in the form of a Beta distribution

$$f_\pi(\pi) \propto \pi^{a-1}(1-\pi)^{b-1} \tag{4.24}$$

with parameters (a, b) yields a posterior pdf of the same Beta form

$$\begin{aligned} f_\pi(\pi|\mathbf{x}) &\propto [\pi^{a-1}(1-\pi)^{b-1}][\pi^m(1-\pi)^{n-m}] \\ &\propto \pi^{a+m-1}(1-\pi)^{b+(n-m)-1} \end{aligned} \tag{4.25}$$

but with updated parameters, $a' = a + m$ and $b' = b + (n - m)$.

More detailed discussions of conjugate distributions are presented by Jeffreys (1983), Raiffa (1968), Zellner (1971), and Box and Tiao (1992). Conjugate distributions for common likelihoods are given in Table 4.4.

4.4 Inferences from Sampling

Attention now turns to the use of Bayes' Theorem in drawing inferences from specific sampling processes, specifically in estimating the parameters of probability distributions from observations. Following the question of inference is that of forecasting: how is Bayes' Theorem used to establish probability distributions over unobserved outcomes? We consider only select cases; more complete results are given in Table 4.4 and in Appendix A.

4.4.1 Binomial sampling

To review the continuing example earlier in the chapter, the Binomial process with parameter π is

$$f(m|n, \pi) = \binom{m}{n} \pi^m (1-\pi)^{n-m} \tag{4.26}$$

in which m is the number of successes out of n observations; the mean is $E[m] = n\pi$ and $Var[m] = (n\pi)(1-\pi)$. Thus the likelihood of the parameter π is

$$L(\pi|m, n) \propto \pi^m (1-\pi)^{n-m} \tag{4.27}$$

The non-informative prior distribution of π is

$$f_\pi(\pi) \propto \pi^{-1/2}(1-\pi)^{-1/2} \tag{4.28}$$

although one often sees the naïve prior, $f_\pi(\pi) \propto$ constant, used in the literature. The informative, conjugate prior distribution of π is of the Beta form with parameters (a, b),

$$f_\pi(\pi) \propto \pi^{a-1}(1-\pi)^{b-1} \tag{4.29}$$

and the posterior pdf on π is of the same form, but with parameters, $a' = a + m$ and $b' = b + (n - m)$

$$\begin{aligned} f_\pi(\pi|\mathbf{x}) &\propto [\pi^{a-1}(1-\pi)^{b-1}][\pi^m(1-\pi)^{n-m}] \\ &\propto \pi^{a+m-1}(1-\pi)^{b+(n-m)-1} \end{aligned} \tag{4.30}$$

4.4.2 Poisson sampling

The Poisson process with parameter λ is of the form

$$f_N(n|\lambda) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \tag{4.31}$$

in which n is the number of occurrences; the mean is $E[n] = \lambda t$ and $Var[n] = \lambda t$. Thus the likelihood of the parameter λ for a set of observations $\boldsymbol{n} = \{n_1, \ldots, n_k\}$ in periods $\boldsymbol{t} = \{t_1, \ldots, t_k\}$ is

$$L(\lambda|n) \propto \prod_i (\lambda t_i)^{n_i} e^{-\lambda t_i} \tag{4.32}$$

The non-informative prior distribution of λ is

$$f_\lambda(\lambda) \propto \lambda^{-1} \tag{4.33}$$

The informative, conjugate prior distribution of π is of the Gamma form with parameters (a, b)

$$f_\lambda(\lambda|a, b) = \lambda^{a-1} \exp(-b\lambda) \tag{4.34}$$

and the posterior pdf on π is of the same form, but with parameters, $a' = a + \Sigma n$ and $b' = b + \Sigma t$,

$$f_\lambda(\lambda|\mathbf{n}, \mathbf{t}) \propto \lambda^{(a+\Sigma n_i)-1} e^{-(b+\Sigma t_i)\lambda} \tag{4.35}$$

As an example, consider that $n = 5$ events have been observed in $t = 10$ years. The uniform prior pdf on λ^{-1} and the posterior pdf are shown in Figure 4.6. The modal (most probable) value of λ is found by setting the derivative of Equation (4.35) to zero, giving $\hat{\lambda} = 0.4$. The outcome that the mode is located at 0.4 rather than the average of the sample (events/year = 0.5) derives from the prior pdf on λ. For the case of uniform prior pdf on λ, the posterior mode is 0.5, the same as the mode of the likelihood function. No matter how large the sample size, there will always remain a $1/\lambda$ difference in the posterior pdf between the cases of the two prior pdf's.

4.4.3 Exponential sampling

The Exponential process with parameter λ is of the form,

$$f_x(x|\lambda) = \lambda e^{-\lambda x} \tag{4.36}$$

in which x is a continuous variable with domain [0,∞]; the mean is $E[x] = 1/\lambda$ and $Var[n] = 1/\lambda^2$. Thus the likelihood of the parameter λ is

$$L(\lambda|\mathbf{x}) \propto \prod_i \lambda e^{-\lambda x_i} \tag{4.37}$$

for a set of observations $\mathbf{x} = \{x_1, \ldots, x_k\}$. The non-informative prior distribution of λ is

$$f_\lambda(\lambda) \propto \lambda^{-1} \tag{4.38}$$

The informative, conjugate prior distribution of π is of the Gamma form with parameters (a, b),

$$f_\lambda(\lambda|a, b) = \lambda^{a-1} \exp(-b\lambda) \tag{4.39}$$

and the posterior pdf on λ is of the same form, but with parameters, $a' = a + k$ and $b' = b + \Sigma x_i$,

$$f_\lambda(\lambda|n) \propto \lambda^{(n+a)-1} \exp(-(b + \Sigma x_i)\lambda) \tag{4.40}$$

4.4.4 Normal sampling

Perhaps the most common sampling encountered in geotechnical practice is that from Normally distributed properties. The Normal process with parameters (μ, σ) is of the form

$$f_X(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \tag{4.41}$$

in which x is a continuous variable with domain $(-\infty, +\infty)$; the mean is $E[x] = \mu$ and $Var[x] = \sigma^2$. Thus the likelihood of the parameters (μ, σ) is

$$L(\mu, \sigma|x) \propto \frac{1}{\sigma^n} \prod_i \exp\left\{\frac{-(x_i-\mu)^2}{2\sigma^2}\right\} \tag{4.42}$$

for a set of observations $\mathbf{x} = \{x_1, \ldots, x_n\}$. The non-informative prior distribution of (μ, σ) is

$$f_{\mu,\sigma}(\mu, \sigma) \propto \sigma^{-1} \tag{4.43}$$

The informative, conjugate prior distribution of (μ, σ) is of the Normal-Inverted Gamma form with parameters (m', d', n', v'),

$$f(\mu, \sigma|m', d', n', v') \propto \left[\exp\left\{-\frac{1}{2\sigma^2}n'\right\}(\mu - m')^2\right]\left[\frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}d'v'\right\}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}v'-1}\right] \tag{4.44}$$

That is, the prior distribution of μ conditional on σ is Normal, and the marginal prior distribution of σ is Inverted-Gamma. The posterior pdf of (μ, σ) is of the same form, with parameters (Raiffa and Schlaifer 1968),

$$\begin{aligned} m'' &= \frac{1}{n''}(n'm' + nm) \\ d'' &= \frac{1}{\nu''}[(\nu' d' + n' m'^2) + (\nu d + n m^2) - n'' m''^2] \\ n'' &= n' + n \\ \nu'' &= \nu' + \nu + 1 \end{aligned} \tag{4.45}$$

in which

$$\begin{aligned} m &= \frac{1}{n}\sum_i x_i = \overline{x} \\ d &= \frac{1}{\nu}\sum_i (x_i - m)^2 \\ \nu &= n - 1 \end{aligned} \tag{4.46}$$

For the data of Figure 4.7, this yields the joint posterior pdf shown as a contour plot in Figure 4.8. The ovals in the figure show the contours of the pdf.

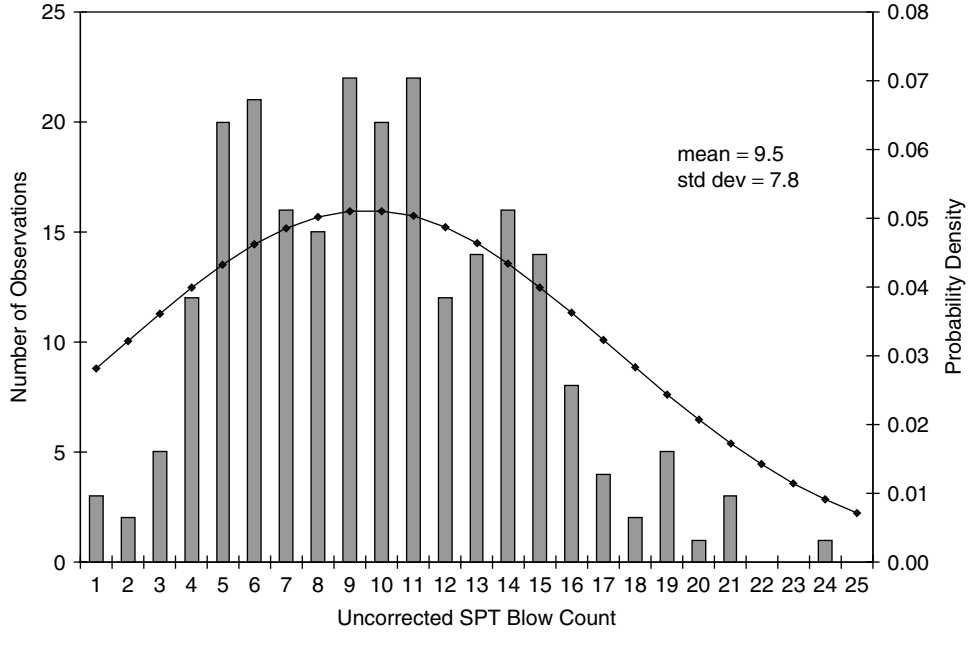


Figure 4.7 Comparison of SPT blow count data with the best-fitting Normal distribution. Data set has broader 'shoulders' than the Normal pdf and less populated tails (from Baecher 1987a).

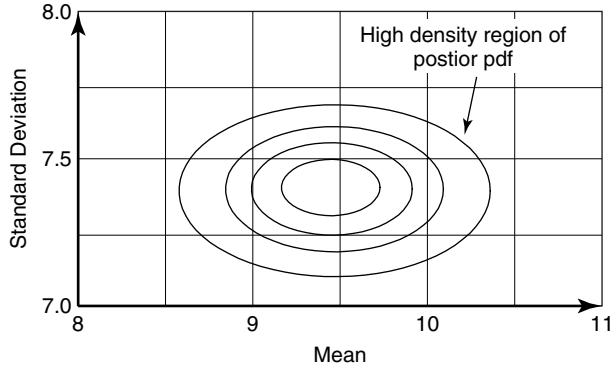


Figure 4.8 Posterior pdf of the Normal mean μ and standard deviation σ based on the data from Figure 4.7.

4.4.5 Predictive distribution

The inferences from sample observations discussed above lead to posterior pdf's on the parameters of the underlying processes. Often, however, we are not so concerned with uncertainty in the parameters as we are with the uncertainties in future outcomes of the process itself. For example, from the historical record of flood or earthquake events, we use Equation (4.35) to infer a pdf on the rate of occurrence, λ. The question becomes, what uncertainty does this imply about the number of future events, and what time duration can we expect between events?

Common practice is to integrate out the parametric uncertainty to obtain a *predictive probability distribution*. For example, the pmf of n marginal of uncertainty in λ is found by integrating the model forecast over the uncertainty in the parameter (Aitchison and Brown 1969)

$$f_n(n) = \int_\lambda f_n(n, \lambda)\, d\lambda = \int_\lambda f_n(n|\lambda)\, f_\lambda(\lambda)\, d\lambda \tag{4.47}$$

Thus, presume that we have observed $\boldsymbol{n} = \{n_1, \ldots, n_k\}$ numbers of events in k unit time periods. From Equation (4.35), starting with a non-informative prior on λ, the posterior pdf on λ is Gamma with parameters $(\Sigma n_i, k)$, and thus from Equation (4.31), the predictive pmf on the future number of events per unit of time, n, is

$$\begin{aligned} f_N(\mathbf{n}) &= \int_\lambda f_N(\mathbf{n}|\lambda)\, f_\lambda(\lambda|\mathbf{n})\, d\lambda \\ &\propto \int_\lambda \left(\frac{\lambda^n e^{-\lambda}}{n!}\right) (\lambda^{\Sigma n_i - 1} e^{-(k+1)\lambda}) d\lambda \end{aligned} \tag{4.48}$$

which, upon integrating, becomes the Negative-Binomial pmf (Aitchison and Dunsmore 1975):

$$f_N(n|\Sigma n_i, k) = \binom{n + \Sigma n_i - 1}{\Sigma n_i} \left(\frac{1}{k+1}\right)^n \left(1 + \frac{1}{k+1}\right)^{\Sigma n_i} \tag{4.49}$$

Note that the predictive distribution may combine uncertainty due to natural variability, for example, the number of flood occurrences in a period of time, with uncertainty due

to limited knowledge, that is, the uncertainty in the rate parameter. If more than one aspect of natural variability depends on the same uncertain parameter (e.g. number of occurrences in two separate periods, n_1 and n_2) then forming the predictive distribution of each in isolation from the other may mask an implied correlation caused by their shared dependence on the same uncertain realization of λ.

A point of caution in dealing with predictive distributions is that, if the conditional distribution $f_X(x|\theta)$ changes rapidly within the interval of interest of θ, then reliance is being placed on the precise shape of the pdf of θ within that region. While the resulting predictive pdf is theoretically correct, errors in the specification of $f_\Theta(\theta)$ can result in large errors in $f_X(x)$.

4.5 Regression Analysis

It is often the case that soil properties or other variables are related to one another, as illustrated in Figure 4.9, in which a gradual increase in maximum past pressure in a Gulf of Mexico clay formation is observed with depth. Since the undrained strength of the clay increases with pre-consolidation pressure (Ladd and Foott 1974), this information can be used in a reliability analysis to model the increasing strength of the clay as one moves deeper in the deposit. To do so, however, necessitates a quantitative relationship between depth and maximum past pressure.

In Chapter 3 we introduced the notion of a correlation coefficient as a measure of the degree to which two variables are (linearly) related. Here, the correlation coefficient between depth and maximum past pressure is $\rho = 0.41$. What we need, however, is a mathematical relationship that captures this correlation and allows a probabilistic prediction to be made of maximum past pressure for a known depth. This is the purview of regression analysis.

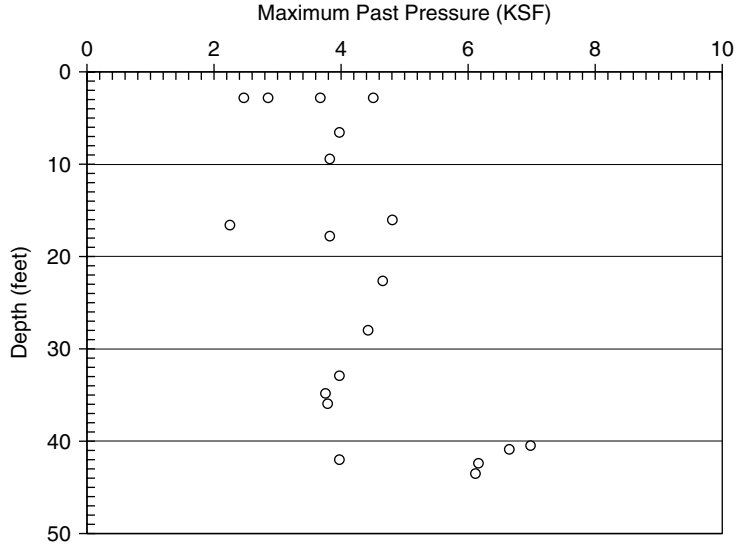


Figure 4.9 Maximum past pressure in an over consolidated Gulf of Mexico clay as a function of depth (Baecher and Ladd 1997).

Regression analysis was introduced by Sir Francis Galton (1822–1911), also the discoverer of fingerprinting and inventor of the silent dog whistle, to describe the relationship of hereditary characteristics between parents and children, the most often cited being that of the heights of fathers and their sons (Gillham 2001). The term *regression* was adopted because Galton observed that, on average, the characteristics of the offspring regress toward the overall population average, tending to be less extreme than the characteristics of the parents. That is, tall fathers tend to breed taller than average sons, but not as tall as they. The reverse is true for short fathers.

4.5.1 Linear regression model

The univariate linear regression model has the form

$$y_i = \beta_1 + \beta_2 x_i + u_i \tag{4.50}$$

in which $\boldsymbol{y} = \{y_1, \ldots, y_n\}_i$ are dependent observed variables, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ are independent variables, $\boldsymbol{\beta} = \{\beta_1, \beta_2\}$ are scalar constants usually called *regression parameters* (or, *intercept* and *slope*), and $\boldsymbol{u} = \{u_1, \ldots, u_n\}$ are random errors. The assumptions are made that, (1) the $\boldsymbol{u}$ are Normally and independently distributed with zero-mean and a common variance, σ^2; and (2) the $\boldsymbol{x}$ are either fixed, deterministic variables, or they are random variables independent of $\boldsymbol{u}$ (see, e.g. Zellner 1971).

The Likelihood of the observations $\boldsymbol{y}$ conditioned on $\boldsymbol{x}$ and the parameters of the regression model is

$$\begin{aligned} L[\mathbf{y}|\mathbf{x}, \beta_1\beta_2\sigma] &= \prod_{i=1}^{n} N[\mathbf{y}|\mathbf{x}, \beta_1\beta_2\sigma] \\ &\propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_1 + \beta_2 x_i)^2\right] \end{aligned} \tag{4.51}$$

From Bayes' Theorem

$$f(\beta_1, \beta_2, \sigma|\mathbf{x}, \mathbf{y}) \propto f(\beta_1, \beta_2, \sigma) L[\mathbf{y}|\mathbf{x}, \beta_1\beta_2\sigma] \tag{4.52}$$

in which $f(\beta_1, \beta_2, \sigma|\mathbf{x}, \mathbf{y})$ is the posterior pdf of the regression parameters, and $f(\beta_1, \beta_2, \sigma)$ is the prior pdf. If a non-informative prior is adopted,

$$f(\beta_1, \beta_2, \sigma) \propto 1/\sigma \tag{4.53}$$

the posterior pdf of the regression parameters becomes

$$f(\beta_1, \beta_2, \sigma|\mathbf{x}, \mathbf{y}) \propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_1 + \beta_2 x_i)^2\right] \tag{4.54}$$

Integrating over the uncertainty in σ yields the marginal posterior pdf on the regression coefficients β_1 and β_2

$$f(\beta_1, \beta_2|\mathbf{x}, \mathbf{y}) = \int_0^\infty f(\beta_1, \beta_2, \sigma|\mathbf{x}, \mathbf{y})\, d\sigma$$
$$\propto [v\ s^2 + n(\beta_1 - \hat{\beta}_1)^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\Sigma x_i + (\beta_2 - \hat{\beta}_2)^2 \Sigma x_i^2]^{-n/2} \tag{4.55}$$

in which, $v = n - 2$, and

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$
$$\hat{\beta}_2 = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\Sigma(x_i - \overline{x})^2}$$
$$s^2 = v^{-1}\Sigma(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \tag{4.56}$$

where $\overline{y} = n^{-1}\Sigma y_i$, and $\overline{x} = n^{-1}\Sigma x_i$. Equation (4.55) is in the form of a Student-t distribution (Appendix A), which means that inferences about the coefficients β_1 and β_2 can be made from tabulated versions of the Student-t pdf. The marginal distributions of β_1 and β_2 are (Zellner 1971),

$$f(\beta_1|\mathbf{x}, \mathbf{y}) \propto \left[v + \frac{\Sigma(x_i - \overline{x})^2}{s^2 \Sigma x_i^2 / n} (\beta_1 - \hat{\beta}_1)^2 \right]^{-(v+1)/2} \tag{4.57}$$

$$f(\beta_2|\mathbf{x}, \mathbf{y}) \propto \left[v + \frac{\Sigma(x_i - \overline{x})^2}{s^2} (\beta_2 - \hat{\beta}_2)^2 \right]^{-(v+1)/2} \tag{4.58}$$

and the marginal distribution of σ is (Zellner 1971)

$$f(\sigma|\mathbf{x}, \mathbf{y}) \propto \frac{1}{\sigma^{v+1}} \exp\left(-\frac{v s^2}{2\sigma^2} \right) \tag{4.59}$$

which is of the Inverted-Gamma form. The derivation here follows Zellner (1971), but similar results are also found elsewhere (Box and Tiao 1992; O'Hagan 1994; Press *et al.* 1992; Sivia 1996; Zellner 1997). For the maximum past pressure data of Figure 4.9, the joint posterior pdf of β_1 and β_2 is shown in Figure 4.10; and the marginal posterior pdf of σ is shown in Figure 4.11. Note the strong dependence of β_1 and β_2. Since both parameters depend on the same observational data, they are not independent.[3] The same is true for σ in its relation to of β_1 and β_2, although this cannot be seen in the marginal distribution. Informative priors for the case of linear Normal regression are discussed by Zellner and by Box and Tiao.

4.5.2 Predictive distribution of y|x

The predictive distribution of maximum past pressure at a particular depth is found by integrating the conditional pdf with given regression coefficients over the uncertainty in

[3] The same correlation arises in estimates of the Mohr–Coulomb strength parameters, c, ϕ, which are negatively correlated when estimated from measurements.

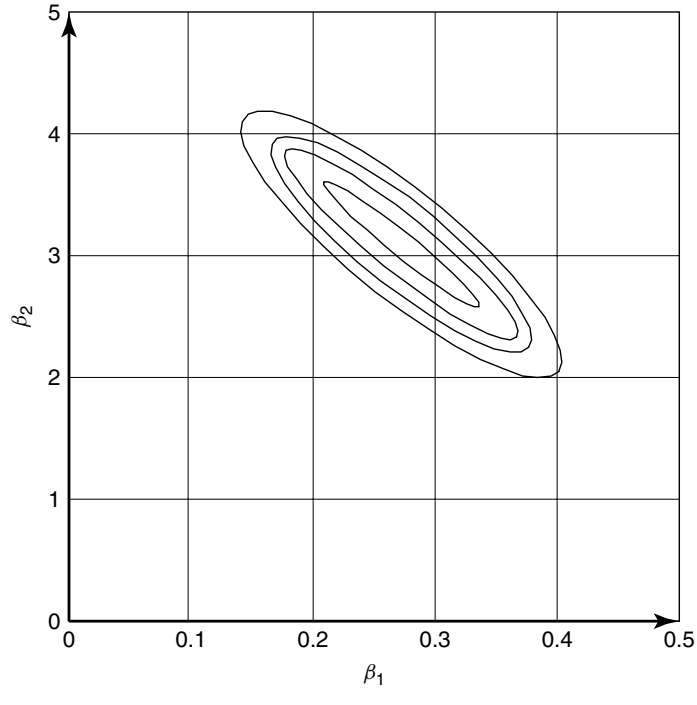


Figure 4.10 Joint posterior pdf of β_1 and β_2 for the maximum past pressure data of Figure 4.9.

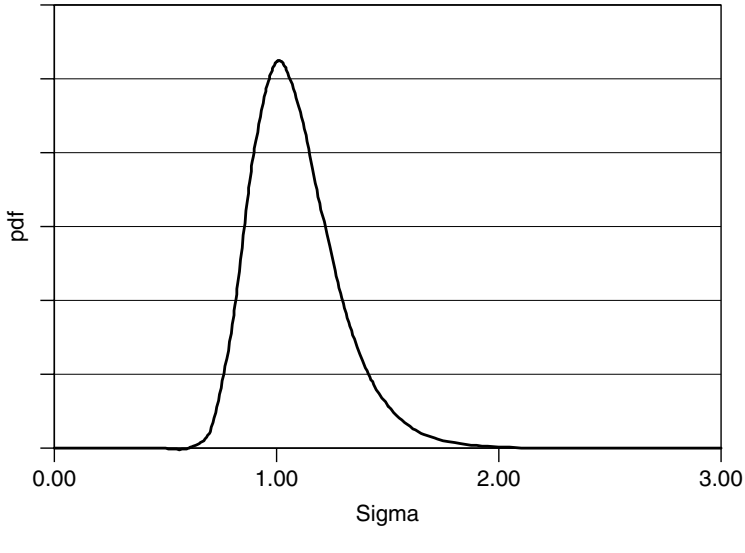


Figure 4.11 Marginal posterior pdf of σ for the maximum past pressure data of Figure 4.9.

the regression coefficients

$$f(y_i|x_i) = \iiint N[y_i|x_i, \beta_1, \beta_2, \sigma] f(\beta_1, \beta_2, \sigma)\, d\beta_1 d\beta_2 d\sigma \tag{4.60}$$

which has the form of the Student-t distribution with $\nu = n - 2$ degrees of freedom (Broemeling 1985; O'Hagan 1994; Zellner 1971):

$$\frac{y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_i)}{s\left(\frac{1}{n} + \frac{x_i}{S_{xx}} + 1\right)^{1/2}} \sim t_v \tag{4.61}$$

in which $S_{xx} = \Sigma(x_i - \overline{x})^2$. The predictive distribution on the mean of y for a given x is also Student-t, with mean equal $(\hat{\beta}_1 + \hat{\beta}_2 x_i)$, spread equal $s(1/n + x_i/S_{xx} + 1)^{1/2}$, and ν degrees of freedom. Aitchison and Dunsmore (1975) also discuss the problem of updating an estimate of regression relationships (in their case, 'calibrations') to account for site specific information. Figure 4.12 presents results for the problem of maximum past pressure.

4.5.3 Normal multiple regression

Results for the general Normal multiple regression model involving multiple independent variables are presented in Zellner and Box and Tiao, but are only summarized here. The general model for Normal multiple regression is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{4.62}$$

in which $\boldsymbol{y}$ is the vector of observations, as before, $\boldsymbol{X}$ is an $(n \times k)$ *design matrix* for the n observations made against k independent variables, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of regression

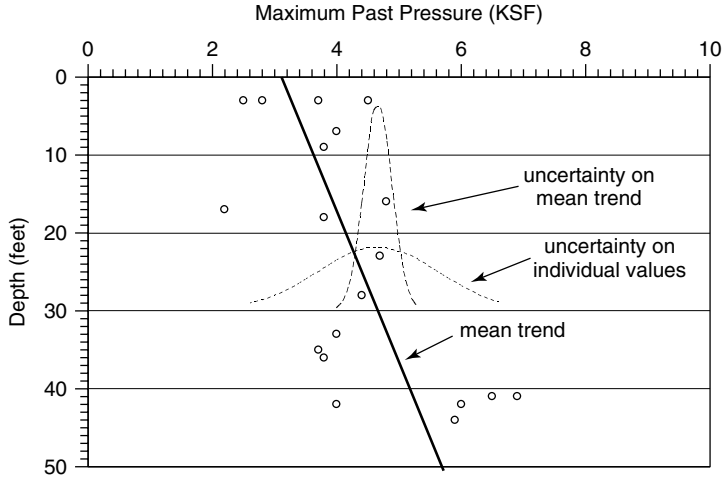


Figure 4.12 Predictive distributions on the mean maximum past pressure at dept 30 feet, and on the local (point value) of maximum past pressure at 30 feet.

coefficients, and $\mathbf{u}$ is an $(n \times 1)$ vector of independent, Normally distributed error terms of zero mean and common variance, σ^2.

The following results are offered without proof, which is to be found in a number of places (e.g. O'Hagan 1994; Press *et al.* 1989; Zellner 1971). The marginal posterior pdf of $\boldsymbol{\beta}$ starting from a non-informative prior, $f(\boldsymbol{\beta}, \sigma) \propto 1/\sigma$, is

$$f(\beta|\mathbf{y}, \mathbf{X}) \propto \{v\ s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t \mathbf{X}^t \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\}^{-n/2} \tag{4.63}$$

which is multivariate Student-t. The marginal posterior pdf on σ under the same assumption on prior is the same as in Equation (4.44). For a vector of as yet unobserved values of $\boldsymbol{y}$ at locations or conditions $\boldsymbol{X}$, the predictive distribution of $\boldsymbol{y}$ (i.e. marginal of uncertainty in $\boldsymbol{\beta}$, σ) is multivariate Student-t with mean,

$$E[\mathbf{y}] = \mathbf{X}_0 \hat{\boldsymbol{\beta}} \tag{4.64}$$

in which $\boldsymbol{X}_0$ is the matrix of future independent variables, and $\hat{\boldsymbol{\beta}}$ is the vector of mean regression coefficients; and covariance matrix,

$$\begin{aligned} \Sigma &= E\{[\mathbf{y} - \hat{\mathbf{y}}][\mathbf{y} - \hat{\mathbf{y}}]^t\} \\ &= \frac{v\ s^2}{v-2}[\mathbf{I} + \mathbf{X}_0(\mathbf{X}^\mathbf{t}\mathbf{X})\mathbf{X}_0{}^\mathbf{t}] \end{aligned} \tag{4.65}$$

in which ν is the number of degrees of freedom, $\mathbf{I}$ is the identity matrix, $\mathbf{X}_0$ is the matrix of future independent variables, and $\mathbf{X}$ is the matrix of x values from the original data set upon which the regression was based.

4.6 Hypothesis Tests

In the modern literature, Abraham Wald (1950) is credited with observing that statistical estimates and hypothesis tests are each a type of decision and should be approached from that point of view. Wald's work built on the then recent advances in game theory introduced by von Neumann and Morgenstern (1947). This led, especially in the Bayesian literature, to a broad treatment of both estimation and hypothesis testing under an umbrella of *statistical* (or sometimes, *Bayesian*) *decision theory*.

There are many excellent introductory and advanced level introductions to statistical decision theory from a Bayesian perspective: Lindley (1965; 1985), Raiffa (1968), Schlaifer (1978), Aitchison (1970), Pratt *et al.* (1995), and Clemen and Reilly (2001) provide introductory presentations; while De Groot (1970), Raiffa and Schlaifer (2000), Berger (1993), and French and Raios Insua (2000) provide more advanced introductions. We will treat the general topic of statistical decision analysis in Chapter 5; this section of Chapter 4 addresses the more narrow topic of hypothesis testing.

4.6.1 Historical background

As with probability theory, there are different schools of thought on hypothesis testing. Following Raiffa (1968), it is convenient to associate these with their principal proponents: Fisher, Neyman-Pearson, Wald, and Savage (Baecher 1972; 1983a).

Fisher (Fisher and Bennett 1971) used statistics to evaluate hypotheses on the state of nature, while relegating action decisions to judgment. Thus, Fisherians are primarily concerned with testing *null hypotheses* and with the probability of rejecting the null hypothesis when it is true.[4] Decisions rely on maximum likelihood, and probabilities are not placed on the state of nature.

Consider deciding on a foundation treatment to reduce seepage. For simplicity, presume experience suggests that the natural permeability of the deposit is either about 10^{-4} or about 10^{-5} cm/sec, depending on alternate soil property characteristics about which there is little *a priori* evidence. The best decisions, if we knew the permeability, are shown in Table 4.5. Treatment A is more expensive but more thorough than Treatment B. Which treatment should we chose?

Fisher would proceed by evaluating the likelihood function of a set of experimental observations (i.e. soil property measurements), $\mathbf{z}$, which for a given state of nature θ equals the conditional probability of observing $\mathbf{z}$. He would then conclude that the best estimate of the real state of nature is that which maximizes $L(\theta|\mathbf{z}) = f_{\mathbf{z}}(\mathbf{z}|\theta)$. The decision of which treatment to use, however, is left to judgment.

Neyman and Pearson (1933) held that experimentation and accepting or rejecting a null hypothesis are 'action problems.' When considering alternate strategies for accepting or rejecting hypotheses, each having the same probability of incorrectly rejecting a true null hypothesis (i.e. a so-called, *type I error*), one should favor those decision rules with the lowest probability of incorrectly accepting a false null hypothesis (i.e. a *type II error*). The tool for making this comparison is the *power function*, which describes the conditional probability of a type II error resulting from the use of that decision rule, given the state of nature.

Two power functions are shown in Figure 4.13 for two statistical rules for testing the hypothesis that there is no difference between the population of soil specimens sampled from the site and a population having an average permeability of 10^{-5} cm/sec. Assume that the rules were chosen to have the same probability of rejecting this null hypothesis when it is in fact true (type I error). The power functions describe the probability of accepting

Table 4.5 Consequence table for the foundation treatment

Treatment option	Worth	
k =	10^{-4}	10^{-5}
Treatment A	−1000	−100
Treatment B	−300	−300

[4] The *null hypothesis* is that which is the subject of statistical tests of significance. Typically, this is the hypothesis that an apparent difference between the populations is due to sampling or experimental error and that there is no actual difference between the populations. The OED (Brown 2002) attributes the first use of the term to Fisher (Fisher 1935): "The two classes of results which are distinguished by our test of significance are those which show a significant discrepancy from a certain hypothesis; namely, in this case, the hypothesis that the judgments given are in no way influenced by the order in which the ingredients have been added; and results which show no significant discrepancy from this hypothesis. This hypothesis is again characteristic of all experimentation. We may speak of this hypothesis as the 'null hypothesis', and it should be noted that the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation."

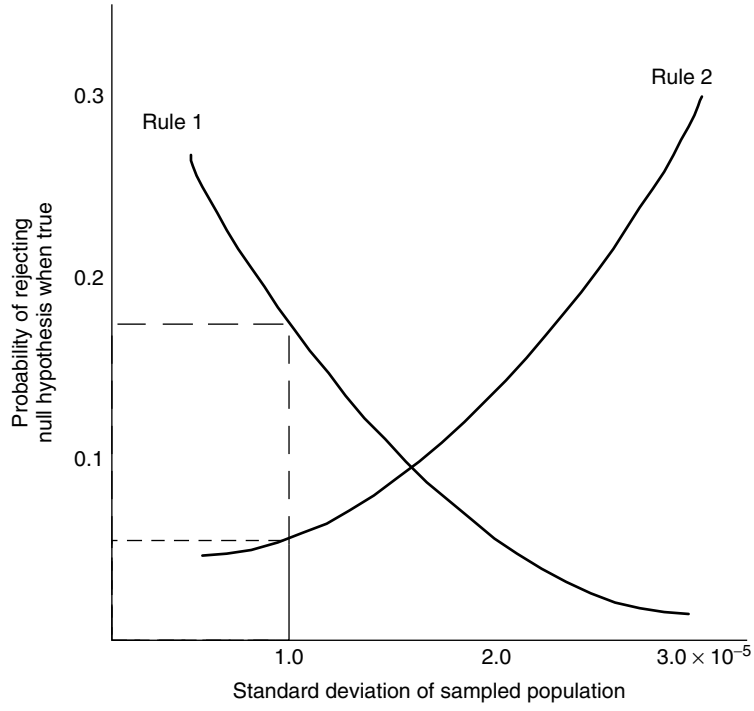


Figure 4.13 Power functions of two rules for deciding between alternative foundation characteristics.

the hypothesis when it is actually false (type II error). For the sake of illustration, this latter probability is assumed to depend on the value of the population variance. If the true standard deviation were actually 1.0×10^{-5}, then rule 1 would have a probability of a type II error of about 0.18, and rule 2 of about 0.06. The choice between the rules would depend on the range within which the true variance is thought to lie. The 'best' strategy is the one that has the universally lowest power function. One can usually obtain better error control over the entire parameter space only by increasing the cost of experimentation. Probability assessments are conditional and never assigned to states of nature; decisions are based on judgment.

Wald (1950) expanded the solution to the decision problem by adding the costs and worth of consequences directly into the equation. However, as in the Fisher and Neyman-Pearson schools, Wald did not place probabilities on states of nature. Wald first evaluates the worth of a proposed rule under each possible state of nature (Table 4.5). Then, he proposed using a *minimax* rule, which is to minimize the maximum loss. This would lead to treatment option B, since the maximum loss associated with that option for any state of nature is -300, which is a smaller loss than the -1000 associated with option A should the worst state of nature obtain. Alternative decision criteria have also been proposed for this approach.

Wald defines *admissible solutions* as those for which no other option exists with a greater worth under all possible states of nature. For example, a hypothetical option C with worth $\{-500, -500\}$ would not be admissible since B has a greater worth no matter

the true state of nature. The decision therefore is reduced to choosing among admissible solutions. Although the rationality of restricting consideration to admissible solutions cannot be questioned, the choice of secondary strategy may not be consistent with one's preferences and aversion to risk. For example, if one assigned a probability to $k = 10^{-4}$ that was exceedingly small, the minimax criterion would still lead to option B; yet, one might be willing to risk this small probability of a large loss to save a non-negligible cost difference between treatments.

Everything to do with statistical decision making changes under a degree-of-belief view, because the stricture of not placing probabilities on states of nature is relaxed. The concept that probability and action are inseparable was introduced in the philosophy literature by Ramsey and Braithwaite (1931) (although foreshadowed in the writings of Peirce (1998) and James (1907)) and leads to the use of subjective probability and utility in decision making. *Utility* might be thought of as subjective worth, differing from direct monetary value by its inclusion of personal values and attitudes toward risk (Raiffa 1968).

In *The Foundations of Statistics*, Savage (1954) develops a theory of idealized behavior under uncertainty predicated upon Ramsey's view of probability and upon the theory of utility developed by von Neumann and Morgenstern (1947). Beginning from a set of axioms describing what is meant by a consistent set of preferences (utilities). Savage shows that an ideally rational person should make decisions on the basis of expected utility. Having already entered subjective probability into the analytical formulation, Savage would evaluate the expected utility of taking each of several actions and elect that with the largest (Table 4.6).

4.6.2 Bayes' decision framework for hypothesis testing

Bayesian theory views all hypothesis tests and parameter estimates as decisions. The structure of these decisions involves the following (Figure 4.14):

- A space of actions, $\mathbf{A} = \{a_1, \ldots, a_m\}$;
- A space of states of nature, $\mathbf{\Theta} = \{\theta_1, \ldots, \theta_k\}$;
- A sample space of observations, $\mathbf{z} = \{z_1, \ldots, z_n\}$;
- Likelihood functions, $L(\theta_i|z_j)$, over all i, j; and
- Prior probabilities over, $f(\theta_i)$, over all i;
- Utilities of the outcomes, $\mathbf{U} = \{u(a_i, \theta_j)\}$, over all i, j;

The problem is to choose the best action within $\mathbf{A}$, given sample information, $\mathbf{z}$, and the utilities, $\mathbf{U}$. The Bayesian approach does so by using the likelihood functions, $L(\theta_i|z_j)$,

Table 4.6 Expected utilities for treatment options given probabilities on the states of nature. The change point between treatments is for $p(k = 0.0001)$ of 0.22

Treatment option	Utility		Expected utility
Permeability	1.E-04	1.E-05	
Probability	*0.10*	*0.90*	
Treatment A	−1000	−100	−190
Treatment B	−300	−300	−300

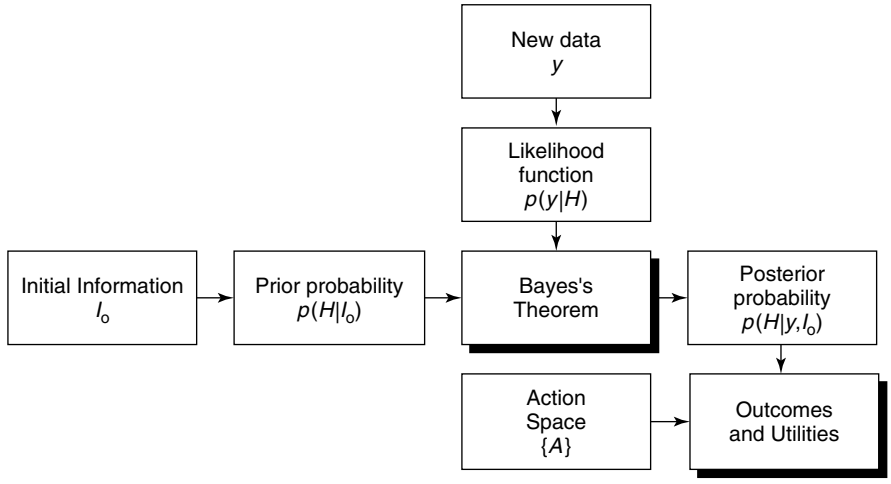


Figure 4.14 Process of revision of probabilities combined with decision making.

and Bayes' Theorem to update the *a priori* probabilities over the states of nature, $f(\theta_i)$, to *a posteriori* probabilities, $f(\theta_i|\mathbf{z})$, and then to maximize the expected value of **U** over **A** for these probabilities.

The action space, **A**, is the set of alternative hypotheses under consideration. For example, in testing the hypothesis that the average permeability of subsurface soils at a site in question is $k = 10^{-4}$ vs. $k = 10^{-5}$, the alternative actions (or hypotheses) are

$$A = \begin{cases} a_1\text{:}accept\ that\ k = 10^{-4}\ cm/s \\ a_2\text{:}accept\ that\ k = 10^{-5}\ cm/s \end{cases} \tag{4.66}$$

Note in the Bayesian sense, that a hypothesis can never be tested without there being at least one alternative hypothesis.

The utilities of the outcomes are a function of the consequences, in the present case, the costs shown in Table 4.6. For the moment, we will take the utilities simply to be equal to the cost outcomes, and return to the question of non-linear utility and the attitudes toward risk it implies in Chapter 5. In some cases, as for example in using least-squares estimates, the utilities are implicitly taken to be quadratic loss. In principle, at least from a Bayesian view, the implicit structure of utilities and prior probabilities can always be back-calculated from a set of decisions actually taken.

The situation sometimes arises in which, rather than arriving at a yes-or-no decision among two hypotheses (the problem is extendible to more than two alternatives), the outcome of the hypothesis test is a set of probabilities on the hypotheses. In this case we might seek a loss function defined on the assigned probability, of the form

$$u(\pi, \theta) = \begin{cases} \ell_1(\pi) \text{ if } \theta_1 \\ \ell_2(\pi) \text{ if } \theta_2 \end{cases} \tag{4.67}$$

in which π is the probability assigned to hypothesis H_1, and $\ell_1(\pi)$ is the corresponding loss if θ_1 obtains, and $\ell_2(\pi)$ if θ_2 obtains. This problem arises in subject probability

assessment, where one wishes to establish *scoring rules* that reward subjects for honest reporting (Speltzer and Stael von Holstein 1975). Thus, a reward structure is established that reduces the fee paid to the subject by $\ell_1(\pi)$ if θ_1 occurs, and $\ell_2(\pi)$ if θ_2.[5]

4.6.3 Test of a point hypothesis

A *finite action problem* is one in which the set of alternative actions contains a finite number of discrete alternatives. For example, the choice above between two average permeabilities is a finite action problem, as is the case that follows, in which the choice is between accepting an historical defect rate or deciding that the rate differs from the historical.

Consider a case in which n observations, $\mathbf{z} = \{z_1, \ldots, z_n\}$, are made from a population having probability density $f(z|\boldsymbol{\theta})$, in which $\boldsymbol{\theta}$ is a vector of unknown parameters. From these observations we wish to draw conclusions about some hypothesis, H_0. For example, we may be testing the puncture resistance of specimens of filter fabric and recording the number defects. In this case, the z_i are ones (defects) or zeros (no defect), and the relevant statistics are the number of defective specimens $m = \Sigma z_i$ and the total number n. The process is $f(z|\boldsymbol{\theta})$ and is Binomial, with θ the fraction of failures.

We wish to test the hypothesis that this rate of defects has a certain value, say the rate experienced over the project to date, against an alternative hypothesis that the rate is something different. Thus, the hypotheses are

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned} \tag{4.68}$$

Presuming no prior knowledge, we set the prior probabilities of the two hypotheses equal; that is, $Pr(H_0) = Pr(H_1) = 1/2$. We then apply Bayes' Theorem to calculate a posterior probability on the null hypothesis, $Pr(H_0|\mathbf{z})$.

The likelihood of θ_0 given $\mathbf{z}$ is simply $L(\theta_0|\mathbf{z}) = f(\mathbf{z}|\theta_0)$, is the Binomial pmf. What, however, is the likelihood when $\theta \neq \theta_0$? It depends on the value of θ_0. Thus, we have to

[5] Typically, the extremes of the scoring rule are set as $\ell_1(1.0) = 0$ and $\ell_2(0) = 0$; and the $\ell_1(\pi)$ is taken to be a continuous decreasing function of π, while $\ell_2(\pi)$ is taken to be a continuous increasing function of π. If the subject's actual probability of θ_1 is p, the expected score is

$$Eu(\pi, \theta) = p\ell_1(\pi) + (1 - p)\ell_2(\pi)$$

We wish this to be maximized for $\pi = p$, thus if we differentiate with respect to π and recognize that the maximum is desired to occur for $\pi = p$, the following condition results:

$$\pi\ell_1'(\pi) + (1 - \pi)\ell_2'(\pi) = 0$$

Any scoring rule that satisfies this condition for π between zero and one is said to be a *proper scoring rule*, one example of which is the quadratic loss function:

$$u(\pi, \theta) = \begin{cases} \ell_1(\pi) = (1 - \pi)^2 \\ \ell_2(\pi) = \pi^2 \end{cases}$$

which is why this function is so common in subjective assessment.

specify some pdf, $f_{\theta \neq \theta_0}(\theta)$, over the value θ might have when it is not θ_0. The obvious assumption is to say that it can be anything, that is, that it has a uniform pdf, or $f_{\theta \neq \theta_0}(\theta) = k$, where k is a constant. Then, by Bayes' Theorem,

$$\Pr(H_0|\mathbf{z}) = \frac{\Pr(H_0)\, f(\mathbf{z}|\theta_0)}{\Pr(H_0)\, f(\mathbf{z}|\theta_0) + \Pr(H_1) \int\limits_{\theta \neq \theta_0} f(\theta)\, f(\mathbf{z}|\theta_0)\, d\theta} \tag{4.69}$$

which, after substitution becomes (Berger 1993),

$$\Pr(H_0|\mathbf{z}) = [1 + {\theta_0}^{-m}(1-\theta_0)^{m-n} B(m+1, n-m+1)]^{-1} \tag{4.70}$$

in which $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha, \beta)$ is the Beta function.

Consider the null hypothesis that the historical defect rate of 5% applies to the new filter fabrics as well. Presume that $n = 20$ tests are performed among which there are $m = 2$ defects. This is about twice the 5% defect rate. Testing the null hypothesis $H_0 : \theta = \theta_0$ against the alternate and calculating the posterior probability by Equation (4.68) leads to a posterior probability on H_0 of $Pr(H_0|m = 2, n = 20) = 0.83$. So, the evidence leads us to assign a posterior probability of more than 0.8 to the hypothesis of the historical defect rate of 5% even though the observed rate among 20 specimens was twice that. Had $n = 100$ specimens been tested with a failure rate of 10%, the posterior probability on the null hypothesis would drop to about 0.75. The reason these posterior probabilities are still high is that the sample sizes are statistically relatively small.

4.6.4 Altering the posterior probability of an hypothesis for non-uniform prior

The calculation leading to the above results for the posterior probability of a null hypothesis H_0 given the evidence from observations was predicated on a non-informative prior, that is, $Pr(H_0) = Pr(H_1) = 1/2$. It is relatively easy to adjust this value for a non-uniform prior, say, the prior subjective probability of another observer, simply by rearranging Bayes' Theorem. If $\Pr_{\text{non-informative}}(H_0|\mathbf{z})$ is the posterior probability of the null hypothesis calculated using a non-informative prior, and $\Pr_{\text{subjective}}(H_0)$ is a different prior probability on H_0, then the following adjustment provides the posterior probability of the null hypothesis associated with the different prior:

$$\Pr(H_0|\mathbf{z}) = \frac{1}{\left[1 + \left(\dfrac{1}{\Pr_{\text{subjective}}(H_0)} - 1\right)\left(\dfrac{1}{\Pr_{\text{non-informative}}(H_0|\mathbf{z})} - 1\right)\right]} \tag{4.71}$$

For example, a skeptical observer might conclude *a priori* that the probability was only 0.2 (one in five) that the historical defect rate applies to the new fabric. This observer's posterior probability would then be about 0.55. Conversely, a willing observer might conclude *a priori* that the probability was as much as 0.75 (three in four) and that the historical defect rate applies to the new fabric. This observer's posterior probability would then be about 0.93.

4.7 Choice Among Models

Another common problem in hypothesis testing is the comparison among models. The models might be, say, alternative probability distribution shapes for a set of data (i.e. are the data better fit by a logNormal or a Gamma pdf?), or the models might be alternative geomechanics models (e.g. different constitutive relations).

As a generalization, suppose there are k alternative models, $\{M_1, \ldots, M_k\}$. For each model there is some density function of the set of observations, $f_i(\mathbf{z}|\boldsymbol{\theta}_i)$, but with unknown parameters, $\boldsymbol{\theta}_i$. The subscript refers to the model. The *a priori* density function over the parameters is, $f(\boldsymbol{\theta})$, which could be non-informative (i.e. uniform in some sense). Then, for any given model, the marginal pdf of the observations, $\mathbf{z}$, over the possible values of the parameters, is

$$f_i(\mathbf{z}) = \int f(\theta_i) f(\mathbf{z}|\theta_i)\, d\theta_i \tag{4.72}$$

and the *a posteriori* pdf of the parameters for each model, given the data is

$$f_i(\theta_i|\mathbf{z}) = \frac{f(\theta_i) f(\mathbf{z}|\theta_i)}{\int f(\theta_i) f(\mathbf{z}|\theta_i)\, d\theta_i} \tag{4.73}$$

If the *a priori* probabilities of the models, respectively, are $Pr(M_i)$; then applying Bayes' Theorem, the *a posteriori* probabilities, given the data, $\mathbf{z}$, are

$$\Pr(M_i|\mathbf{z}) = \frac{\Pr(M_i) f_i(\mathbf{z})}{\sum_i \Pr(M_i) f_i(\mathbf{z})} \tag{4.74}$$

Consider the hypothesis test that a set of n data is modeled with a Normal distribution *vs.* a Uniform distribution. Berger (2001) shows that the marginal pdf's of the observations for these two distributions (among others), are

$$f(\mathbf{z}|M_N) = \frac{\Gamma[(n-1)/2]}{2\pi^{(n-1)/2}\sqrt{n}\left[\sum_j (z_j - \overline{z})^2\right]^{(n-1)/2}} \tag{4.75}$$

$$f(\mathbf{z}|M_U) = \frac{1}{n(n-1)[z_{\max} - z_{\min}]^{(n-1)}} \tag{4.76}$$

in which the subscripts, M and U, refer to the Normal and Uniform distributions.

As an example, $n = 31$ data points were generated from a [0, 1] Uniform random number generator by adding sets of four values each. By the Central Limit Theorem, as the number of variables in the sum increases, the pdf should approach Normality. On the other hand, the unsummed variables are by definition Uniform. The resulting statistics were: $n = 31$, $\overline{z} = 2.02$, $\Sigma(z_j - \overline{z})^2 = 3.10$, $z_{\max} = 2.87$, and $z_{\min} = 1.06$. Thus, the respective marginal pdf's become, $f(\mathbf{z}|M_N) = 7.16\text{E-}10$ and $f(\mathbf{z}|M_U) = 2.00\text{E-}11$. The *a posteriori* probability for the Normal hypothesis by Equation (4.82) is then, $Pr(M_N) = 0.97$. Repeating the experiment for sums of two uniform variables led to $n = 31$, $n = 31$, $\overline{z} = 1.08$,

$\Sigma(z_j - \overline{z})^2 = 4.60$, $z_{\max} = 1.74$, and $z_{\min} = 0.16$. For this case, the posterior probability for the Normal hypothesis was reduced to $Pr(M_N) = 0.60$. For the case of one Uniform variable (by definition, generated from a Uniform pdf), the corresponding statistics were, $n = 31$, $\overline{z} = 0.39$, $\Sigma(z_j - \overline{z})^2 = 2.12$, $z_{\max} = 0.96$, and $z_{\min} = 0.03$; and the corresponding *a posteriori* probability for the Normal hypothesis was, $Pr(M_N) = 2 \times 10^{-5} \rightarrow 0$, all of which is to be expected.

5 Risk, Decisions and Judgment

What do we mean by risk, and how can decisions be made better by attempting to quantify the probabilities and potential consequences of undesired events? This chapter begins by examining the potential hazards that people face every day and the various ways that risk is defined. Attention then turns to how knowing something about the probabilities of potential outcomes and magnitudes of consequences can be used to help us make decisions more systematically and thoughtfully. Finally, the chapter looks at some of the strategies that engineers use to make decisions, and to the question of what is engineering judgment.

5.1 Risk

In the course of daily life, as well as engineering practice, we routinely encounter situations that involve some event that might occur and that, if it did, would bring with it some adverse consequence. We might be able to assign probability to the occurrence of the event and some quantified magnitude or cost to the adversity associated with its occurrence. This combination of uncertain event and adverse consequence is the determinant of risk:

$$Risk = (probability, consequence) = (p, c) \tag{5.1}$$

In common usage, the word *risk* has a variety of meanings.[1] In engineering practice the definition is usually more narrow. Kaplan and Garrick (1981) write that, to assess risk, three things need to be defined: a scenario, a range of consequences, and a probability of the event's leading to the consequences. Building on this, Bedford and Cooke (2001)

[1] **risk** *n*. 1. The possibility of suffering harm or loss; danger. 2. Factor, thing, element, or course involving uncertain danger; a hazard: 3(a) The danger or probability of loss to an insurer. (b) The amount that an insurance company stands to lose. 4(a) The variability of returns from an investment. (b) The chance of nonpayment of a debt. 5. One considered with respect to the possibility of loss (Merriam-Webster 1998).

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

write that risk analysis attempts to answer three questions: What can happen? How likely is it to happen? Given that it occurs, what are the consequences?

Thus, in engineering, risk is usually defined as comprising:

- A set of scenarios (or events), $E_i, i = 1, \ldots, n$;
- Probabilities associated with each element, p_i; and
- Consequences associated with each element, c_i.

The quantitative measure of this risk might be defined in a number of ways, and, indeed, different industries quantify risk in different ways. For example, in the insurance industry, *risk* is commonly defined as the monetary value of the insured casualty (Bühlmann 1970). So, if an insurance company writes a earthquake damage policy to insure a building against, say, total loss in the event of an earthquake, then the risk is the total value of the insured building:

$$Risk = (consequence) = (c) \tag{5.2}$$

In public health, *risk* is commonly defined as the probability that an adverse effect will occur due to some agent or activity (Alaszewski *et al.* 1998). So, if a large number of people are exposed to some pathogen, the public health risk is the fraction that become sick or are in some way adversely affected:

$$Risk = (probability) = (p) \tag{5.3}$$

In engineering contexts, *risk* is commonly defined as the product of probability and consequence, or expressed another way, risk is taken as the expectation of adverse outcome:

$$Risk = (probability \times consequence) = (pc) \tag{5.4}$$

This is the way we will use the term *risk*. When more than one event may lead to an adverse outcome, Equation (5.4) is extended to be the expectation of consequence over that set of events:

$$Risk = \sum\nolimits_i p_i c_i \tag{5.5}$$

in which p_i is the probability of consequence c_i.

5.1.1 Acceptable risks

In engineering, as in other aspects of life, lower risk usually means higher cost. Thus, we are faced with the question, "how safe is safe enough," or "what risk is acceptable?"

Starr and Whipple (1980) note, "implicit in the term, acceptable risk is, 'acceptable to whom?"' Governmental approval is usually required to establish the legitimacy of methods for making risk analysis decisions, but public consensus is needed to make them satisfactory. In the United States, the government acting through Congress has not defined acceptable levels of risk for civil infrastructure, or indeed for most regulated activities. The setting of 'reasonable' risk levels – or at least the prohibition of 'unreasonable' risks – is left up to regulatory agencies, such as the Environmental Protection Agency, Nuclear Regulatory Commission, or Federal Energy Regulatory Commission. The procedures these

regulatory agencies use to separate reasonable from unreasonable risks vary from highly analytical to qualitatively procedural.

People face individual risks to health and safety every day, from the risk of catching a dread disease, to the risk of being seriously injured in a car crash (Table 5.1). Society faces risks that large numbers of individuals are injured or killed in major catastrophes (Table 5.2). We face financial risks every day, too, from the calamities mentioned to the risk of losses or gains in investments. Intuition does not necessarily yield accurate estimates of these risks (Figure 5.1).

Some risks we take on voluntarily, like participating in sports or driving an automobile. Others we are exposed to involuntarily, like a dam failing upstream of our home or disease

Table 5.1 Average risk of death to an individual from various human-caused and natural accidents (US Nuclear Regulatory Commission 1975)

Accident type	Total number	Individual chance per year
Motor Vehicle	55,791	1 in 4,000
Falls	17,827	1 in 10,000
Fires and Hot Substances	7,451	1 in 25,000
Drowning	6,181	1 in 30,000
Firearms	2,309	1 in 100,000
Air Travel	1,778	1 in 100,000
Falling Objects	1,271	1 in 160,000
Electrocution	1,148	1 in 160,000
Lightning	160	1 in 2,500,000
Tornadoes	91	1 in 2,500,000
Hurricanes	93	1 in 2,500,000
All Accidents	111,992	1 in 1,600

Table 5.2 Average risk to society of multiple injuries or deaths from various human-caused and natural accidents (US Nuclear Regulatory Commission 1975)

Type of event	Probability of 100 or more fatalities	Probability of 1000 or more
Human-Caused		
Airplane Crash	1 in 2 yrs.	1 in 2000 yrs.
Fire	1 in 7 yrs.	1 in 200 yrs.
Explosion	1 in 16 yrs.	1 in 120 yrs.
Tonic Gas	1 in 100 yrs.	1 in 1000 yrs.
Natural		
Tornado	1 in 5 yrs.	very small
Hurricane	1 in 5 yrs.	1 in 25 yrs.
Earthquake	1 in 20 yrs.	1 in 50 yrs.
Meteorite Impact	1 in 100,000 yrs.	1 in 1 million yrs.

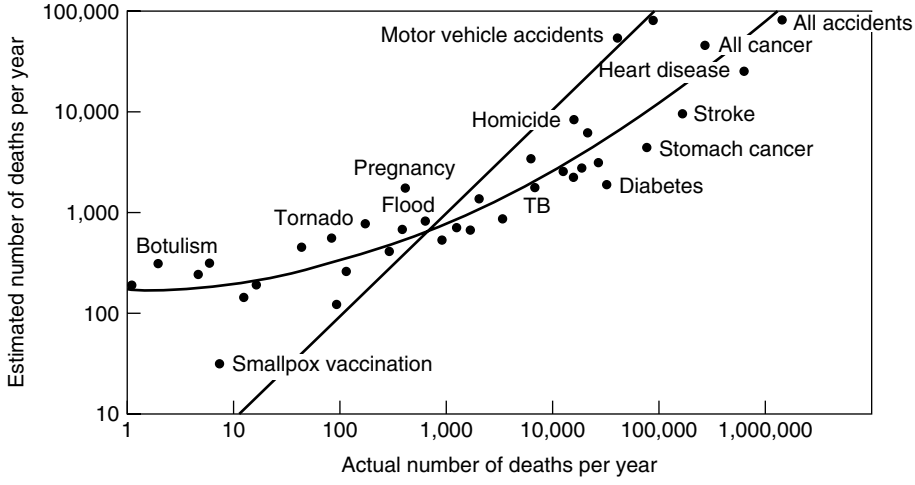


Figure 5.1 Comparison of perceived with actual risk for common hazards and diseases (Lichtenstein *et al*. 1982).

due to air pollution. Just being alive in the US or Europe carries a risk of dying of about 1.5×10^{-6} per hour. Some risk analysts consider this number, about 10^{-6}, a baseline to which other risks might be compared (Stern *et al*. 1996).

For voluntary risks, the individual uses his or her preferences to decide whether or not to accept the risk. It has long been observed, for example, that wage rates in dangerous occupations tend to rise with level of risk (Figure 5.2). In mining, accident rates are roughly proportional to the third-power of wages (Starr 1969). This balancing of personal risk and benefits may not be quantitative, but the implicit trade-offs people make have been back-calculated to obtain Figure 5.3. Risk-benefit rates for voluntary risks in the US are somewhat higher than 10^{-6} per hour, and for involuntary risks, roughly 1000 times lower.

Starr arrives at four conclusions about acceptable risk: (1) the public is willing to accept 'voluntary' risks roughly 1000 times greater than 'involuntary' risks; (2) statistical risk of death from disease appears to be a psychological yardstick for establishing the level of acceptability of other risks; (3) the acceptability of risk appears to be proportional to the third-power of the benefits; and (4) the societal acceptance of risk is influenced by public awareness of the benefits of an activity, as determined by advertising, usefulness, and the number of people participating. The exactness of these conclusions has been criticized (Otway and Cohen 1975), but the insight that acceptable risk exhibits regularities is important.

5.1.2 Risk perception

People view risks not only by whether those risks are voluntary or involuntary, or by whether the associated benefits outweigh the dangers, but also along other dimensions (Table 5.3). Over the past twenty years, researchers have attempted to determine how average citizens perceive technological risks.

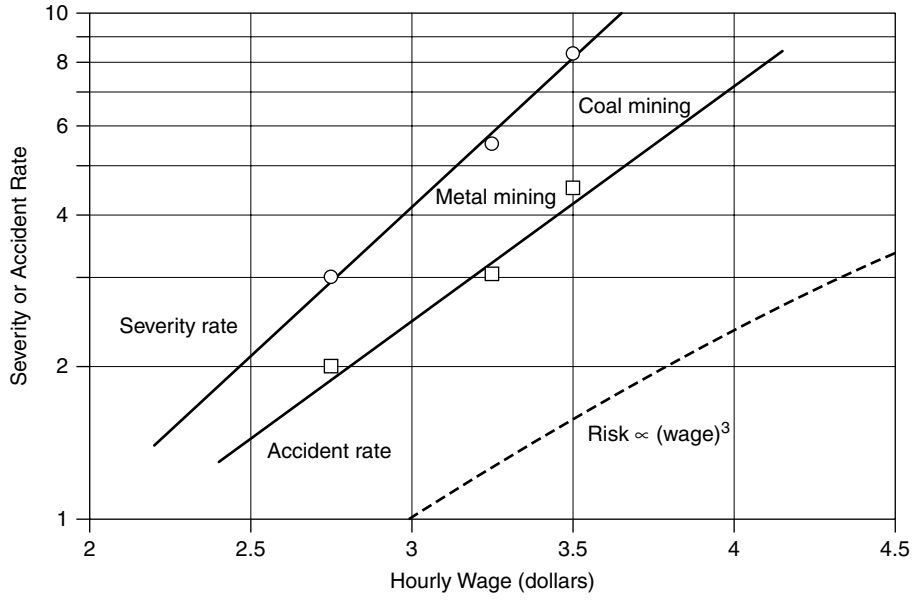


Figure 5.2 Mining accident and accident severity rates plotted against wage incentive. (Reprinted with permission from Starr, C. 1969 'Social benefit versus technological risk.' *Science* **165**: 1232–1238. Copyright 1969 American Association for the Advancement of Science.) As noted by the author, the relation between accident or accident severity rates and wage compensation is roughly proportional to the third power of the wage.

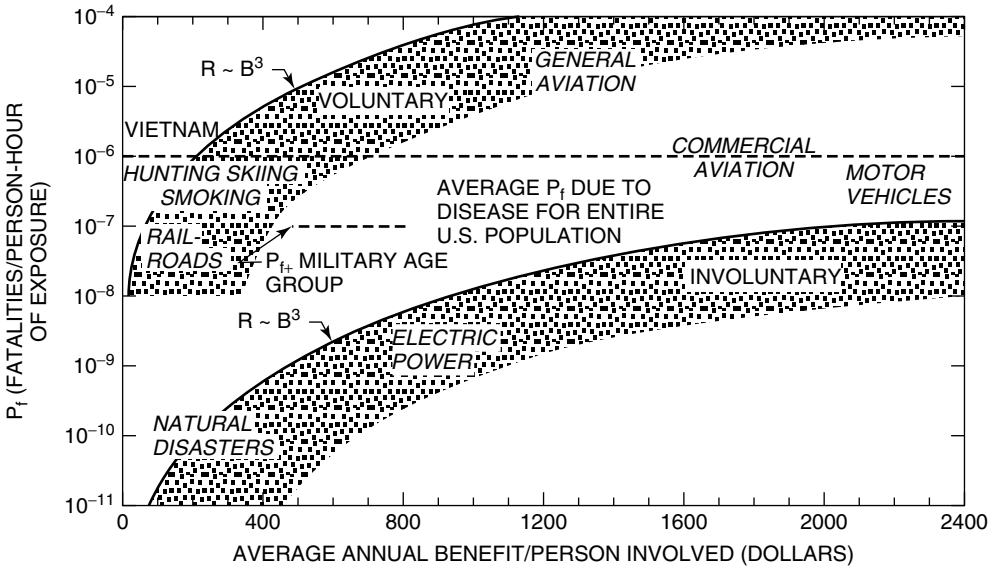


Figure 5.3 Risk (R) plotted against benefit (B) for various kinds of voluntary and involuntary exposures. (Reprinted with permission from Starr, C. 1969 'Social benefit versus technological risk.' *Science* **165**: 1232–1238. Copyright 1969 American Association for the Advancement of Science.)

Table 5.3 Separation of risk perceptions along two factor dimensions

Factor I: controllable vs. uncontrollable	
Controllable	Uncontrollable
Not dread	Dread
Local	Global
Consequences not fatal	Consequences fatal
Equitable	Not equitable
Individual	Catastrophic
Low risk to future generations	High risk to future generations
Easily reduced	Not easily reduced
Risk decreasing	Risk increasing
Voluntary	Involuntary
Factor II: observable vs. unobservable	
Observable	Unobservable
Known to those exposed	Unknown to those exposed
Effect immediate	Effect delayed
Old risk	New risk
Risk known to science	Risk unknown to science

A common approach has been to use multivariate analysis to reveal an implied structure of preferences over various risks. An interesting finding is that, whereas risk analysts tend to categorize risk by annual fatalities or economic impacts, normal people tend to categorize risk by the potential for catastrophe, controllability, threat to future generations, familiarity, equity, level of understanding of the risk, and a host of other less quantitative factors.

Various models have been proposed to represent these findings. For example, Figure 5.4 suggests a separation along the factors of dread or uncontrollability of the risk and of unknown-ness or unobservability. Better understanding of the way people perceive risk may help in planning projects and in communication. The public's perception of risk is arguably more subtle than the engineer's, and quickly becomes political (Douglas and Wildavsky 1982).

5.1.3 F-N Charts

An approach to describing risk popularized by Wash 1400 (USNRC 1975) – although Farmer had earlier used the approach in the UK (Bedford and Cooke 2001) – is to plot the exceedance probability of risks against their associated consequence (Figure 5.5). Mathematically, these are complementary cumulative probability density functions:

$$CCDF_x(x_0) = 1 - F_x(x_0) = \Pr[x > x_0] \tag{5.6}$$

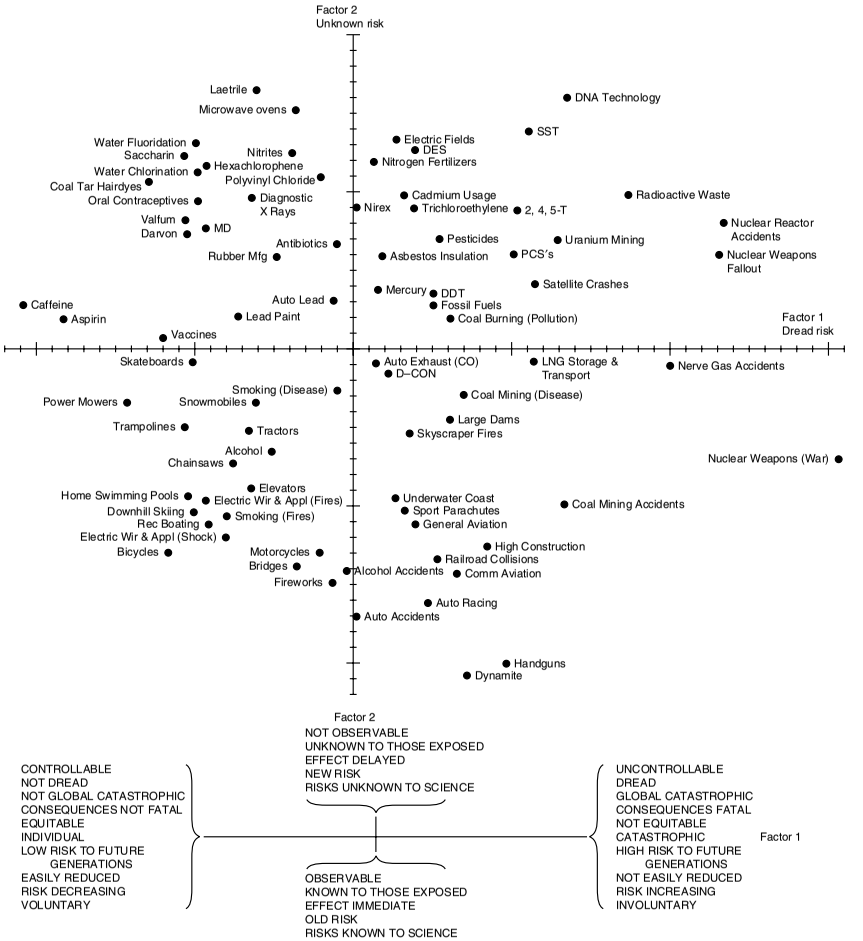


Figure 5.4 Factor analysis of 81 hazards plotted against 'controllable: uncontrollable' and 'observable: not-observable'. (Reprinted with permission from Starr, C. 1969 'Social benefit versus technological risk.' *Science* **165**: 1232–1238. Copyright 1969 American Association for the Advancement of Science.)

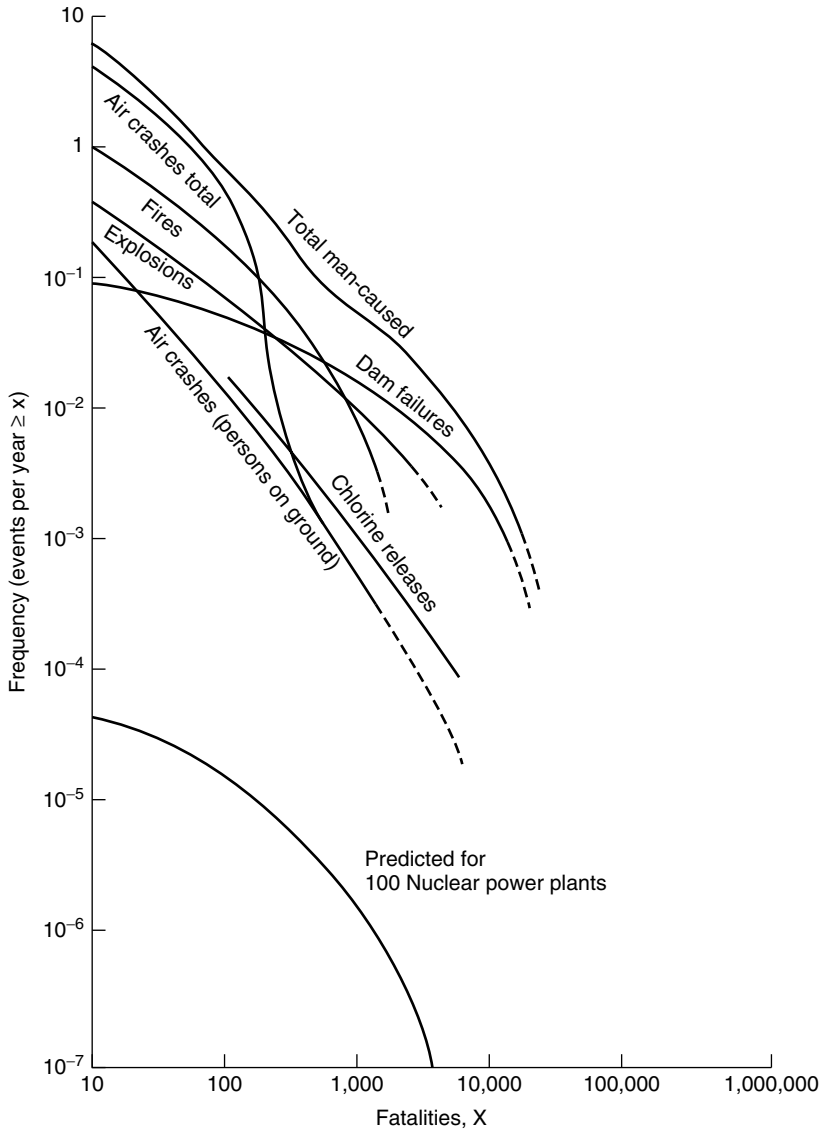


Figure 5.5 Annual risk in the US due to a variety of natural and man-made hazards (US Nuclear Regulatory Commission 1975).

in which $F_x(x_0)$ is the cumulative density function (CDF). Figure 5.6 is a similar plot showing fatalities caused by dam failures.

Such curves are sometimes referred to as *F-N curves*, in that they summarize the relation between frequencies and number of lives lost, or some other undesired consequence. They have become a convenient way for comparing risk associated with different facilities or with different design alternatives. Chapter 22 discusses a project for which F-N charts were used to compare seismic hazards facing a refinery to averaged risks facing other civil works (Figure 5.7).

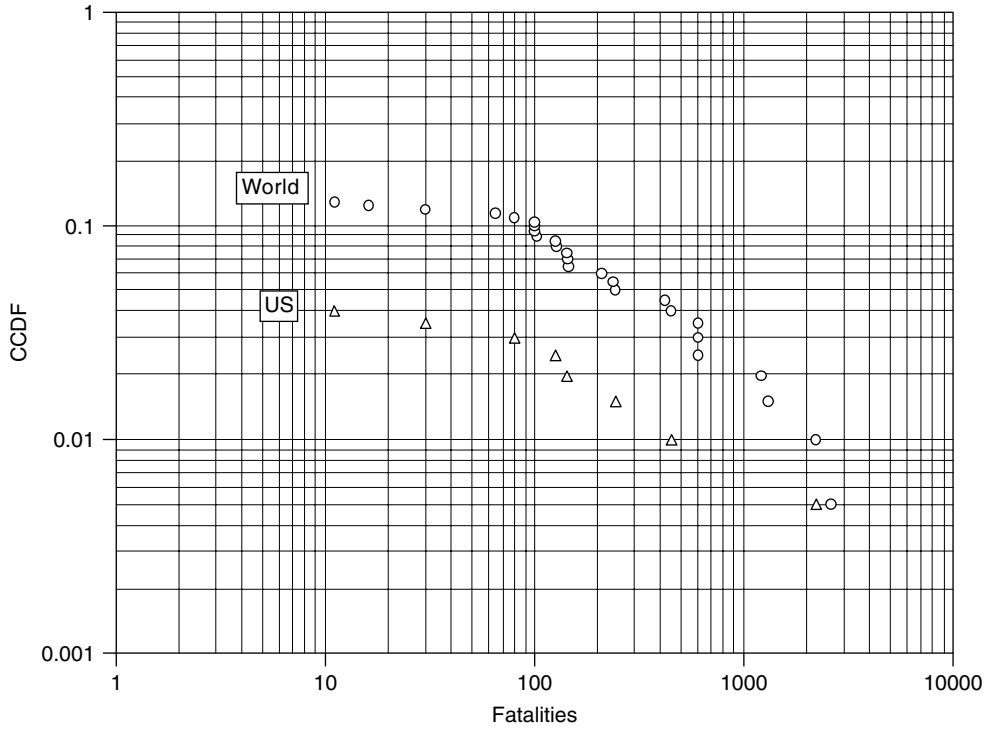


Figure 5.6 Historical frequency of fatalities due to dam failures from 1800 through 2000 (Baecher and Christian, unpublished manuscript).

European government safety agencies have been forward thinking in developing quantitative risk guidelines for public works. In the UK, the Health and Safety Executive published guidelines for acceptable risk as early as 1982 (HSE 1982). These were published as quantified guidelines in the early 1990s (HSE 1991; HSE 1992) and were later modified and adopted by the Hong Kong Government Planning Department (1994), as shown in Figure 5.8. Similarly, the Dutch government published risk guidelines (Figure 5.9) in the late 1980s (Versteeg 1987).

The slope of the lines dividing regions of acceptability expresses a policy decision between the relative acceptability of low probability/high consequence risks and high probability/low consequence risks. The steeper the boundary lines, the more averse is the policy to the former. Note that, the boundary lines in the Hong Kong guidelines are twice as steep (in log-log space) as the slopes in the Dutch case. Note also that, in the Hong Kong case there is an absolute upper bound of 1000 on the number of deaths, no matter how low the corresponding probability.

Decision theorists have noted that the use of F-N charts is not without inconsistencies. Specifically, one can conceive of two individual risks, each of which is acceptable under a guideline, yet when the two are probabilistically combined, the combination is not (Bedford and Cooke 2001; Evans and Verlander 1997; Vrijling *et al.* 1995). In geotechnical practice, however, the guidance derived from F-N curves is usually viewed as qualitative.

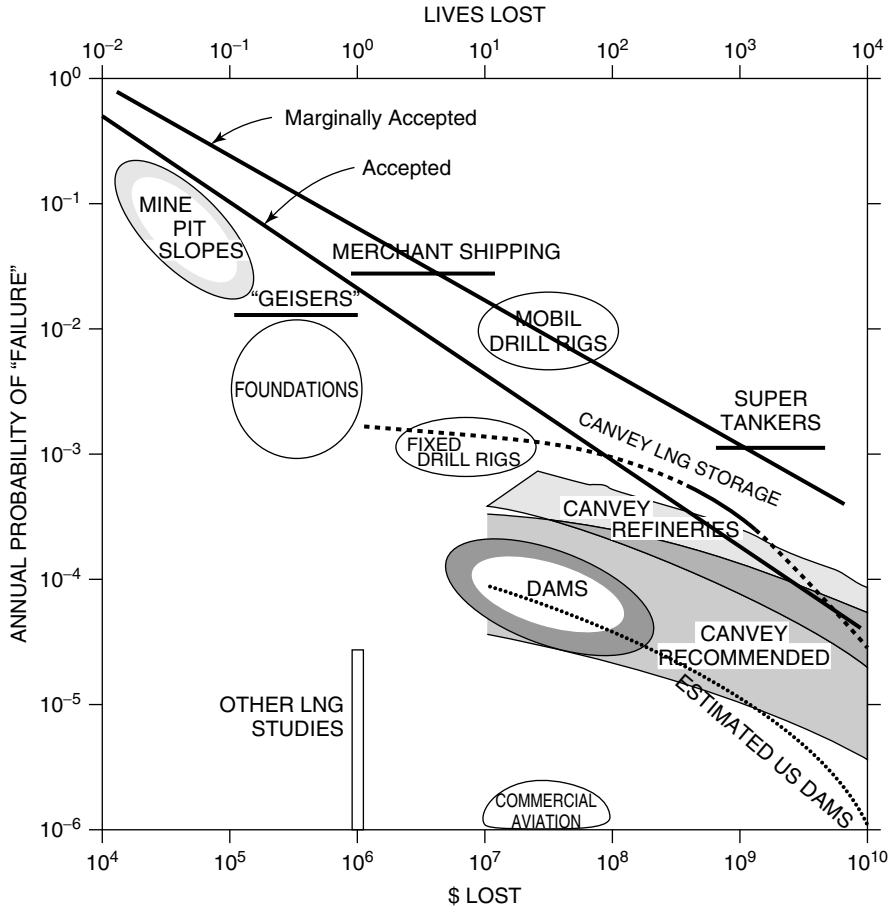


Figure 5.7 F-N chart showing average annual risks posed by a variety of traditional civil facilities and other large structures or projects (Baecher 1982b).

In 1993, British Columbia Hydro and Power Authority began a program to develop risk criteria for flood and earthquake safety of its dams (Salmon 1998; Salmon and Hartford 1995a; Salmon and Hartford 1995b; Salmon and von Ehn 1993). At the time, the company operated 49 dams, including large structures at Mica (240 m), Revelstoke (175 m), and WAC Bennett (185 m) (Vick and Stewart 1996). In pursuit of this goal, company engineers called upon the emerging development of F-N curves by the European safety agencies for use as standards in setting facility risks. ANCOLD followed suit and published quantitative risk guidelines in 1994 (Figure 5.10). The US Bureau of Reclamation has used F-N charts in comparing safety risks within a portfolio of dams (Figure 5.11).

5.2 Optimizing Decisions

Rather than judge risk by comparing probabilities and consequences to simple thresholds, we can also identify best decisions by optimizing over some objective function. For

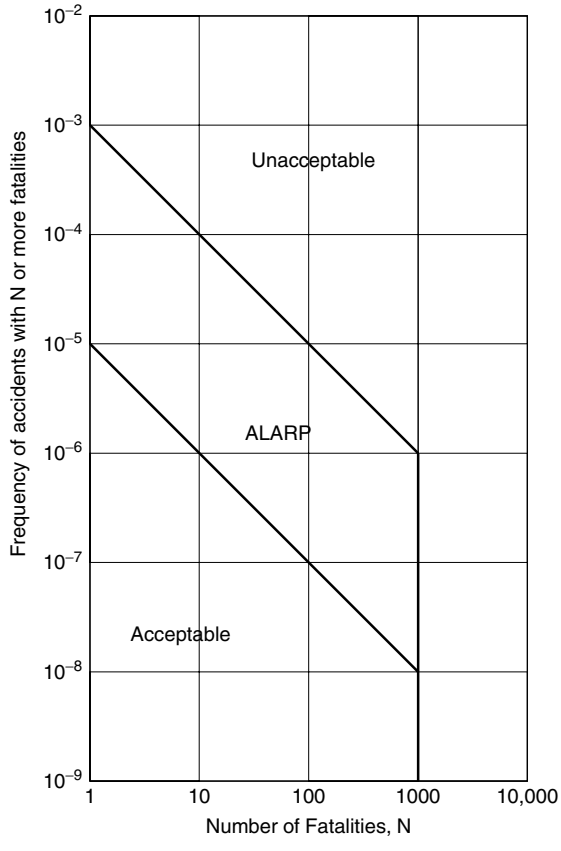


Figure 5.8 Acceptable, societal risk guidelines for Hong Kong Planning Department.

example, we could pick the risky alternative that minimized expected monetary loss or some other function.

5.2.1 Decision trees

Consider the decision whether to use spread footings or driven piles for the abutment of a bridge. The advantage of footings is cost. The disadvantage is risk. The footings may settle excessively, necessitating repairs. Thus, while footings are less expensive to begin, they may be more expensive in the long run. Table 5.4 gives the engineer's estimates of these costs and probabilities. Which alternative is better?

A *decision tree* (Figure 5.12) is a graph for organizing what we know about a decision (Raiffa 1968). The choice is between adopting footings *vs.* a deep foundation, shown as a *decision (box) node* at left. From that node emanate two decision *branches*. If we use footings, the cost for design and construction is \$1 million, shown as a *gate*. If we use piles, the cost is \$3 million.

If we use footings, either the footings perform satisfactorily or they do not (i.e. they settle excessively). These uncertain events emanate from a *chance (circle) node*, each having some associated probability. Say, the chance the footings settle excessively is

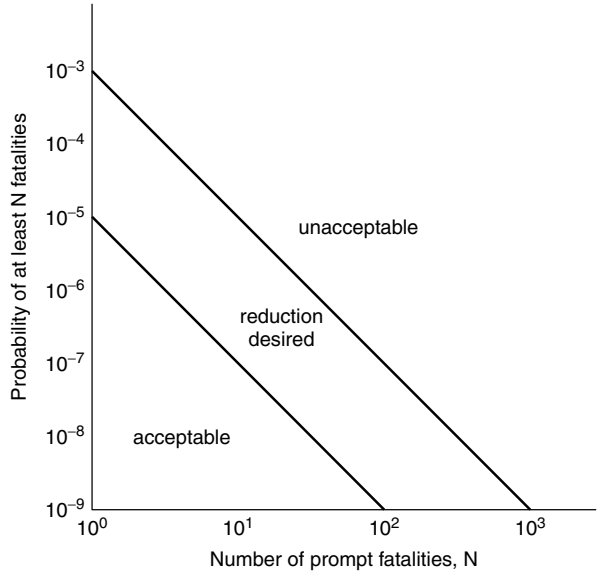


Figure 5.9 Netherlands government group risk criterion (Versteeg 1987).

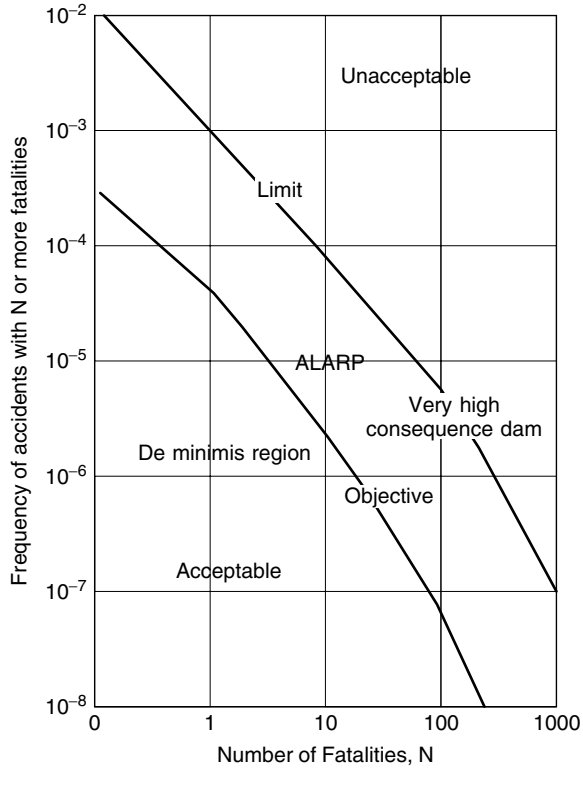


Figure 5.10 Risk guidelines from ANCOLD (1994).

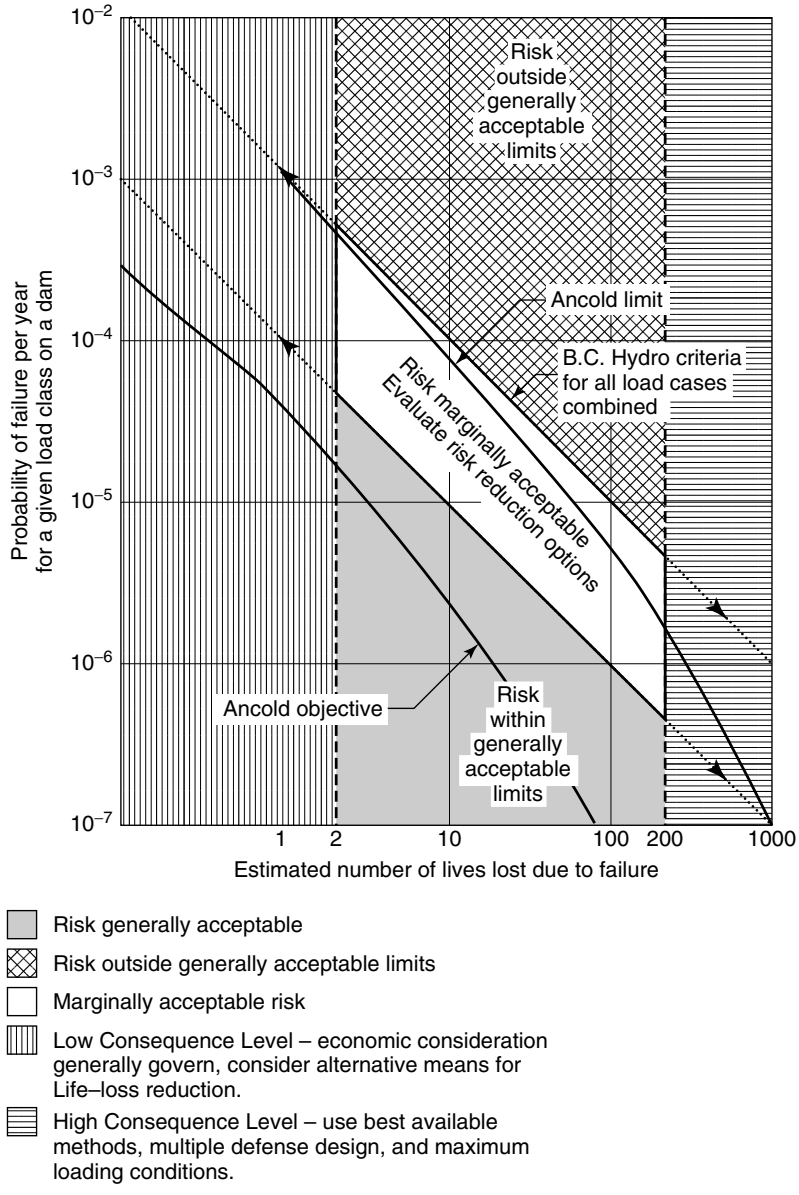


Figure 5.11 F-N chart used by US Bureau of Reclamation on Nambe Falls Dam, New Mexico, to compare risks faced by the existing structure with recommended limits. (Von Thun, J. L., 1996, 'Risk assessment of Nambe Falls Dam,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, Madison, WI, pp. 604-635, reproduced by permission of the American Society of Civil Engineers.)

assessed to be 0.2, and the chance they do not is 0.8. If the footings do settle excessively, the total cost to the project is the initial cost plus the cost of repair, or, \$1 m + \$5 m = \$6 m. On the other hand, if the footings perform satisfactorily, the total cost to the project is only the initial cost, \$1 m.

Table 5.4 Probability cost matrix for shallow and deep foundations under a bridge structure

Foundation type	Initial cost	Probability of large settlement	Possible repair costs
Shallow footings	\$1 million	0.2	\$5 million
Driven piles	\$3 million	0.01	\$3 million

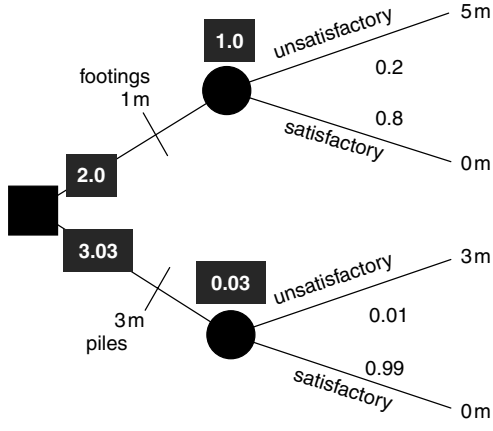


Figure 5.12 Simple decision tree for choice of bridge foundation.

The expected cost is the total cost of each possible outcome weighted by its respective probability. For excessive settlement, this is $(0.2)\$6\,\text{m} + (0.8)\$1\,\text{m} = \$2.0\,\text{m}$. A similar calculation could be made for the pile alternative, which yields, $(0.01)\$6\,\text{m} + (0.99)\$3\,\text{m} = \$3.03\,\text{m}$. So, from an expected cost basis, the footings are superior, because their expected cost (shown above the corresponding node) is lower.

Say the decision is made to use footings but then the soil conditions turn out to be unfavorable, settlement is excessive, and repairs must be made (Table 5.5). The total

Table 5.5 Outcomes of foundation decision

Shallow Foundations		
Soil Conditions	Probability	Cost
'good'	0.8	\$1 m
'poor'	0.2	\$6 m
Deep Foundations		
Construction Conditions	Probability	Cost
'good'	0.01	\$3 m
'poor'	0.99	\$6 m

cost actually turns out to be \$4 m, which is more than the piles would have cost. Note that neither of the potential outcomes of selecting footings leads to the expected cost: the favorable outcome is less, and the unfavorable outcome is more. Was this a 'bad' decision? No, at least not given the objective. Given the state of information at the time the decision was made, footings were the best decision. While the outcome was unfavorable, the decision was the right one.

5.2.2 Expected value of perfect information (EVPI)

Faced with the uncertainty in this example, about whether the soil conditions are good enough to use footings, one's reaction might be not to make a decision immediately, but to gather more information first. *In situ* load tests cost \$50,000, but let us presume that they provide essentially perfect information on whether the soil conditions are suitable for footings.

Without the tests, the analysis leads us to use footings. Thus, if the tests confirm that the soil conditions are good, the decision to use footings would not change. The money spent on the tests would have been wasted because the decision would be the same. On the other hand, if the tests indicate that the soil conditions are poor and that footings would settle excessively, then the decision would be to use piles. We would have avoided the excess cost of repairs, and saved \$2 m (less the cost of the tests). The tests would be a real bargain, even at \$50,000.

The decision tree can be expanded to include this new decision, whether to perform load tests (Figure 5.13). This decision to test or not has the same functional form as any other decision and is represented in the tree as a decision node. Since we make this decision about testing – and find out the result – prior to making the decision about foundation type, its node appears first in the tree. If we decide to test, we incur an up-front cost of \$50,000, then we find out whether the soil conditions are 'good' or 'poor.' Thus, if the outcome of the test is 'good' soil conditions, the probability that footings would settle excessively and lead to the need for repairs becomes zero, and they are the best choice, with a corresponding cost of \$1 m. Conversely, if the outcome is 'poor' soil conditions, the probability that footings would settle excessively and lead to the need for repairs becomes one, and piles are the best choice, with a corresponding cost of \$3 m. Given a 0.8 probability that the load test will indicate 'good' soil conditions, the expected cost is, $(0.8)\$1\,\text{m} + (0.2)\$3.03\,\text{m} = \$1.4\,\text{m}$. This means that, the perfect information provided by the test lowers the overall expected cost from \$2.0 m to \$1.4 m, and thus the *expected value of perfect information* (EVPI) is \$0.6 m. That is, the value of obtaining perfect information is \$0.6 m, and any cost up to that amount would be reasonable to pay to obtain the information.

This can also be argued in a reverse way, minimizing the *regret* that would be suffered by using a less good foundation scheme. There is an 0.8 chance that the soil conditions are 'good,' in which case the results of the load test do not alter the choice of foundation type. On the other hand, there is an 0.2 chance that the soil conditions are 'poor,' in which case the results of the load test do alter the choice of foundation type, and rather than adopting shallow foundations and then incurring \$5 m of repairs, the pile foundation is used and \$3.0 m is saved. In this case the expected regret would be, $(0.8)\$0\,\text{m} + (0.2)\$3\,\text{m} = \$0.6\,\text{m}$, which is the same as the difference between the best choice with and without perfect information, weighted by the probability of the two states of nature.

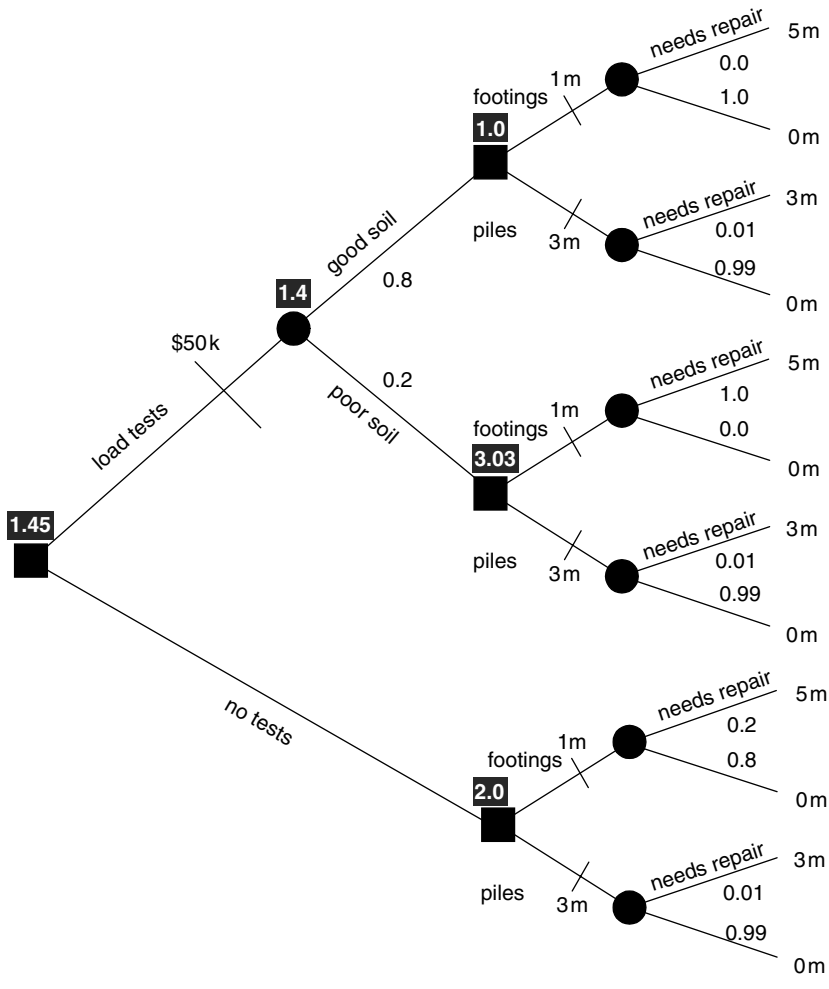


Figure 5.13 Decision tree including the option of performing a load test before deciding between footings or piles.

5.2.3 Expected value of sampled information (EVSI)

In most cases, of course, the information obtained from testing is not perfect. The results merely increase or decrease the probabilities associated with the possible states of nature. For example, rather than performing expensive load tests, we could elect to place less expensive, but simultaneously less diagnostic, borings. Table 5.6 shows an assessment of what the borings might show, given the actual soil conditions.

If the actual soil conditions are 'good,' the assessment is made that the borings might correctly show 'favorable' conditions with probability 0.85, but incorrectly show 'unfavorable' with probability 0.15. The latter is akin to a 'false-positive' in clinical testing. If the actual soil conditions were 'poor,' the corresponding probabilities of 'favorable' and 'unfavorable' are assessed as 0.3 and 0.7, respectively. The values at the LHS show

Table 5.6 Probabilities of possible outcomes of borings

		Actual soil conditions	
		'Good'	'Poor'
Outcome of borings	Pr	Pr = 0.8	Pr = 0.2
Favorable	0.74	0.85	0.3
Unfavorable	0.26	0.15	0.7

the marginal probabilities of the borings returning 'favorable' and 'unfavorable' results, respectively, given the conditional probabilities in the body of the table.

Given that the borings indicate 'favorable,' the (posterior) probability of 'good' soil conditions can be calculated by Bayes' Theorem (Chapter 2) as,

$$\Pr[good|favorable] = \frac{\Pr[good]L[favorable|good]}{\Pr[good]L[favorable|good]\Pr[poor]L[favorable|poor]} \tag{5.7}$$

Similarly the probability of 'good' conditions, given 'unfavorable' indications, can be calculated by Bayes' Theorem. These results are shown in Table 5.7, and the corresponding probabilities are entered into the decision tree of Figure 5.14. The *expected value of sample information* (EVSI) is the difference between the expected value of the decision using the information and that without, or EVSI = \$2.0 − \$1.82 = \$0.18 m.

So, the expected value of the borings, which lead not to perfect information but only to a change in the probabilities of 'good' and 'poor' soil conditions, is still positive, but smaller than the ESPI. Given that the borings cost but \$10,000, they are a good value compared to the expected \$180,000 they save.

This is an exceptionally simple example. More complex decisions involve more complex trees, but the calculation of these trees is straightforward. Examples in the geotechnical literature are many (Angulo and Tang 1996; Baecher 1984a; Einstein and Baecher 1983; Einstein *et al*. 1978; Liao *et al*. 1996). Related examples in oil and gas exploration are also common (Grayson 1960; Harbaugh *et al*. 1977; Kaufman 1963; Newendrop 1975).

5.2.4 Utility theory

Expected monetary value may be a reasonable decision criterion for governments or large corporations, but if the potential consequences of decisions are large relative to the asset position of a decision maker, then other considerations enter the picture. The

Table 5.7 Probabilities of actual conditions given indications in borings

		Posterior probability of: 'good' soil conditions	Posterior probability of: 'poor' soil conditions
Outcome of borings	Pr	Prior Pr = 0.8	Prior Pr = 0.2
"Favorable"	0.74	0.92	0.08
"Unfavorable"	0.26	0.46	0.54

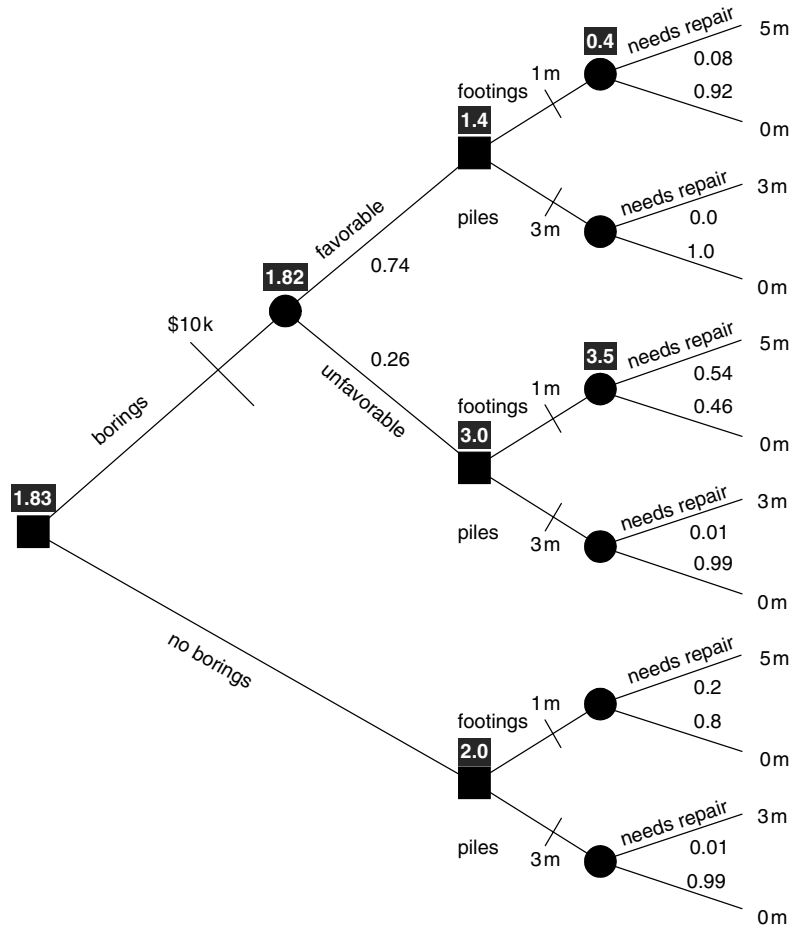


Figure 5.14 Decision tree including the option of placing borings before deciding between footings and piles.

Figure 5.15 Despite the higher EMV of the right-hand bet most people would choose the bet on the left.

most important of these is *risk aversion*, the propensity to treat large losses or gains as different from simply the multiple of so many smaller ones. That is, to an individual decision maker the loss of, say, $100,000 may be valued as different from – and usually greater than – simply one hundred times the loss of $1000. Most individual decision makers, and many corporate decision makers, are more than proportionally averse to large losses (Figure 5.15).

The presumption of *utility theory* is that no matter how complex or subtle an individual's preferences over the consequence of a decision, people do have such preferences, and their preferences are revealed through the decisions they make. Thus, one can posit a function, called a *utility function*, which is optimized in making decisions. Subject to assumptions on coherence, this utility function can be shown to enjoy well-behaved mathematical properties, as first suggested by Ramsey and Braithwaite (1931).[2]

Let C_1, C_2, and C_3 be three outcomes along an individual attribute scale. Let the preference ordering among these outcomes be

$$C_1 \succ C_2 \text{ and } C_2 \succ C_3 \tag{5.8}$$

where the symbol $\succ$ means, 'is preferred to.' By transitivity, $C_1 \succ C_3$. The utility function $u(C_i)$ will have the same ordering

$$u(C_1) > u(C_2) > u(C_3) \tag{5.9}$$

From a measurement theory point of view, this function $u(C_i)$ need only be invariant up to a positive linear transformation. That is, the scale and origin of the function can vary, and the function still serves its intended purpose. It is common practice to assign the utility of the least preferred outcome a value of zero and of the most preferred a value of one, but this is not necessary.

The *expected utility hypothesis* says that there exists a function $u(C_i)$ such that the expectation of the function over uncertainties in the consequences C_i is a logically consistent criterion for decisions. That is, a choice between the sure outcome C_2 and a lottery yielding C_1 with probability p and C_3 with probability (1-p) can be made on the basis of a comparison of $u(C_2)$ against the expected value, $\{pu(C_1) + (1 - p)u(C_3)\}$.

The question is, how can this utility function for an individual be inferred from past decisions? This can be done by reversing the normal order of thinking, asking not which lottery involving known probabilities of uncertain outcomes one prefers, but rather, what probability in such a lottery would make one change one's preference? Consider a monetary lottery offering a p chance of winning \$1000 *vs.* the complementary $(1 - p)$ chance of winning nothing. Set $u(0) = 0$ and $u(1000) = 1.0$. Then, the expected utility of the lottery is

$$Eu = (1 - p)u(0) + pu(1000) = p \tag{5.10}$$

What value of p would cause one to be indifferent between the lottery and a sure outcome of \$100? If p were close to 1.0, most people would take the lottery; but, if p were 0.1 or 0.2, or even 0.4, many people would take the sure \$100. Presume that the answer $p_0 = 0.25$ is the point of indifference. Figure 5.16 shows the beginnings of a utility function incorporating this datum,

$$\begin{aligned} u(100) &= (1 - p_0)u(0) + p_0u(1000) \\ &= (1 - 0.25)u(0) + (0.25)u(1000) = 0.25 \end{aligned} \tag{5.11}$$

[2] The formal concept of utility functions, and of the hypothesis of expected utility maximization, is treated in specialized books on the topic (Fishburn 1964; Fishburn 1970; Keeney and Raiffa 1976; De Groot 1970), and in general texts on statistical decision theory (Pratt *et al.* 1995).

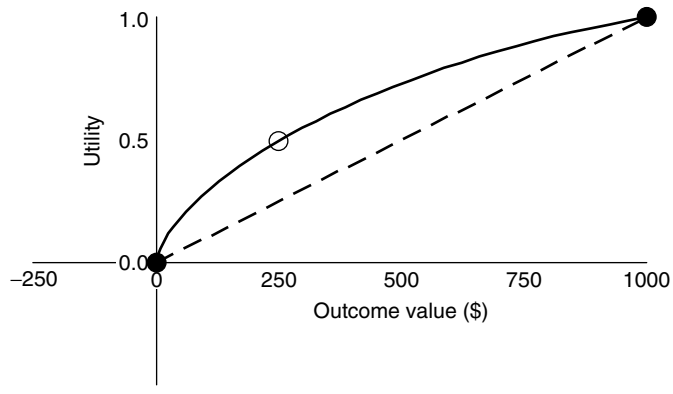


Figure 5.16 Uniattribute utility function over monetary outcome.

Given these relations, the value of $u(C_2)$ must lie between $u(C_1)$ and $u(C_3)$, which is satisfied for any p. This experimental question of finding the certainty equivalent for a lottery by setting a deterministic outcome, here \$100, and then asking, what probability p_0 would cause indifference, could also be structured in the reverse by asking, for a given p_0 what value of certain outcome would cause indifference.

We can now specify a new lottery involving a deterministic outcome within the interval [0,100], asking what new value of probability p_1 would cause indifference; and then specifying another new lottery involving an outcome within [100,1000] and assessing a probability p_2, and so forth, until the shape of the utility function between zero and 1000 is sufficiently approximated.

5.2.5 Risk aversion

In the present case, as in most commonly encountered cases, the utility function for money is concave. The function displays diminishing marginal returns. For a concave function $u(x)$ on the interval (a, b), and for a random variable X defined within the interval, for which the expectations of X and $u(X)$ exist, Jensen's inequality (De Groot 1969) holds that $E[u(X)] \leq u(E[X])$. Thus, a person exhibiting a concave utility function over money will always choose a fixed gain x in the interval (a, b) over a lottery involving a random gain in (a, b) with an expectation $E[X] = x$. This can be seen from the dashed line in Figure 5.16 which is the locus of all linear combinations of $u(0)$ and $u(1000)$ for values of p.

Grayson (1960) experimentally measured the utility functions for money of wildcat oil and gas operators in Texas in the late 1950's, an example of which is shown in Figure 5.17. These functions typically display risk aversion, despite the reputation of wildcatters as risk-prone. The curves are gently concave over positive outcomes, and plunge steeply concave over negative outcomes. This risk aversion behavior is what motivates the insurance industry. An insurance company is a large entity for which individual losses on policies are comparatively small and within the linear range of the corporate utility function. On the other hand, such losses are not small to the individual insured, and thus the policy holder's utility within the pertinent interval may be highly concave. Thus, by Jensen's inequality, the individual policy holder is willing to pay a premium larger than the actuarial value of the risk to avoid the disutility associated with a loss. At the same

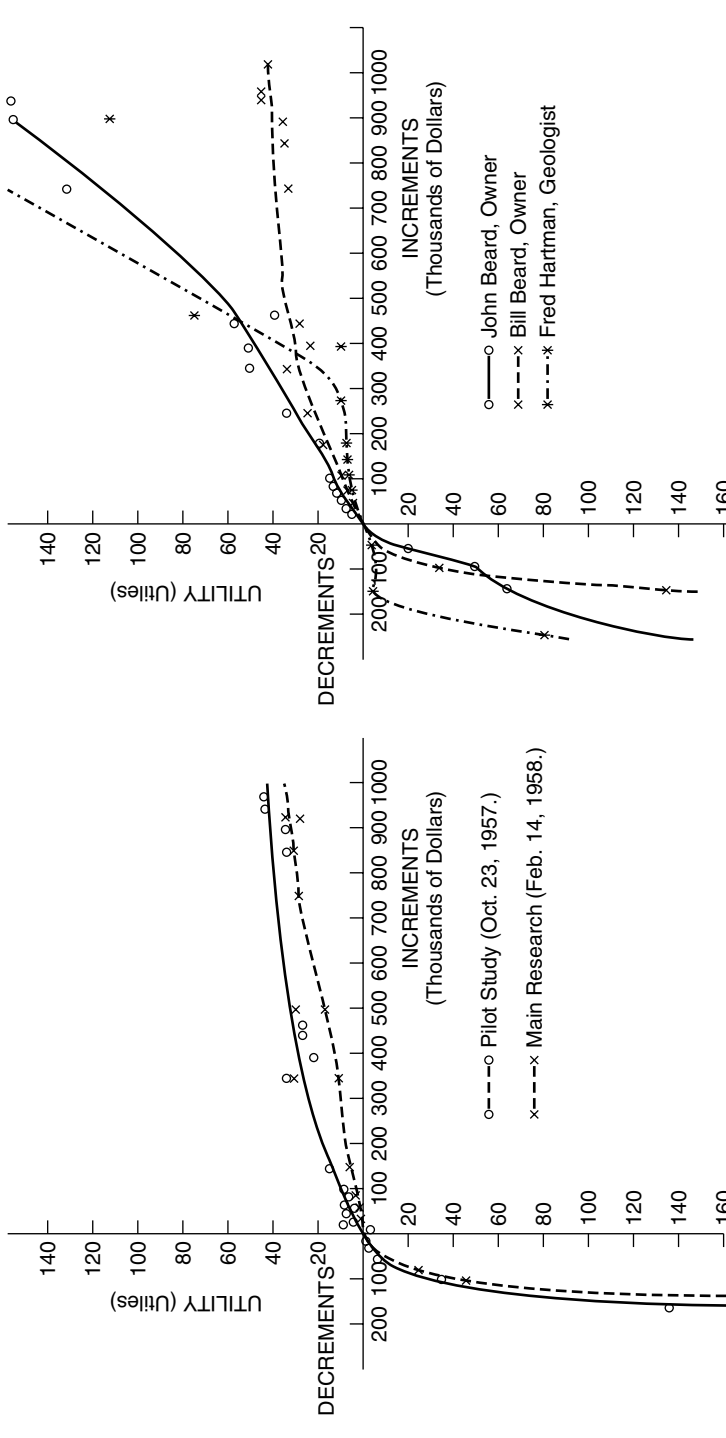


Figure 5.17 Example utility curve for money empirically measured by Grayson (1960). (Reproduced by permission of Harvard Business School).

time, the insurance company values the lottery at the actuarial value, and thus generates a profit from the transaction, and both sides benefit.[3]

The theory of multi-attribute utility over monetary and non-monetary attributes has advanced significantly over the past thirty years, and is well beyond the scope of this book. Early work on utility theory within the context of economic games was provided by von Neumann and Morgenstern (1944). More recent work on multi-attribute utility has been contributed by Keeney and Raiffa (1976).

5.3 Non-optimizing Decisions

Not all decisions need be optimized, indeed intuitive decisions tend to the acceptably good rather the absolutely best. This section considers alternative decision criteria, including consistency, the observational approach, and some other suggested approaches.

5.3.1 Consistent decisions

The James Bay Project involved large, low-head hydropower reservoirs of large extent, requiring long lengths of low embankment dykes on soft clays. These soft clays would not support dykes of the height needed, so a staged approach was adopted. The plan was for low, 6 m embankments to be built, and the foundation clays allowed to consolidate before heightening to 23 m. In some parts of the site, the 6 m dykes were all that was required, and would became the final embankments.

Initially, the plan called for the 6 m dykes to be designed with a modest factor of safety against instability, perhaps $FS = 1.5$; but the 12 m and 23 m dykes, due to their greater size and cost of failure, would be designed 'more conservatively,' with $FS = 1.75$ to 2.0. In the end, these choices – so obvious to engineering judgment – proved inconsistent. Reliability analysis showed that the 6 m dykes were much less reliable than the 12 m and 23 m dykes, and consequently needed to be designed to a higher FS to make the risks about the same. These calculations are detailed in Chapter 14.

For both the low and the high dykes, the largest contributor to uncertainty was the spatial variability of the properties of a lacustrine clay that underlay a marine clay. For the higher dykes, however, the failure surfaces over which this variability was averaged were large. For the low dykes, the failure surfaces were smaller. Variability in the clay properties enjoyed less averaging and thus less variance reduction. Another factor affecting uncertainties in clay properties was that, for the low dykes, undrained clay strengths were correlated to cone penetration resistance and a limited number of laboratory tests. For the high dykes, they were correlated to normalized soil engineering properties (Ladd and Foott 1974). The latter led to considerably smaller variances.

The combined result was that the variances in predictions of FS for the 23 m and 12 m dykes were substantially lower than for the 6 m dykes. However, the consequences of failure of the larger dykes were correspondingly larger, and the design team thought it prudent to set their target probabilities of failure to be lower (Table 5.8). The consequent

[3] Another perspective is that the insurance company views risk as a relative frequency while the insured views it as a degree of belief.

Table 5.8 Consistent factors of safety for the James Bay dykes

Design Case	Nominal Pf	Consistent FS (Calculated)	Design FS (Recommended)
Single stage, H = 6 m, circular failure surface	0.01	1.63	1.6
Single stage, H = 12 m, circular failure surface	0.001	1.53	1.5
Multi-stage, H = 23 m, wedge failure surface	0.0001	1.43	1.4

choices of a design $F = 1.6$ for the 6 m dykes and $F = 1.4$ for the 23 m dykes may or may not be 'optimal,' but they are consistent in risk.

5.3.2 Terzaghi's observational approach

Since Terzaghi, geotechnical engineering has been characterized by a special approach to uncertainty: *the observational approach* (Peck 1969). What exactly this approach comprises is subject to debate, but everyone agrees it is fundamental to modern practice.

The observational approach as described today seems not to appear in Terzaghi's writings prior to Terzaghi and Peck (1948). It was rather Terzaghi's contemporaries, especially Bjerrum, Casagrande, and Peck, who brought it to the attention of the profession. Bjerrum (1960) described the approach as well as anyone,

> The conventional way of dealing with the prevailing uncertainties is to prepare the design on the basis of the most pessimistic interpretation of the available data. Terzaghi's procedure is different. He commonly designed his dams on what the outsider might consider optimistic assumptions, but at the same time he made elaborate and carefully planned provisions for detecting during construction whatever differences might exist between the assumed and the real subsoil conditions. Furthermore, he worked out in advance his plans for action to be carried out in the event that significant errors in the original assumptions were detected. Hence, he could not be taken by surprise.

Peck (1969) comments that it was not until late in the writing of Soil Mechanics in Engineering Practice that Terzaghi "realized the undercurrent of observation common to all his work," and organized his concepts in what he at first called, the 'learn-as-you-go' method. Peck cites an unpublished introduction written by Terzaghi, that includes the following.[4]

[4] In a well-known paper near the end of his life, Terzaghi (1961) wrote, "Soil engineering projects, such as dams, tunnels, and foundations, require a vast amount of effort and labor securing only roughly approximate values for the physical constants that appear in the [design] equations. The results of the computations are not more than working hypotheses, subject to confirmation or modification during construction. In the past, only two methods have been used for coping with the inevitable uncertainties: either adopt an excessively conservative factor of safety, or make assumptions in accordance with general, average experience. The first method is wasteful; the second is dangerous. A third method is provided that uses the experimental method. The elements of this method are 'learn-as-you-go': Base the design on whatever information can be secured. Make a detailed inventory of all the possible differences between reality and the assumptions. Then compute, on the basis of the original assumptions, various quantities that can be measured in the field. On the basis of

1. Base the design on whatever information can be secured.
2. Make a detailed inventory of all the possible differences between reality and the assumptions.
3. Then compute, on the basis of the original assumptions, various quantities that can be measured in the field. [. . .]
4. On the basis of the results of such measurements, gradually close the gaps in knowledge and, if necessary, modify the design during construction.

The observational approach has had broad impact on geotechnical engineering, even to the point that its principles form the basis for environmental policy on field measurements and data quality objectives for hazardous waste site investigations (Wallace 1995; US Environmental Protection Agency 2000). The approach also foreshadowed the move toward adaptive management of natural resources (Holling 1980).

Angulo and Tang (1996) address a groundwater sampling problem at the Savannah River Site of the US Department of Energy using a probabilistic version of the observational approach. The study focuses on the continuity of geological formations affecting radionuclide transport. As a basis for the study, a probabilistic site model is developed and updated as site characterization data become available. At any point in the project the full weight of current information can be used to project the outcomes of subsequent exploration and development stages, and to make decisions accordingly.

A schematic profile of the site is shown in Figure 5.18. The lowermost aquifer is the Congaree Formation, a well-sorted marine sand. Above lies the Green Clay, generally thought to be an aquitard. The Green Clay is mostly continuous, 2–3 m thick, but with offsite discontinuities. Above lie the McBean, Tan Clay, Barnwell Formations. These form a surface aquifer that is of most concern.

The authors consider the possibility of a window in the aquitard allowing groundwater communication from the upper aquifer to the Congaree. The decision tree of Figure 5.19 shows possible outcomes of an initial exploration plan with respect to detecting a window. If a window is found, the sub-tree labeled I shows the forecast performance of two subsequent monitoring schemes. If a window is not found, then sub-tree II applies. In this way, the decision tree was used as a form of accounting scheme, by which possible outcomes of exploration and monitoring activates were projected and corresponding best decisions identified for each possibility. The decision tree became an analytical device for keeping track of an observational approach.

5.3.3 Satisficing, muddling through, and other styles of decision making

The decision theory literature is abundantly provided with alternatives to traditional optimization, some similar in spirit to the observational approach. This section provides a few points of entry into this literature, without surveying it in any detail.

Perhaps no one is more associated with non-optimizing approaches to decision making than Herbert Simon. Simon (1957) introduced the notion of *satisficing*. Satisficing is a decision criterion that seeks to reach an outcome that is good enough rather than best, a satisfactory decision. Simon argued that it is rational in many situations to arrive at a good rather than best result, that is easier, quicker, and cheaper to find. Finding the best

---

the results of such measurements, gradually close the gaps in knowledge, and if necessary modify the design during construction."

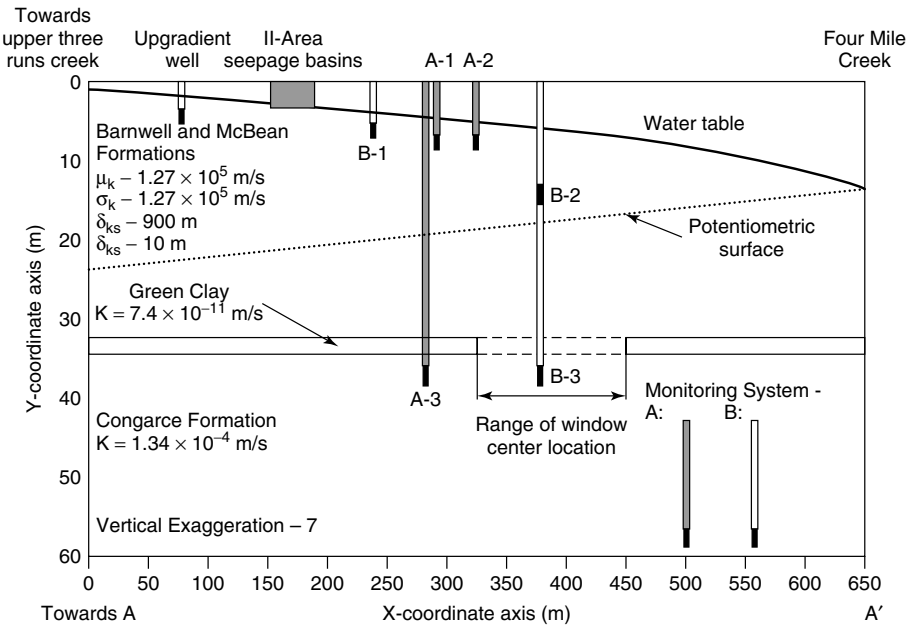


Figure 5.18 Profile of the Savannah River Plant site, showing details used in the analysis of Angulo and Tang. (Angulo, M. and Tang, W. H., 1996, 'Groundwater monitoring system design using a probabilistic observational method for site characterization,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, Madison, WI, pp. 797–812, reproduced by permission of the American Society of Civil Engineers.)

result may even be infeasible with finite effort. This is especially the case, Simon argued, when there are multiple, competing objectives.

Lindblom (1959; 1979), in addressing public policy decisions, proposed a 'science of muddling through,' in contrast to what he called the rational-comprehensive method, or what we would call optimization. He held that optimization "assumes intellectual capacities and sources of information that men simply do not possess [. . ..]," requiring clarity of objective, explicitness of evaluation, comprehensiveness of overview, and quantification of values that are usually not possible in any but small-scale problems.

In contrast, he proposed a method of successive limited comparisons, or 'muddling through.' This method is characterized by, building out from current situations, a step-by-step process, objectives based on consensus, limited analysis, and alternate theories being reduced by successive comparisons. The step-by-step process bypasses excessive calculations, reduces time needed to define goals, and limits comprehensive comparisons between the current situation and the changes that result from incremental changes. An advantage, it is argued, of incrementalism is that one can recover errors and mistakes in a one-step-at-a-time process. Also, decisions evolve in mutually agreed increments to meet the current needs and objectives, constantly changing and moving toward the goal.

5.4 Engineering Judgment

People have a great deal of confidence in human intuition. This is not limited to geotechnical engineers, although the geotechnical literature makes a great to do about *engineering*

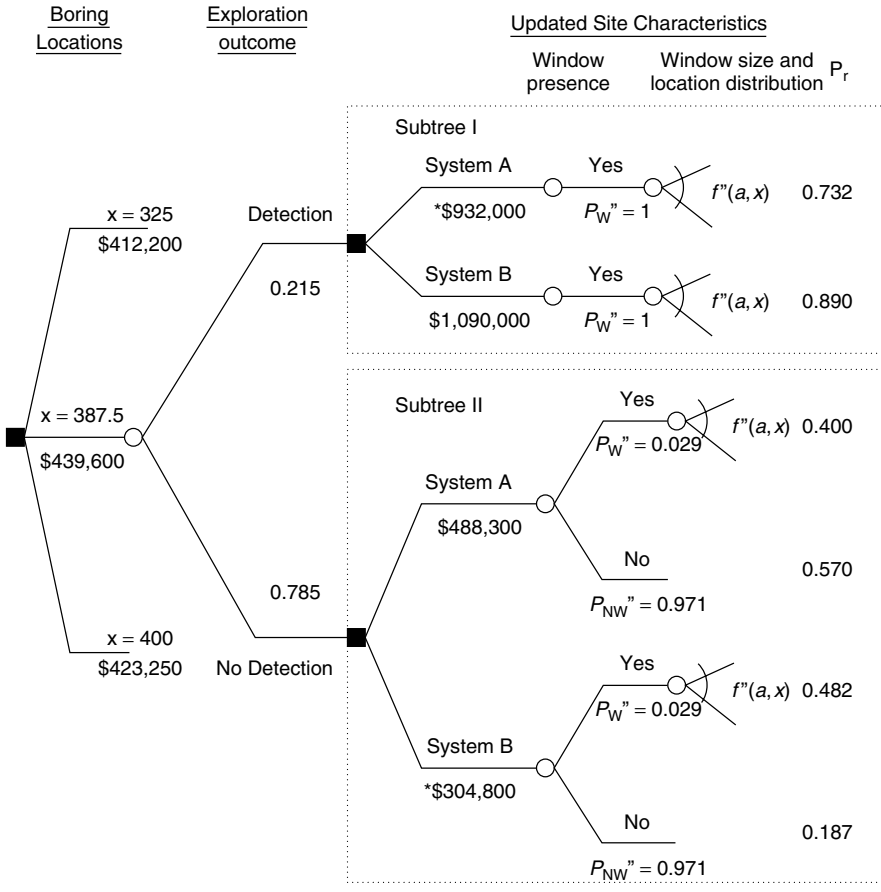


Figure 5.19 Decision tree for groundwater monitoring at the Savannah River site, based on observational approach. (Angulo, M. and Tang, W. H., 1996, 'Groundwater monitoring system design using a probabilistic observational method for site characterization,' *Uncertainty in the Geologic Environment, GSP No. 58, ASCE*, Madison, WI, pp. 797–812, reproduced by permission of the American Society of Civil Engineers.)

judgment with little discussion of what the notion means. There seems tacit agreement that readers understand the notion, but a cursory review of the geotechnical literature gives little comfort.

To some writers, engineering judgment is the basic tenet of responsible professional decision making, without which even the term *professional* is suspect (Peck 1980; Vick 2002); to others, engineering judgment little more than guesswork, a flimsy substitute for logic and reasoning (Hartford 2000). Although too ambitious a task for the current volume, better understanding of the nature and limitations of engineering judgment would be helpful in geotechnical practice.

While engineering judgment is raised to transcendent heights within the geotechnical community, it is often questioned by policy makers and the public. Despite the benefits provided society by modern constructed facilities, adverse environmental consequences and other unfavorable impacts of those same facilities are much on people's minds. To an

extent, engineers are blamed for these adverse consequences, and the paternalistic sentiment embedded in the notion of engineering judgment – that professionals know best – is out of fashion. In recent years, US federal government agencies such as the Nuclear Regulatory Commission (NRC) have even discouraged the use of the term engineering judgment in their deliberations (US NRC 1999).

Even though public opinion polls consistently rank engineers among the most trusted occupational groups, that same public increasingly questions whether engineering judgment can be relied on to answer questions of public safety. Perhaps safety can be improved by more public involvement, more regulatory oversight, and more explicit proof of engineers' contentions. Perhaps engineering truth is no more than a social construct, lacking objective validity, the way deconstructionist theory in recent years has viewed science itself (Hacking 1999). This is to say that, the elevated position to which engineering judgment is raised by the geotechnical community is by no means echoed by the rest of society.

The role of engineering judgment in routine design, for example in selecting pile driving criteria or in dimensioning a retaining wall, is relatively accepted; but its role in making decisions with potentially catastrophic consequences, for example in dam safety or environmental remediation, is another matter. The public and regulatory authorities that represent public interests require more than assurances of confidence from experts to put fear to rest. Modern history has examples of technological failures about which engineering experts had expressed great confidence beforehand: the Tay Bridge, Titanic, Saint Francis Dam, Malpasset Dam, the Challenger, among others. To say to the public – or even to owners and clients – 'trust us,' is to miss the mark by a wide margin.

So, what is engineering judgment, how does it relate to decision making and risk management in geotechnical practice, and how does it relate to probabilistic risk and reliability?

5.4.1 Problem solving and cognition

In a book dedicated to this topic, Parkin (2000) notes that "judgment is informed by experience, expertise, reasoning, and analysis." Such judgment may be brought to bear quickly, or only after a long period of deliberation. It may be the result of one person's considerations, or of a collaborative process involving many people, possibly with differing professional expertise.

We know a great deal about the psychological factors that come into play when people attempt to estimate risks subjectively. This understanding has developed over the past 30 years in the work of Tversky, Kahneman, Slovic, and many others (Chapter 21). We similarly know something about the psychology of judgment, although the present volume is not the place to delve into this topic. This understanding has developed in the field of social judgment theory growing out of the work of Brunswik (1969) and others, out of various fields of philosophy, for example in the work of Nozick (1993) on rationality, and to some extent in the growing literature of knowledge management, for example in the work of Prusak (Davenport and Prusak 1998), and others.

The field of social judgment theory views judgment as the ability to arrive at conclusions based on attributes of an object or situation. Margolis (1987) identifies seven progressive levels of cognitive function, which Parkin suggests may be thought of as different ways that people process experience and reach decisions on actions. These are:

(1) simple feedback, (2) pattern recognition, (3) learning, (4) choice, (5) intuitive judgment, (6) reasoning, and (7) calculation.

Simple feedback is involuntary stimulus-response of the nervous system; driving a car and hand-eye coordination involve simple feedback. *Pattern recognition* is the integration of stimuli from the natural world into recognized configurations; butterflies migrate thousands of miles based on patterns of temperature, wind speed, and other things. *Learning* is the process by which a lasting change in potential behavior occurs as a result of practice or experience. The pattern of cues is remembered and abstracted from the past; in response to the stimulus of a cabinet door opening and a can of food being opened, a pet dog will run to its bowl. *Choice* arises when multiple stimulus-response relations have been learned; the dog may flee danger or stand and fight when someone tries to take its bowl away.

Feedback, pattern recognition, learning, and choice are cognitive process we share with many other animals. Judgment, reasoning, and calculation are not. *Intuitive judgment* has to do with being able to see with the mind's eye what might happen if specific actions are taken in the external world and thereby to engage in mental simulation to find an answer to a problem. Pattern recognition is brought within the process of cognition. Data, experience, representations, and patterns are compared within the mind until a fit is found with the problem at hand. Margolis says that this is not the sort of thing that "very plausibly requires logic, calculation, following out of formal rules, or even verbal reasoning of the most informal sort." It is the sort of thing that is at the center of craftsmanship, and even our cousins in the primate world evidence judgment.

Reasoning and calculation lie beyond intuitive judgment, and are the stuff of human intelligence. *Reasoning* is the explicit use of language to facilitate judgment. It is among the reasons that language skills are so important to engineering. Without language, there is no reasoning, and thus reasoning is a human development of relatively recent vintage, since language is a development of the late Pleistocene. Language provides the abstract symbolic framework within which people move beyond intuitive judgment. Syntax, the rules whereby words or other elements of language are combined to form grammatical sentences, is the logical structure within which the "language equivalent of physical reality," to use Parkin's phrase, can be manipulated as one might manipulate an engineering model. Semantics, the meanings and classifications of what words signify, provides the elements to be manipulated. So, reasoning is a rationally structured way to draw conclusions about the world and to chose among alternate courses of action. The emphasis here is on *rationally structured* – in contrast to *intuitive* – which is the substance of judgment.

It is a small step conceptually from reasoning to calculation. *Calculation* replaces the informal logic of reasoning with the formal logic of mathematics. Semantic names are replaced by abstract variables, and the disciplines of symbolic logic and mathematics are brought to bear on uncovering relationships or patterns and reaching conclusions. This is the 'engineering modeling' with which so much of the university curriculum is obsessed. For most people, there is little intuitive about calculation, and calculation and reasoning are tightly bound.

5.4.2 Cognitive continuum theory

The cognitive basis of intuitive judgment is poorly understood. Brunswik's model, which has been the basis of later work by Hammond (1996), Cooksey (1996), and others – and

which has been cited by Parkin, Vick (2002), and others – is based on perception, specifically the perception of attributes which he calls, *cues*. Such cues are statistically related to objects and situations based on experience. When faced with a new object or situation, the individual perceives a relatively limited number of cues (Brunswik speculates that the number seldom exceeds seven) from among the almost limitless possibilities, and from these draws conclusions. The cues tend to be complex; thus they many not be interpreted in the same way each time they are perceived, or they may be perceived differently each time. Different people, presumably, perceive and interpret cues in different ways, presumably place different weights on them, and presumably combine cues in different ways, and thus may come to different conclusions about the same object or situation. Brunswik posits that most intuitive judgments can be modeled as weighted averages of cues and that the redundancy among cues ensures that conclusions are not highly sensitive to the exact values of the weights, suggesting a robustness of judgment.

Hammond and Cooksey combined Brunswik's model of cues and intuitive judgment with reasoning and calculation to form cognitive continuum theory. Cognitive continuum theory holds that intuitive judgments should be evaluated by the *correspondence* between the weighted average of the cues perceived about an object or situation, on the one hand, and the critical attributes of the real object or situation they reflect, on the other. If these two correspond, then the judgment is said to be valid. In contrast, reasoning or calculation should be evaluated by the *coherence* of the model produced. If the parts of the model form an internally consistent totality, then the reasoning or calculation is said to be valid. The correspondence of this logically sound model to physical reality is of secondary importance.

Cognitive continuum theory further holds that people do not – or cannot – simultaneously think in both correspondence mode (intuitive judgment) and coherence mode (reasoning and calculation) but rather flip back and forth between these two cognitive processes. People form intuitive judgments, then subject those judgments to reasoning and calculation, and then take the results back into an intuitive correspondence mode, and so on, and so on. Hammond calls this, 'quasi-rational cognition.' In solving a difficult problem, one might first look to hunches, intuitive guesses, or premonitions, and then subject whatever arises to analytical thought. When the analysis becomes bogged down, one might go the other way and seek hunches about the analysis. This is something akin to Karl Popper's (1968) hypo-deductive view of the scientific method, in which an hypothesis is developed intuitively but then tested deductively.

Parkin notes that, in practical applications, analytical cognition is more highly accurate on average than is intuitive judgment but sometimes can be wildly inaccurate. This is unsurprising. When underlying assumptions are more or less correct, analytical cognition can be both accurate and precise; but, when those same assumptions are incorrect, the conclusions based on analysis can be widely off. Large errors were sometimes made in the analytical mode, but research suggests that they are less frequent in the intuitive mode (Hammond 1996).

The confidence that people, both professionals and the public, place in a prediction appears to be related to the degree of internal consistency that manifests itself in the arguments or model underlying the prediction. For example, this has been observed in legal proceedings among other places, where juries seem to reconcile conflicting evidence by constructing stories based on prior experiences (Pennington and Hastie 1993), the more complete and consistent the story is, the more it tends to be believed. Thus,

because intuitive judgments often lack the clearly stated internal consistency of analytical cognition, they are sometimes underestimated in empirical value.

5.4.3 The place of engineering judgment

Among other things, this book embraces two apparently contradictory positions. First, applications of probability to engineering problems yield the most satisfactory results when probability is based on a degree of belief. That is, the most important aspects of probability are fundamentally subjective. Secondly, engineers, along with other experts, are usually excessively optimistic about their knowledge of important parameters and especially about the uncertainty in that knowledge. In fact, however, the two are not contradictory. Much of our uncertainty is subjective, and we are often not very good at estimating that uncertainty. One of the things that distinguishes a professional in any field from a dilettante is the degree to which he or she is able to assimilate experience and extrapolate from it to new situations. In effect, what is important is not the amount of experience but the amount of *evaluated* experience.

Properly applied, engineering judgment reflects accumulated and evaluated experience. The term should not be used as an excuse for avoiding analysis or investigation. As the quotations in the earlier sections reveal, Terzaghi's success depended in large part on a willingness to examine a problem deeply before rendering an opinion. Engineers who have invoked 'engineering judgment' without thought when asked to justify an opinion have done themselves and the profession a disservice and have tended to put the phrase engineering judgment into disrepute.

Part II

6 Site Characterization

This second section of the book deals with many issues of modeling, site characterization, and drawing inferences about geological conditions. All of the enterprises revolve around statistical reasoning and the weight of information in data. They have to do with the logic of decision making in allocating exploration resources and deciding when enough information has been gathered for design.

A site characterization program is a plan of action for obtaining information on site geology and for obtaining estimates of parameters to be used in modeling engineering performance. In other words, site characterization is the enterprise of exploring and investigating a site to gain sufficient information about the geometry and material properties of local geological formations that rational engineering decisions can be made. Some of those engineering decisions have to do with the best use of the site, some have to do with remedial measures needed before a project can be built, and some have to do with selecting soil or rock engineering parameters for modeling and design.

6.1 Developments in Site Characterization

Nearly every textbook on geotechnical engineering or engineering geology starts with a discussion of the enterprise and strategies of site characterization. Surprisingly – given the centrality of site characterization to all the analyses and predictions that geotechnical and geological engineers make – these discussions are notably brief. The third edition of the classic *Soil Mechanics in Engineering Practice* (Terzaghi *et al.* 1996) devotes two and a half of its 550 pages to the subject, before moving off into details of drilling rigs and test apparatuses. Taylor (1948) devotes three pages; Lambe and Whitman (1969), none at all. Perhaps the sense is that site characterization is a wholly intuitive process, dependent on 'engineering judgment', but not analytical thinking, and as such there is little to discuss regarding strategy and inference. As Underwood (1974) notes and Dowding (1979) expands upon, there is "an apparent hope for some new magical development that will fill in the gaps between a few poorly sampled, widely spaced and often poorly logged borings;" what Dowding calls, new 'gigits' to substitute for adequate information.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

The engineering geology literature is somewhat more generous in its discussions of site characterization approaches (e.g. Legget and Hatheway 1988), and specialty volumes do exist on the topic (e.g. Hvorslev 1949), but even they tend to dwell on equipment rather than logic and inference. There have also been several specialty conferences on site characterization and geotechnical uncertainty, which provide greater, if still limited, consideration of strategy (e.g. Dowding 1979; Saxena 1982; Shackelford *et al.* 1996, Robertson and Mayne 1998).

6.1.1 Minor geological details

In 1929, Terzaghi presented a classic paper at the New York meeting of the American Institute of Mining and Metallurgical Engineering entitled, 'Effects of minor geological details on the safety of dams,' in which he wrote (Terzaghi 1929),

> 'Minor geological details' refer to features that can be predicted neither from the results of careful investigations of a dam site nor by means of a reasonable amount of test borings (*sic*). They include such items as the exact position and the variations in width of fissures passing through the rock beneath a dam foundation, the shape and the local variations of the permeability of minor seams, of coarse sand and gravel contained in fine-grained alluvial valley fills, and similar features of minor geological importance.

One might add, "of minor geological importance but major geotechnical importance." Two dam sites might be to outward appearance almost identical in their geology; yet one might harbor adverse details too minor to detect with reasonable exploration effort, while the other might not. A dam constructed at the latter may perform well, but a dam at the former might prove disastrous. Terzaghi went on to ask, "To what extent can an opinion based on analogy be depended on in the field of dam foundations?"

In Chapter 12 we show that the probability of finding geometrically small details in the geological subsurface is negligible, absent unthinkably large investments in people, equipment, and time. The probability of finding geological features even on the scale of meters in width is seldom large enough to give one confidence that none have been missed. In this case, the importance of applying probabilistic methods to the question of site exploration may lie mostly in dispelling the sometimes intuitive notion that one has any real chance of detecting small features.

The implication of this state of affairs as reflected in Terzaghi's writing is the need for painstaking observation, both of site conditions as they are revealed during construction and of engineering performance as it manifests itself in the as-built structure. This led to the development of what Peck later called *the observational approach* (Peck 1969, 1980; Peck *et al.* 1984). The observational approach is now seen as a hallmark of geotechnical engineering, and it has even been adopted outside the field, for example, in environmental engineering (Wallace 1995). The *adaptive management* approach widely used for environmental problems is similar in spirit and application to the observational method. We return to the observational method later in this chapter.

6.1.2 Exploration strategy in the oil, gas, and mineral industries

It should hardly surprise us that the oil, gas, and mineral industries have invested heavily in attempting to deal rationally with exploration strategies. The exploration expenditures in

these industries are exceptionally large, and the rewards of finding economically sized oil pools or mineral deposits even larger. The minerals industry has, for the most part, taken a traditional statistical sampling approach to geological exploration. This led to a swell of book length publications in the 1960s and 1970s on various aspects of relative frequentist methods applied to mineral exploration and earth science issues more generally (e.g. Agterberg 1974; Davis 1986; Koch and Link 1970; Miller and Kahn 1962) and also to establishment of the journal, *Mathematical Geology*.

In Europe, starting in the 1960s, a group of researchers at the Ecole des Mines de Paris around Georges Matheron began developing a theory of random fields that came to be called *geostatistics*. This name is somewhat misleading, because the method it describes is not the equivalent of its cognate in the life sciences, *biostatistics,* but refers rather to a specific modeling approach to spatial variation. Also, the approach is probabilistic rather than statistical, at least within the original French school, although more recent North American and English work has tended to blur this distinction (*viz.* Cressie 1991). Nonetheless, the name has stuck and now also describes a large literature of work that began with mining applications (Matheron 1971) but now finds application to many fields, including environmental pollution (Ohlson and Okland 1998), groundwater flow (Marsily 1986), satellite imagery (Blodgett *et al.* 2000), paleobotany (McCartney 1988), and even geotechnical engineering (Soulie *et al.* 1990).

The development of exploration strategy in the oil and gas industry has progressed along a somewhat different path than that in the minerals industry. This work has been driven more by decision theory and Bayesian probability, starting from the pioneering PhD theses by Grayson and Kaufman under Raiffa at Harvard Business School (Grayson 1960; Kaufman 1963). Later contributors include Harbaugh *et al.* (1977), Lerche (1992; 1997), Newendrop (1975), as well as many others. Today, sophisticated software packages and other tools are available to assist in the use of decision theory and statistics to oil and gas exploration, and large annual conferences focus on the topic (SPE 2001).

6.2 Analytical Approaches to Site Characterization

Ideally, we would like to have perfect knowledge of site conditions, but, in reality, resources are limited. Expenditures must be commensurate with both the scope of the proposed project and with the potential consequences of using imperfect information to make decisions. Generally, we try to obtain as much information as possible about four aspects of site geology:

- The geological nature of deposits and formations,
- Location, thickness, and material composition of the formations,
- Engineering properties of formations as they may affect facility performance, and
- Groundwater level and its fluctuations.

The information resulting from site characterization may be distinguished, in large measure, as having either to do with geometry or with material properties. Either can be important, and uncertainties in either geometry or material properties combines with uncertainties in the other, as Dowding (1979) points out concerning the stability of a rock slope with a mylonized shale interface (Figure 6.1). The factor of safety for slope

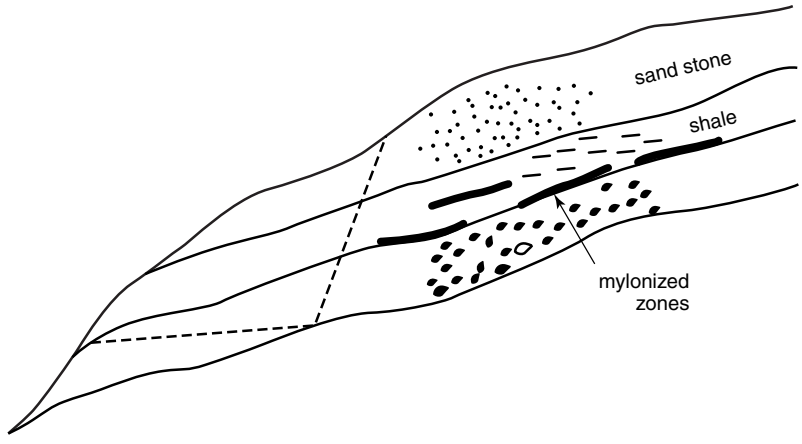


Figure 6.1 Idealized stability – mylonite problem. (Dowding, C. H., 1979 'Perspectives and challenges of site characterization,' *ASCE Conference on Site Characterization and Exploration*, Evanston, IL, pp. 10–38, reproduced by permission of the American Society of Civil Engineers.)

stability depends both on the extent of the mylonization as well as on the properties of the shale where the mylonization occurs. In this case, it was estimated that the mylonization was about 40% of the potential shear surface; Dowding accurately points out that this 40% estimate is numerically just as crucial to the calculation of factor of safety as is the residual strength value assigned to the material.

Here, both the connected extent of a critical joint surface and the shear properties of the interface of the joint complement one another in leading to stability or instability: connected length multiplies with frictional resistance to yield total resistance. Terzaghi *et al*. (1996) argue that information on geometry comes first and that precise information on material properties is important only if the geometry is reasonably well known. This, of course, is the substance of Terzaghi's argument that in specialized domains such as dam construction the influence even of minor geometries can control performance.

6.2.1 Exploration and investigation

When we write *exploration* in the present discussion, we mean the process of gaining sufficient information upon which to found important hypotheses that will later be used to explain the detailed measurements that are made (Figure 6.2). For example, exploration involves interpretations of regional geology and geological history, and it involves deciding whether significant anomalies should be expected in site formulations. Exploration is a process of induction and hypothesis formation. On the other hand, when we write *investigation*, we mean the process of collecting the detailed measurements upon which models and calculations can be based. For example, investigation involves statistical summaries of strength or permeability test data, and it involves mapping the stratigraphy or other geometric structures at the site. Investigation is a process both of deduction and of statistical inference.

Obviously, both exploration and investigation are needed for site characterization. However, the development of hypotheses is not something usually modeled with probabilistic and statistical methods. Peck (1969, 1980) has said that site characterization requires

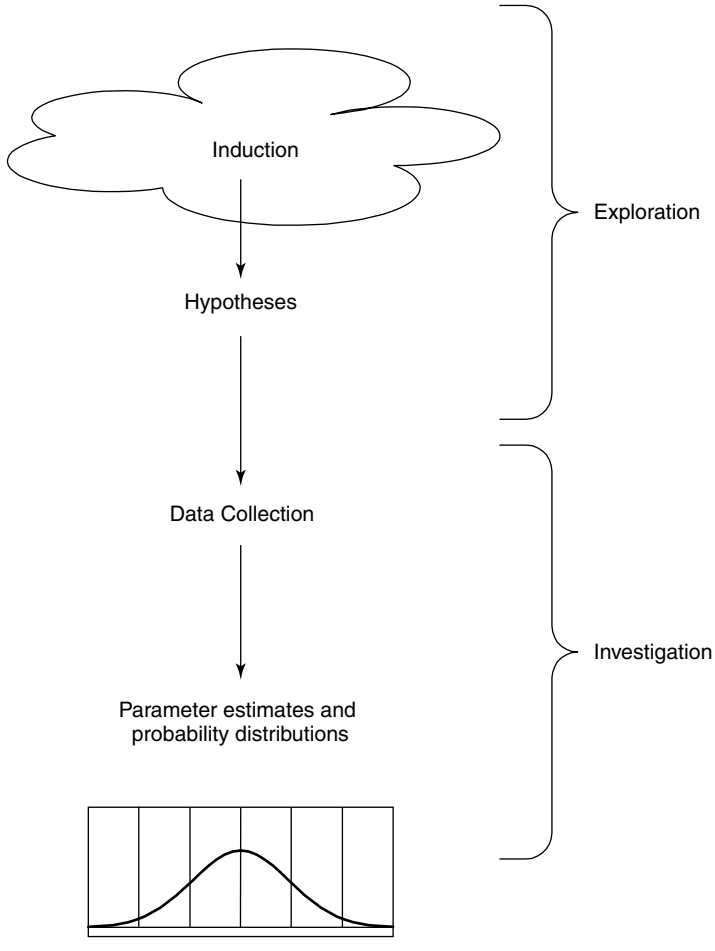


Figure 6.2 Nature of reasoning in site characterization, and the relationship of exploration and investigation to the logic of inference.

familiarity with soil mechanics but knowledge of geology. That is, he places emphasis on geology, not mechanics, because it is the awareness of earth history and the acquaintance with case stories elsewhere that enable the engineer or geologist to 'explain' what is observed within a larger framework. This framework is a set of hypotheses.

Once a set of hypotheses has been generated by which to explain the observations made in exploration, quantitative methods of analysis and inference – that is, probability, statistics, and decision theory – can be brought to bear on developing models of site geometry and random process descriptions of soil or rock parameters. This latter activity is amenable to modeling and analysis. It is the principal subject of this chapter.

6.2.2 Tasks in site characterization

Site characterization is usually organized in three stages, described variously as reconnaissance, preliminary investigation, and detailed investigation (Figure 6.3). The findings of

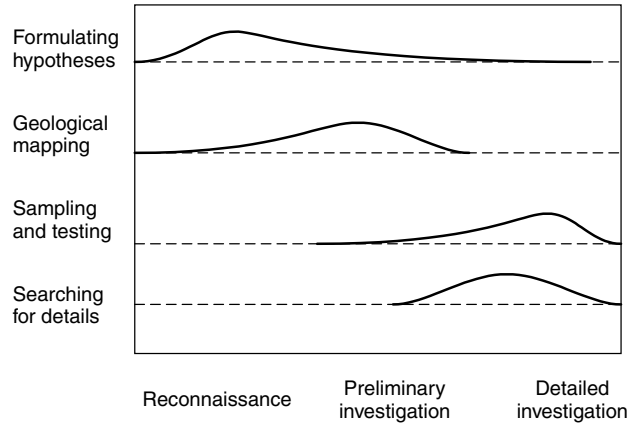


Figure 6.3 Traditional phases of site characterization and the various tasks that are undertaken in relation to each.

each stage are integrated with project considerations and design objectives to identify the sorts of information to be gathered in the subsequent phase and to identify the principal unknowns that need to be investigated.

Reconnaissance involves a general review of local and regional geology with the purpose of obtaining qualitative estimates of (1) the geological nature of site formations, (2) the engineering properties of major formations, and (3) the possibility of adverse geological details. This typically starts with a study of published documents, for example geological and topographic quadrangles, airphotos, and records of nearby construction. Good practice dictates that the documentary record be augmented by on-site inspection. The products of reconnaissance are qualitative hypotheses. As Dowding (1979) notes, many times this reconnaissance phase is not pursued as vigorously as it should be, and, as a result, the hypotheses to be subsequently investigated by sampling and testing may not be sufficiently well posed.

Preliminary investigation intends to collect the first quantitative information at the site and, by so doing, to extend the reconnaissance phase by confirming qualitative hypotheses and moving into a first set of quantitative hypotheses. The location and descriptive geometry of major formations are determined by a limited number of borings, by field mapping, and perhaps by geophysical surveys. Initial quantitative estimates of engineering properties are made, and the type and possible locations of geological anomalies are suggested. On projects involving issues of rock mechanics, joint surveys are typically begun.

Detailed investigation seeks to confirm the quantitative hypotheses generated in preliminary investigation, particularly about conditions thought to be controlling for facility performance. Accurate information is sought on geometry and material properties through detailed field mapping, comprehensive boring programs, and additional geophysical surveys. Specimens are taken for laboratory testing, and in-situ tests are made to complement estimates of material properties made on specimens. Anomalous details are searched for in earnest, with the objective of either finding such features or of reducing their chance of existing undetected to a tolerable level.

Most of the quantitative results of site characterization involve properties measured at points (or over small volumes) and separated from one another in space. The continuity of

strata or zones with site formations can only be inferred. It is the engineer's responsibility to interpret these data and to draw conclusions about site conditions between observations. These conclusions rest primarily on the hypotheses about site geology generated during exploration and investigation, not on the specific numerical results of laboratory and field testing of physical properties. The importance of the sampling and test results lies in the extent to which they lend support to the hypotheses.

Rowe (1972) suggested in the Twelfth Rankine Lecture that we categorize projects into three groups with respect to site characterization requirements. There are those projects, both important and risky, where the complexity of local geology necessitates extensive investigation and design decisions require a great deal of subsurface information (Class A). These are the sorts of projects – dams, large underground openings, and major and sensitive structures – for which the observational method is also appropriate. Next, there are those projects where the relationship between a modest project risk and tolerable uncertainties in site conditions makes the question of how much exploration and investigation is needed difficult to answer (Class B). These are possibly the largest in number. Finally, there are those projects, of routine use and low risk, where relatively straightforward site conditions require little investigation (Class C). These are the routine, smaller projects for which the costs of exploration planning and extensive testing are not justified by reduced project risk.

6.3 Modeling Site Characterization Activities

The logical process of modeling site characterization involves the following steps (Figure 6.4):

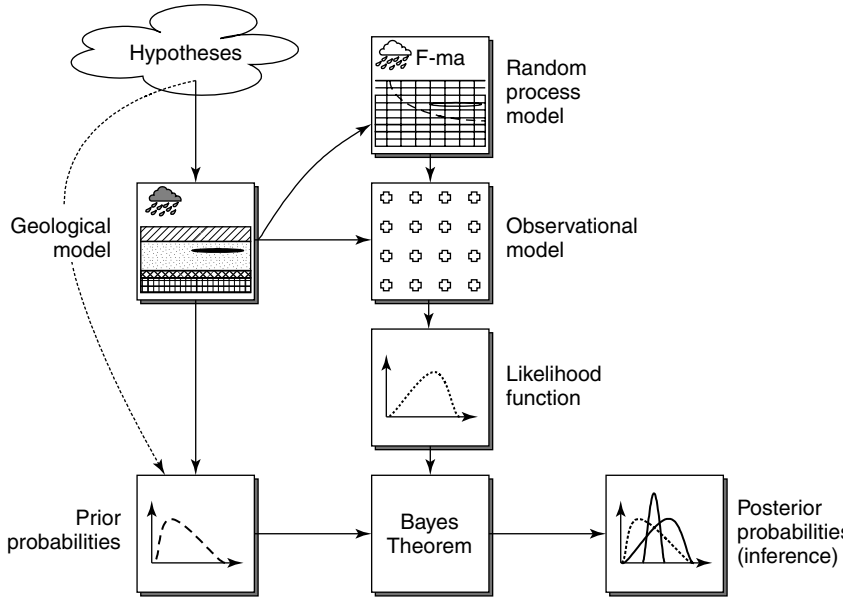


Figure 6.4 Modeling site characterization activities and inferences.

1. Develop hypotheses about site geology.
2. Build a random process model based on the hypotheses, which describes site geology and engineering parameters for design.
3. Make observations (i.e. measurements) in the field or laboratory.
4. Perform statistical analysis of the observations to draw inferences about the random process model (i.e. about the geometry of site formations and about the values of soil or rock engineering parameters).
5. Apply decision analysis to optimize the type, number, and location of observations.

Developing hypotheses is an undertaking for which mathematics and probability provide limited insight, relying as they do on the inductive act of conceiving explanations and building orderly structures within which the narrowly specific observations one makes at the site might be organized to paint a coherent picture of site geology and history. This act is not a *logical* process in the sense of deductive inference, nor is it illogical; it simply does not have to do with the formalities of logical reasoning. It is an intuitive act of judgment.

In site characterization, the random process models we use are usually models of spatial variation. Based on the language of probability theory, these models describe how soil or rock structures and properties may vary in space. Since a sufficiently large number of observations is never available with which to characterize precisely the spatial variation of site geology, we use these random process models of spatial variation to interpolate what might exist among the locations were we to have made the observations.

6.3.1 Probability is in the model not the ground

It is important to point out that the presumed randomness in this analysis is a part of the model(s) we employ, not part of the site geology. The assumption is not being made that site geology is in some way random. A convenient analogy, as used in Chapter 3, is the deck of playing cards. Once the deck has been shuffled and in play, the (spatial) order of the cards in the deck is not random. The order is unknown to the players, but the order is determined. Yet, probability theory models are useful in making decisions with respect to the play. Site characterization is much the same, only more complex. Once a formation has been deposited or otherwise formed, and has come down to us through the vagaries of geological time, the spatial distribution of structure and material properties is fixed. We are ignorant of that spatial distribution, but the distribution is not random. Unlike the deck of cards, however, a geological formation cannot be reshuffled to generate a 'new' trial, so the reality is a unique instance. Matheron (1989) discusses this point in considerable depth.

We can think about these random process models of spatial variation in one of two ways. The most common way is to presume that they are frequency (i.e. aleatory) models of natural variation, for example, in the same way that a flood-frequency relation is a frequency model of natural variation. The model of spatial variation reflects an inherently random process – just like the shuffling of a deck of cards – that has generated the site. Metaphorically, we consider an arbitrarily large number of sites generated in exactly the way this one has been, and then we use probability theory as does the card player. Of course, in reality there has not been a large number of similar sites; there is only one site. So, the philosophical logic of this approach is suspect, but in practice, the results

are useful. Unlike the card player, we get only one shuffle of the deck. This is the view adopted by the overwhelming majority of risk and reliability practitioners and of workers in the allied field of geostatistics.

A second way to think about these random process models of spatial variation is to presume they are models of our information about a fixed but spatially varying realization. Such an approach is unusual in the literature, but not unknown (*viz.* Gelb *et al.* 1974; Matheron 1989).

In all risk and reliability analysis we deal with two distinct categories of uncertainty: *natural variation* and *knowledge uncertainty*. As we discussed at greater length in Chapter 2, this distinction is not part of nature, but part of our modeling of nature. We take the philosophical position that, at the most primitive level, all uncertainty has to do with limited knowledge, not inherent randomness. This is a Laplacian view of the world. Nonetheless, it is convenient to structure our models as if some fraction of the uncertainty we deal with has to do with irreducible randomness and then to use statistical methods to draw inferences about the models applied to that natural variability. We do this in seismic hazard analysis, we do this in flood hazard estimation, and we do this in site characterization.

6.3.2 Centrality of prior probability

Uncertainty in site characterization is subjective, because it arises from limited knowledge. Attempts to make this uncertainty objective are never likely to be successful. The most important results of site characterization are hypotheses. These are generated inductively, given *a priori* support by subjective reasoning, and confirmed to differing and imperfect degrees by observations made at the site. This confirmation rests on the logic of Bayes' Theorem:

$$P(H|data) \propto P(H)L(H|data) \tag{6.1}$$

in which the posterior probability attributed to an hypothesis, H, given a set of data, is proportional to the product of the prior probability of H and the likelihood of H given the data. The likelihood of H is the conditional probability of observing the data that we actually did observe, were the hypothesis H true. If it is probable that these data would be observed, then the hypothesis is further confirmed; if it is improbable, then the hypothesis is less confirmed. We see here the import of De Finetti's (1972) well-known saying (Chapter 2), "data never speak for themselves." The data tell us only how to modify a degree of belief held prior to their expression into a logically consistent degree of belief after their expression. We have experience and knowledge of geology that cause us to suspect conditions not directly manifest in exploration records and to correlate physical behavior with the results of geological processes. The uncertainties we associate with these suspicions and correlations cannot be objectively evaluated from the records of exploration.

We see in Equation (6.1) the centrality of the Likelihood function in attributing probabilities to hypotheses about site conditions. Bayesian inference is built around the likelihood of hypotheses given the observations made. However, much of the frequentist school of statistical estimators and confidence limits – at least that part building on the work of Fisher – is similarly based on maximum likelihood estimation and the desirable sampling properties of those estimators.

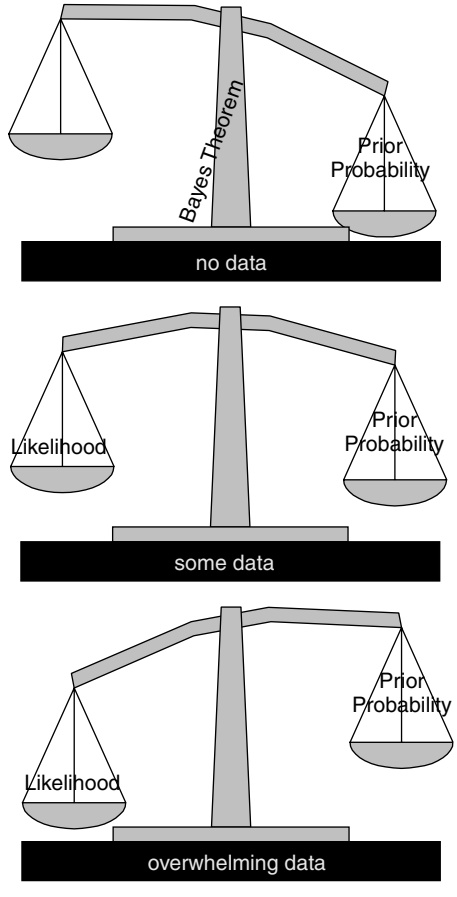


Figure 6.5 Bayes' Theorem serves to balance the 'weight of evidence'.

The weight we give hypotheses, as expressed by their probability, exists in balance with the likelihood of data as if on a scale, and Bayes' Theorem is that scale (Figure 6.5). In the early phases of site characterization we have few data, and essentially all of our confidence or degree of confirmation in hypotheses comes from our prior, subjective probabilities. As data accumulate, the confidence or degree of confirmation in hypotheses starts to be influenced by the observations as manifested in the Likelihood function. Some hypotheses are given more credibility while others are given less. As data become overwhelming (seldom the case in a real site characterization program, of course), the weight of observational evidence comes to dominate and the influence of the prior recedes.

6.4 Some Pitfalls of Intuitive Data Evaluation

Statistical modeling and analysis is a relatively new set of tools in the history of site characterization, which leads some to suggest that the enterprise of geotechnical engineering has gotten along fine without these methods, so why introduce them now? The intuitive

methods of engineering judgment, based on extensive experience with site conditions, should do just as good a job – maybe better – than mathematical methods. Surprisingly, however, even trained statisticians are easily led astray using intuition to interpret sampling observations based simply on inspection. Interestingly, the errors people make in intuitively interpreting data are remarkably similar from one person to another, and seem to exist somewhat independently of professional training (Kahneman *et al.* 1982).

6.4.1 Sample variation

When measurements are made in the laboratory or field they exhibit scatter. These measurements might more generically be called *observations* to include things other than instrument readings. A set of observations is typically called a *sample*. If the sample comprises a large number of observations, the data scatter among the observations tends to exhibit regularity. That is, the scatter within a sample and from one sample to another tends to display regular patterns, and over the years statisticians have learned to categorize those patterns and to use them to draw inferences about the population from which the sample comes, the *parent population*. Patterns of scatter within an individual sample are interpreted against what is known of the probable scatter among samples to make estimates about the parent population.

For convenience, the scatter within a sample or from one sample to another is described by a frequency distribution or histogram, and this in turn can be summarized by its low order moments. The most useful of these are the first two moments, the *mean* or arithmetic average of the observations, and the *variance* or mean-square variation about the mean. Such mathematical functions of the sample observations are said to be *statistics* of the data, or alternately, *sample statistics*, and form the basis for making inferences about the parent population.

The *law of large numbers*, a fundamental principle of statistical theory, implies that, as a sample becomes larger, the statistical properties of the sample become ever more to resemble the population from which the sample is taken.[1] The operative phrase here is, "as the sample becomes larger." For example, one is often interested in the average value of some parameter or performance measure in the field, across all the elements of the population that may not have been observed in the sample. If the number of observations in a sample is large, it seems reasonable, and the law of large numbers confirms, that one might use the sample average of the set of observed values as indicative of the population average in the field. But, what about the case where the sample is not large, as is almost always the case in geotechnical practice?

The law of large numbers says that variations of the statistics of a sample about their counterparts in the parent population become ever smaller as sample size increases, but for small samples these variations can be large. Presume we take many samples each containing n test specimens from the same soil layer, and for each sample we calculate the sample mean of tests performed on the n specimens, m. The values of m across the many samples themselves exhibit scatter, and could be plotted as a histogram. This distribution of the sample mean, or of any other sample statistic, is called the *sampling distribution*. The sampling distribution is the frequency distribution of some sample statistic over

[1] The Law of Large Numbers is more specific and limited than this colloquial interpretation (Feller 1967), but the practical implication is quite broad. See, also, Maistrov (1974) for an historical treatment.

repeated sampling. The theoretical variance of the scatter among the means of each of many samples of size n is $Var(m) = \sigma^2/n$, where σ^2 is the actual variance within the layer. The standard deviation is $s_m = \sigma/\sqrt{n}$.

The coefficient of variation of soil properties measured in the field can be as large as 100%, although values of 30–50% are more common (Kulhawy and Trautmann 1996; Phoon and Kulhawy 1996). Thus, if ten (10) tests are made, the variation of their sample average about the (unknown) layer average would have a standard deviation of 10–16%. Since, under very general assumptions, the sampling distribution of the mean is approximately Normal, the range for which one would be comfortable bracketing the population mean is, say, 2–2.5 standard deviations, or in this case between ±20–40% of the best estimate. In other words, there is considerable uncertainty in the inference of even average soil properties, when reasonable sample sizes are taken into account. There is, of course, even more uncertainty about inferences of soil properties at specific locations within a soil deposit.

6.4.2 Representativeness

Despite the fact that sampling variations can be large – and in geotechnical practice, they *are* large – there is an intuitive tendency to treat sample results as representative of – or similar to – the population from which they are taken. Most people believe intuitively that samples should reflect the essential characteristics of the population out of which they arise, and thus the converse, that essential characteristics of the population should mimic those of the sample. People's intuition tells them that a sample should be similar to the population from which it comes, but that is only true in the limit, as sample sizes become large. This leads to errors. Speaking strictly, *representativeness* is a property of sampling plans, not of samples. A sampling plan is representative of the population being sampled if every element of the population has an equal chance of affecting the (weighted) properties of the sample (Cochran 1977), and from this one speaks of 'representative sampling.' A sample, in contrast can never be representative; it is a unique collection of particular elements within the population, and each such collection has different properties.

Take, for example, the string of sample outcomes deriving from six tosses of a coin, {H,H,H,H,H,H}. Most people intuitively think of this string as less likely to occur than, say, the string, {H,T,T,H,T,H}, even though each has the same probably: $(1/2)^6$. This is akin to the Gambler's Fallacy that if *heads* has not appeared in some time, it is overdue and should occur with increased probability. Intuition tells us that the sample should represent the population, that is, be similar to the population in salient aspects, and in short runs as well as long. In this case, the sample should have about the same number of heads as tails, and the sequence of heads and tails should be 'random,' that is, erratic. That this is a misperception is obvious to anyone who thinks about it, but our intuition tells us otherwise.

The same thing is true of samples of geotechnical observations. We presume them to be representative of the geotechnical population out of which they arise. The averages within the sample ought to be about the same as the averages in situ. The variability of observations ought to be about the same as the variability *in situ*. Spatial patterns of variation among the observations ought to mirror spatial patterns *in situ*. All of these things are true in the limit, but for small samples they are compromised by sampling variability, and they may be profoundly untrue. Small samples of the size typical in geotechnical

practice seldom display the salient properties of the population; the variability among sample outcomes is simply too great.

6.4.3 Overconfidence

The intuitive belief in representativeness leads people to believe that important characteristics of a population should be manifest in every sample, no matter the sample size. Yet, we know from statistical theory that this is not true: small samples exhibit large variation from one to another. This leads people to put too much faith in the results of small numbers of observations and to overestimate the replicability of such results. If tests are repeated, people have unreasonably high expectations that significant results will be replicated. Thus, the ten (10) observations of field performance are made, and one is surprised that the next set of ten yields a 30% difference in average results. A person's typical response is not to ascribe this difference to perfectly normal statistical variation but to seek a *cause* for the difference. The engineering literature is filled with well-intentioned attempts to explain away differing sample results, when in fact, such explanations are totally unnecessary. The explanation is normal sampling variation.

A corollary to this belief in the representativeness of small samples is the overconfidence even scientifically trained people place in their inferences or estimates of unknown quantities. In a famous early study, Alpert and Raiffa (1982) demonstrated that when people are asked to place 25–75% or 5–95% confidence bounds on estimates of unknown quantities, the true values of the quantities being estimated fall outside the assessed bounds considerably more often that the nominal 50% or 10%, respectively. Often more than half the real values fall outside 5–95% confidence bounds people estimate. This result has been replicated in another early study by Folayan *et al*. (1970) involving engineers' estimates of the properties of San Francisco Bay Mud, and by Hynes and Vanmarcke (Hynes and Vanmarcke 1976) involving predictions of embankment height at failure for the MIT I-95 Test Embankment.

As reliability analysis becomes increasingly important to geotechnical practice, it is sometimes suggested that an expedient way of assessing the standard deviation of an uncertain quantity is by eliciting the maximum and minimum bounds one could conceive the quantity having, and then assuming that this range spans a certain number of standard deviations of variation, typically, $\pm 3\,\sigma$. The reasoning is that for a Normal variate, ± 3 standard deviations spans 99.75% of the variation. However, if people are overconfident of their estimates of uncertain quantities – which we know them to be – then people will frequently be surprised in practice to find their maximum and minimum bounds exceeded. Thus, the 'six-sigma' rule is unconservative, and possibly quite significantly. This can also be seen in Figure 6.6, in which the expected range of sample values, $r_n = |x_{\max} - x_{\min}|$, for a Normal variate is plotted as a function of sample size. Even for samples as large as $n = 20$, the range expected in a sample is less than four standard deviations. The reciprocal of this expected range, in fact, makes a useful estimator of standard deviation, and one with known sampling properties (Snedecor and Cochran 1956).

6.4.4 'Law of small numbers'

In a series of papers in the 1970s, Tversky and Kahneman described the systematic difference between the way people perceive probability and the way statistical theory

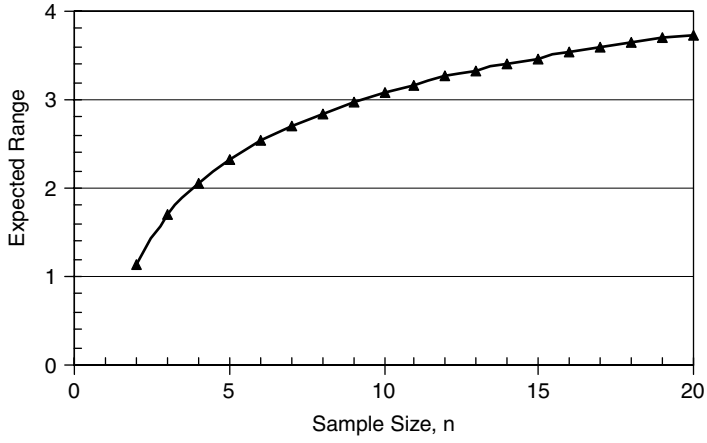


Figure 6.6 Expected range of Normal sample in standard deviation units (Snedecor and Cochran 1956).

operates (Kahneman and Tversky 1982a; Kahneman and Tversky 1982b; Kahneman and Tversky 1982c; Tversky and Kahneman 1971; Tversky and Kahneman 1974). That body of work, and the studies that followed, are sometimes referred to as the *heuristics and biases* school of thought on subjective probability (*viz.* Morgan and Henrion 1990).

This work emphasizes that the use of representativeness to judge probabilities is fraught with difficulty, because it is not affected by factors that should influence judgments of probability. Important among these are the overconfidence described above, a disregard for base rates (*a priori* probabilities), and ignorance of common regression effects. The concept that observers presume samples to be representative of the population seems benign, but it leads to serious errors of judgment in practice. Tversky and Kahneman dubbed this, 'The Law of Small Numbers,' which states that the Law of Large Numbers applies to small numbers as well.

This overlooking of sample size occurs even when a problem is stated so as to emphasize sample size, and in many different contexts. Consider, for example, a question in flood hazard damage reduction. A river basin is analyzed in two different ways to assess levee safety. In the first case, the river is divided into 10 mile (16 km) long reaches; in the second, the river is divided into 1 mile (1.6 km) long reaches. Would the average settlements among the levee reaches be more variable in the first case, the second case, or about the same? Of an admittedly unscientific sample of 25 engineers, seven said the first (more variation among long reaches), six said the second (more variation among short reaches), and 12 said about equal. Clearly, the long reaches have the least variation among their average settlements, because they are larger samples.

6.4.5 Prior probabilities

A second manifestation of representativeness is that people tend to overlook background rates and focus instead on the likelihood of the observed data when drawing conclusions. Sometimes representativeness leads people to place undue importance on sample data (because they "should be similar to the population"), and in so doing ignore, or at

least downplay, prior probabilities (the latter sometimes referred to as *base-rates* in the heuristics and biases literature). As a simple example, in risk analyses for dam safety a geologist might be asked to assess the probability that faults exist undetected in the bottom of a valley. Noting different rock formations on the adjoining valley walls, he or she might assign a high probability to faulting, because of the association of this condition with faulting, in spite of the fact, say, that the base-rate of faulting in the region is low. The two sources of evidence, prior probability and likelihood, should each influence the *a posteriori* probability (Equation (6.1)), but intuition leads us to focus on the sample likelihood and, to some extent, ignore the prior probability.

6.5 Organization of Part II

The second part of this book looks at a number of specific aspects of how we can formally model site characterization activities and draw statistical conclusions from the observations made in exploration and investigation. It begins with a review of the variability of typical soil and rock engineering properties encountered in site characterization, starting with index properties and then turning attention to strength, deformation, and permeability. It then considers the issues involved in classifying site conditions into identifiable groups and mapping those groups over space. Next, it discusses the spatial variation of engineering properties within formations that are judged to be 'homogeneous.' Finally, the section concludes with a discussion of the search for anomalous geological details.

7 Classification and Mapping

Among the most basic tasks of site characterization is to map local and regional geological formations, create vertical profiles (or cross sections), and infer the continuity and homogeneity (or lack thereof) of important deposits. This mapping attempts to divide the three-dimensional site into coherent formations or strata; identify, interpolate, and characterize structural features that form the boundaries among formations and strata; and relate these formations or strata to regional geological processes and trends. Conventionally, this has been done by observing the geological nature of a finite number of locations and inferring contacts among formations.

We, as a profession, have a great deal of experience with geological mapping. Nonetheless, probability theory and statistics have been relatively little used in rationalizing how we do this mapping, and the undertaking is often viewed as more art than science. As a consequence, there are fewer tangible results in the literature on quantitative mapping than on other aspects of quantitative site characterization. This is a shame, and it is likely to change, because increasing use of remote sensing, global positioning, and other sensing technologies has generated an abundance of data on geology that could feed more rational and efficient approaches to mapping. Many of the topics considered in this chapter find equal applicability to regional geological mapping as to local site mapping.

Probability theory and statistics can aid this process of geological mapping in at least two ways. First, they can be used to make inferences about the geology of unobserved parts of a site (or region) that are more powerful than those based on intuition, and those inferences can be associated with measures of precision and accuracy. Second, they can be used to optimize the way exploration effort is allocated across a site or among different means of collecting data, balanced against competing investments in sampling for material properties or finding geological anomalies. In principle, they should be able to inform the process of site characterization on "how much mapping effort is enough," but as we will see, this question is more difficult to answer than one might hope because the consequences of uncertainty in the products of mapping programs are often difficult to measure and sometimes even ill-defined.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

The development of a comprehensive theory of quantitative mapping is a big undertaking, but the rewards are enticing. Such a theory would afford measures of potential confidence to be placed on projections, techniques for economizing expensive allocations of effort, and facility for rational design decisions based on mapped information. Such a theory might also usher in the use of machine constructed geological maps, including favorable subsequent allocations of effort as output. Haralick and Dinstein (1971) have developed automated procedures for clustering geographical phenomena and producing area classification maps directly from raw satellite photographs.

7.1 Mapping Discrete Variables

The attributes portrayed on geological maps are of two types: discrete and continuous. This chapter treats the former. Continuous attributes are considered in Chapters 9 through 11. A common geological map, like those provided by the USGS in the U.S. or Royal Geological Survey in the UK, assigns to each location at a site one of several discrete classes of attribute, for example, formation name or lithology. Geotechnical maps typically assign a soil type or name to each location, which itself may be a compound description combining many different physical attributes, for example, a "silty soft green clay with traces of organics." But this compound description describes a soil judged to be in pertinent respects the same, no matter where it exists at the site. Some limited subset of these locations have actually been directly observed, while most locations at the site have been interpolated or extrapolated from this subset of direct observations, and thus these are to one degree or another uncertain classifications.

Discrete maps are constructed by observing and classifying geology at a finite number of locations. Sometimes these observed locations are in the two-dimensional plane of the ground surface, and sometimes in the three-dimensional space of the subsurface. Sometimes the observations are made at points, sometimes within small areas. In principle, this dimensionality of the observations and whether the observations are point values or area averages should make no difference to our modeling of the enterprise. From this set of observations, classifications are inferred for all the other, unobserved locations at the site. The way people typically make these classifications at unobserved locations (i.e. construct the full map) is very much an inductive process. People recognize patterns in the site geology. Based on these patterns, and based on the individual's knowledge of geology and experience at other sites, he or she will interpret and infer geological structures to exist, and on the basis of these inferred structures, extend the perceived patterns into maps. Those perceived patterns may or may not actually exist at the site, and, even if they are accurately interpreted, the extrapolations and interpolations across the site may be subject to error.

All of this is difficult to model mathematically. The difficulty can be judged by considering a parallel problem out of another discipline: automated face recognition with computers. Given the photograph of some unidentified person, from which measurable features can be abstracted, such as eye color and position, the height and width of the nose, and so forth, whose face is this? This is a problem of *pattern recognition*. Pattern recognition is a problem of classification. Then, given that we have made a guess at whose face this is, what does the rest of the body look like, and does he or she have the right of access to this facility? This is a problem of extending the pattern, sometimes

referred to as *pattern reconstruction*. Pattern reconstruction is a problem of interpolation or extrapolation. A great deal more work appears in the various engineering and mathematical literatures on pattern recognition, which has been fairly well-researched in computer science and artificial intelligence, than on pattern reconstruction, which is more inductive and thus subjective.

7.2 Classification

The first step in mapping (Figure 7.1) is to classify individual observations, each within one of a number of discrete categories. As a simple example, soil cores from a split spoon are classified with respect to soil type, and subsequently these classified data are used to create a soil profile showing the locations, depths, and continuities of soil strata underlying a site (Figure 7.2). In the case of soil type, the classifications are made by visual inspection and are seemingly straightforward. This is not the case when the borders among classes are fuzzy, when indirect measurements such as geophysical signals are used, or when multi-attribute data are used to classify observations. In these cases a more robust approach than simple inspection is needed, and this approach is statistical classification analysis.

7.2.1 Pattern recognition

Statistical *classification analysis*, sometimes called statistical *pattern recognition*, is a well-developed set of techniques that have been widely used in biology, archeology,

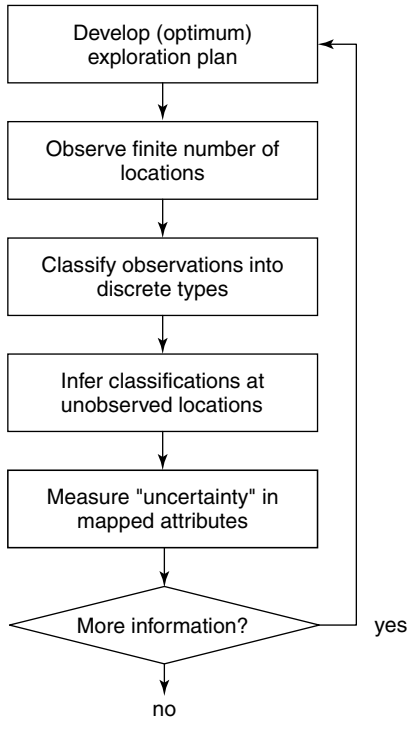


Figure 7.1 Logic flow chart for discrete mapping of the type used in developing geological maps.

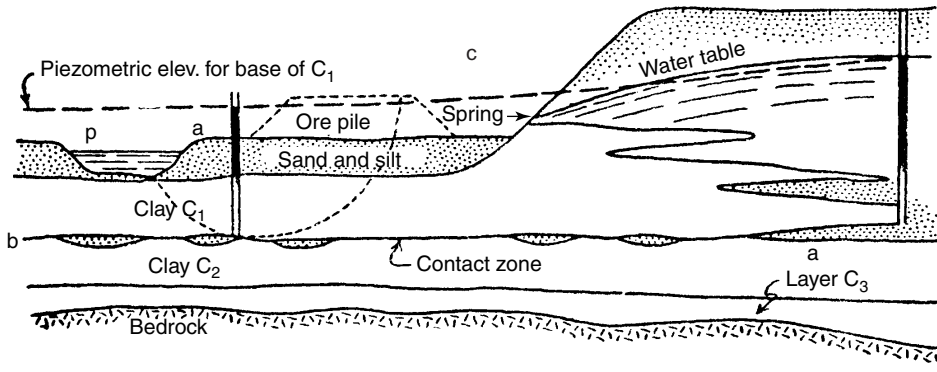


Figure 7.2 Soil profile constructed from individual soil cores classified as to soil type, and 'diagram illustrating the writer's hypothesis concerning the origin of the artesian pressure.' (Terzaghi, K., 1960. *From Theory to Practice in Soil Mechanics: Selections from the Writings of Karl Terzaghi, with Bibliography and Contributions on his Life and Achievements*. New York, John Wiley & Sons, © John Wiley & Sons 1960, reproduced with permission.)

computer science, and other fields for separating observations into discrete classes. For example, skeletal remains of fossilized fishes might be classified into species type on the basis of a suite of geometric measurements of length, height to width ratio, jaw length, and so on. The frequency distributions of any one of these attributes may overlap from one species to another, but taken together, the set of measurements allows individual specimens to be assigned to a species (i.e. class) with the minimum error rate. These same statistical techniques have found application to geotechnical engineering in the question of classifying soil profiles subject to seismic ground shaking as being either, (A) prone to liquefaction, or (B) not prone to liquefaction, based on SPT blow counts, peak ground acceleration, and fines content (Christian and Swiger 1975, 1976; Liao *et al*. 1988).

Some classification tasks are obvious and may be accomplished by inspection, for example, logging a core run visually by rock type. Others are not so obvious and require analytical or statistical techniques, for example, logging a drill hole using geophysical logs. The classification problem in geological mapping is this: given measurements of a limited number of physical properties on a particular specimen or observed location, to which geological class should the specimen or location be assigned such that some decision criterion is optimized (e.g. the most likely class, the class maximizing some expected utility, or so forth)? Figure 7.3 shows data from a geophysical core in a sedimentary formation consisting principally of sandstones, shales, and limestones. Assigning stratification by lithology to this run based only on the geophysical properties is an example of a (statistical) classification problem.

The first step in classification is extracting *features*, sometimes also called *factors*, to form the basis for deciding to which class an entity belongs. These are the measured values that are used to make the classification decision. In the example of the archeological fishes, these features were length, height-to-width ratio, and jaw length. In the liquefaction example, these features were SPT blow count or CPT value, estimated peak ground acceleration at depth, earthquake magnitude, and fines content. In the geophysical profile example of Figure 7.3, these features were self-potential and electrical resistivity. The features can be any measure against which the alternate classes have frequency distributions that differ from one class to another and therefore provide discriminatory

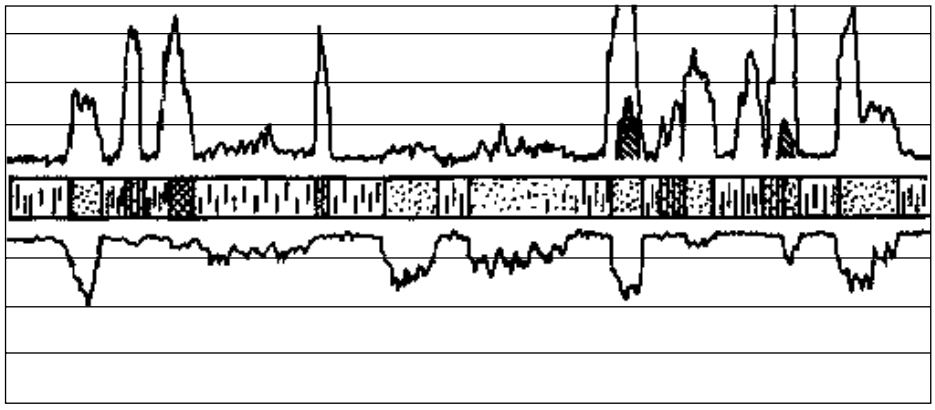


Figure 7.3 Geophysical core in a sedimentary formation consisting principally of sandstones, shales, and limestones. The problem is to classify the lithology of each depth interval based on the geophysical properties alone. (Krumbein and Sloss 1963.)

information for judging the class to which an entity likely belongs. The approach is a traditional Bayesian comparison of likelihoods: the greater the Likelihood function for one class relative to another class, based on the actually observed feature data, the greater the probability of the class being correct for the entity on which the feature measurements were made.

7.2.2 Bayesian classification

Let there be n classes to one of which an entity might belong, $w_1, \ldots, w_n$; and m factor measurements upon which the classification is to be made, $\mathbf{x} = \{x_1, \ldots, x_m\}$, in which the vector $\mathbf{x}$ is the set of measurements. The prior probability mass function (pmf) of the entity belong to class w_i is $p(w_i)$, for $i = 1, \ldots, n$, and may be uniform if no prior information is available on classification. The Likelihood (*i.e.*, conditional probability) of the measured vector $\mathbf{x}$ given the class w_i is $L(w_i|\mathbf{x}) = Pr\{\mathbf{x}|w_i\}$. Then, by Bayes' Theorem, the posterior pmf that the entity belongs to class w_i given that the set of measurements $\mathbf{x}$ has been made, follows as,

$$p(w_i|\mathbf{x}) \propto p(w_i)L(w_i|\mathbf{x}) \tag{7.1}$$

in which the normalizing constant of the proportionality is $\Sigma_i p(w_i)L(w_i|\mathbf{x})$. So, the basis of classification is a simple decision under uncertainty: classify an entity as one of the n classes based on the posterior pmf of its membership (Equation (7.1)), such that the appropriate utility function or other criterion of cost and loss in the case of misclassification is optimized.

If the utility (e.g. cost) of each type misclassification is the same, minimizing the probability of misclassification maximizes expected utility. This probability is minimized by assigning the entity to the most probable class (e.g. $max_i\ \{p(w_i|\mathbf{x})\}$). Since the normalizing term in Equation (7.1) is a constant, this criterion is the same as $max_i\ \{p(w_i)L(w_i|\mathbf{x})\}$, or $max_i\ \{ln\, p(w_i) + ln L(w_i|\mathbf{x})\}$, the latter being convenient when the likelihood function is exponential, as it often is in practice. If the utility of each type misclassification is not

the same, but can be expressed as the function $u(i, j)$, which is the utility of deciding that the class of the entity is i when the correct class is j, then the Bayesian optimal classification decision is $max\ E(u) = max_i\{p(w_i|\mathbf{x})u(i, j)\}$. This is a standard statistical decision problem (Figure 7.4).

Consider an example from oil exploration. Shoestring sands are elongated sandstone deposits formed by the burying of near-shore sand bars or of ancient streams. The mode of deposition is important to reconstructing subsurface maps because the plan view shapes of the two deposits differ from one another: bar deposits more resemble low-order polynomials, whereas stream deposits are more sinusoidal and therefore more difficult to project. Several features might be used to differentiate bar deposits from stream deposits, but for simplicity we consider only one geometric feature, the ratio of arc length along known locations of the body to the linear separation of the end points (Figure 7.5), which might be called the 'straightness index' or SI.

From empirical data on shoestring sands, the frequency distributions of straightness index could be evaluated as shown schematically in Figure 7.6. Let the classes be

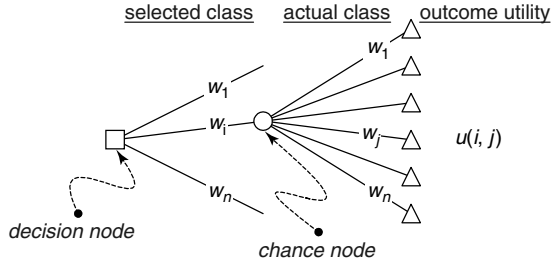


Figure 7.4 Decision tree for classification problem. A class is selected from the n classes, $w_1, \ldots, w_n$, and depending on the actual class, a utility $u(i, j)$ is achieved. The 'best' classification is that which maximized the expected value of the utility over the pmf of the w_i.

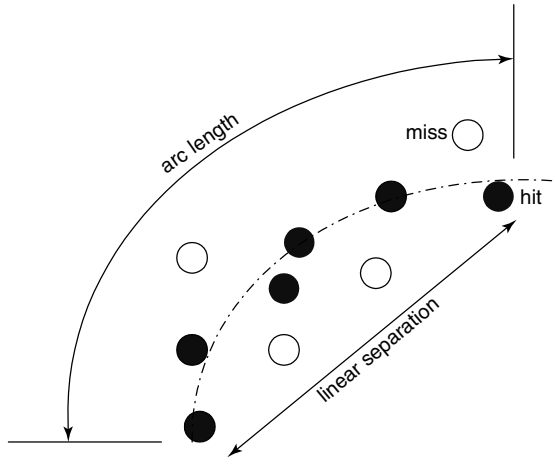


Figure 7.5 Definition of 'straightness index' or *SI* for shoestring sand deposit: *SI* = arc length/linear separation. Filled circles show borings intersecting sand deposit; open circles show borings not intersecting sand deposit. Dashed line shows supposed centerline of sand deposit.

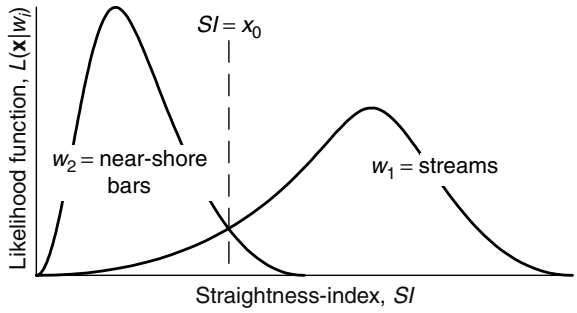


Figure 7.6 Hypothetical likelihoods of straightness index for near-shore bar deposits and for stream deposits.

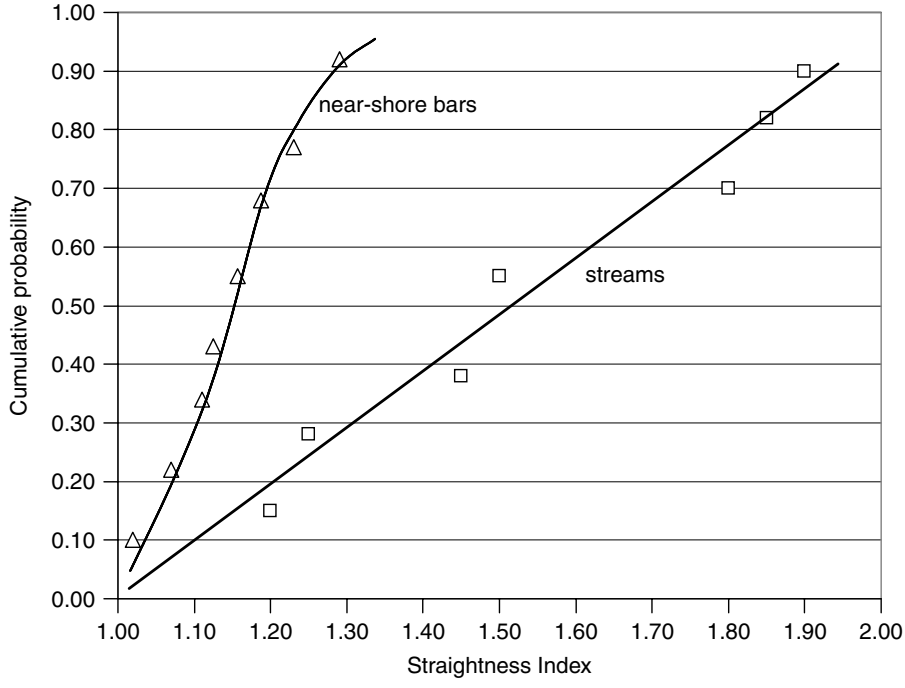


Figure 7.7 Probability distributions (cdf) for straightness index for a limited sample of shoestring sand deposits from near-shore bars and streams (Baecher 1972).

w_1 = stream deposit, and w_2 = near-shore bar deposit. Following our classification procedure, assuming the utilities $u(i, j)$ to be constant, the SI of the formation in question is measured and the formation assigned to class w_1 or w_2 depending on which class maximizes Equation (7.1). Assuming the prior probabilities equal, the decision rule reduces to "assign the formation to w_1 if $SI > x_0$, or to w_2 otherwise."

Figure 7.7 shows empirical data collected for a limited sample of bar and channel type deposits (Baecher 1972). Let w_1 and w_2 be defined as before, and assume

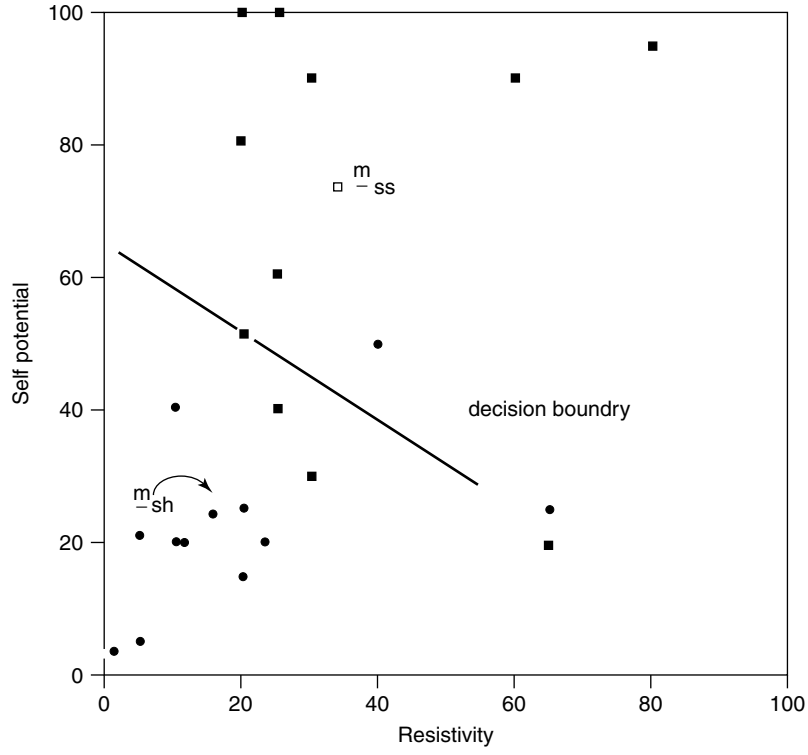


Figure 7.8 Calibrating data for downhole geophysical measurements to classify sandstone and shale with respect to the factors self-potential and resistivity (LeRoy 1951).

$p(w_1) = p(w_2) = 1/2$. Then from Equation (7.1),

$$\begin{aligned} p(w_2|SI = 1.12) &= \frac{p(w_2)p(SI = 1.12|w_2)}{p(w_1)p(SI = 1.12|w_1) + p(w_2)p(SI = 1.12|w_2)} \\ &= 0.83 \\ p(w_1|SI = 1.12) &= 1 - 0.83 = 0.27 \end{aligned} \tag{7.2}$$

Were another shoestring deposit in the same vicinity and of about the same age as the Silverville known to be bar-type, the prior probabilities might no longer be uniform. Assume, $p(w_1) = 1/4$ and $p(w_2) = 3/4$; then

$$\begin{aligned} p(w_2|SI = 1.12) &= 0.94 \\ p(w_1|SI = 1.12) &= 1 - 0.94 = 0.06 \end{aligned} \tag{7.3}$$

which is stronger than with uniform priors, as would be expected.

Posterior probabilities of classifications (Equation (7.1)) depend on the factor or sets of factors used to make the classification decision. Some sets lead to greater deviations of the expected likelihood ratio, $L(w_i|\mathbf{x})/L(w_j|\mathbf{x})$, than do others. A set of factors that leads to large expected deviations is said to have high resolution. One that leads to small expected

deviations is said to have low resolution. Other things equal, factors having high resolution are favored to those having low resolution, because they lead to lower probabilities of misclassification and fewer factor measurements when classification is performed sequentially. The use of multiple factors increases resolution but also increases cost. Sometimes, instead of using the best subgroup of factors from a large number of possible factors, *dimensionality reduction* is used. This simply refers to forming linear combinations of the original factors, thereby reducing their number to a convenient size (Duda *et al*. 2001). The linear discriminant function used in traditional frequentist sampling theory (Kendall *et al*. 1994) is a dimensionality reduction.

7.3 Discriminant Analysis

When the number of items to be classified becomes large, evaluating Equation (7.1) for each of them becomes infeasible. If, however, a map of factor space (the space over which **x** is defined) could be partitioned into regions of similar 'best' classification, then classification could be accomplished simply by noting where the vector **x** plots in this partitioned space. This is the underlying concept of *discriminant analysis*. Discriminant analysis is a widely used statistical tool in many branches of science (Huberty 1994; McLachlan 1992; Rao and Toutenburg 1999).

7.3.1 Bayesian discriminant analysis

Consider the geophysical log of Figure 7.3. Here is a case where a pre-analysis of the best classification of rock type could be made over the space of self-potential and electrical resistivity and the borings automatically logged from the digital records. The borings penetrate interbedded sandstones and shales. The question is how to use the observations of the factor vector,

$$\mathbf{x} = \{\textit{self-potential, electrical resistivity}\} \tag{7.4}$$

to discriminate best between the rock types.

Given sample (i.e. calibration) data from logs for which a physical core was recovered, we first need to evaluate the conditional probability densities of the factors the two lithologies. Much of the difficulty in determining the factors' conditional probability densities rests with the assumptions made at the outset. We can simplify the task by assuming the factors to be statistically independent. The validity of this assumption is sometime questionable, but it reduces the two-dimensional problem to two one-dimensional problems. If we also assume that the density functions are members of particular families of distributions (e.g. Normal, Beta, etc.), the problem reduces to estimating the parameters of the distribution.

A common assumption in practice is that the marginal distributions of the factors (i.e. the marginal Likelihood functions)

$$x_1 = \textit{self-potential} \sim N(\mu_1, \sigma_1) \tag{7.5}$$

$$x_2 = \textit{electrical resistivity} \sim N(\mu_2, \sigma_2) \tag{7.6}$$

are independently Normal, with means μ_1 and μ_2, and variances σ_1^2, and σ_2^2, respectively. The joint pdf of $\mathbf{x} = \{x_1, x_2\}$ is then

$$f(x_1, x_2) \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{7.7}$$

where MN is the multi-Normal distribution, with (vector) mean $\boldsymbol{\mu} = \{\mu_1, \mu_2\}$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \tag{7.8}$$

The parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, being presumably different for the alternative classes, sandstone and shale. The likelihood for an observation, $\mathbf{x}$, is the value of this joint density function at $\boldsymbol{x}$, and the posterior pmf over {shale, sandstone} is found by Equation (7.1).

If the prior pmf is Uniform (i.e. $p(x_1) = p(x_2) = 1/2$) and the utilities of misclassification are constant over i, j, then the classification rule based on the 'most probable class,' (i.e. $max_i\{p(w_i|\mathbf{x})\}$) is to maximize the kernel of the joint Likelihood by minimizing the term in the exponent

$$\text{if} \begin{cases} (\mathbf{x}-\boldsymbol{\mu}_{\text{sh}})^t \boldsymbol{\Sigma}_{\text{sh}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\text{sh}}) < (\mathbf{x}-\boldsymbol{\mu}_{\text{ss}})^t \boldsymbol{\Sigma}_{\text{ss}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\text{ss}}), & \text{then 'shale'} \\ \text{otherwise,} & \text{then 'sandstone'} \end{cases} \tag{7.9}$$

in which the subscript indicates the class to which the statistical parameters apply. Calibrating data from LeRoy (1951) were used to estimate marginal density parameters for the distributions of Equations (7.5) and (7.6). The results, in the relevant units, were $\boldsymbol{\mu}_{\text{sh}} = \{15.3, 24.4\}$, $\boldsymbol{\mu}_{\text{ss}} = \{33.3, 73.7\}$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 243 & 0 \\ 0 & 433 \end{bmatrix} \tag{7.10}$$

Combining these parameter estimates with the decision rule, Equation (7.9) gives the linear rule (partition) shown in Figure 7.8. This is a line passing through the point $(1/2)(\boldsymbol{\mu}_{\text{sh}} + \boldsymbol{\mu}_{\text{ss}})$, and normal to $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_{\text{sh}} - \boldsymbol{\mu}_{\text{ss}})$ (Rao and Toutenburg 1999). Strata whose factor vectors plot above and to the right of the decision boundary are classified as sandstone; those to below and to the left are classified as shale. The error rates for this classification rule can be found from the volumes under the respective pdf's of $\mathbf{x}$ on the misclassified side of the partition. That is, the error rate for misclassifying shale as sandstone is the volume under the joint pdf for $\boldsymbol{x}$ given shale above and to the right of the line. Breitung (1994) provides algorithms and approximations for calculating such volumes.

Relaxation of the assumptions of Normality and common covariance matrix, and treatment of other variants such as non-linear and non-parametric discrimination are treated by McLachlan (1992), among others. The use of sequential methods in classification problems has received considerable attention, but is beyond the present scope. Fu (1968a, 1968b, 1970) discusses the theory of sequential classification as applied to optical character recognition, and McLachlan (1992) presents a more current survey of statistical

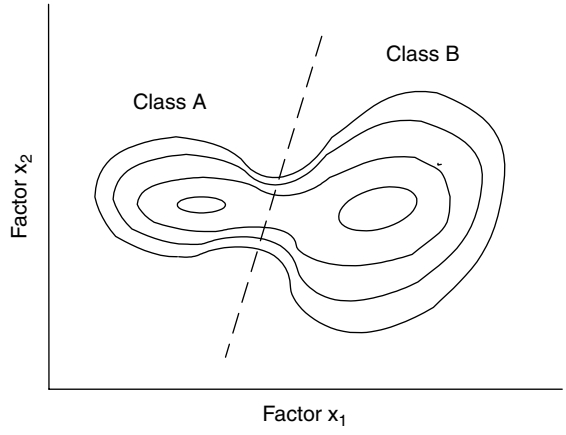


Figure 7.9 In one or two dimensions the determination of clusters within factor space may be a simple matter of inspection. This is not the case for higher dimensions.

pattern recognition. Baecher (1972) discusses the use of sequential methods in geological exploration.

In practice, the population parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ must themselves be estimated from the sample, and for this, a straightforward inference via Bayes' theorem is used (Chapter 4) to develop a obtain a joint pdf $f(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i |$ data). This pdf can be used to form a predictive pdf over the classification equations (Aitchison and Dunsmore 1975). Duda *et al*. (2001) treat this problem in more detail.

7.3.2 Frequentist discriminant analysis

Discriminant analysis within the frequentist school is conceptually similar to the Bayesian treatment, except that (1) prior probabilities on class membership are generally disallowed, and (2) in considering the statistical issues attending imprecisely known likelihood functions for the respective classes, probabilities are placed on the repeated sampling activity of estimating the discriminant function (i.e. the dividing boundary) rather than on the probabilities of correctly or incorrectly classifying a particular observation.

Confusion sometimes arises in comparing the Bayesian and frequentist approaches to discriminant analysis. This comes about because discriminant analysis deals with two types of probability. The discriminant (i.e. classification) problem itself is formulated as a probabilistic model, much as, say, Bernoulli trials are a probabilistic model. Two questions arise upon which Bayesian and frequentist approaches to the problem differ. This first is, given that we know the relevant pdf's of the class populations (i.e. moments of the distributions), what statements can be made about the probable class membership of a unique observation. Within the Bayesian school, the answer is clear: assess a prior pmf, informative or not, and apply Bayes' Theorem. Within the frequentist school the answer is the same if a prior pmf is known, which it usually is not; otherwise, the answer is to express the uncertainty of classification as error rates over many trials, conditional on the true but unknown classification.

The second point of confusion is the treatment of statistical uncertainty in the parameters or distributional form of the pdf's describing the classes. Since we usually have only

limited numbers of observations from which to infer these pdf's, the estimates of moments of the respective distributions are uncertain. The Bayesian approach treats this statistical uncertainty just as it does the statistical uncertainty in a Bernoullian model: it expresses the parameter uncertainty in pdf's and then forms a predictive distribution by integrating over the parameter uncertainty. The frequentist approach treats this statistical uncertainty, also, just as it does the statistical uncertainty in a Bernoullian model: it considers the sampling variability of possible sets of observations and then assesses best estimators and confidence intervals.

As in the Bayesian example, the goal is to establish a mathematical equation, using the factor values as arguments, that in some 'best' way discriminates among populations. The common approach has been to consider discrimination as related to regression analysis, in that a (usually) linear function of the factor values,

$$V = b_0 + b_1 x_1 + \cdots + b_n x_n = \mathbf{b}^t \mathbf{x} \tag{7.11}$$

is sought that best distinguishes among populations. This value V is sometimes called the *discriminant score*. Following Fisher's original work (Fisher 1936), the regression coefficients are taken to be values that maximize the distance between the respective means of two populations, chosen to make the ratio of between-group variance relative to the within-group variance as large as possible (Figure 7.9). The discriminant score is found by multiplying the discriminant weights associated with each factor by the corresponding factor values, and summing over the thus weighted factors of Equation (7.11), creating a projection as shown in Figure 7.10.

The original solution to the two-class discriminate analysis problem is due to Fisher, who developed the method in dealing with biological taxonomy. Assuming, as in earlier subsections, that the class means and covariance matrices are $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, respectively, and that the latter have common value, $\boldsymbol{\Sigma}$, Fisher suggested maximizing the ratio of the

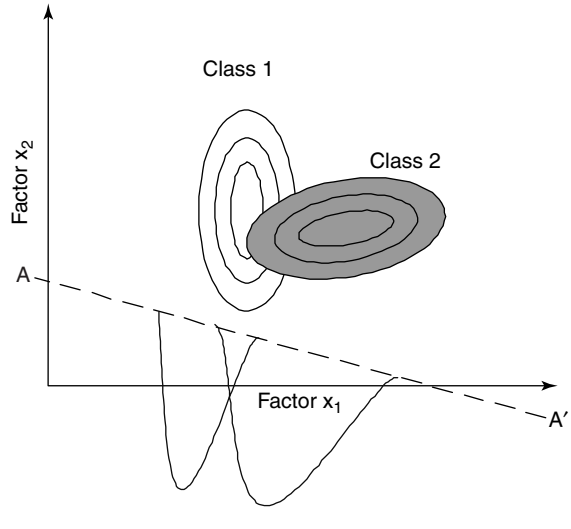


Figure 7.10 Illustration of the two-class discriminant analysis.

squared distance between class means to the within class variance

$$\Delta = \frac{(\mathbf{b}^t \boldsymbol{\mu}_1 - \mathbf{b}^t \boldsymbol{\mu}_2)}{\mathbf{b}^t \boldsymbol{\Sigma} \mathbf{b}^t} \tag{7.12}$$

over the weights $\mathbf{b}$. The formulation makes no assumptions on the distributional forms of the underlying classes, other than that they have a common covariance matrix, $\boldsymbol{\Sigma}$. The solution is to set $\mathbf{b}$ proportional to $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

This result can also be obtained from an approach originally due to Wald (1944) in which the probability ratio between the pdf's of two Normally distributed classes with common covariance matrix is set to equal a constant. This leads to the following decision rule for assigning an observation to class 1:

$$\text{if} \begin{cases} \mathbf{x}^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) > k, & \text{then Class 1} \\ \text{otherwise,} & \text{then Class 2} \end{cases} \tag{7.13}$$

in which k is a constant, usually, $k = 0$. Note, the first term in Equation (7.13) is just, $\mathbf{b}^t[\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)]$, Fisher's *linear discriminant function*. Building on Fisher's original work, which does not rest of parametric distributional assumptions for the pdf's of the respective classes, we should expect that discrimination according to Equation (7.13) is reasonably robust, which it is. Also, as should be expected from Figure 7.10, the solution is not unique, only the ratios among the weights are. Anderson (1984) provides greater detail on the derivation of this and related mathematical results, while Everitt and Dunn (2001) provide a more tutorial introduction.

Using this rule, the probabilities of misclassification of observations in class 1 and class 2, respectively, are (Anderson 1984)

$$\begin{aligned} P(2|1) &= \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2}(z - \tfrac{1}{2}\alpha)^2/\alpha\right) dz \\ &= \int_{-\infty}^{(c-1/2\alpha)/\sqrt{\alpha}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) dy \end{aligned} \tag{7.14}$$

$$\begin{aligned} P(1|2) &= \int_{c}^{\infty} \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2}\left(z + \frac{1}{2}\alpha\right)^2/\alpha\right) dz \\ &= \int_{(c+1/2\alpha)/\sqrt{\alpha}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) dy \end{aligned} \tag{7.15}$$

in which $\alpha = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. The advantage of the transformed integrals is that they can be evaluated from tables of the Normal distribution.

In practice, the population parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ must themselves be estimated from the sample, and for this, the sample mean and sample covariance matrix are used. The statistical sampling properties of these estimators and the resulting sampling properties of the estimates of the discriminant boundary are discussed by Kendall *et al.* (1994).

7.3.3 Categorizing liquefaction data

One problem of discriminant analysis and classification that has seen a good deal of attention in the geotechnical literature is that of liquefaction. When a saturated, fine, loose

sand is subjected to vibratory motions, such as those caused by an earthquake, the sand can liquefy, losing almost all its shear strength and possibly flowing like a dense fluid. Structures supported on such liquefying soil may fail catastrophically, and, indeed, the literature contains many references to liquefaction either during or immediately following earthquakes (Richter 1958). The Niigata earthquake of 1964 was notable for dramatic photographs, many times reprinted, of completely intact apartment buildings capsized and overturned due to liquefaction beneath their shallow foundations. The question is: how to decide whether a soil formation is subject to liquefaction in the event of an earthquake? This is the question addressed using statistical methods of discriminant analysis and classification.

In deciding whether a formation is subject to liquefaction, one normally considers two things: the level of ground shaking to which the formation is likely to be subject (i.e. the 'load'), and the dynamic resistance of the soil to the stresses induced by that ground shaking (i.e. the 'resistance'). The load is related to the intensity of ground shaking, which in turn is usually measured by peak ground surface acceleration of the design earthquake. Other measures could be used, but this is the most common. To compare field observations with measurements taken in the laboratory under controlled conditions, the peak ground acceleration may be modified by a stress averaging factor, usually taken to be 0.65, which relates the irregular time history of earthquake acceleration to the steady cyclic loading used in laboratory tests.

The resistance is related to soil density, fines content, and other factors. This is usually empirically associated to Standard Penetration Test (SPT) blow count. These SPT blow counts were converted into apparent relative densities by the relationships of Gibbs and Holtz (1957) or others or otherwise manipulated to remove the effects of different overburden pressures at different depths. In recent years, the data sets using SPT to predict liquefaction potential have been augmented by, and compared to, data based on Cone Penetration Test (CPT) and various geophysical measurements, yet SPT remains the most common *in situ* measurement upon which resistance to liquefaction is predicted.

In 1971, based on a series of studies conducted during the late 1960's (Seed and Idriss 1967; Seed and Lee 1966; Seed and Peacock 1971), Seed and Idriss (1971) published a famous paper reporting empirical data on observed liquefaction and non-liquefaction events that occurred during or after a large number of earthquakes, comparing those observations against cyclic shear stress ratio (CSR) and dynamic resistance, in turn a function of modified SPT blow count expressed as relative density, D_R (Figure 7.11). This created a two-dimensional chart showing two clusters of data, the 'black dots' and 'white dots' in the figure, corresponding, respectively, to the liquefied and non-liquefied cases. The authors separated the two clusters by visual inspection to yield a discriminant line for classifying future cases. This line errs to the side of minimizing the number of liquefaction data misclassified while allowing the error rate for misclassifying non-liquefaction cases to rise. That is, this is not the equi-probable discriminant curve, but a conservative one biased to enlarge the area of factor space allocated to 'liquefaction.'

Christian and Swiger (1975, 1976) analyzed the same liquefaction problem using linear discriminant analysis on a data set of 39 cases, overlapping but not exactly the same as the data set used by Seed and Idriss. Because the various sites represented in the data set had different depths of overburdened and different elevations of the water table, the peak ground surface acceleration values were multiplied by the ratio of the total vertical stress at the critical point to the effective vertical stress at the same point. This ratio was

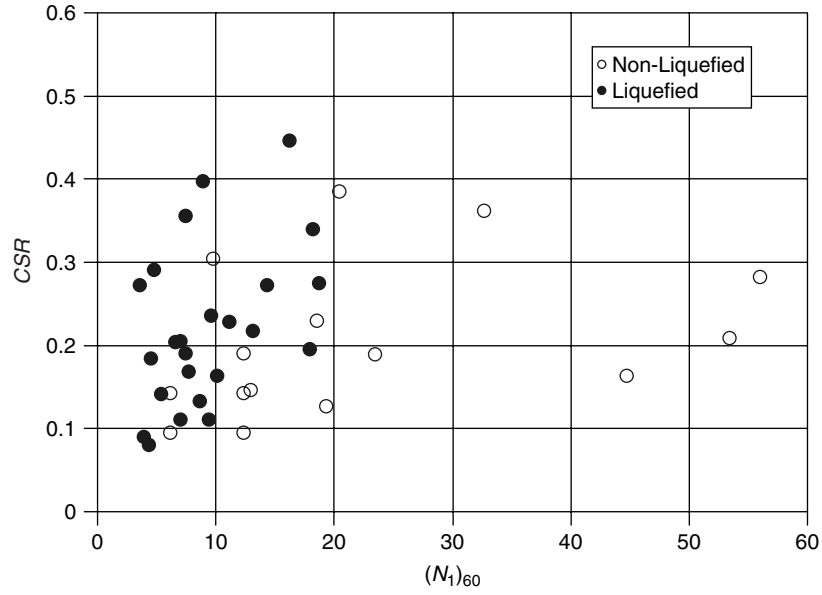


Figure 7.11 Empirical data showing observed cases of liquefaction and non-liquefaction plotted against dynamic stress and cyclic resistance (after Seed and Idriss 1971 and Christian and Swiger 1975). (Christian, John T., and Swiger, William F., 1975, 'Statistics of liquefaction and SPT results', *Journal of the Geotechnical Engineering Division*, reproduced by permission of the American Society of Engineers.)

thought to account approximately for the effect of overburden above the saturated material being analyzed. This ratio had a maximum value of 2.0, if the ground water level was at the surface, and a minimum of 1.0. The averaging factor of 0.65 was not included. Soil densities were taken as 'apparent relative densities' using the Gibbs and Holtz relations.

Probability paper plots of the $\mathbf{X} = \{ln(CSR), (N_1)_{60}\}$ data, separated into liquefaction and non-liquefaction cases are shown in Figure 7.12, suggesting that the data are well-modeled by logNormal distributions. Sample statistics for the two data sets are shown in Table 7.1, suggesting also that the covariance matrices for the two data sets are similar. The discriminant function value V was then calculated by Equation (7.13), yielding the regression form

$$V = -1.4964 \ln A + 4.8165 \ln D_r - 21.6221 \tag{7.16}$$

The value of V is chosen from the probability that a new point actually belonging to class 2 will seem to fall on the class 1 side of the line, e.g. a case of actual liquefaction having data that would indicate no liquefaction. This, in turn, in found from Equation (7.15). The procedure is to select a desired value of the probability of mislabeling and to invert Equation (7.15) to obtain the corresponding value of V.

Figure 7.13 shows the 20% probability line obtained from a discriminant analysis based on Christian and Swiger's data (note, on a logarithmic grid this is a straight line). The averaging factor of 0.65 has been re-introduced for this plot for compatibility with other similar plots.

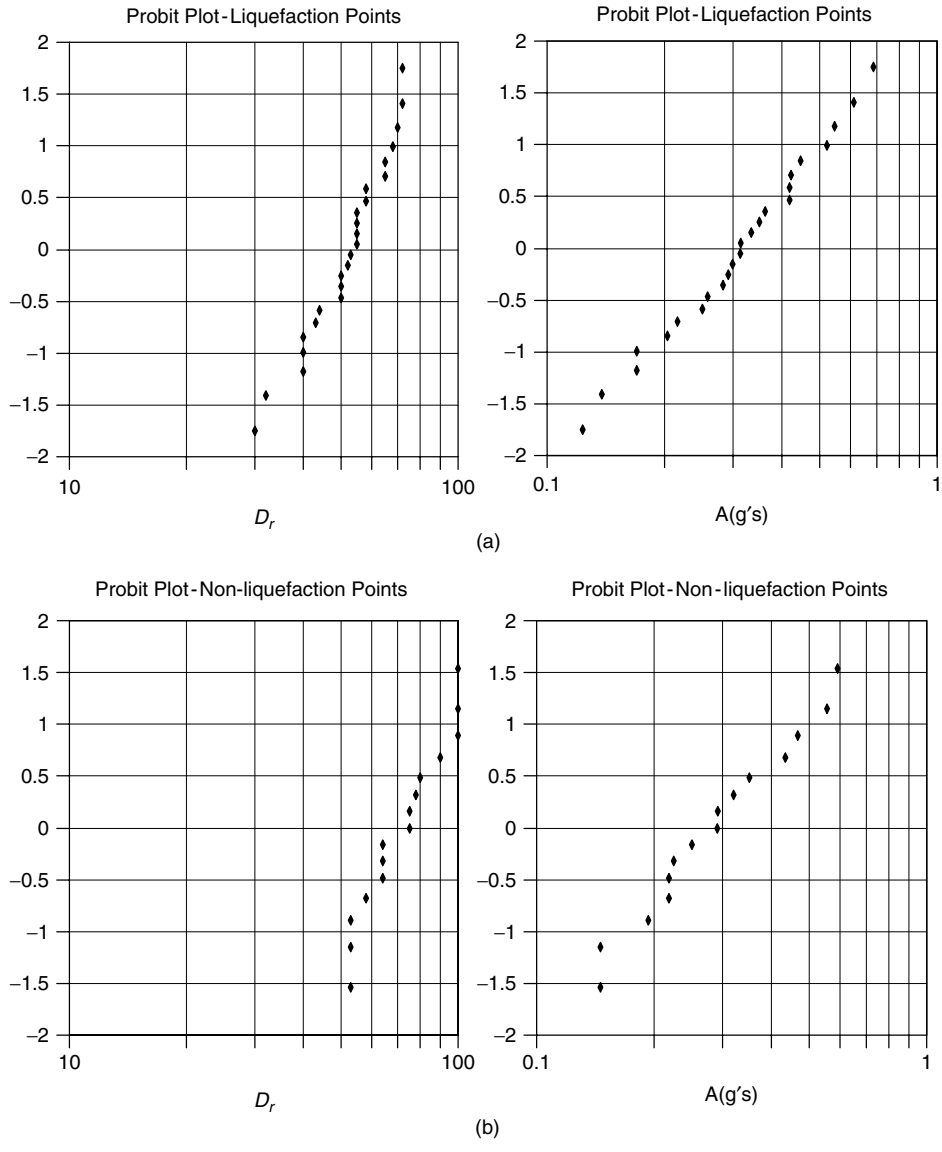


Figure 7.12 Cumulative probability plots of (a) liquefied and (b) non-liquefied data, suggesting logNormality of the frequency distributions (after Christian and Swiger 1975). (Christian, John T., and Swiger, William F., 1975, 'Statistics of liquefaction and SPT results', *Journal of the Geotechnical Engineering Division*, reproduced by permission of the American Society of Engineers.)

7.3.4 Logistic regression

Another approach to classifying observations among classes is *logistic regression*. This approach has become popular in recent years because it requires fewer assumptions than traditional discriminant analysis and therefore is considered more robust. For the multi-Normal case with common covariance matrix, linear logistic regression gives the same

Table 7.1 Sample statistics for liquefied and non-liquefied data of Figure 7.12 (after Christian and Swiger 1975). (Christian, John T., and Swiger, William F., 1975, 'Statistics of liquefaction and SPT results', *Journal of the Geotechnical Engineering Division*, reproduced by permission of the American Society of Engineers.)

		Category 1	Category 2
Description (1)	All data (2)	No Liquefaction (3)	liquefaction (4)
Mean in ln A	−1.20982	−1.26815	−1.17336
Equivalent A, in g	0.298	0.281	0.309
Mean in ln D_R	4.07164	4.27580	3.94404
Equivalent D_R, as a percentage	58.7	71.9	51.6
Variance of ln A	0.19517	0.18511	0.20617
Standard deviation	0.44178	0.43024	0.45406
Coefficient of variation	−0.36516	−0.33927	−0.38697
Variation of Ln D_R	0.08161	0.05460	0.05742
Standard deviation	0.28567	0.23367	0.23962
Coefficient of variation	0.07016	0.05465	0.06076
Covariance of ln A, ln D_R	0.04096	0.05525	0.04666
Correlation coefficient	0.32456	0.54957	0.42885

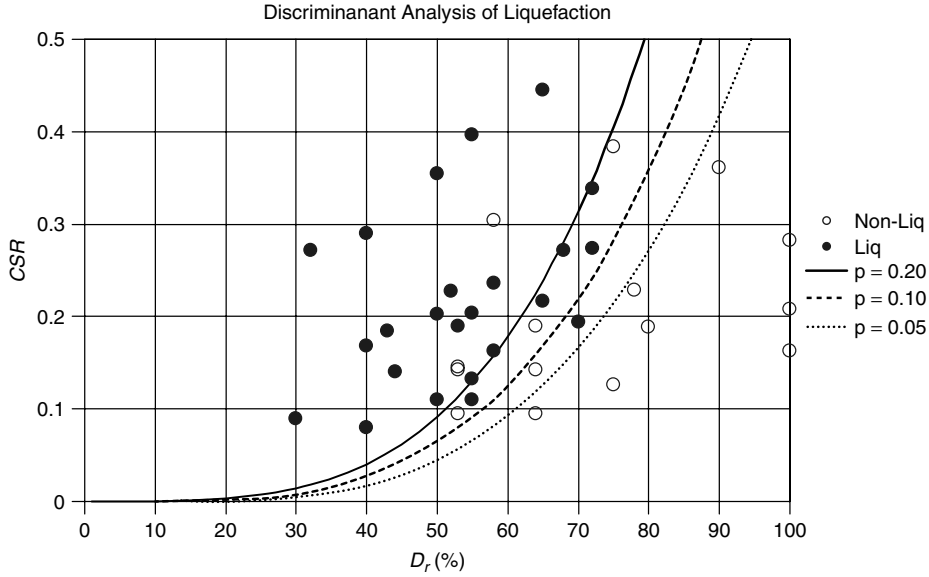


Figure 7.13 Discriminant analysis of liquefaction using Christian and Swiger data updated to $(N_1)_{60}$. Curve is for probability 0.2. (Christian, John T., and Swiger, William F., 1975, 'Statistics of liquefaction and SPT results', *Journal of the Geotechnical Engineering Division*, reproduced by permission of the American Society of Engineers.)

result as traditional discriminant analysis (Goldstein and Dillon 1978). Logistic regression is conceptually related to the Bernoulli process of Chapter 3. A Bernoulli process consists of n repeated trials, each of which is classified as a 'success' or 'failure.' The probability

of success, π, remains constant across trials, and outcomes of the trials are independent of one another. The number of successes (or failures) in n trials follows a Binomial pdf.

Consider a simple mechanical device with some empirical failure rate, π, denominated in probability of failure per use. We can use observed numbers of failures, x, in given numbers of tests, n, to estimate π from the Binomial distribution

$$f(x|n) = \binom{x}{n} \pi^x \pi^{n-x} \tag{7.17}$$

in which x is the number of failures in n tests. What if the failure rate depended on the temperature at which the device operated, with higher failure rates associated with higher temperatures? We could divide the empirical data into bins by temperature, and then individually estimate π for each bin, and, in effect, develop a functional relationship between failure rate and temperature, $\pi = \mathrm{g}(T)$. This is essentially what logistic regression analysis attempts to do, but by forming a regression model through which the probabilities are approximated. This model has the form

$$\pi = g(\mathbf{X}) \tag{7.18}$$

in which π is probability, and $\mathbf{X} = \{X_1, \ldots, X_n\}$ is a vector of parameters upon which the probability depends.

The first thing to note, of course, is that probabilities range from zero to one, while regression analysis fits a model to variables that vary from $-\infty$ to $+\infty$. Thus, the probabilities – for us, the classification probabilities – are transformed by forming the *log odds-ratio*, usually called, the *logit*,[1]

$$Q = \ln\left(\frac{\pi}{1-\pi}\right) \tag{7.19}$$

to give values that vary within $(-\infty, +\infty)$. The logit or log odds-ratio is set equal to a linear regression of the parameters $\mathbf{X}$

$$Q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{7.20}$$

in which the parameters $\mathbf{X} = \{X_1, \ldots, X_n\}$ are the factors influencing the regressed probability, and the scalar weights $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_n\}$ are regression coefficients estimated from data. Thus

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{7.21}$$

From which it follows, by taking the exponential of each side, that

$$\pi = \frac{1}{1 + \exp(-\mathbf{x}^T \boldsymbol{\beta})} \tag{7.22}$$

[1] The etymology of the term, *logit* is often traced to the 1944 paper by Berkson in the *Journal of the American Statistical Association, XXXIX*: 361, in which the author says, "Instead of the observations q_i we deal with their logits $l_i = ln(p_i/q_i)$. [*Note*] I use this term for $ln\ p/q$ following Bliss, who called the analogous function which is linear on x for the normal curve 'probit'."

The term $\mathbf{x}^T\boldsymbol{\beta}$ is called the *linear predictor*, but this regression could also be, for example, polynomial.

A quick look at the shape of a logistic regression curve (Figure 7.14) suggests why it is so useful for the problem of estimating probabilities. It is non-linear, it asymptotically approaches zero at the low end and unity at the high end, and the coefficients can be estimated using standard maximum likelihood techniques of regression analysis, which are well known.

Estimating the parameters of logistic regressions is slightly more complex than the corresponding estimates of linear discriminant functions in that simple moment estimators do not suffice. The usual manner is by means of maximum likelihood estimators, which work for both Bayesian and frequentist approaches. The likelihood function for the vector of parameters $\boldsymbol{\beta}$, given a set of m independent observations, $\mathbf{X} = \{x_1, \ldots x_i, \ldots x_m\}$, is the product of the respective probabilities of the observations, given the parameters

$$\begin{aligned} L[\boldsymbol{\beta}|x_1, x_2, \ldots, x_m] &= \prod_{j=1}^{k} f_X(x_j|\boldsymbol{\beta}) \\ &= \prod_{j=1}^{k} [P_1(x_j|\boldsymbol{\beta})]^{y_j} [P_2(x_j|\boldsymbol{\beta})]^{1-y_j} \\ &= \prod_{j=1}^{k} \left[\frac{1}{1+\exp(-\mathbf{x}^t\boldsymbol{\beta})}\right]^{y_j} \left[1 - \frac{1}{1+\exp(-\mathbf{x}^t\boldsymbol{\beta})}\right]^{1-y_f} \end{aligned} \tag{7.23}$$

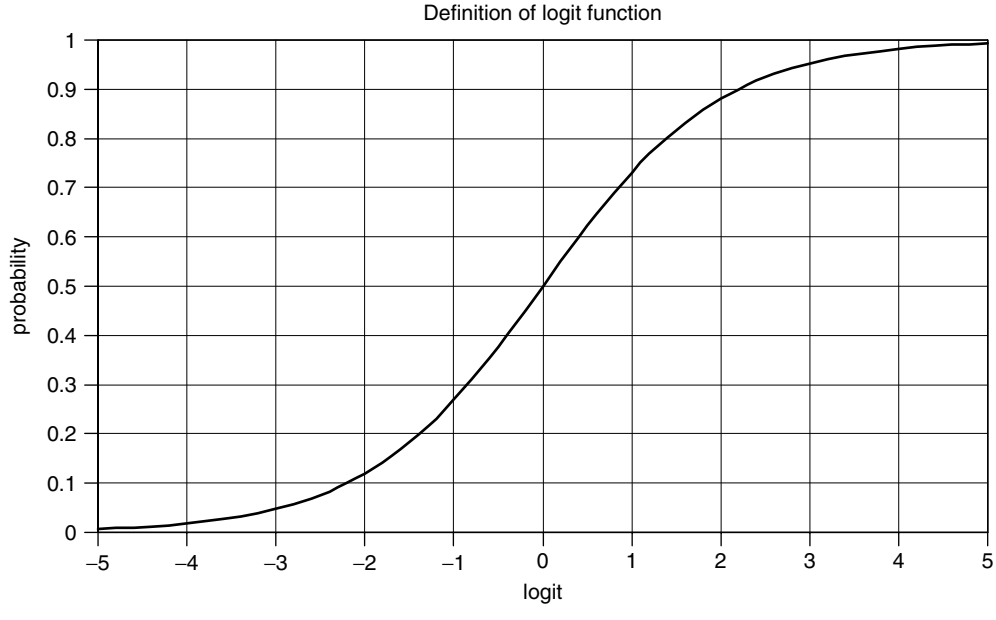


Figure 7.14 Typical shape of a logistic regression relationship.

where y_j, $j = 1, \ldots, k$, is a dichotomous variable (0 or 1), depending on whether the point j is a member of class 1 or of class 2. Equation (7.23) can be optimized by direct enumeration (e.g. numerically in a spreadsheet), particularly if m is of modest size. For larger sample sizes, however, it may be more convenient to work with the logLikelihood function

$$\log L = \sum_{j=1}^{k} [y_j \ln P_1(\mathbf{x}_j|\boldsymbol{\beta}) + (1 - y_j) \ln P_2(\mathbf{x}_j|\boldsymbol{\beta})] \tag{7.24}$$

in which the subscript on the vector of factor values, $\mathbf{x}_j$, refers to the j$^{\text{th}}$ observation. The maximum likelihood estimators are then found by maximizing $\log L$ over $\boldsymbol{\beta}$.

Logit coefficients need to be interpreted with care, as they are not as intuitive as, say, common regression coefficients, to which most of us are accustomed. Logit coefficients are not the rate of change in class probability with change in x_i; they are the rate of change in log odds-ratio with change in x_i.[2] Thus, e^{β_i} is the effect of changes in x_i on the odds-ratio. For example, if $e^{\beta_i} = 2$, then a unit change in x_i makes an event twice as likely. Odds ratios equal to unity mean there is a equal chance that a event or its complement will occur; and negative coefficients imply odds ratios less than one.

Following the early work of Christian and Swiger (1975; 1976) on applying discriminant analysis to a relatively limited set of liquefaction observations, several studies have used logistic analysis, as well as other techniques, based on larger databases (Cetin *et al.* 2002; Haldar and Tang 1979; Haymsfield 1999; Ho and Kavazanjian 1986; Juang *et al.* 2002; Liao *et al.* 1988; Loh *et al.* 1995; Tang and Angulo 1996; Veneziano and Liao 1984; Youd *et al.* 2001). Obviously, the results are dependent on the details of the data used. For present purposes it is interesting to compare the results of discriminant analysis and logistic regression for the same liquefaction data. The original Christian and Swiger data were updated to use $(N_1)_{60}$ rather than $\ln D_R$, and both methods were applied to the same database. Figure 7.15 shows the curves for 20% probability. They are close throughout and nearly the same in the region of the data points.

For this case the probability of mislabeling if the deposit does not liquefy is 0.70. If we assume a non-informative prior (i.e. the prior probability of liquefaction is 0.5 reflecting ignorance) and one obtains data falling below and to the right of the 20% line, the posterior probability of liquefaction based on Bayes' Theorem is 0.22. Figure 7.16 shows the posterior probabilities for the full range of priors when discriminant lines with probabilities of 0.1, 0.2, and 0.3 are used.

[2] Interestingly, log odds-ratios also play an important role in Bayesian estimation, especially in the assessment of subjective probabilities (Chapter 21), since Bayes' Theorem can easily be restated as a relationship among posterior and prior log odds-ratios, and corresponding log Likelihood ratios,

$$\frac{f(a|b)}{f(b|a)} = \frac{f(a)L(b|a)}{f(b)L(a|b)}$$

$$\log\left(\frac{f(a)}{f(b)}\right) = \log\left(\frac{L(b|a)}{L(a|b)}\right)$$

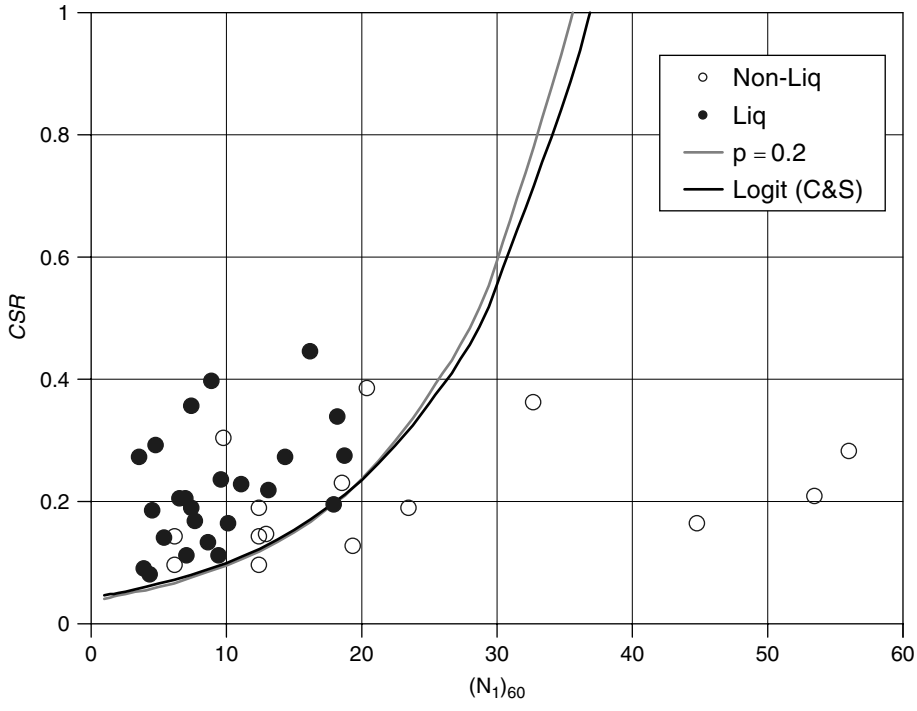


Figure 7.15 Discriminant analysis of liquefaction using Christian and Swiger data. Logistic curve is for probability 0.2; it falls almost identically on the discriminant curve for probability 0.2.

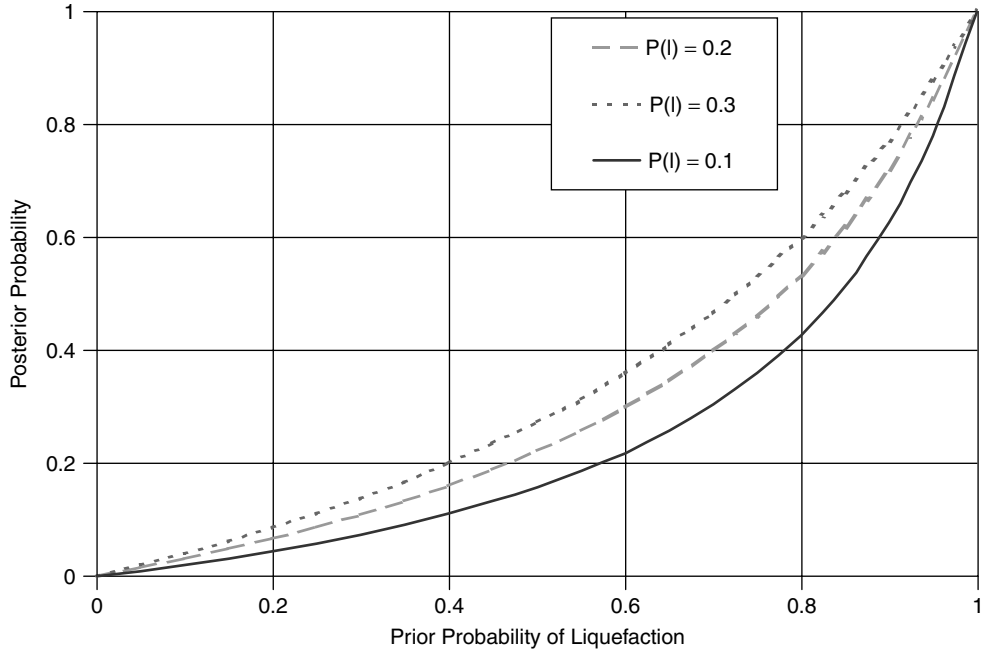


Figure 7.16 Bayesian updating based on discriminant lines with probabilities of 0.1, 0.2, and 0.3 based on liquefaction data used by Christian and Swiger.

Both discriminant analysis and logistic regression are widely used in classification studies, although the latter has gained favor in recent years due, in part, to the looser assumptions it requires and the broader array of situations it can handle. If the classes to be decided among are Normally distributed with the same covariance matrix, then both discriminant analysis and logistic regression yield the same answer, and the discriminant analysis is arguable easier to use. In many applications, however, the factoring variables, **X** may be non-Normal, or at least one of the x_i is a qualitative variable (thereby eliminating Normality). Under these circumstances, logistic regression is the preferred method. Further discussion of the relative properties and uses of discriminant analysis and logistic regression is provided by Press and Wilson (1978) and by Dillon and Goldstein (1984).

7.3.5 Clustering

When one encounters a new problem, the number of classes into which data should be grouped and their locations in factor space are generally unknown. In one or two dimensions the number of classes can be adequately determined by inspection of the data plot (Figure 7.9). With computer visualization, this is increasingly becoming the case for 3D factor spaces as well. Unfortunately, for higher dimensions, visual inspection is of little use.

One approach to this clustering problem is to project data against lines, planes, or volumes in which visual inspection is useful, and to isolate modes or clusters within those lower-dimension spaces. The question then becomes which lines, planes, or spaces to project the data against to obtain the greatest discriminatory benefit? Figure 7.17 shows this conundrum. Projecting against the line B-B' provides little benefit, while projecting against A-A' provides nearly the best discrimination.

The way to find the best projections is to seek those lines, planes, etc. in which the projected data have the maximum variance (this can be thought of as the lines, planes, etc. of maximum moment of inertia of the data). Let there be k observations in the data set, and let the ith factor value of the jth observation be $x_{i,j}$. The location, s_j, of the jth observation when projected into a line S is the scalar product of **x** and **a**, where **a** is the vector along line S

$$s_j = \mathbf{x} \cdot \mathbf{a} = \sum_{i=1}^{m} x_{ij} a_i \tag{7.25}$$

and m is the dimension of the factor space (e.g. number of factors measured on each observation). The mean locations for the entire data set are

$$E[s_j] = \frac{1}{k} \sum_{i=1}^{k} s_j = \sum_{i=1}^{k} E[x_j] a_i \tag{7.26}$$

in which $E[s_j]$ is the expected value of the ith component of the observed factor vectors. The variance along S is proportional to

$$Var[s_j] \propto \sum_{i=1}^{k} (s_i - E[s])^2 \tag{7.27}$$

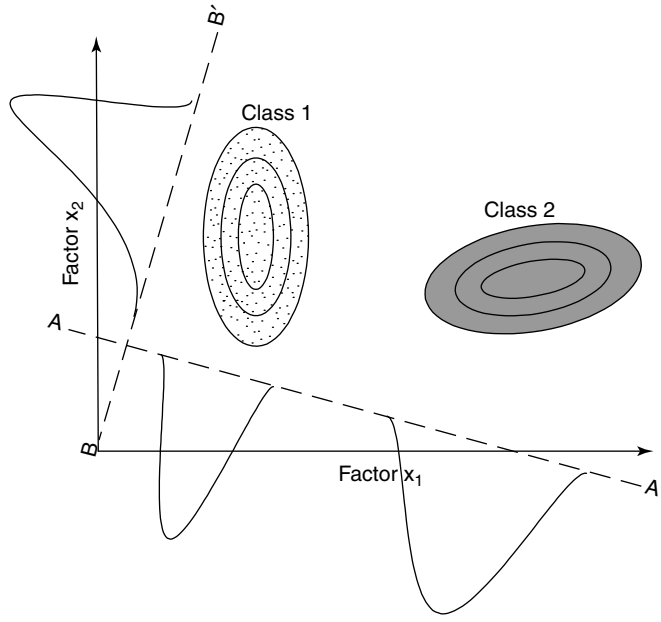


Figure 7.17 Projecting data into well-chosen lower-dimensional spaces allows visual inspection to be used in clustering classes of data.

which can be made a large as desired by allowing the a_i to be large. If, however, we seek to maximize Equation (7.27) subject to the constraint that Σa_i^2 is small, the criterion becomes

$$\max \left[\frac{\sum_{i=1}^{k} (s_i - E[s])^2}{\sum_{i=1}^{m} a_i^2} \right] \tag{7.28}$$

and the problem reduces to finding the eigenvector associated with the largest eigenvalue of the matrix (Mattson and Dammann, 1965)

$$\{V_{ij}\} = \sum_{r=1}^{k} (x_{ir} - E[x_i])(x_{jr} - E[x_j]) \tag{7.29}$$

which, in turn, is proportional to the data covariance matrix.

The algorithm for sequential clustering a data set into multiple classes is to apply repeatedly this eigenvector calculation: (1) compute the largest eigenvalue of the data set covariance matrix and determine the eigenvector corresponding to this eigenvalue; (2) project the multivariate data set onto the eigenvector to obtain a one-dimensional density distribution; (3) partition the data set at any nodes in the projected density which approach zero probability density; (4) if no such nodes exits, repeat the process using the eigenvector corresponding to the second largest eigenvalue (and if all eigenvectors are considered with no partitioning appearing, stop and conclude that only one class exists);

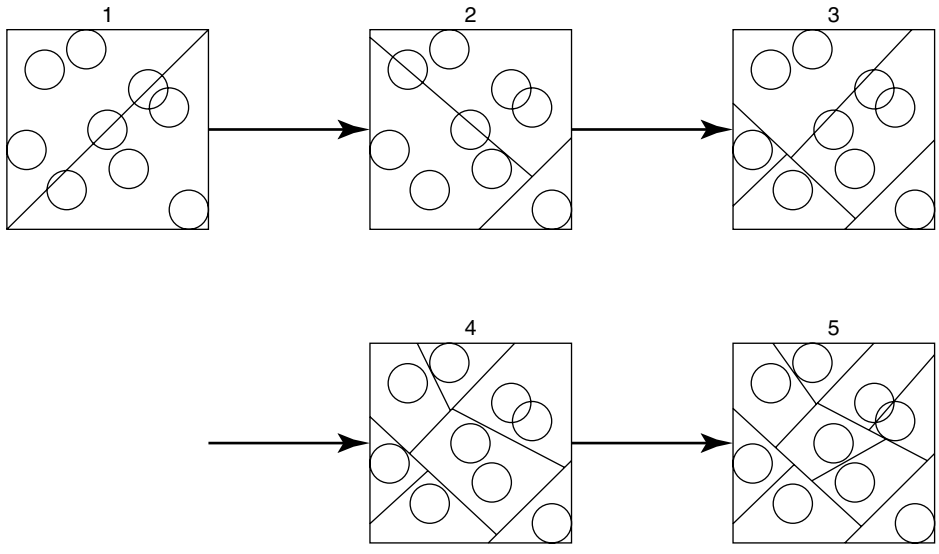


Figure 7.18 Sequentially clustered factor data in multiple classes using the algorithm of Mattson and Dammann (1965).

(5) to partition the data set further, isolate the data in each partitioned class and repeat the procedure for each (Figure 7.18).

7.4 Mapping

Geological mapping seems more 'art' than 'science,' but the rise of remote sensing and other data intensive technologies has given rise to the need for and experience with automated procedures for mapping. Clearly, from a reliability perspective, the uncertainties of site geometry may be equally important to those of material properties, as discussed by Dowding (1979).

Geological mapping assigns each point at the site to a geological class, which in turn may be a formation, a stratigraphic bed, a lithological type, or some other discrete unit. Thus, one might think of mapping as the estimation of a discrete field, analogous to the continuous fields discussed in Chapters 8 through 10, but for which the random process generates a discrete rather than continuous variable at each point in space. In geostatistics there is a newly-minted term for this, as for most things; this is called an *indicator* field (Matheron 1971), and is discussed in more detail in Chapter 11. The estimation in mapping leads to a probability mass function (pmf) at each point rather than to a probability density function (pdf) as would be the case with a continuous field. This outcome is a set of ordered n-tuples, one at each point in space describing the probabilities of possible geological classes.

7.4.1 Probabilistic mapping

A probabilistic mapping scheme based on uncertainties in knowledge of site geology should satisfy four requirements: (1) it should be simple to conceive and use, (2) one

should be able to assign subjective prior probabilities to its components, (3) it should allow probabilistic rules for inferring the classifications of unobserved locations, and (4) the uncertainties in mapping should be directly useful in design decisions.

The use of discrete stochastic fields (Bernoulli fields in the two-color case) for mapping satisfies these four requirements: They are reasonably well-known in the stochastic process literature, pmf's (possibly subjective) may be assigned *a priori* to individual locations, inference (statistical estimation) rules can be constructed for interpolation and extrapolation from known locations, and uncertainties may be directly interfaced with design decisions. Design decisions can be related to the probabilities of composition in regions influenced by a proposed project, and these probabilities in turn can be read directly from a map of the type described. Such a map includes the probabilities of unlikely as well as likely composition. Other schemes, such as assigning each point to the most probable class and associating with it a probability of misclassification, generally do not have this facility of directly informing the reliability assessment.

7.4.2 Discrete stochastic fields

Consider a site composed of only two geological materials – a two-class, or *black and white*, region. We can assign to each point at the site an ordered pair $\mathbf{p} = \{p_1, p_2\}$ whose first term is the probability that the point belongs to class 1 and whose second term is the probability that the point belongs to class 2, where $p_1 + p_2 = 1$ (all points must be classified as either class 1 or class 2). The terms of this pair are the elements of a pmf of class membership for the point in question, defined as

$$p(\mathbf{c}) = p(c_1, c_2) \tag{7.30}$$

in which $\mathbf{c} = \{c_1, c_2\}$ are the two classes (e.g. black and white), and $p(\mathbf{c})$ is the pmf. For observed points the pmf would either be (1,0) or (0,1), assuming no observation error. For unobserved points the pmf could be anything between these extremes, subject to the constraint that $p_1 + p_2 = 1$. The dimension of this assignment can be expanded to any required size, in which case one would assign an ordered n-tuple $p(\mathbf{c}) = p(c_1, c_2, \ldots, c_n)$ rather than ordered pair (and, in principle, could include a term representing the probability that the location belongs to an unenumerated or 'unknown' class).

Unless the function that assigns pmf's to locations can be approximated by some continuous mathematical function, discretizing the site into elemental areas or pixels (sometimes called *quadrats* in the geography literature) of homogeneous classification is a convenience. These elements may be made as small as needed to adequately satisfy the assumption of homogeneous classification, and this approximation is no worse here than in other discretized methods of engineering analysis.

The classification pmf at a given point or within a given element might be estimated on the basis of two considerations: The first is the classification of near-by elements; the second is the presence of trends in local geology. Figure 7.19 illustrates these proximity and trend considerations. Figure 7.19(a) shows a clear case in which proximity considerations dominate. All the points around the question mark point are of a single class, and thus it seems reasonable that the unobserved point would more likely than not be of the same class. Figure 7.19(b) shows a somewhat clear case in which trend considerations dominate. Although the unobserved point has an equal number of each class as its

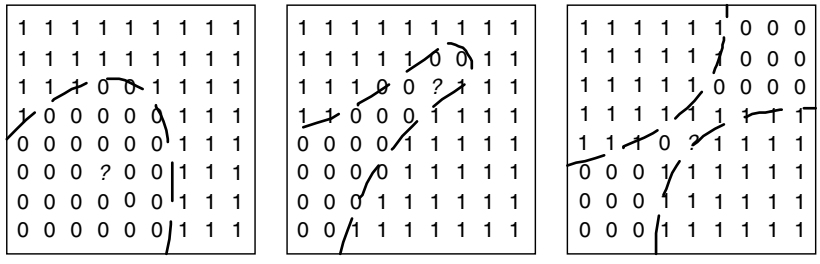


Figure 7.19 Influence of proximity and trend considerations in mapping discrete fields (here a two-class or black-and-white field).

immediate neighbors, the trending shape of the formation suggests one class over another. Figure 7.19(c) shows a case in which proximity and trend considerations conflict. It is not clear to which class to assign the highest probability for the unobserved point. Theories exist for modeling each of these two considerations, proximity and tends, individually, but not yet for combining them in a single model.

7.4.3 Proximity rules

Simple *nearest neighbor rules* appear frequently in civil engineering practice (e.g., assigning areas of influence to rainfall measurements), but they usually specify only the most probable classification. The typical nearest-neighbor rule would be: "Assign to any unobserved location the class to which the nearest observed point belongs." This leads to the familiar Thiessen polygon maps so common in hydrology, ore reserve estimation, and many other applications (Dingman 2002; Koch and Link 2002; Okabe 2000). We want more detailed inferences than this. At a minimum, we want a rule that specifies the most probable class and its probability. However, better would be a rule that specifies the entire pmf over possible classes for unobserved locations.

Proximity rules are based on the assumption that locations close together are more likely to be of the same class than are locations far apart. Switzer (1967) and Matérn (1947) have suggested the use of a function related to autocorrelation (i.e. spatial correlation) that specifies the probability that a location will be of class i, given that an observation at distance d is also of class i. The value of this function would equal 1.0 at $d = 0$, and the prior probability at $d = \infty$. Matérn has shown that the function is convex in the vicinity of $d = 0$, and Switzer has proposed the form

$$p_{i|i}(d) = p_i + (1 - p_i)e^{-\alpha d} \tag{7.31}$$

in which p_i is the prior probability of class i, or the spatial frequency of the class in a frequency sense, and α is a decay constant related to the autocorrelation function of the discrete field (Figure 7.20). Equation (7.31) assumes isotropic autocorrelation (i.e. correlation depends on separation distance but not direction), but can be expanded to account for anisotropy in a form such as

$$p_{i|i}(d) = p_i + (1 - p_i)e^{-\sqrt{\alpha^2 d_x^2 + \beta^2 d_y^2}} \tag{7.32}$$

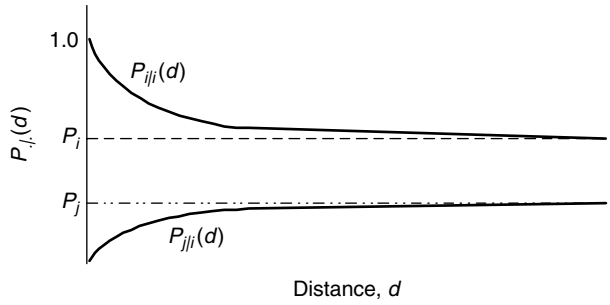


Figure 7.20 Switzer (1967) proximity rule for discrete mapping.

in which α and β are directional decay constants, and d_x and d_y are separation distances parallel to the principal axes of an anisotropic correlation structure (i.e. preferred regional directions in the geology).

Since the terms of the pmf must sum to unity, the sum of the complementary functions $P_{j|i}(d)$ for $j \neq i$ is related to the original decay function by

$$\sum_j p_{j|i}(d) = 1 - p_{i|i}(d) \tag{7.33}$$

and the two-color map (Figure 7.20)

$$p_{j|i}(d) = 1 - p_{i|i}(d) \tag{7.34}$$

Much of the earlier work has only considered the partitioning, not the determination of probabilities, and has only considered the influence of nearest neighbors, but proximity rules can be extended to include all sample data by application of Bayes' Theorem. Given r sampled locations, $z_1, \ldots, z_r$, each at some distance $d(z_j)$ from an unexplored location, the posterior probability of class i at an unexplored location is

$$p(i|\mathbf{z}) \propto p(i)L(\mathbf{z}|i) \propto p(i)\Pi_{j=1}^{r} L(z_j|i) \tag{7.35}$$

presuming the observations to be independent, which of course they are not, because they are spatially situated with respect to one another as well. This assumption of independence appears not to have a large effect on the resulting inference; however, it can be easily relaxed by assigning a covariance matrix among the observations the terms of which are related to the separations among observations. An example is shown in Figure 7.21.

The proximity rule, can only be considered an empirical model whose parameters are determined for the site in question. Several authors have discussed the possibility of developing random process models that would predict the form of Equation (7.31), but no realistic model has yet appeared. A random process model is one that describes the generating mechanism for a particular physical phenomenon in probabilistic terms, from which is described the theoretically 'correct' stochastic behavior of the phenomenon. This is in contrast to an empirical model that simply fits a convenient, smooth analytical function to observed data with no theoretical basis for choosing the particular function. Lack of a random process model need not concern us as long as empirical data are

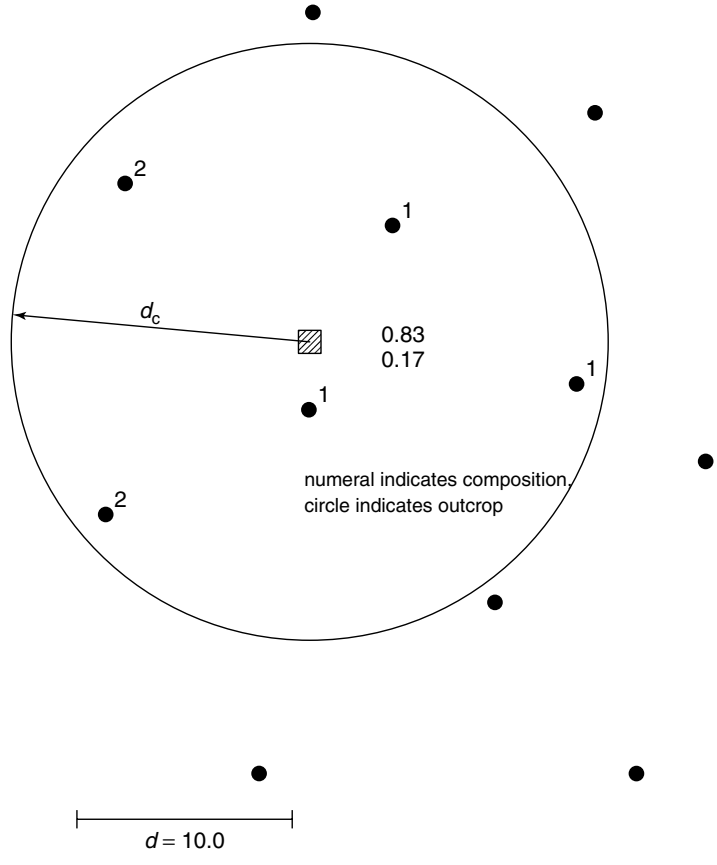


Figure 7.21 Example of proximity rule applied to a two-class site.

available from which to construct a function; however, in a given situation we may find need to exchange Switzer's exponential form for some other.

We have assumed in our discussion that autocorrelation and other properties of the process (e.g. the relative frequency of different classes of materials) are stationary. This means that Equation (7.31) is the same for every location at the site (i.e. it depends only on separation distance and direction, not location), which in certain contexts may prove to be a poor assumption. One suspects the density of certain geologic bodies may be greater in some regions of a site than In others (i.e. bodies may be 'clumped'). For example, altered bodies may occur in clumps due to the structural nature of the region. Non-stationarity would cause the parameters (and potentially even the shape) of Equation (7.31) to depend on location.

Estimation of the parameters of the proximity model can be made by maximizing the joint Likelihood function of the data.

$$\max_{p_i,\alpha} L(p_i, \alpha|\mathbf{z}) = \prod_{u=1}^{k} \prod_{v=i+1}^{k} [p_i + (1 - p_i)e^{-\alpha d_{u,v}}]^{y_{u,v}} [1 - (p_i + (1 - p_i)e^{-\alpha d_{u,v}})]^{y_{u,v}} \tag{7.36}$$

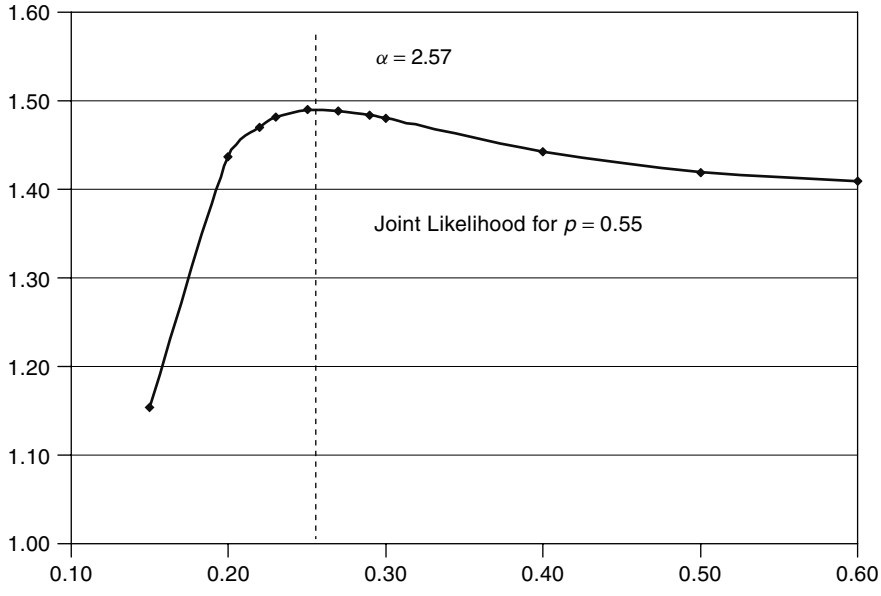


Figure 7.22 Numerical optimization of joint likelihood function for the observed data of Figure 7.21 using moment estimator of p_i as the starting point.

in which $y_{u,v}$ is a Boolean variable equaling zero when the classes of u and v are the same, and unity otherwise. This maximization is easily obtained numerically, starting from a moment estimator of $\hat{p}_i = n_i/N$, in which n_i is the number of class i observations and N is the total number. For the data of Figure 7.21, numerical optimization of Equation (7.36) is shown in Figure 7.22.

7.5 Carrying Out a Discriminant or Logistic Analysis

Both discriminant and logistic analyses are techniques for developing an expression that separates data into two categories; these might be liquefaction versus non-liquefaction, success versus failure, or any other discrete division into two groups. The raw data consist of a set of observations that fall into one category or the other. Both analyses work best when there are approximately equal numbers of points in each category. Each point is described by two or more variables. In the case of an analysis of liquefaction, the categories would cases where liquefaction was observed and cases where it was not.

The analyst's first task is to compile a set of data points approximately evenly distributed between the two categories. Next the analyst must determine the functional form in which each variable will enter the analysis; that is, are the values of the data to be used directly, or are some functions of the values to be used? For example, in the case of the liquefaction analysis we could consider CSR and $(N_1)_{60}$ directly, but we might consider ln CSR to be a better variable. Such choices will depend on the physics of the problem at hand and judgment.

After establishing the functional forms to be used, the analyst prepares a table containing columns of the values for each of the data points. Further work is greatly simplified if

the table is sorted into two blocks, say first the non-liquefaction cases and then the liquefaction cases.

7.5.1 Discriminant Analysis

Once the table of data points has been prepared, it is necessary to determine whether the functional forms of the data to be used in the discriminant analysis satisfy the necessary assumptions. First of these is that the functional form of the data should be Normally distributed. There are analytical statistical tests for this, but the most satisfying is usually to plot the data on probability paper or in a quantile-quantile (Q-Q) plot. Each variable in each category and in the data as a whole should be plotted. Thus, for the liquefaction case, six plots are required: ln (CSR) and $(N_1)_{60}$ for each of the two categories and ln (CSR) and $(N_1)_{60}$ for the entire data set. If the data seem to conform to a Normal distribution, the analysis proceeds to the next step. If they do not, it helps to try a different functional form, but, if nothing avails, discriminant analysis is not appropriate for this data set.

The next step is to compute for each of the two categories and for the data as a whole the mean of each variable, the variance of each variable, and the covariance between all the variables. The variances and covariances for the two categories should be close, but they do not have to be equal.

The notation adopted is that the variables are listed in a vector $\mathbf{X}$, whose components are the functional forms of the m individual variables $\{X_1, X_2, \ldots, X_m\}$. In the liquefaction case $\mathbf{X} = \{\ln(CSR), (N_1)_{60}\}$. Then we define

$$\overline{\mathbf{X}}^{(1)} = \{\, E[X_1] \quad E[X_2] \quad \cdots \quad E[X_m]\} \; category\ 1 \tag{7.37}$$

$$\overline{\mathbf{X}}^{(2)} = \{\, E[X_1] \quad E[X_2] \quad \cdots \quad E[X_m]\} \; category\ 2 \tag{7.38}$$

$$\Sigma = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_m] \\ Cov[X_1, X_2] & Var[X_2] & \cdots & Cov[X_2, X_m] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_1, X_m] & Cov[X_2, X_m] & \cdots & Var[X_m] \end{bmatrix} \tag{7.39}$$

The covariance matrix $\mathbf{\Sigma}$ is best calculated as the weighted average of the covariance matrices of the two categories, the weights being proportional to the number of data points in each category.

The discriminant function V is then calculated by

$$V = \left[\mathbf{X} - \tfrac{1}{2}\left(\overline{\mathbf{X}}^{(1)} + \overline{\mathbf{X}}^{(2)}\right)\right] \cdot \Sigma^{-1} \cdot \left(\overline{\mathbf{X}}^{(1)} - \overline{\mathbf{X}}^{(2)}\right) \tag{7.40}$$

When the matrix multiplication is carried out, an expression of the following form results:

$$V = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{7.41}$$

The βs are all determined; the task is now to find V.

The value of V is chosen from the probability that a new point actually belonging to category 2 will seem to fall on the category 1 side of the line, e.g. a case of actual

liquefaction having data that would indicate no liquefaction. This, in turn, depends on a parameter α defined as

$$\alpha = \left(\overline{\mathbf{X}}^{(1)} - \mathbf{X}^{(2)}\right)^T \cdot \sum^{-1} \cdot \left(\overline{\mathbf{X}}^{(1)} - \mathbf{X}^{(2)}\right) \tag{7.42}$$

The probability is then

$$P[1|2] = \int_{\frac{V+0.5\alpha}{\sqrt{\alpha}}}^{\infty} \frac{1}{\sqrt{2\pi\alpha}} \exp\left[-\frac{t^2}{2}\right] dt \tag{7.43}$$

which is simply the complementary CDF of the standard Normal distribution. Note that if we were to compute $P[2|1]$, V would be the negative of Equation (7.41), and the equation corresponding to Equation (7.43) would be

$$P[2|1] = \int_{-\infty}^{\frac{-V+0.5\alpha}{\sqrt{\alpha}}} \frac{1}{\sqrt{2\pi\alpha}} \exp\left[-\frac{t^2}{2}\right] dt \tag{7.44}$$

The procedure is then to select a value of the probability of mis-labeling and to invert Equation (7.43) to obtain the value of V. This then is inserted into Equation (7.41) to obtain an equation relating the variables at the desired probability level.

7.5.2 Logistic Regression

Logistic regression finds values of the coefficients in

$$Q_L = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{7.45}$$

which is similar to Equation (7.41) except that Q_L is related to the probability P_1 of being in category 1 (say liquefaction) by

$$Q_L = \ln\left(\frac{P_1}{1 - P_1}\right) \tag{7.46}$$

The notation Q_L has been used in liquefaction analysis, but other notation may apply to other fields. It follows that

$$\begin{aligned} P_1 &= \frac{1}{1 + \exp(-Q_L)} \\ P_2 &= 1 - P_1 = \frac{1}{1 + \exp(Q_L)} \end{aligned} \tag{7.47}$$

The likelihood function for a set of parameters $\boldsymbol{\beta}$, given a set of observations is the product of the probabilities of the observations, given the parameters:

$$\begin{aligned} L[\boldsymbol{\beta}|x_1, x_2, \ldots, x_m] &= \prod_m f_X(x_i|\boldsymbol{\beta}) \\ &= \prod_{i=1}^{m} [P_1(x_i)]^{Y_i} [P_2(x_i)]^{1-Y_i} \end{aligned} \tag{7.48}$$

where Y_i is 1 or 0 depending on whether point i is in category 1 or 2. Equation (7.48) means that, if the β's were known, one would evaluate the likelihood by working point by point, evaluating Q_L for each point from Equation (7.45), using Equations (7.47) to evaluate P_1 or P_2, depending on which category the point belongs in, and multiplying the results. It is easier to work with the logarithm of the likelihood:

$$\log L = \sum_{i=1}^{m} [Y_i \ln P_1(x_i) + (1 - Y_i) \ln P_2(x_i)] \tag{7.49}$$

The problem is then to select the $\boldsymbol{\beta}$'s so that the log likelihood is maximized.

The following computational procedure implements this analysis:

1. Set up initial estimates of the $\boldsymbol{\beta}$'s in some locations in a spreadsheet or mathematical analysis software.
2. For each point set up an expression for Q_L based on the variable $\boldsymbol{\beta}$'s.
3. For each point falling in category 1 set up an expression for P_1 based on the first of Equation (7.49), and for each point falling in category 2 set up an expression for P_2 based on the second of Equation (7.49).
4. Using the solving capabilities in the software, maximize the sum of the expressions in step 3 by iterating on the $\boldsymbol{\beta}$'s.

8 Soil Variability

Measurements of soil properties along with detailed observation are what differentiated the emerging discipline of soil mechanics of the early twentieth century from traditional groundworks as practiced since at least the time of the great projects of Egypt and the Near East long before the common era. These measurements suggested considerable variability in soil properties, not only from site to site and stratum to stratum, but even within apparently homogeneous deposits at a single site.

In recent years, our ability to measure the properties of natural soils has expanded rapidly, although it might be argued that our corresponding abilities to understand what many of these measurements mean for engineering predictions, or to deal with the variability among our measurements, have not kept up. As mentioned in Chapter 6, the trend in site characterization research seems to be to seek the next magical device for unlocking the secrets of *in situ* soil properties, rather than attempting to deal more analytically with the data we already collect (Dowding 1979). This chapter summarizes some of the things we know about the variability of natural soils and how that variability can be captured in probabilistic descriptions for use in reliability or risk analysis.

The Mexico City soil profile illustrated in Figure 8.1, taken from Lambe and Whitman (1979), suggests the variability often observed in comparatively homogenous soil formations. The profile where this boring was taken consists mostly of a thick, soft lacustrine clay layer, from about 6–33 m in depth. The clay is overlain by a thin layer of silt and, over that, a thick layer of fill. Thin sand lenses appear throughout the clay, with thick layers of sand and silt beneath. The measured properties of the clay are more or less uniform with depth but have coefficients of variation from about 20% to as much as 50%, depending on the property. The strength parameters are least variable, while the water contents are most variable. This latter observation should not be surprising, for strengths are usually measured on much larger specimens that are water contents or soil densities.

The sandy soil profile of Figure 8.2, well-known to soil engineers as it appeared in the first edition of Terzaghi and Peck (1948), displays considerably more erratic stratification and property data. Whereas the Mexico City profile comprises a single thick layer, the Chicopee profile contains many layers and lenses or inclusions. The soil permeabilities vary dramatically from layer to layer and lens to lens.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

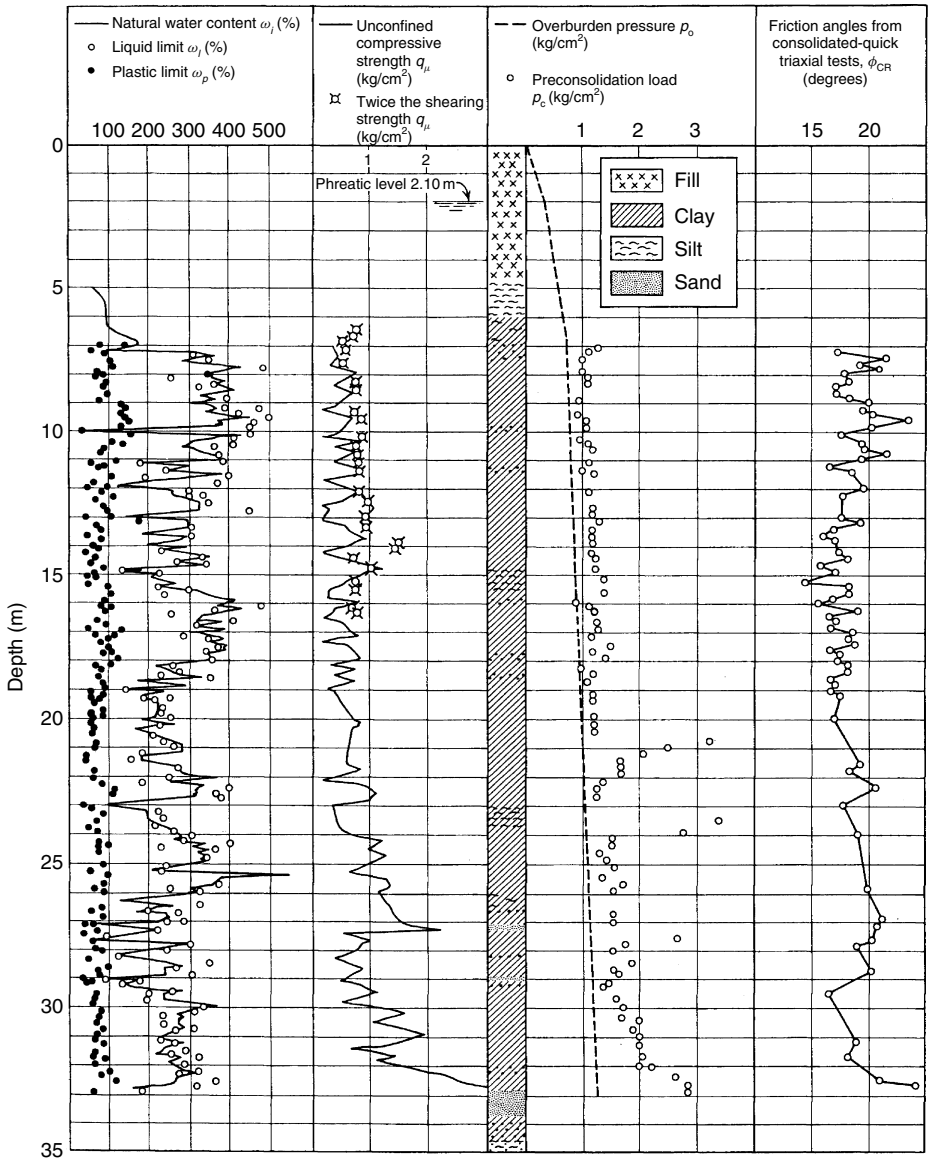


Figure 8.1 Mechanical properties of clays of the Valley of Mexico at a typical spot of the City. (Lambe, T. W. and Whitman, R. V., 1969. *Soil Mechanics*. New York, John Wiley & Sons, © John Wiley & Sons 1969, reproduced with permission.)

The variability in soil properties encountered on any project is inextricably related to the particular site and to a specific regional geology. It is neither easy nor wise to apply typical values of soil property variability from other sites in performing a reliability analysis. One should no sooner do this than apply 'typical friction angles' or 'typical undrained strength ratios' from other sites. The discussion of soil variability in this chapter intends to suggest conditions that might be encountered in practice. The range of variabilities from one site

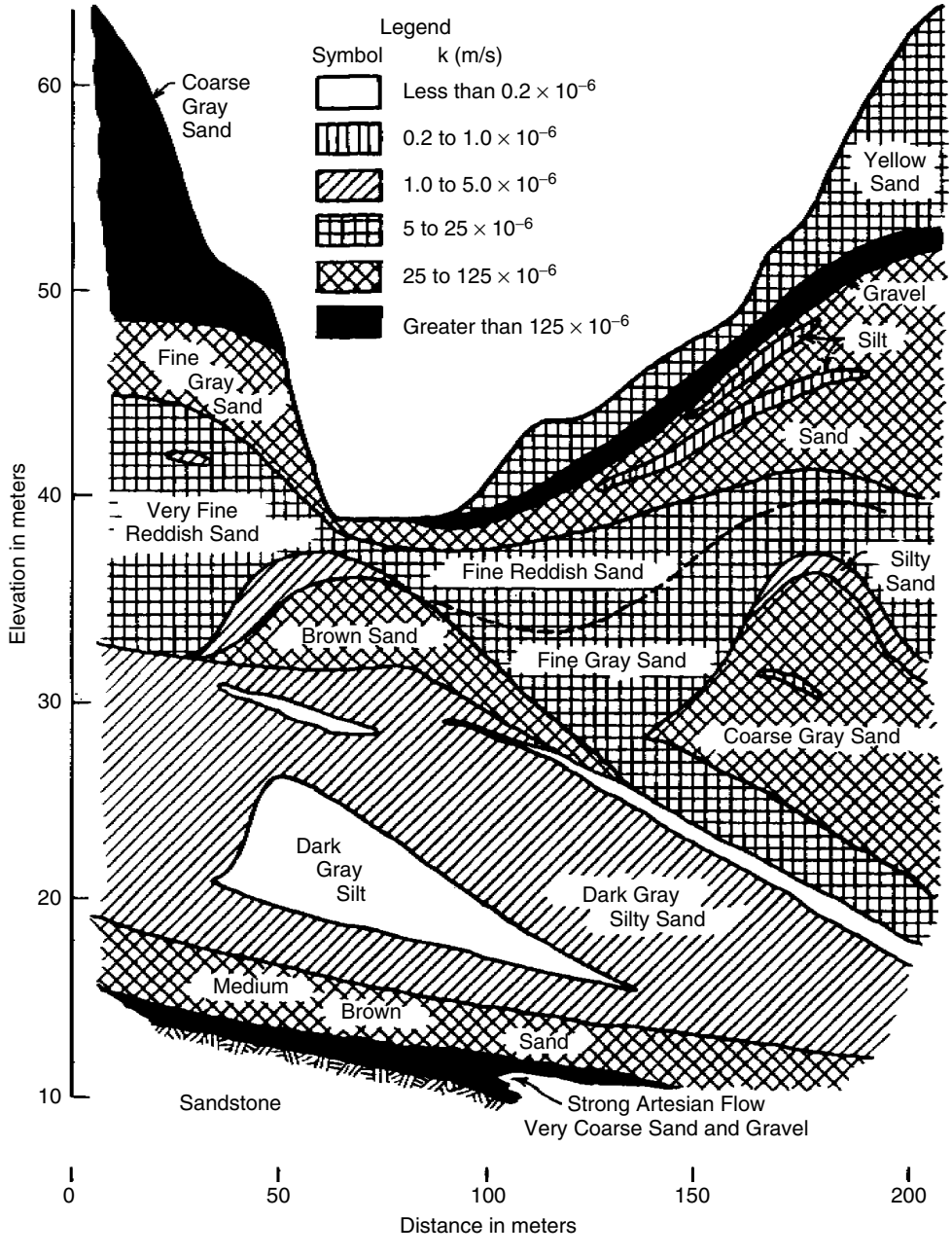


Figure 8.2 Permeability profile of relatively homogeneous glacial delta deposit near Chicopee, Massachusetts. (Terzaghi, K., Peck, R. B. and Mesri, G., 1996. *Soil Mechanics in Engineering Practice*. 3rd ed., New York, John Wiley & Sons, © John Wiley & Sons 1996, reproduced with permission.)

to another is great, and the data one finds reported in the literature often compound real variability with testing errors. In practice, soil variability measures should always be based on site-specific data.

We begin this chapter by discussing the variability in some elementary soil properties. Much of the description of the properties themselves is well known to engineers trained in soil mechanics, but it may not be familiar to people working in other fields. Therefore, we ask the indulgence of those versed in soil mechanics as we review some basic material.

8.1 Soil Properties

Soils have some properties, such as color, that are useful for differentiating one soil from another, other properties, such as compressibility, that are useful for predicting mechanical behavior, and yet others, such as electrical resistivity, that are useful (to geotechnical engineers at least) only in specialized situations.

In this chapter we consider three categories of physical properties: (1) index and classification properties, (2) mechanical properties of intact specimens of soil and rock, and (3) *in situ* properties of soil masses. We use the term *soil mass* to describe the complex structure of geological materials and their discontinuities that makeup and control the physical behavior of geological formations as nature created them in the field.

The extent to which data on these various properties vary is suggested by Table 8.1, taken from Lee *et al*. (1983), in which the authors have compiled reported coefficients of variation for a broad variety of soil properties. The ranges of these reported values are sometimes quite wide and can only be considered as suggestive of conditions on any specific project. Lacasse and Nadim (1996) have also published a table of suggested coefficients of variation for various soil properties (Table 8.2).

Table 8.1 Coefficients of variation for soil engineering test (Lee *et al*. 1983)

Test	Reported *COV*(%)	Source	Standard	Remarks
Absorption	25	1	25	
Air voids	16–30	4,16	20	
Angle of friction (sands)	5–15	2,3,20,21	10	
Angle of friction (clays)	12–56	2,20,21	–	Wide variation
Bitumen content	5–8	4,16	6	
CBR	17–58	1,4	25	
Cement content	13–30	4,5	20	
Clay content	9–70	1,2,14,15,22	25	
Cohesion (undrained) (clays)	20–50	2,20	30	
Cohesion (undrained) (sands)	25–30	2	30	
Compaction (OMC)	11–43	1,6,12	20,40	Lower value, clay sails; higher value, sands and gravels

Table 8.1 (*continued*)

Test	Reported *COV*(%)	Source	Standard	Remarks
Compaction (MDD)	1–7	2,4,6,12,22	5	
Compressibility	18–73	2,15,21,22	30	
Consolidation coefficient	25–100	2,15,22	50	
Crushing value	8–14	1		
Density (apparent or true)	1–10	1.2,12,20,22	3	
Elastic modulus	2–42	17,18	30	
Elongation	18–33	1	25	
Flakiness	13–40	1,6	25	
Flow (bitumen)	11–15	4.16	15	
Grading (means)	9–31	I	25	
Grading (slopes)	19–37	1	30	On a *Rosin–Rammler* plot
Levels (e.g. pavement)	50–300	7,11	200	
Linear shrinkage	57–135	1,6	100	Refers to gravel and crushed rock; will be lower for soils
Liquid limit	2–48	2,4.10,12.14,15,20,22	10	
Los Angeles abrasion	31	1	30	
Moisture content	6–63	4,5,9,15,20,21,22	15	Is gamma distributed
compaction, "OMC"				
Permeability	200–300	2	300	
Plastic limit	9–29	2,12,14,15,20	1	
Plasticity index	7–79	1,4,6,10,12,15,20	30,70	Lower value clay soils; Higher value sandy, gravelly soils
Gravelly soils				
Sand content	1–43	2,4.5,8,10,15	20	
Sand equivalent	25	1	25	
Specific gravity	see	Density		
Stability (bitumen)	13–19	4	20	
Standard Penetration	Test	27–85	21,22	30
Sulphate soundness	92	1	100	Is gamma distributed
Tensile strength	15–29	18	20	
Thickness (concrete pavement)	3–4	4	4	
Thickness (AC pavement)7–16	4,13	15		

(*continued overleaf*)

Table 8.1 (*continued*)

Test	Reported *COV*(%)	Source	Standard	Remarks
Thickness (stabilization work)	11–13	4,6	15	
Unconfined compressive strength	6–100	17,19,22	40	
Void ratio	13–42	2,4,15,21	25	

1. (Ingles and Noble 1975)
2. (Lumb 1974)
3. (Hoeg and Murarka 1974)
4. (Kuhn 1971)
5. (Ingles and Metcalf 1973)
6. (Leach 1976)
7. (Auff 1978)
8. (Kennedy and Hudson 1968)
9. (Wu 1974)
10. (Mitchell 1993)
11. (Murphy and Grahame 1976)
12. (Sherwood 1970)
13. (Hoede-Keyser 1970)
14. (Minty *et al.* 1979)
15. (Corotis *et al.* 1975)
16. (Cranley 1969)
17. (Otte 1978)
18. (Kennedy 1978)
19. (Morse 1971)
20. (Singh 1971)
21. (Schultze 1975)
22. (Stamatopoulos and Kotzias 1975)

Table 8.2 Coefficients of variation of different soil properties for summary data from (a) Lacasse and Nadim (1996), and (b) Lumb (1974)

(a) Soil property	Soil type	pdf	Mean	*COV*(%)
Cone resistance	Sand Clay	LN	*	*
	Clay	N/LN		
Undrained shear strength	Clay (triaxial)	LN	*	5–20
	Clay (index Su)	LN		10–35
	Clayey silt	N		10–30
Ratio S_u/σ'_{vo}	Clay	N/LN	*	5–15
Plastic limit	Clay	N	0.13–0.23	3–20
Liquid limit	Clay	N	0.30–0.80	3–20
Submerged unit weight	All soils	N	5–11 (kN/m3)	0–10
Friction angle	Sand	N	*	2–5
Void ratio, porosity, initial void ratio	All soils	N	*	7–30
Over consolidation ratio	Clay	N/LN	*	10–35
(b) Soil property				
Density	All soils			5–10
Voids ratio	All soils			15–30
Permeability	All soils			200–300
Compressibility	All soils			25–30
Undrained cohesion (clays)	All soils			20–50
Tangent of angle of shearing resistance (sands)	All soils			5–15
Coefficient of consolidation	All soils			25–50

8.2 Index Tests and Classification of Soils

Index and classification tests usually pertain to descriptive properties of soils rather than to mechanical properties. *Descriptive properties* include such things as color, unit weight, water content, and grain size distribution; whereas *mechanical properties* refer to strength, deformability, and permeability. In addition to descriptive and mechanical properties which are of most interest to geotechnical engineers, soils also possess thermal, electromagnetic, and other properties of interest in special circumstances or to specific scientific enterprises. For example, the thermal conductivity of soils is of interest to someone wishing to store liquefied natural gas in underground openings, and electrical resistivity would be of interest to the geophysicist conducting regional magnetic surveys.

Index and classification properties are those easily measured attributes useful in categorizing soils and possibly in making rough forecasts of mechanical properties based on correlations to those simple measures. The most common index and classification properties of soils are:

- Field identification properties, including color and odor (which are not further considered here).
- Bulk properties related to density, including water content, specific gravity, void ratio, relative density, unit weight, and specific gravity.
- Classification properties, including grain size distribution and index limits.

Detailed discussion of how such tests are performed can be found in a variety of references (e.g. Lambe 1951).

8.2.1 Bulk properties

Bulk properties based on density and water content are related to one another through volumetric formulae, for which the well known representation of Figure 8.3 is common to most introductory texts. *Porosity*, n, of a soil is the ratio of the volume of voids (air plus water) to the total volume ($n = V_v/V_t$). *Void ratio, e*, is the volume of voids to volume of solids ($e = V_v/V_s$). *Degree of saturation, S*, is volume of water to the volume

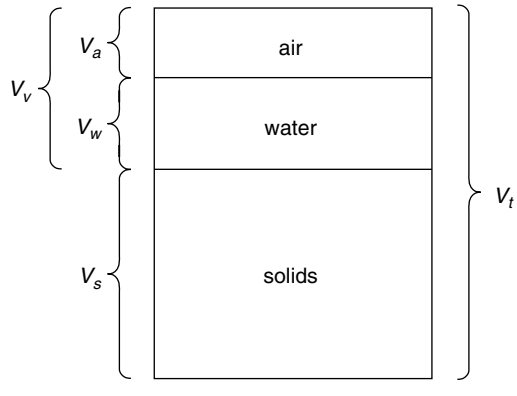


Figure 8.3 Volumetric relationship of soil constituents (after Taylor 1948).

of voids ($S = V_w/V_v$), expressed as a percentage. *Water content, w*, is the ratio of the weight (not volume) of water to the weight of solids ($W = W_w/W_s$), also expressed as a percentage. Note that, because the total volume is the sum of the volumes of air, water, and solids, the properties of porosity, void ratio, degree of saturation, and water content are interrelated by formulas such as, $n = e/(1 + e)$, which can be found in introductory texts (Lambe and Whitman 1969). Furthermore, this simple view of soil components is only an abstraction. Some water may be bound to the soil minerals and may or may not be driven off by heating. Pore water may have dissolved solids such as salt (NaCl) or calcium chloride (CaCl) that precipitate during heating. These and many other such considerations are beyond the present discussion but can be found in the literature (Hillel 1998; Lambe and Whitman 1969; Sowers and Sowers 1979; Taylor 1948; Terzaghi *et al.* 1996).

Index properties based on bulk properties are summarized in Table 8.3. Unit weights for the total specimen, water, and solids are formed from the ratios of weights to corresponding volumes. Specific gravities are formed by the ratios of total, water, and solid unit weights to the unit weight of water at Standard Temperature and Pressure (STP). Because soil-forming minerals (quartz, feldspars, clay minerals) have similar unit weights, most soils have specific gravities of solids between 2.65 and 2.85 (Taylor 1948). The exceptions are soils high in organic content, calcareous soils, and soils containing less common dense minerals such as ferromagnesians.

The terms 'very dense,' 'dense,' 'loose,' and the like, relate in principle to void ratio, but are commonly measured by or correlated to *in situ* mechanical tests such as the Standard Penetration Test (SPT) or Cone Penetration Test (CPT), and generally describe the soil mass rather than a soil specimen. For example, Seed and Lee (1966) describe soil deposit with SPT blow counts greater than some number of blows per foot (bpf) as 'very dense' for the purpose of predicting liquefaction behavior. This description is not uniquely related to void ratio as factors other than e, such as particle angularity, also influence blow count.

Unit weight and water content are perhaps the two bulk properties that most often turn up in quantitative calculations of engineering performance. Given the ease of measuring

Table 8.3 Index properties based on weights and densities of soils

Property	Saturated sample	Unsaturated sample
Volume of solids	$\dfrac{W_s}{G\gamma_w}$	
Volume of water	$\dfrac{W_w}{\gamma_w}$	
Volume of gas	0	$V - (V_s + V_w)$
Volume of voids	$\dfrac{W_w}{\gamma_w}$	$V - \dfrac{W_s}{G\gamma_w}$
Total volume	$V_s + V_w$	measured
Porosity	$\dfrac{V_v}{V}$	
Void ratio	$\dfrac{V_v}{V_s}$	

these properties, a great number of data on them appear in project reports and publications. Lee *et al*. (1983) report (Table 8.1) that the coefficient of variation of the water (moisture) content for clay soils varies from negligible up to about 60%, although most sites appear to display considerably less than this upper bound. There is some evidence that sandy soils display less variability in water content than do clay soils (Fredlund and Dahlman 1971), but everything depends on local site conditions. Unit weight appears to display considerably less variability than do most other soil properties, usually with COV's less than 10% (Table 8.1). This is perhaps not surprising. First, soil minerals mostly have specific gravities within a narrow range; second, the procedure for measuring unit weight is reasonably precise, thus the measurement error is small.

8.2.2 Classification properties

The *Atterberg limits* are index properties for cohesive soils, relating water content to simple mechanical properties (Atterberg 1913). The limits consist of three values, the liquid limit, w_l, plastic limit, w_p, and shrinkage limit, w_s. These tests are hardly sophisticated, but they have proved of value over time as a consistent way of comparing and classifying cohesive soils.

The *liquid limit* is the water content at which a remolded specimen of soil exhibits a certain small shear strength, where this 'certain shear strength' is based on a standard apparatus and test procedure. This test involves placing a specimen of soil in a brass bowl, cutting a standard grove in the soil, and then allowing the bowl to drop from a specified height onto a rubber mat until the grove just closes. The liquid limit, measured as a water content, is attained when 25 blows are needed to close the grove (Lambe 1951). Another test in wide use employs a falling cone whose penetration into the soil measures the strength.

The *plastic limit* is the water content in which the remolded, partially dried soil can be rolled into a thread 1/8 inch (3.2 mm) in diameter on a glass plate before crumbling and thus ceasing to behave plastically. The difference between liquid limit and plastic limit, expressed as a percentage difference in water content, is called the *plasticity index, PI*. The Plasticity index is a calculated quantity, and since it depends both on the liquid limit and the plastic limit, its variance depends on the variances of each measured quantity, and on the correlation between the two measured values:

$$Var(PI) = Var(W_p) + Var(W_l) - 2\,Cov(W_l, W_p) \tag{8.1}$$

in which $Var(\cdot)$ is the variance and $Cov(\cdot)$ is the covariance.

The *shrinkage limit* is the smallest water content that can occur in a completely saturated specimen of soil. As a specimen dries, its total volume decreases. At some water content, further shrinkage of the specimen stops, even though after this point water content continues to decrease as further water is dried out of the specimen. The saturated water content at this smallest volume is called the shrinkage limit.

Typical Atterberg limits data, in this case for Hong Kong clay, are shown in Figure 8.4, where the data have been re-plotted from Zhu *et al*. (2001). It is not unusual for liquid and plastic limits to have COV's of 30% or more, especially since the testing error in Atterberg limits is itself 10% or more. The liquid limit and plasticity index are almost always strongly, positively correlated. Liquid limit and plastically limit tend to be positively correlated, although not so strongly. Data clusters on the plasticity chart of Figure 8.4 for

which the slope is less than 45° display positive correlation between liquid limit and plasticity limit; those for which the slope is greater than 45° display negative correlation. This is because plasticity index equals the simple difference of liquid limit and plastic limit.

Holtz and Krizek (1971), among others, have performed extensive cross correlation analyses of index tests and simple engineering properties. An example of their results is shown as Table 8.4. As is well known, weak correlations exist between many pairs of these parameters – which is the reason that the use of index properties is widespread. For example, the undrained strength of clay is weakly correlated to liquid limit, with correlation coefficients of between about 0.2 to about 0.5, depending on the soil.

Grain size distributions are by traditional practice plotted as Cumulative Frequency Distributions (CDF) or as complimentary cumulative frequency distributions (CCDF) over

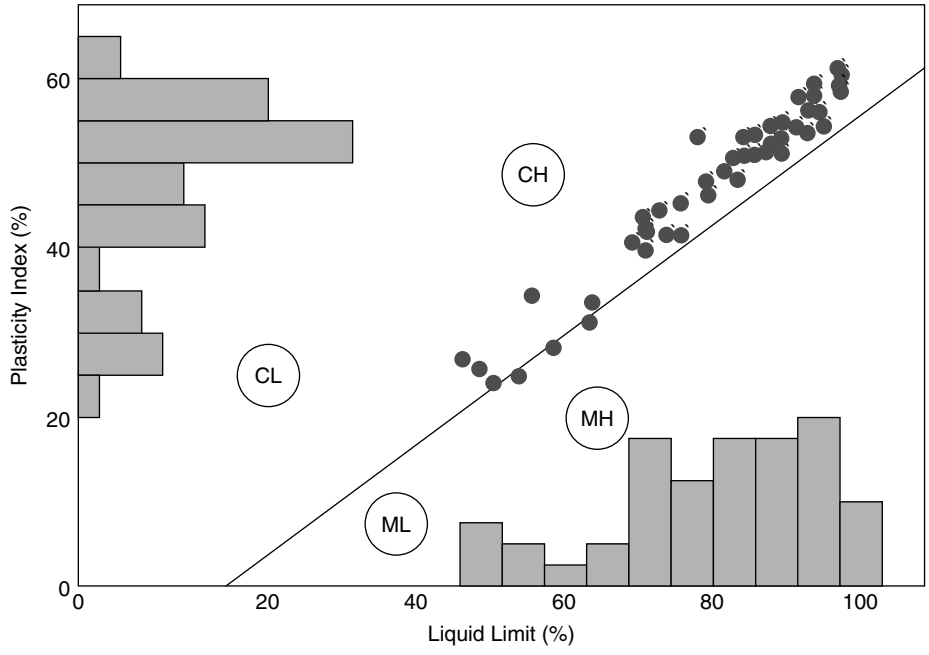


Figure 8.4 Summary of Atterberg limits data for upper marine clan, upper alluvial crust, and lower marine clay (re-plotted from Zhu *et al.* 2001), based on the plasticity chart of Casagrande (1932).

Table 8.4 Correlation matrix among index properties (Holtz and Krizek 1971)

	q_u(tsf)	w_n(%)	LL(%)	PL(%)	%Sand	%Silt	%Clay
q_u(tsf)	1.00	−0.54	−0.30	−0.12	−0.20	−0.29	0.41
w_n(%)	−0.54	1.00	0.67	0.48	−0.15	0.36	−0.19
LL(%)	−0.30	0.67	1.00	0.55	−0.36	0.20	0.10
PL(%)	−0.12	0.48	0.55	1.00	−0.40	0.29	0.06
%Sand	−0.20	−0.15	−0.36	−0.40	1.00	−0.27	−0.55
%Silt	−0.29	0.36	0.20	0.29	−0.27	1.00	−0.64
%Clay	0.41	−0.19	0.10	0.06	−0.55	−0.64	1.00

a logarithmic grain diameter scale. Normal distributions of particle diameters plot as inverted 'S' shapes, starting at 1.0 at the upper left and decreasing monotonically to 0.0 at the lower right. Curves with flat portions are said to be 'gap-graded,' in that they lack one or more ranges of grain size, yielding bimodal or multimodal pdf's.

With a simple form for the grain size CDF, for example, a Normal, logNormal, or Exponential distribution, the shapes of a grain size distribution can be described in a small number of statistical moments, as could any other frequency distribution. In the Normal or logNormal case, the mean and variance suffice; in the Exponential case, the mean alone. Most natural soils, however, have grain size distributions less well behaved. Hazen (1911) proposed using the 0.1 and 0.6 fractiles of the grain size distribution, denoted D_{10} and D_{60}, as summary measures, rather than, say, the mean and variance. These two fractiles carry less information about the frequency distribution then do statistical moments, but they have been found useful in characterizing the suitability of soils for particular uses. Also, Hazen found that the permeability (or hydraulic conductivity in modern parlance) of sand filters could be estimated roughly by the statistical correlation of the form, $k = 100\ D_{10}^2$, in which k is permeability or hydraulic conductivity in cm/s, and D_{10} is in cm. Hazen called D_{10} the 'effective size' for the distribution. The log scale of diameters is used because it nicely reflects the relative gradation of a set of particles, irrespective of average size. That is, similar proportional distributions of particle sizes produce similar curves simply displaced horizontally.

8.3 Consolidation Properties

When soil is placed under stress, its volume decreases as mineral grains compress, rearrange their positions, or both. If the soil is saturated, this volumetric compression corresponds to a net outflow of pore water, and the soil is said to *consolidate*. Consolidation strains can be large, causing displacement of serious magnitude within the soil mass. In this section we consider only cases in which the applied stresses are of such magnitude or of such proportion as not to threaten shear failure of the soil.

The above description concerns the consolidation behavior of saturated soil; the consolidation behavior of unsaturated soils under compression loading is more complex, in that pore gasses (e.g. air) are compressible. In this case, compression of the soil may correspond to a net outflow of pore water, plus a net outflow of pore gases, plus a net compression of the pore gas. This more complex case of unsaturated soil consolidation is not further considered.

As stress increments are added to the saturated soil, the relative incompressibility of the pore water compared to the mineral skeleton leads to the octahedral stress increment being borne by the water alone. However, since the water cannot sustain shear stress, deviator stresses are borne by the mineral skeleton alone. Consolidation is the slow process by which outflow of pore water due to the pore pressure gradient from inside the specimen or stressed volume to the outside causes a gradual decrease in pore water pressure and a corresponding transfer of stress from the water to the mineral skeleton. As stress builds upon the mineral skeleton, the latter compresses due to the combined action of particle deformation and particle rearrangement, and the macroscopic volume of the specimen decreases. This definition of consolidation is sufficiently broad to apply to the three-dimensional as well as the one-dimensional case, the latter being the case routinely measured in the laboratory.

Consolidation properties in the laboratory are usually measured in the one-dimensional oedometer or consolidometer test. The primary test results are reported as a relationship between void ratio, e, total vertical stress, and time (Figure 8.5 from Terzaghi *et al*. 1996). The specimen is loaded one-dimensionally with zero lateral strain.

Above some level of applied stress, the change in void ratio, e, is approximately linear with increasing log σ_v. The slope of this relationship is called the *compression index*, C_c,

$$e = e_0 - C_c \log_{10} \frac{\sigma}{\sigma_0} \tag{8.2}$$

in which e_0 is the initial void ratio at the initial vertical stress σ_0, and C_c is the compression ratio. The rebound upon unloading may be expressed in a similar fashion, in which the slope of the rebound curve is the *rebound index*, C_s. As a specimen is first loaded, it follows a recompression slope C_r approximately similar to C_s. The supposition is that, at some earlier point in time, the soil had been subjected to a one-dimensional stress about as large as that at the point of inflection of the loading curve. Thus, the initial branch of the curve reflects recompression, while the steeper curve above the point of inflection reflects first-time or virgin compression.

The transition between recompression and virgin compression in the consolidation curve is not sharp, but gradual. The more the specimen may have been disturbed from its initial state in the ground or from its condition upon initial loading, the more gradual this transition becomes. Casagrande (1936) among others proposed a standard graphical

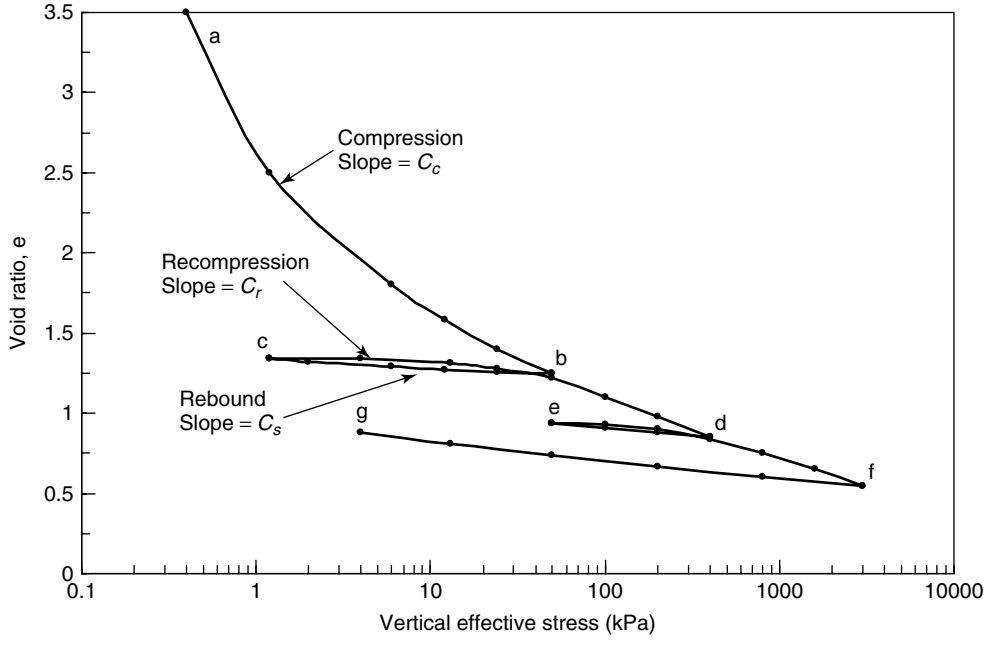


Figure 8.5 Relationship between the end of primary void ration and log effective vertical stress for a clay after sedimentation in 35 gm/l sodium chloride solution. (Terzaghi, K., Peck, R. B. and Mesri, G., 1996. *Soil mechanics in Engineering practice*. 3rd ed., New York, John Wiley & sons, © John Wiley & sons 1996 reproduced with permission.)

Table 8.5 Representative values of the variability in consolidation parameters, expressed as coefficient of variation in percent

Parameter	Soil	Reported *COV* (%)	Source
C_c, C_r	Bangkok Clay	20	Zhu *et al.* (2001)
	Various	25–50	Lumb (1974)
	Dredge Spoils	35	Thevanayagam *et al.* (1996)
	Gulf of Mexico Clay	25–28	Baecher and Ladd (1997)
c_v	Ariake Clay	10	Tanaka *et al.* (2001)
	Singapore Clay	17	Tanaka *et al.* (2001)
	Bangkok Clay	16	Tanaka *et al.* (2001)

procedure for estimating a unique stress σ_{vm} at which the change from recompression to virgin compression takes place. This stress, in principle, is the *maximum past pressure* that the soil has experienced in its loading history. At the very least, the Casagrande or similar procedures provide a standard way of specifying this point.

When the incremental stresses are first applied, the octahedral increment is taken up by increased pore water pressure, and the specimen deforms only under the action of increased shear stress. As pore pressure dissipates, stresses are transferred to the mineral skeleton and the soil begins to deform volumetrically. At first, the rate of deformation is small, but the changes soon become approximately linear in log-time. The slope of this linear portion of the e vs. log t curve is defined as the *coefficient of consolidation*, c_v.

Table 8.5 shows representative values of COV of consolidation parameters for various soils. These values are somewhat smaller than reported by Lee *et al.* (Table 8.1), ranging only up to about 50%, but the data are not numerous.

8.4 Permeability

The hydraulic conductivity of soils, also referred to as Darcy permeability, or just permeability, is that property of soil describing the rate at which water flows through a unit cross section of soil mass under a unit gradient of pore pressure. This parameter, k, is used by means of Darcy's law to model seepage, and by extension, the pore pressure fields that are inextricably related to seepage.

Darcy's law relates pore pressure gradients and water flows by the simple relationship

$$q = kiA \tag{8.3}$$

in which q is the fluid flux or *discharge* measured in volume per time, i is the pore pressure gradient measured in change of pressure head per distance, A is the cross-sectional area of flow, and k is a constant of proportionality, called *hydraulic conductivity* or *permeability*. Thus, Darcy's law is similar in structure to the linear heat flow law.

The permeability of soils, k, varies over orders of magnitude, from values on the order of 10^{-12} cm/sec, or less, for clays, to nearly unity for gravels. Similarly, the coefficient of variation of k for a specific soil can also be large. For example, Lumb (1974) reports, COVs of 200 to 300% for k; Benson (1993) reports values of 50 to 200%, but with some soils reaching 500 or even 700%. As a result, common practice is to express permeability

Table 8.6 Statistics of the hydraulic conductivity at 57 sites. (Benson, C. H., 1993, 'Probability distributions for hydraulic conductivity of compacted soil liners,' *Journal of Geotechnical and Geoenvironmental Engineering*, reproduced by permission of the American Society of Engineers.)

Site	*n* (cm/s)	mean *k* (cm/s)	*COV* k	Skew in K	Kurtosis in K	Skew test	Kurtosis test	PPCC test
1	46	3.30E-09	94	0.60	2.20	P	P	F
2	27	4.50E-09	129	0.10	3.03	P	P	P
3	36	1.30E−08	151	0.74	2.70	P	P	F
4	33	9.36E−09	139	1.84	9.79	F	F	F
5	13	4.70E−08	55	0.24	2.10	P	P	V
6	31	8.20E-08	42	0.98	7.50	F	F	P
7	31	5.OOE-08	34	0.18	2.30	P	P	P
8	115	7.70E−09	133	0.35	2.96	P	P	P
9	15	3.OOE-08	88	0.34	2.41	P	P	P
10	15	4.40E-08	27	0.89	3.43	P	P	P
11	184	1.29E−06	92	1.09	1.61	F	F	F
12	73	8.40E−07	177	0.05	1.93	P	F	P
13	120	1.OOE-05	767	2.53	9.84	F	F	F
14	25	4.20E−08	49	0.46	2.45	P	P	P
15	15	2.50E-08	91	0.16	1.56	P	F	P
16	24	2.12E−08	79	0.067	2.01	P	P	P
17	64	2.30E-08	164	0.67	2.71	P	P	F
18	2(1	9.OOE−09	120	0.05	1.70	P	F	P
19	42	6.11E−07	507	1.97	7.71	F	F	P
20	18	1.40E−08	62	0.63	2.45	P	P	P
21	24	9.50E−09	101	1.14	2.47	F	P	F
22	44	4.50E−08	52	0.38	2.43	P	P	P
23	33	6.20E−08	86	0.53	3.29	P	P	P
24	57	7.20E-08	64	0.68	3.70	F	P	P
25	84	5.91E−08	53	1.56	3.31	F	P	F
26	112	3.40E−08	113	0.39	2.37	P	P	P
27	8	2.43E−08	210	1.18	3.18	F	P	F
28	31	6.00E-08	161	0.60	3.20	P	P	P
29	16	2.45E−08	88	0.50	2.55	P	P	F
30	27	8.OOE-08	99	0.59	4.36	P	P	F
31	10	2.87E−09	143	0.90	3.35	P	P	P
32	34	5.30E-08	89	0.52	2.37	P	P	P
33	14	5.20E−08	46	0.07	2.15	P	P	P
34	30	5.62E−08	118	0.13	2.04	P	P	P
35	22	1.17E-08	163	1.44	4.77	F	F	F
36	18	1.22E-08	142	0.70	3.62	p	p	P
37	26	2.30E-08	107	0.33	2.47	P	P	P
38	28	4.45E−08	278	2.32	8.91	F	F	F
39	18	1.40E−07	134	0.42	3.06	P	P	P
40	40	1.60E-08	157	0.89	3.04	F	P	F
41	34	2.30E-08	60	0.02	3.26	P	P	P
42	10	2.40E−08	157	1.74	4.86	F	F	F
43	87	5.OOE−08	56	1.53	4.46	F	F	F
44	17	2.40E−08	63	0.69	2.31	P	P	P
45	40	9.10E−09	114	0.44	3.34	P	P	P
46	82	6.38E09	166	0.12	3.56	∼	P	P
47	102	3.90E-08	77	−0.64	2.75	F	P	F
48	39	5.50E−08	151	0.42	2.80	P	P	P
49	16	7.1OE−08	90	0.15	2.58	P	P	P
50	13	6.50E−08	138	0.20	1.82	P	P	P
51	22	1.97E−08	84	0.49	2.32	P	P	P
52	26	1.97E−08	121	0.31	2.65	P	P	P
53	205	9.90E−09	100	0.48	4.63	F	F	F
54	49	8.90E−09	47	0.58	2.43	P	P	F
55	45	8.90E−09	46	0.13	3.50	P	P	P
56	34	1.50E−08	129	1.99	10.3	F	F	F
57	53	9.50E−09	33	0.17	3.30	P	P	P

in $\log_{10}$, and to model its variability for a given soil formation by logNormal probability distributions (Marsily 1986). This is done in both geotechnical practice (Ang and Tang 1975) and hydrogeology (Kitanidis 1997).

Benson (1993) studied the probability distributions of permeabilities of compacted clay liners for landfills and reports the results shown in Table 8.6. As can be seen, the coefficients a variation range from a low of 27% to high of 767%, but with all but two of the 57 data sets showing COVs under 300%. A more interesting result is, however, that Benson finds that the (two-parameter) logNormal pdf only passes common goodness-of-fit tests (from frequentist theory) for about 2/3 of the data sets (all significance tests were set at the $\alpha = 0.05$ level). For the other approximately 1/3, either the skews of the data sets are larger than would be predicted by the logNormal model, or the kurtosis is too great, or both.

Benson speculates that the physical cause for this lack of good fit in many of the data sets could be, "(1) mixtures of micro- and macro-scale flows; (2) undetectable trends in hydraulic conductivity; and (3) measurement errors. [...] The cumulative effect of the mixtures is a distribution of hydraulic conductivity with characteristics that change from site to site and even within a site." He goes on to investigate other candidate pdf forms, including the three-parameter logNormal, generalized Extreme Value, inverse Gamma, and three-parameter Gamma. These are all families of distributions that are common in the stochastic hydrology literature. The paper concludes that either the three-parameter logNormal or the generalized Extreme Value pdf provides the best fit to the widest number of cases. It is not entirely surprising that these models would provide more flexible fit to data, as they both enjoy three parameters in contrast to the two parameters of the traditional logNormal distribution.

The appropriate conclusion to draw from these well-formulated studies is not immediately clear. The logNormal pdf enjoys widespread use in hydrogeology and appears to have been a serviceable model. Similarly, Benson's study is restricted to a narrow class of materials, namely artificially placed compacted clay liners. The study suggests that good practice is to remain skeptical of the choice of pdf model and to verify the model choice against data wherever possible.

8.5 Strength Properties

The downhole measurements of standard penetration test blow counts (SPT) and cone penetration resistance (CPT) are widely used in practice and return large numbers of data. As a result, a great deal of statistical work on *in situ* data has been reported in the literature. Phoon and Kulhawy (1996) have summarized a number of these studies. Table 8.7 shows results for various downhole test devices, including in addition to SPT and CPT, the field vane (FVT), dilatometer (DMT), and pressure meter (PMT). The coefficients of variation for all these tests range to large values, in part because these test data many times reflect small-scale variations in soil properties, and in part because the tests themselves may introduce significant measurement error.

Baecher *et al*. (1983) summarize strength and other geotechnical data on tailings from 41 copper, uranium, and gypsum mines in north America, Africa, and Japan. Table 8.8 summarizes within site variability for SPT and CPT measurements, linearly de-trended using regression analysis to remove systematic variations with depth. Kulhawy and Trautmann (1996) have analyzed the probable measurement error in the same types of *in situ*

Table 8.7 Coefficient of variation for some common field measurements (Phoon and Kulhawy 1999)

Test type	Property	Soil type	Mean	Units	*COV*(%)
CPT	q_T	Clay	0.5–2.5	MN/m^2	<20
	q_c	Clay	0.5–2	MN/m^2	20–40
	q_c	Sand	0.5–30	MN/m^2	20–60
VST	s_u	Clay	5–400	kN/m^2	10–40
SPT	N	Clay & Sand	10–70	blows/ft	25–50
DMT	A Reading	Clay	100–450	kN/m^2	10–35
	A Reading	Sand	60–1300	kN/m^2	20–50
	B Reading	Clay	500–880	kN/m^2	10–35
	B Reading	Sand	350–2400	kN/m^2	20–50
	I_D	Sand	1–8		20–60
	K_D	Sand	2–30		20–60
	E_D	Sand	10–50	MN/m^2	15–65
PMT	p_L	Clay	400–2800	kN/m^2	10–35
	p_L	Sand	1600–3500	kN/m^2	20–50
	E_{PMT}	Sand	5–15	MN/m^2	15–65
Lab Index	w_n	Clay and silt	13–100	%	8–30
	W_L	Clay and silt	30–90	%	6–30
	W_P	Clay and silt	15–15	%	6–30
	PI	Clay and silt	10–40	%	–[a]
	LI	Clay and silt	10	%	–[a]
	γ, γ_d	Clay and silt	13–20	KN/m^3	<10
	D_r	Sand	30–70	%	10–40; 50–70[b]

[a]COV = (3–12%)/mean.

[b]The first range of variables gives the total variability for the direct method of determination, and the second range of values gives the total variability for the indirect determination using SPT values.

Table 8.8 Variabilities of *in situ* test results for SPT and CPT in 41 mine tailings deposits (Baecher *et al*. 1983)

Commodity	SPT (b/ft)		CPT (psi)	
	Mean	*COV*	Mean	*COV*
Copper	28	39	304	76
Uranium	14	52	–	–
Gypsum	20	56	161	52

measurements and report total COV's due to various measurement errors (equipment, procedure, random noise) of between 10% and 25% for all but SPT, and COV's of up to 45% for SPT (Table 8.9). This would suggest that as much as half of the variability among *in situ* measurements could arise from random testing errors. This result is consistent with other reports in the literature (e.g. Hilldale 1971).

The variability among laboratory measurements of effective friction angle, $\overline{\phi}$, is considerably less than that among *in situ* tests. First, greater care is usually taken with laboratory tests than with *in situ* tests, and, second, the specimen quality for laboratory tests is almost

Table 8.9 Uncertainty estimates for five common *in situ* tests. (Kulhawy, F.H. and Trautmann, C.H., 1996) 'Estimation of *in-situ* test uncertainty.' *uncertainty in the Geological Environment*, Madison, WI, ASCE: 269–286, reproduced with permission of the American Society of civil Engineers

Test	Coefficient of Variation, COV (%)					
	Acronym	Equipment	Procedure	Random	Total	Range
Standard Penetration Test	SPT	5[c]–75[d]	5[c]–75[d]	12–15	14[c]–100[d]	15–45
Mechanical Cone Penetration Test	MCPT	5	10[e]–15[f]	10[e]–15[f]	15[e]–22[f]	15–25
Electric Cone Penetration Test	ECPT	3	5	5[e]–10[f]	8[e]–22[f]	5–15
Vane Shear Test	VST	5	8	10	14	10–20
Dilatometer Test	DMT	5	5	8	11	5–15
Pressuremeter Test, Pre-bored	PMT	5	12	10	16	10–20[g]
Self-Boring Pressuremeter Test	SBPMT	8	15	8	19	15–25[g]

[a]COV(total) = $\{\text{COV(equipment)}^2 + \text{COV(procedure)}^2 + \text{COV(random)}^2\}^{1/2}$.
[b]because of limited data and judgment involved in estimating COVs, ranges represent probable magnitudes of field test measurement error.
[c,d]Best to worst case scenarios, respectively, for SPT.
[e,f]Tip and side resistances, respectively, for CPT.
[g]Results may differ for p_0, p_f, and p_1, but data are insufficient to clarify this issue.

Table 8.10 Variabilities of laboratory-measured $\overline{\phi}$ for various soils

Soil type	COV	Source
Various soils	9	(Lumb 1966)
Clay	40	(Kotzias *et al*. 1993)
Alluvial	16	(Wolff 1996)
Sands	2–5	(Lacasse and Nadim 1996)
Tailings	5–20	(Baecher *et al*. 1983)

Table 8.11 Variabilities of laboratory-measured $\overline{\phi}$ for five (5) mine tailings deposits (Baecher *et al*. 1983)

Commodity	*n*	Mean	Standard Deviation	*COV*	Skew (β_1)	Kurtosis (β_2)	pdf
Copper	26	30.4	2.56	0.08	0.0015	2.17	Beta/Uniform
Copper	16	35.7	2.90	0.08	0.750	3.99	Beta/Gamma
Copper	21	36.2	4.35	0.12	0.100	3.01	Normal
Uranium	23	35.8	6.04	0.17	0.202	3.02	Normal
Gypsum	19	41.6	5.80	0.14	0.146	2.68	Normal

always better than for *in situ* tests. As a general rule, the most intact and least disturbed section of a boring is selected for laboratory testing, creating an implicit sampling bias that reduces variability.

Table 8.10 shows reported variabilities in laboratory measured values for $\overline{\phi}$ for a variety of soils. The range of variabilities is large, and it is not clear what one can conclude from these results. Baecher *et al*. (1983) report variabilities of $\overline{\phi}$ for three types of mine tailings (Table 8.11) along with third and fourth moment information which allows the data to be plotted on Pearson distribution charts (more below).

8.6 Distributional Properties

The preceding sections have concerned themselves with the coefficients of variation of soil data. The empirical probability distributions characterizing those observations are typically reported in histograms, but analytical forms are also available with which to model sample distribution functions, and such models are of convenience both for statistical inference and for engineering modeling.

Certain analytical functions that play a central role in statistical theory and data analysis are discussed in Chapter 3. The more common and useful of these for fitting empirical data are summarized here. The most common analytical distribution functions are those of the general exponential form

$$f_X(x) \propto \exp\{a + bx + bc^2)x^d \tag{8.4}$$

in which *a, b, c*, and *d* are constants. The better known distributions having this form are the Normal, logNormal, Exponential, and Gamma.

The *Normal distribution*, recognized by its characteristic bell-shape, is the most common of all distributions. It is observed in sample data with such frequency that Galton (Gillham 2001), in his early work on the distribution of features in human populations, coined its common name, at least for the English literature. In non-English literatures the distribution is more commonly called the *Gaussian*, in honor of Gauss's original proof of the *central limit theorem*, which says that variables composed of the sum of independent perturbations necessarily tend toward Normal distributions as the number of perturbations becomes large. Thus, through the central limit theorem there exists theoretical justification for the widespread use of the normal form, and for the central position occupied by this form in statistical sampling theory. Stigler (1999) observes that the term 'Gaussian' is a misnomer, as De Moivre discovered the distribution early in the seventeenth century.

The *logNormal distribution* describes the distribution of a variable, the logarithm of which is normally distributed. Thus, the logNormal is closely related to the Normal, and by an extension of the central limit theorem one can show that the logNormal distribution describes a variable formed by the product of independent perturbations as the number of perturbations becomes large.

The *Exponential distribution*, sometimes called Negative Exponential, is a one-parameter function and is arguably the simplest of common distributions. While this distribution is often observed in geometric data, as, for example, the spacings among rock joint traces in outcrop, it is not commonly encountered in strength, deformation, or flow data. Theoretical arguments can be made that certain types of data should be

exponentially distributed, for example spacings between random events in time or space, but in the general case its use is primarily one of convenience.

The *Gamma distribution* is positively skewed, as is the logNormal, and although derived from theoretical arguments pertaining to discrete measurements, its use with continuous measurements is often based on its similarity to the logNormal and its greater convenience.

Only a limited number of non-exponential forms are commonly used with geotechnical data. Perhaps the most publicized of these is the *Beta distribution*, which has been advocated principally due to its flexibility and the upper and lower bounds of its domain, matching the upper and lower bounds for such geotechnical variables as friction angle (Harr 1987). The Beta distribution is a four-parameter pdf and is thus flexible. It often approximates empirical data quite well because of this fact. On the other hand, one must be mindful that, due to the large number of parameters required to specify the beta distribution, the degrees of freedom in fitting empirical data are reduced. Thus, with limited data sets the statistical uncertainty in estimated parameters increases. With large data sets, the number of degrees of freedom is seldom a problem, but many geotechnical data sets are not large. Since the Beta distribution is defined over a segment of the measurement axis rather than the entire axis, upper and lower bounds on x must be estimated or fixed *a priori*. This can be a difficult task.

Given the large number of analytical forms for distribution functions, attempts have been made to develop systems of distributions to bring order to the taxonomy of functions. The principal attempt in this direction is due to Pearson (Ord 1972), and his chart of distributional types is now widely used. Other systems of distributions have also been proposed, and are discussed by Ord.

The Pearson family of distributions comprises the solutions to the differential equation

$$\frac{df_X(x)}{dx} = \frac{(x-a)\,f_X(x)}{b_o + b_1 x + b_2 x^2} \tag{8.5}$$

in which a, b_0, b_1, and b_2 are constants. Among others, these solutions include the Normal, logNormal, Gamma, Exponential, and Beta distributions. They also include other common or useful distributions, for example, *Student's t distribution*. Members of the Pearson family are usually identified by study of their low-order moments, specifically those of order one through four (related to the mean, standard deviation, skewness, and kurtosis). These moments may be combined in the two statistics (Ord 1972), $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$ in which μ_i is the ith central moment of the distribution (Chapter 3), which can then be used to distinguish among members of the Pearson family. The simplest use of the β_1 and β_2 statistics is through the *Pearson Diagram* (Figure 8.6), with which distributional forms may be identified by inspection.

Systems of distributions other than Pearson's have been studied, primarily by representing frequency functions as series expansions, or by considering the transformations of frequency functions to common shapes (e.g. to Normal distributions). These are briefly listed for reference. Discussions of these families can be found in Ord (1972) or Kendall and Stuart (1977). The principal series expansion systems of distributions are: (1) the Chebyshev–Hermite polynomials, based on polynomial multipliers of the error function integral; (2) Edgeworth's Type A series, and the (3) Gram–Charlier Type A series, each based on series of normal integrals or their derivatives; (4) the Tetrachoric function series, due to Pearson; and (5) Charlier's Type B series, based on derivatives of the Poisson pdf. The principal transformation systems are (1) polynomial transformations to Normality,

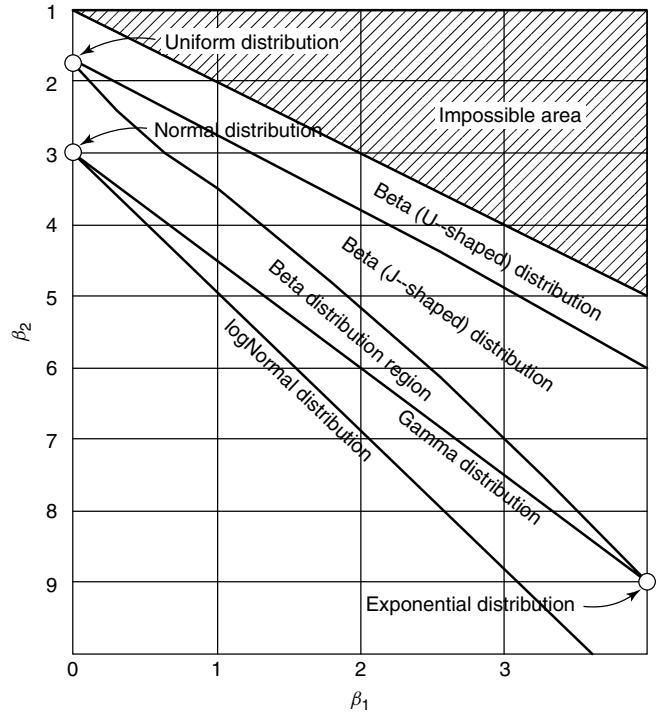


Figure 8.6 Pearson diagram for distinguishing members of the Pearson family of distributions from the low-order moments of data (Ord 1972).

and (2) general (non-polynomial) transformations to Normality. In each case a frequency function is categorized by the nature of the transformation that changes it to a Normal distribution.

Figure 8.7 shows Pearson charts of CPT data for copper tailings collected in 21 borings at the Chingola Mine, Zambia. The plot to the left shows raw measurements without detrending for depth. The plot to the right shows the same data with a linear (or multi-linear) trend with depth removed. That is, the plot to the right is for residuals off the trend. The raw data spread widely, because no trend has been removed, and are well-modeled by uniform or Beta pdf's (Figure 8.8). The residual data are more clustered about their respective means (i.e. trends) are well-modeled in most cases by Normal pdf's, or where not, then by logNormal pdf's (Figure 8.9). Table 8.12 reports best fitting pdf's for various tailings properties at 41 mines.

Many measures of soil strength appear to be well-modeled by Normal distributions (Figure 8.10), or by more flexible distributions of such as the four-parameter Beta, which mimic the Normal within the high probability density region of the distribution. However, this is not the case for all strength data, as illustrated by Figure 8.11, taken from Kaderabek and Reynolds (1981). Most soil strength measurements reflect the influence of moderate to large volumes of soils within which the effects of individual elements add to those of other elements to produce the macroscopic property. Thus, from a random process model view, soil strength is in averaging process, and we should expect the results to display a tendency toward Normal distributions.

The data represented in Figure 8.11 however, do not reflect an averaging process. These data were collected from confined compression tests on brittle specimens of Miami limestone. Rock specimens tested in unconfined compression do not, in fact, fail due to compression stresses, but from lateral tensile stresses in the middle of the specimen (Price

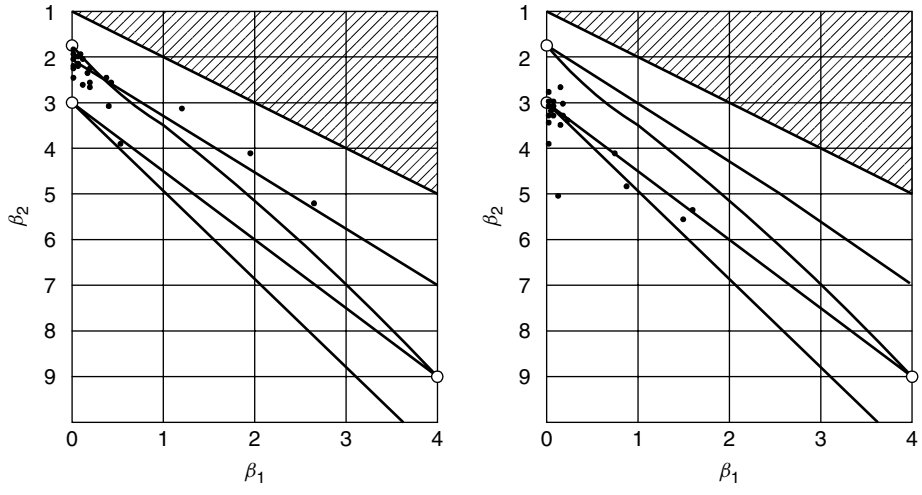


Figure 8.7 Pearson charts of CPT data for copper tailings collected in 21 borings at the Chingola Mine, Zambia, (a) raw (non de-trended) data, (b) residuals off linear trend with depth (Baecher *et al*. 1983).

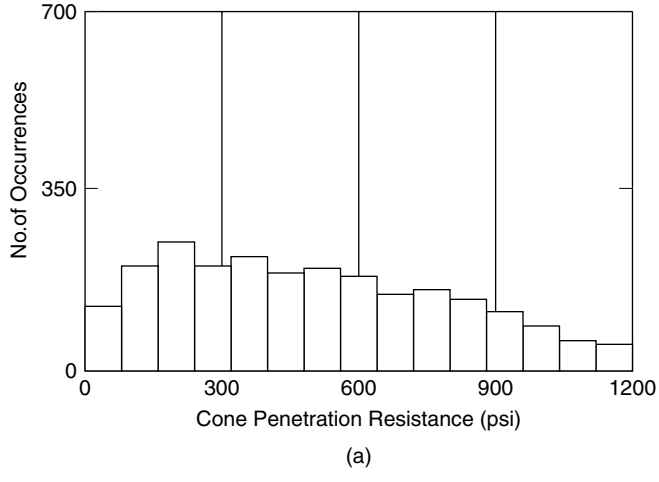


Figure 8.8 Histogram and probability paper plots of raw (non de-trended) CPT data for copper tailings collected in 21 borings at the Chingola Mine, Zambia (Baecher *et al*. 1983). Curves on probability paper plot show best fitting Normal (line), logNormal, Exponential, and Gamma distributions.

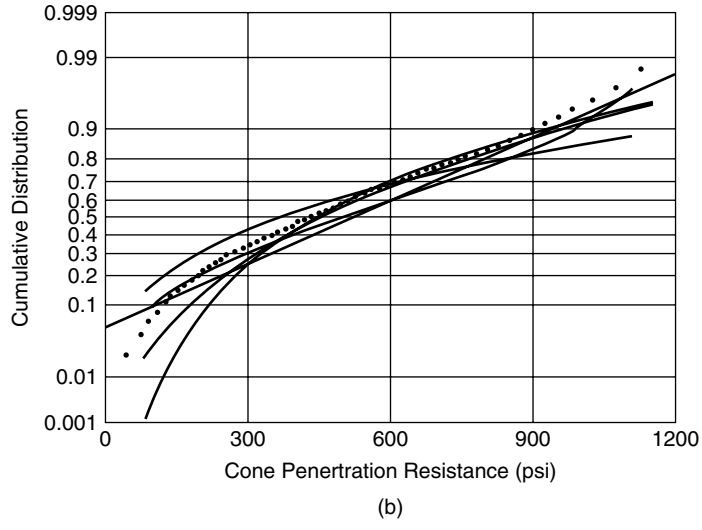


Figure 8.8 (*continued*)

1966). Thus, the failure mechanism reflects extreme values, the most critical flaw, and one should expect the skew that is typical of extreme value phenomena. This can be observed in both the unconfined compression test data shown here, and in splitting tension test data also presented by Kaderabek and Reynolds. Tellingly, however, this skew does not appear in the rock modulus data, which, as should be expected, are closer to Normality.

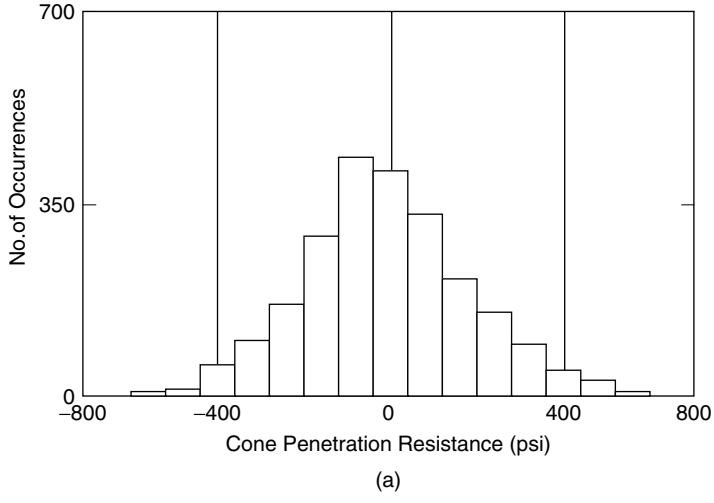
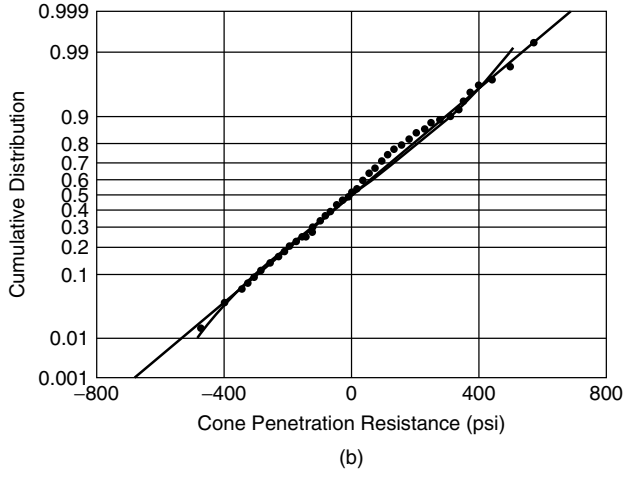


Figure 8.9 Histogram and probability paper plots of residuals off linear trend with depth CPT data for copper tailings collected in 21 borings at the Chingola Mine, Zambia (Baecher *et al*. 1983). Curves on probability paper plot show best fitting Normal (line), logNormal, Exponential, and Gamma distributions.



(b)

Figure 8.9 (*continued*)

Table 8.12 Best fitting pdf's for various tailings properties at 41 mines, distributions passing Kolmogorov-Smirnov goodness-of-fit at 5% level (Baecher *et al*. 1983)

Commodity	Site	Measurement	Raw Data	Residuals
Copper	Chambishi	cone penetration	none	none
	Chibuluma	cone penetration	Beta	Normal
				Beta
	China	SPT	all	Normal
	Chingola	cone penetration	none	none
	Magna	water contents	Normal	Normal
			Beta	Beta
			logNormal	
			Gamma	
		field vane	Normal	Normal
			Beta	Beta
			logNormal	
			Gamma	
		dry density	Normal	none
			Beta	
			logNormal	
			Gamma	
	Mindola	cone penetration	none	none
Uranium	Colorado	SPT	none	none
		triaxial ϕ	Normal	none
			Beta	
			logNormal	
			Gamma	
Gypsum	Texas	SPT	Normal	Normal
			Beta	
			LogNormal	
			Gamma	
		unit weight	Normal	Normal
			Beta	Beta
			logNormal	
			Gamma	

(*continued overleaf*)

Table 8.12 (*continued*)

Commodity	Site	Measurement	Raw Data	Residuals
	La.1	SPT(0 to 50 ft)	Normal Beta logNormal Gamma	Normal Beta
		SPT(55 to 100 ft)	Beta logNormal. Gamma	Normal Beta
		SPT	Beta logNormal Gamma	Normal Beta
	Piney Point	cone penetration	Beta LogNormal	none

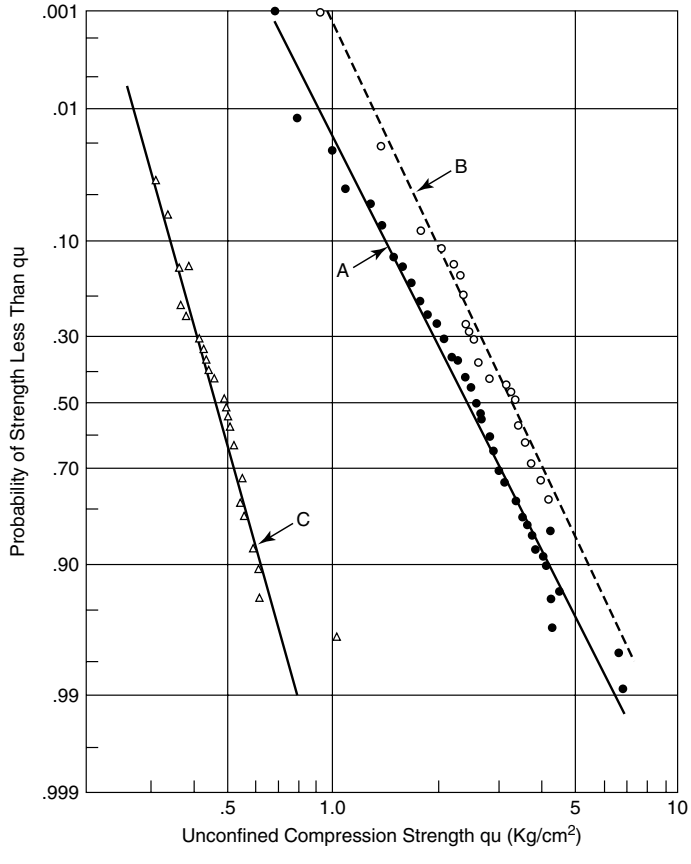


Figure 8.10 Distribution of unconfined compression strength. (Wu, T. H. and Kraft, L. M., 1967. 'The probability of foundation safety.' *Journal of the Soil Mechanics and Foundations Division, ASCE* **93**(SM5), pp. 213-23, reproduced with permission of the American Society of Civil Engineers).

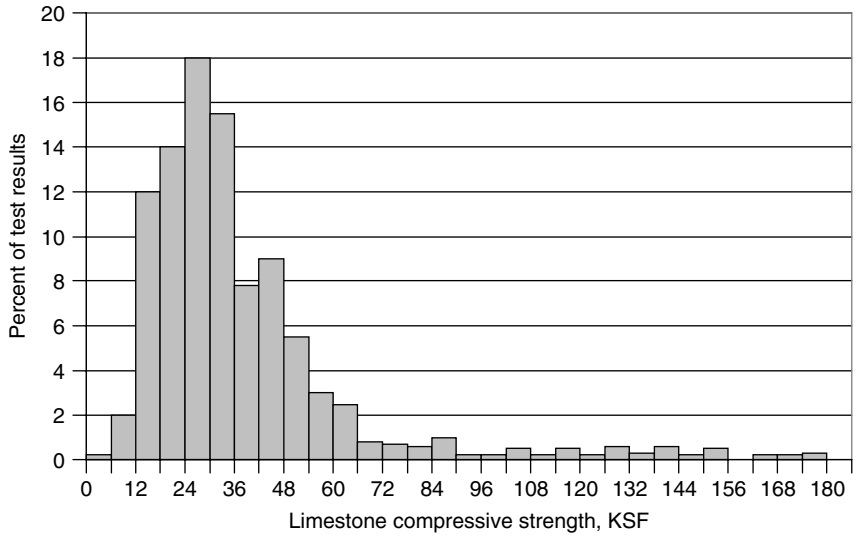


Figure 8.11 Histogram of ultimate unconfined compressive strength for Miami Limestone from 4 inch (100 mm) diamond rock core test. (Kaderabek, T.J. and Reynolds, R.T., 1981, 'Miami limestone foundation design and construction.' *Journal of the Geotechnical Engineering Division, ASCE* **107**(G77): 859–872, reproduced with permission of the American society of civil Engineers.)

8.7 Measurement Error

Measurement error enters the determination of soil properties in at least two ways, through systematic biases in average property measurements, and through random errors or noise. The usual model for these errors is

$$z = bx + e \tag{8.6}$$

in which x is a measured value, b is a bias term, $x.$ is the actual property, and e is a zero-mean independent and identically distributed perturbation. Thus, the error terms are b and e. The bias is often assumed to be uncertain, with mean μ_b and standard deviation σ_b. The IID random perturbation is usually assumed to be Normally distributed with zero mean and standard deviation σ_e.

For a given value of x, the mean and variance of z can be found from a simple expansion of the derived variable (Chapter 3):

$$Var(z) = \mu_x^2 Var(b) + \mu_b^2 Var(x) + \mu_x \mu_b Cov(x, b) + Var(e) \tag{8.7}$$

in which μ_x is the mean of the actual soil property. Usually, the dependence of b on x is ignored, such that covariance *Cov(x,b)* is assumed negligible.

Bias error in measurement arises from a number of reasonably well-understood mechanisms. Sample disturbance is among the more important of these mechanisms, usually causing a systematic degradation of average soil properties along with a broadening of

dispersion. The second major contributor to measurement bias is the phenomenological model used to interpret the measurements made in testing and especially the simplifying assumptions made in that model. For example, the physical response of the tested soil element might be assumed linear, when in fact this is only an approximation; the reversal of principal stress direction might be ignored; intermediate principal stresses might be assumed other than they really are; and so forth. The list of possible discrepancies between model assumptions and the real test conditions is long.

Model bias is usually estimated empirically by comparing predictions made from measured values of soil engineering parameters against observed performance. Obviously, such calibrations encompass a good deal more than just the measurement technique; they incorporate the models used to make predictions of field performance, inaccuracies in site characterization, and a host of other things.

Bjerrum's (1972; 1973) calibration of field vane test results for the undrained strength, s_u, of clay is a good example of how measurement bias can be estimated in practice. This calibration compares values of s_u measured with a field vane against back-calculated values of s_u from large-scale failures in the field. In principle, this calibration is a regression analysis (Chapter 4) of back-calculated s_u against field vane s_u, which yields a mean trend plus residual variance about the trend. The mean trend provides an estimate of μ_b while the residual variance provides an estimate of σ_b. The residual variance is usually taken to be homoscedastic – meaning it has the same variance regardless of the value of x, a common assumption in regression analysis – implying independence of the variation of b and x. This approach to measurement bias is adopted and Chapter 14 in an example of slope stability calculations applied to water retaining dikes on the James Bay project in Quebec.

Random measurement error is that part of data scatter attributable to instrument or operator induced variations from one test to another. This variability may sometimes increase a measurement and sometimes decrease it, but its effect on any one, specific measurement is unknown. Systematic difference between the real value and the average of the measurements is said to be measurement bias, while the variability of the measurements about their mean is said to be random measurement error.

Random measurement errors are ones whose sign and magnitude cannot be predicted; they may be plus or minus. Typically, random errors tend to be small and they tend to distribute themselves equally on both sides of zero. Measurement error is the cumulative effect of an indefinite number of small 'elementary' errors simultaneously affecting a measurement.

Random measurement errors of soil properties are difficult to assess directly because most soil is destructively tested. The standard way to gauge random measurement errors is to make repeated measurements of the same property and then to characterize how much those measurements vary. This is, for example, the way surveyors find the average squared error in transit measurements, and it is the way commercial testing laboratories calibrate instruments. But this cannot be done for measurements of, say, s_u, because the specimen is destroyed in the process of testing.

For tests of soil properties than do not require undisturbed specimens, for example, Atterberg limits, repeated testing is possible. Hammitt (1966) reports findings of a series of comparative laboratory tests conducted by the American Council of Independent Laboratories. In these comparative tests, specimens of three soils, as close to identical as possible, were distributed to different commercial testing laboratories. Each laboratory

Table 8.13 Statistical analysis of soil tests performed on identical soil samples distributed by the American Council of Independent Testing Laboratories (Hammitt 1966)

	Highly Plastic Soil		Medium Plastic Soil		Low Plastic Soil	
Type of test	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
LL	54.3	5.4	32.7	2.3	27	1.7
PL	22.2	3.4	22.4	2.8	23.6	2.4
PI	32	5.7	10.4	3.6	3.8	2.1
Specific Gravity	2.63	0.115	2.66	0.060	2.69	0.054

reported back measurements of liquid limit, plastic limit, plasticity index, and specific gravity. The summarized results are shown in Table 8.13.

The results suggest that the coefficients of variation of random measurement errors for liquid limit and plastic limit range from about 5% to about 15%. The COV of plasticity index should be the rms sum of the COV's of LL and PL, since $PI = LL - PL$, and, as can be seen from the data, this is in fact the case. The COV's for specific gravity measurements are quite small, typically only about 2%. The fact that random measurement errors in Atterberg limits are high should surprise no one who has ever performed such tests.

An indirect way of assessing random measurement error is by making use of the spatial structure of test results in the field. This approach, using the autocorrelation function, manages to get around the problem caused by destructive testing by leveraging the fact that soil properties tend to be spatially correlated, and thus properties at closely spaced locations tend on average to be more alike than do properties at widely separated locations. The decay of this correlation with separation distance can be interpolated back to the origin (zero separation) of the autocorrelation plot to infer measurement error. A more complete description of this approach awaits the introduction of concepts and models of spatial correlation in Chapter 9.

9 Spatial Variability within Homogeneous Deposits

Soils are geological materials formed by weathering processes and, save for residual soils, transported by physical means to their present locations. They have been subject to various stresses, pore fluids, and physical and chemical changes. Thus, it is hardly surprising that the physical properties of soils vary from place to place within resulting deposits. The scatter observed in soil data comes both from this spatial variability and from errors in testing. Each of these exhibits a distinct statistical signature, which can be used to draw conclusions about the character of a soil deposit and about the quality of testing. In this chapter, we consider the spatial variation of actual soil properties.

9.1 Trends and Variations About Trends

In Chapter 8, means and standard deviations were used to describe the variability in a set of soil property data. These are useful measures, but they combine data in ways that mask spatial information. Describing the variation of soil properties in space requires additional tools.

Among the first things engineers do in analyzing data is plot them. This might be a map showing the horizontal extent of the site, or it might be a profile section or fence diagram, showing vertical slices. From such plots, one recognizes trends, anomalous geological conditions, and other features that may affect engineering performance. The same is true of statistical interpretation of data. Indeed, the first rule of statistical analysis might well be: *Plot the data*.

Consider two sequences of measurements shown in Figure 9.1. Presume that each measurement was made at the same elevation, one in each of nine consecutive borings along a line. These two sets of data have the same mean and standard deviation, but reflect different soil conditions. The first data exhibit a horizontal trend; the second are

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

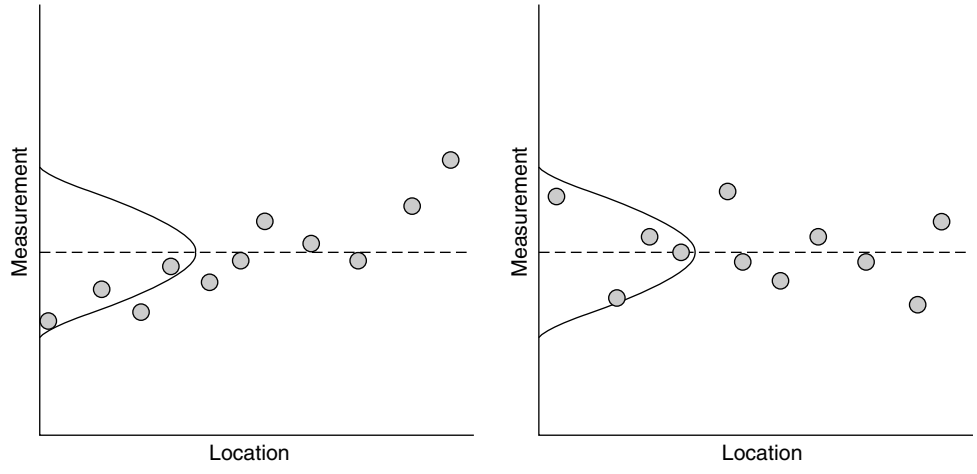


Figure 9.1 The data in the two figures share the same mean and variance, but display different patterns of spatial variation.

erratic. This difference cannot be inferred from the mean and standard deviation alone, for they are the same in both cases. Trend analysis is concerned with distinguishing these two cases.

In principle, the spatial variation of a soil deposit can be characterized in detail, but only after a large number of tests. In reality, the number of tests required far exceeds what can be acquired in practice. Thus, for engineering purposes a simplification is introduced – a model – with which spatial variability is separated into two parts: (i) a known deterministic trend; and (ii) residual variability about that trend. This model is written

$$z(x) = t(x) + u(x) \tag{9.1}$$

in which $z(x)$ is the soil property at location x, where x may be a vector (e.g. location in three dimensions), $t(x)$ is the value of the trend at x, and $u(x)$ is the residual variation. The trend is characterized deterministically by an equation. The residuals are characterized statistically as a random variable, usually with zero mean, and non-zero variance[1]

$$Var[u] = E_z[\{z(x) - t(x)\}^2] \tag{9.2}$$

Rather than characterize soil properties at every point, data are used to estimate a smooth trend, and remaining variations are described statistically. The residuals are characterized statistically because there are too few data to do otherwise. This does not presume soil properties to be random; clearly, they are not. The variance of the residuals reflects uncertainty about the difference between the interpolated trend and the actual value of soil properties at unobserved locations.

[1] Statistical methods of trend analysis usually presume zero-mean residuals, but this is not true of all trend fitting methods. The Galerkin method of fitting polynomial trends, common in finite element analysis, is a counter example.

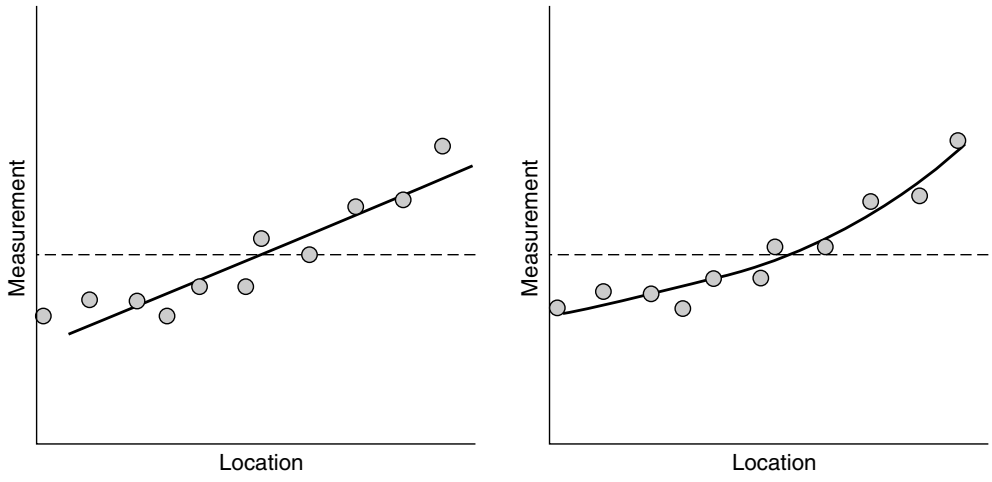


Figure 9.2 Fitting the same data with a line versus a curve changes the residual variance.

The division of spatial variation between trend and residuals is an artifact of how we model soil variability. The data of Figure 9.2 are first modeled using a linear trend. As a result, the residual variation about the trend line has some variance. If, instead, the trend is modeled as a quadratic curve, the trend more closely approximates the data, and the variance of the residual variations is corresponding lower. Carrying this to its extreme, if the trend were modeled using a polynomial one order less than the number of data points, the curve would fit the data perfectly, and the variance of the residuals would be zero. On the other hand, as the flexibility of the trend becomes ever greater, the (statistical) estimation uncertainty in the parameters of the trend rises, eventually approaching infinity as the number of parameters approaches one less than the number of data.

In a world of perfect information, soil properties at every location at the site would be known, and there would be no reason to divide the observed variation artificially. But this is never the case in the real world. Data are limited, and analytical capacity is finite. We model the spatial variation of soil properties using the mathematics of random processes not because soil properties are random, but because our information about those properties is limited. While statistical techniques provide a convenient way to describe what is known about spatial variation, grouping data together may also mask important geological details. The tension between the need to smooth data enough to make them useful in further analyses and the concern that smoothing will eradicate significant details is a salient feature of much engineering work, whether it employs statistical, numerical, graphical, or other tools.

9.1.1 Trend analysis

Trends are estimated by fitting well-defined mathematical functions (i.e., lines, curves, or surfaces) to data points in space. The easiest way to do this is by regression analysis (Chapter 4). For example, Figure 9.3 shows maximum past pressure measurements as a function of depth in a deposit of Gulf of Mexico clay near Mobile. For geological reasons the increase of maximum past pressure with depth is expected to be linear within

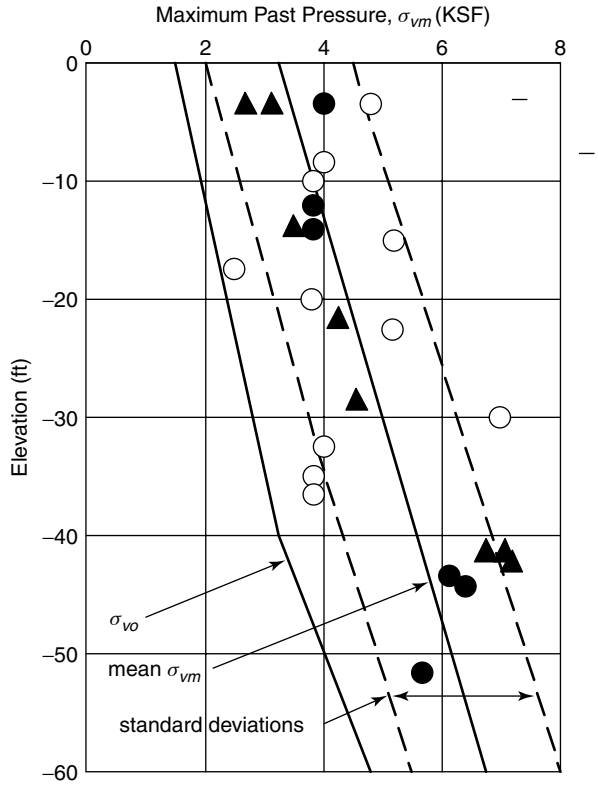


Figure 9.3 Maximum past pressure data as a function of depth (Baecher and Ladd 1997).

this homogeneous stratum. Data from an overlying desiccated crust are not shown. The equation for maximum past pressure, σ'_{vm}, with depth x, is

$$\sigma'_{vm} = t_{\sigma'_{vm}}(x) + u(x) = \beta_0 + \beta_1 x + u \tag{9.3}$$

in which $t_{\sigma'_{vm}}(x)$ is the trend of maximum past pressure with depth, x, β_0 and β_1 are scalar coefficients (intercept and slope), and u = residual variation about the trend taken to be constant with depth.[2] Applying the least squares regression equations of Chapter 4 to the data, thus obtaining the trend parameters minimizing $Var[u] = \sigma^2$, leads to $\beta_0 = 3$ ksf, $\beta_1 = 0.06$ ksf/ft, and $\sigma^2 = 1.0$ ksf, for which the corresponding trend line is shown in Figure 9.3. For data analysis purposes, the trend line $t_{\sigma'_{vm}} = E[\sigma'_{vm}|x] = 3 + 0.06\text{x}$ is the best estimate or mean of the maximum past pressure as a function of depth. This regression analysis can be done using either a frequentist or Bayesian approach, and all the philosophical concerns attending this choice, obviously, apply.

A similar analysis can be made of data in higher dimensions. Figure 9.4a shows a map of the locations of borings within a region of a large office development project underlain

[2] Here we have used x to indicate depth, rather than the more common z, to provide consistency with the generic expressions for multidimensional regression found in the statistical literature, and used in Chapter 4.

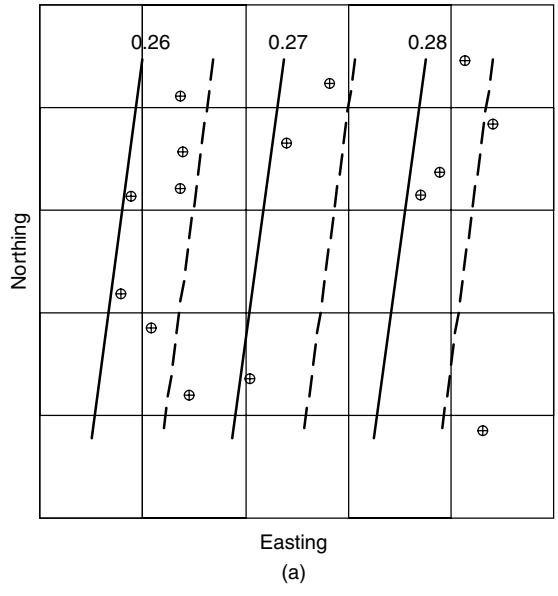


Figure 9.4a Boring locations, showing contour lines of the best-fitting plane to the boring-average shear strength ratio.

by lacustrine clay. Taking the measured undrained strength ratios (S_u/σ'_{vo}) through the clay in each boring, a best fitting planar trend to the spatial variation of the mean is taken as

$$z(x_1, x_2) = \{\beta_0 + \beta_1 x_1 + \beta_2 x_2\} + u(x_1, x_2) \tag{9.4}$$

which may be rewritten in matrix notation as

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{9.5}$$

in which $\mathbf{z}$ is the vector of the n observations $\mathbf{z} = \{z_1, \ldots, z_n\}$, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2\}$ is the $2 \times n$ matrix of location coordinates corresponding to the observations, $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2\}$ is the vector of trend parameters, and $\mathbf{U}$ is the vector of residuals corresponding to the observations.

Minimizing the variance of the residuals $\mathbf{u}(\mathbf{x})$ over $\boldsymbol{\beta}$ gives the best fitting trend surface in a frequentist sense (again, this analysis could also be made in a Bayesian sense):

$$\min_{\beta} \mathrm{Var}[\mathbf{u}] = \min_{\beta} \mathrm{E}[\mathbf{z} - \mathbf{X}\boldsymbol{\beta}]^2 \tag{9.6}$$

which is the common regression surface, described in more detail in Chapter 4.[3] Contours of this planar trend surface are also shown in Figure 9.4a. The corresponding variance

[3] For the general, n-dimensional case, the estimates of the coefficients $\boldsymbol{\beta}$ which satisfy the minimization of the residual variance, $\hat{\boldsymbol{\beta}}$, are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{z}$$

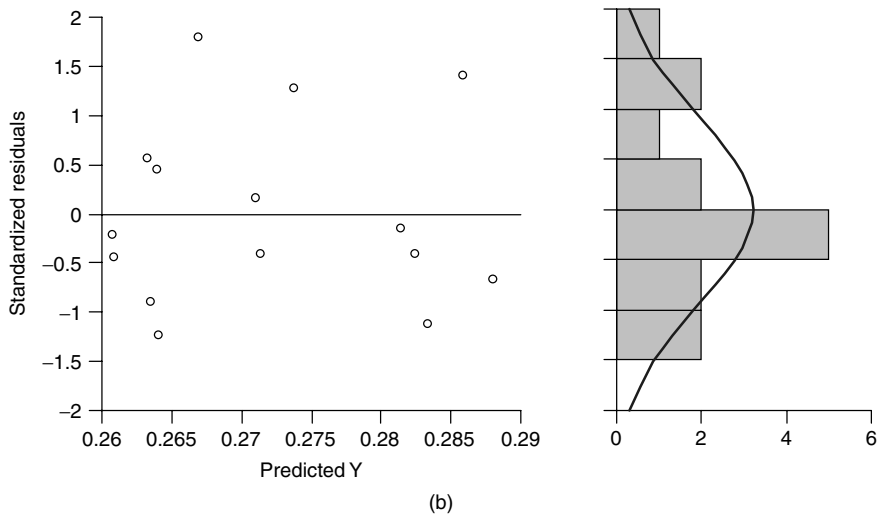


Figure 9.4b Residual variations for the data of Figure 9.4a.

of the residuals about the trend is shown in Figure 9.4b, and the relationship of values predicted by the trend compared to the measurements in Figure 9.4c.

For the quadratic case, the linear expression in Equation (9.4) is replaced by

$$z(\mathbf{x}) = \{\beta_0 + \beta_1 x_1 + \beta_2 {x_1}^2 + \beta_3 x_2 + \beta_4 {x_2}^2 + \beta_5 x_1 x_2\} + u(x_1, x_2) \tag{9.7}$$

and the calculation for $\hat{\boldsymbol{\beta}}$ performed the same way. Because the quadratic surface is more flexible than the planar surface, it fits the observed data more closely, and the residual variations about it are smaller.

Examples of the use of trend surfaces in the geosciences are given by Krumbein and Graybill (1965), Davis (1986), and Agterberg (1974). Examples in the geotechnical literature are given by Wu *et al*. (1996) and Ang and Tang (1975).

9.1.2 Uncertainty in trend parameters: frequentist approach

When trend surfaces are computed, the calculated parameters are those that provide the trend that most closely fits a given set of observations. Had the observations been made in slightly different places, or had the observed values been slightly different because of measurement errors, then the best fitting parameter values would be slightly different.

in which

$$\mathrm{X} = \begin{pmatrix} 1 & \mathrm{x}_{1,1} & \ldots & \mathrm{x}_{\mathrm{k},1} \\ 1 & \mathrm{x}_{1,2} & \ldots & \mathrm{x}_{\mathrm{k},2} \\ \ldots & & & \\ 1 & \mathrm{x}_{1,\mathrm{n}} & \ldots & \mathrm{x}_{\mathrm{k},\mathrm{n}} \end{pmatrix}$$

(sometimes called, the *design matrix* of locations), n is the number of dimensions of the trend model, k is the number of data, and $\mathbf{z}$ is the vector of observed values.

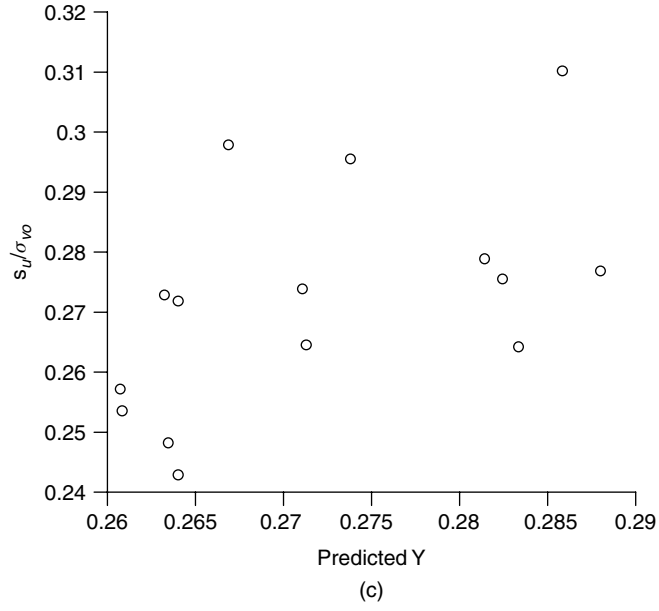


Figure 9.4c Comparison of shear strength ration predicted by best-fitting planar regression to the observed values.

Estimating coefficients of trend surfaces involves statistical error, and the estimates of the parameters differ if the same population is sampled repeatedly.

The probability distribution of the parameters over repeated sampling is said to be the 'sampling distribution' of the parameters. The sampling variances and covariances of the estimated coefficients, $\hat{\boldsymbol{\beta}}$, are summarized in the covariance matrix

$$\Sigma_{\hat{\beta}} = \begin{pmatrix} Var[\hat{\beta}_0] & Cov[\hat{\beta}_0, \hat{\beta}_1] & \dots & Cov[\hat{\beta}_0, \hat{\beta}_n] \\ Cov[\hat{\beta}_1, \hat{\beta}_0] & Var[\hat{\beta}_1] & \dots & \dots \\ \dots & \dots & \dots & \dots \\ Cov[\hat{\beta}_n, \hat{\beta}_0] & \dots & \dots & Var[\hat{\beta}_n] \end{pmatrix} \tag{9.8}$$

which is obtained by propagating residual variances through Equation 9.6:

$$\Sigma_{\hat{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1} Var[\mathrm{u}] \tag{9.9}$$

presuming the residuals to be independent of one another. This assumption of independence of the residuals for spatial data such as those obtained in site characterization is only good if the observations are widely spaced. We return to this theme in the Section 9.3.

For the one-dimensional case of the maximum past pressure trend of Equation (9.3), the sampling variances of the intercept β_0 and slope β_1 are[4]

[4] Matrix notation for finding the sampling variances of trend coefficients is convenient, but for the simple one-dimensional case the solution can also be calculated from the expressions

$$\Sigma_{\hat{\beta}} = \begin{pmatrix} Var[\hat{\beta}_0] & Cov[\hat{\beta}_0, \hat{\beta}_1] \\ Cov[\hat{\beta}_1, \hat{\beta}_0] & Var[\hat{\beta}_1] \end{pmatrix} = \begin{pmatrix} 0.1691 & -0.0002 \\ -0.0002 & 0.0002 \end{pmatrix} \tag{9.10}$$

Again, these are the variances of the regression coefficients in a frequentist sense over repeated sampling, and are shown as standard deviation envelopes on Figure 9.3. From these, the standard deviation of the sampling variation (standard error) of the intercept is 0.411, and that of the slope is 0.0143. The intercept and slope estimates are slightly, negatively correlated, with $\rho = -0.04$.

The common assumption in analyzing trends is that residual variation is Normal (Gaussian); that is, the variations about the trend line or surface are Normally distributed. Usually, this is a reasonable assumption, but, in any event, the statistical properties of the estimators of the parameters of the trend model are robust (i.e. insensitive) to variations from the Normal assumption. In the Normal case, the sampling distributions of the estimators $\hat{\boldsymbol{\beta}}$ of the parameters $\boldsymbol{\beta}$ are themselves jointly Normal, with means equal to the least squares values, and with the covariance matrix of Equation (9.9). Since the covariance matrix of Equation (9.9) itself depends upon the variance of the residuals, the Normality of $\hat{\boldsymbol{\beta}}$ is conditional on the value of σ^2. When σ^2 is unknown (i.e. σ^2 must itself be estimated from the sample), the sampling distribution of $\hat{\boldsymbol{\beta}}$ based on the estimate of σ^2 has the Student-t form (Chapter 4). The Student-t distribution is similar to the Normal distribution, but has thicker tails to account for the added uncertainty of not knowing the true value of σ^2. The covariance matrix of the estimated parameters becomes $n\Sigma_{\hat{\beta}}/(n-k)$, in which n is the number of observations and k is the number of trend parameters. Clearly, for a given n, as the flexibility of the trend, reflected in the number of its parameters, increases, the uncertainty in the estimates of those parameters increases. For $k \to n$, the uncertainty becomes infinite. This result parallels the situation of independent, identically distributed (IID) sampling from a univariate Normal population (Chapter 4). Tests of confidence, significance limits, and other traditional estimation procedures are straightforward. These are summarized in Chapter 4 and in texts on regression analysis (Draper and Smith 1998; Johnson 1960).

In practice, it is common for people to use the sampling variances of Equation (9.9) as if those variances directly expressed uncertainty on the trend parameters $\boldsymbol{\beta}$, that is, as if the variances were the moments of a probability density function (pdf) over $\boldsymbol{\beta}$. This is not the case. The least square estimates and the associated sampling variances of the estimates have to do with the behavior of the estimators in repeated sampling.

$$Var(\beta_0) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 Var[u] = \frac{\sigma^2}{n}$$

$$Var(\beta_1) = \sum_{i=1}^{n} \left[\frac{x_i - \mu_x}{\sum_{i=1}^{n}(x_i - \mu_x)^2}\right]^2 Var(u) = \frac{\sigma^2}{nVar[x]}$$

in which $\sigma^2 = Var[u]$. Note that each variance is inversely proportional to n, the number of observations. The fewer observations, the larger the sampling variance of the coefficients. When σ^2 is unknown but must itself be estimated from the same data, as $\hat{\sigma}^2$, the sampling variances of β_0 and β_1 increase by the factor $n/(n-2)$.

They answer the question, if the same number of observations were made in slightly different ways or in slightly different places, how might the estimates so obtained vary from one another? The least squares approach as presented is a frequentist method, and does not admit probability statements directly on estimated parameters. A Bayesian approach is needed to answer the reverse question, which is, given a set of observations, what is the pdf of $\boldsymbol{\beta}$? Fortunately, under reasonably broad assumptions, the second moments of the Bayesian posterior pdf of $\boldsymbol{\beta}$ are close to or identical with those of Equation (9.9).

9.1.3 Uncertainty in trend parameters: Bayesian approach

The Bayesian approach to trend analysis begins with the same model of Equations (9.1) or (9.5). However, rather than defining an estimator such as the least squares coefficients and then calculating the behavior of those estimators in repeated sampling, the Bayesian approach specifies a pdf on the coefficients of the model $(\boldsymbol{\beta}, \sigma)$, and uses Bayes' Theorem to update that pdf in light of the observed data (for a more detailed discussion, see Chapter 4).

Adopting the trend model of Equation (9.5), let $Var[u] = \sigma^2$, so that $Var[\mathbf{u}] \equiv \boldsymbol{\Sigma} = \mathbf{I}\sigma^2$. The prior probability density function (pdf) of the parameters $(\boldsymbol{\beta}, \sigma)$ is then represented as $f(\boldsymbol{\beta}, \sigma)$. Given a set of observations $\mathbf{z} = \{z_1, \ldots, z_n\}$, the updated or posterior pdf of $(\boldsymbol{\beta}, \sigma)$ is found from Bayes' Theorem, $f(\boldsymbol{\beta}, \sigma|\mathbf{z}) \propto f(\boldsymbol{\beta}, \sigma)L(\boldsymbol{\beta}, \sigma|\mathbf{z})$, in which $L(\boldsymbol{\beta}, \sigma|\mathbf{z})$ is the Likelihood of the data. Assuming, as in the previous section, that variations about the trend line or surface are jointly Normal, the likelihood function becomes

$$L(\boldsymbol{\beta}, \boldsymbol{\sigma}|\mathbf{z}) = MN(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\sigma}) \propto \exp\{-(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\} \tag{9.11}$$

in which $MN(\mathbf{z}|\boldsymbol{\beta}, \sigma)$ is the multivariate Normal distribution having mean $\mathbf{X}\beta$ and covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}\sigma^2$.

Addressing the linear, one-dimensional case for which $\boldsymbol{\beta} = (\beta_1, \beta_2)$, and adopting a non-informative prior, $f(\beta_1, \beta_2, \sigma) \propto 1/\sigma$, the posterior pdf of the regression parameters is (Zellner 1971)

$$f(\beta_1, \beta_2, \sigma|\mathbf{x}, \mathbf{y}) \propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_1 + \beta_2 x_i)^2\right] \tag{9.12}$$

The marginal distributions of the regression coefficients are then,

$$\begin{aligned}
f(\beta_1, \beta_2|\mathbf{x}, \mathbf{y}) &\propto [vs^2 + n(\beta_1 - \hat{\beta}_1)^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\Sigma x_i + (\beta_2 - \hat{\beta}_2)^2\Sigma x_i^2]^{-n/2} \\
f(\beta_1|\mathbf{x}, \mathbf{y}) &\propto [v + \frac{\Sigma(x_i - \overline{x})^2}{s^2\Sigma x_i^2/n}(\beta_1 - \hat{\beta}_1)^2]^{-(v+1)/2} \\
f(\beta_2|\mathbf{x}, \mathbf{y}) &\propto [v + \frac{\Sigma(x_i - \overline{x})^2}{s^2}(\beta_2 - \hat{\beta}_2)^2]^{-(v+1)/2} \\
f(\sigma|\mathbf{x}, \mathbf{y}) &\propto \frac{1}{\sigma^{v+1}} \exp\left(-\frac{vs^2}{2\sigma^2}\right)
\end{aligned} \tag{9.13}$$

in which $\nu = n - 2$, $\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$, $\hat{\beta}_2 = [\Sigma(x_i - \overline{x})(y_i - \overline{y})]/[\Sigma(x_i - \overline{x})^2]$, $s^2 = \nu^{-1}\Sigma(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$, $\overline{y} = n^{-1}\Sigma y_i$, and $\overline{x} = n^{-1}\Sigma x_i$. The joint and marginal pdf's of $\beta 1$, β_2 are Student-t, which means that inferences about β_1 and β_2 can be made from tables of the Student-t pdf.

9.2 Residual Variations

In fitting trends to data, a decision is made to divide the total variability of the data, expressed as a variance, into two parts: one part explained by the trend; the other part reflected in variability about the trend. Residual variations not accounted for by the trend are characterized by a residual variance. For example, the overall variance of the blow count data of Figure 9.5 is 45 bpf^2. Removing a linear trend reduces this total to a residual variance of about 11 bpf^2. The trend explains 33 bpf^2, or about 75% of the spatial variation. Conversely, 25% of the spatial variation remains unexplained by the trend.

Given the procedure by which the trend is fitted, the residuals by definition have zero mean. The assumption made in fitting the trend is that (squared) deviations of the residuals are of equal importance, no matter where in space they occur. Underlying the statistical model for calculating uncertainties in the trend parameters is a similar but stronger assumption that the variance of the residual is probabilistically the same no matter where the residuals occur in space. This is called *homoscedasticity*. This assumption is usually good, but may be relaxed (Johnson 1960). A second assumption that was made in fitting the trend is that the residual variations are independent of one another. That is, knowing the residual variation at one location provides no information about residual variations at other locations, for example, those near by. This assumption is not so good; in fact, it is seldom strictly true for geotechnical data. The trend may explain a large measure of the spatial structure of soil variability, but spatial structure remains within the residuals themselves.

The spatial structure remaining after a trend is removed commonly manifests in correlations among the residuals. That is, the residuals off the trend are not statistically

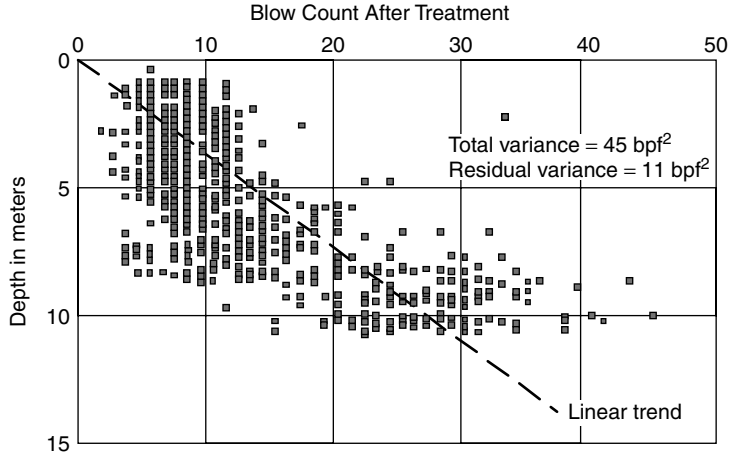


Figure 9.5 SPT blow count data in a silty-sand bay fill (T.W. Lambe Associates 1982).

independent of one another. Positive residuals tend to be clumped together, as do negative residuals. Knowing that an observation at point x_i is above the trend suggests that the observation at a nearby point x_j will, with probability greater than 0.5, also be above the trend at its respective location. The same is true, conversely, for an observation at x_i that would be below the trend. The practical importance of this remaining spatial structure in the residuals is that, for example, low values of soil strength may occur together; thus, the probability of encountering a continuous zone of weakness is greater than would otherwise be predicted.

Figure 9.6 shows residual variations of SPT blow counts measured at the same elevation every 20 m beneath a horizontal transect at a site. The data are normalized (standardized) to zero mean and unit standard deviation. The dark line is a smooth curve drawn through the observed data. The light line is a smooth curve drawn through artificially simulated data having the same mean and same standard deviation, but which are probabilistically independent. Inspection shows the natural data to be smoothly varying, or wavy. The artificial, probabilistically independent data are more erratic.

9.2.1 Autocorrelation

The remaining spatial structure of variation in the residuals can be described by spatial correlation, usually called *autocorrelation*, because it refers to correlations of an individual variable with itself over space (or time, or some other dimension). Formally, autocorrelation is the property that residuals off the mean trend are not probabilistically independent,

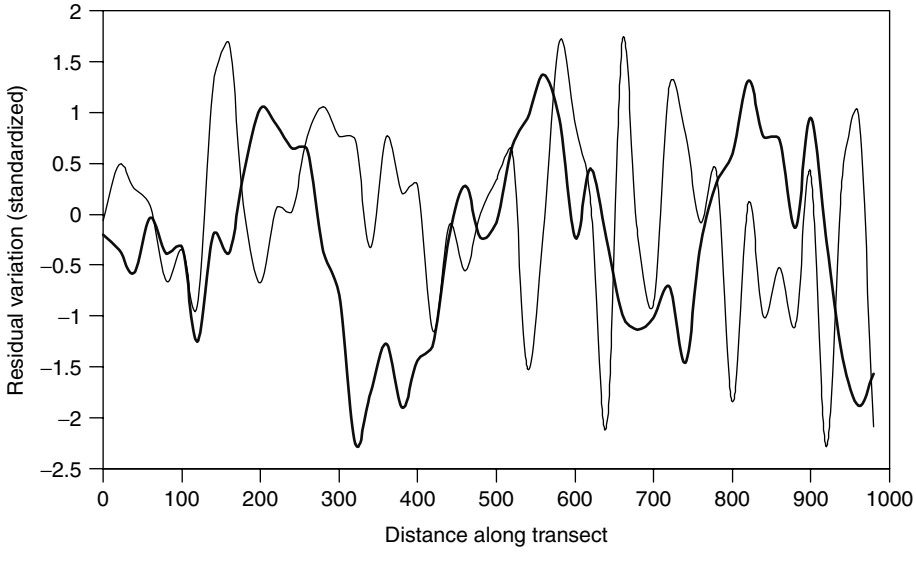


Figure 9.6 Residual variations of SPT blow counts measured at the same elevation every 20 m beneath a horizontal transect at a site. The data are normalized (standardized) to zero mean and unit standard deviation. The dark line is a smooth curve drawn through the observed data. The light line is a smooth curve drawn through artificially simulated data having the same mean and same standard deviation, but which are probabilistically independent.

but display a degree of association among themselves which is a function of their separation in space. This degree of association can be measured by a correlation coefficient, taken as a function of separation distance.

Correlation was introduced in Chapter 4. Correlation is the property that, on average, two variables are linearly associated with one another. Knowing the value of one provides information on the probable value of the other. The strength of this association is measured by a correlation coefficient ρ that ranges between -1, and $+1$. A correlation coefficient $\rho = +1$ means that two residuals vary together exactly. When one is a standard deviation above its trend, the other is a standard deviation above its trend, too. A correlation coefficient $\rho = -1$ means that two residuals vary exactly inversely. When one is a standard deviation above its trend, the other is a standard deviation below its trend. A correlation coefficient $\rho = 0$ means that the two residuals are unrelated to one another.[5] Knowing the realized value of one provides no information about the realized value of the other.

For two scalar variables z_1 and z_2, the correlation coefficient is defined as

$$\rho = \frac{Cov(z_1, z_2)}{\sqrt{Var(z_1)Var(z_2)}} = \frac{1}{\sigma_{z_1}\sigma_{z_2}}\mathrm{E}[(z_1 - \mu_{z_1})(z_2 - \mu_{z_2})] \tag{9.14}$$

in which $Cov(z_1, z_2)$ is the covariance, $Var(z_i)$ is the variance, σ is the standard deviation, and μ is the mean. The two variables might be of different but related types, for example, z_1 might be water content and z_2 might be undrained strength, or the two variables might be the same property at different locations, for examples z_1 might be the water content at one place on the site and z_2 the water content at another place.

The locations at which the blow count data of Figure 9.6 were measured are shown in Figure 9.7. In Figure 9.8 these data are plotted against each other as a function of separation distance, δ. That is, in Figure 9.8a all the data pairs having separation distance $\delta = 20\,\mathrm{m}$ are plotted; then, in Figure 9.8b, all the data pairs having separation distance $\delta = 40\,\mathrm{m}$ are plotted; and so on. The data pairs at close separation exhibit a high degree of correlation, for example, those separated by 20 m have a correlation coefficient of 0.67. As separation distance increases, correlation goes down, although at large separations, where the numbers of data pairs are smaller, there is much statistical fluctuation. The observed correlations as a function of separation distance, δ, are shown in Figure 9.8e. For zero separation distance ($\delta = 0$), the correlation coefficient must equal 1.0. For large separation distances ($\delta \to \infty$), the correlation coefficient typically goes to zero. In between, the autocorrelation of geotechnical data usually falls monotonically from 1.0 to zero, although this need not be the case.

The effect of correlation structure on residual variation can be seen in Figure 9.9 in which four cases are sketched. Spatial variability about a trend is characterized by variance and autocorrelation. Large variance implies that the absolute magnitude of the residuals is large; large autocorrelation implies that the 'wave length' of the variation is long.

An important observation is that the division of spatial variation into a trend and residuals about the trend is an artifact of analysis. By changing the trend model – for example, by replacing a linear trend with a polynomial – both the variance of the residuals

[5] This statement is not strictly true. Correlation is a measure of linear association. Two variables can be deterministically related to one another in a non-linear way and yet have a zero correlation coefficient.

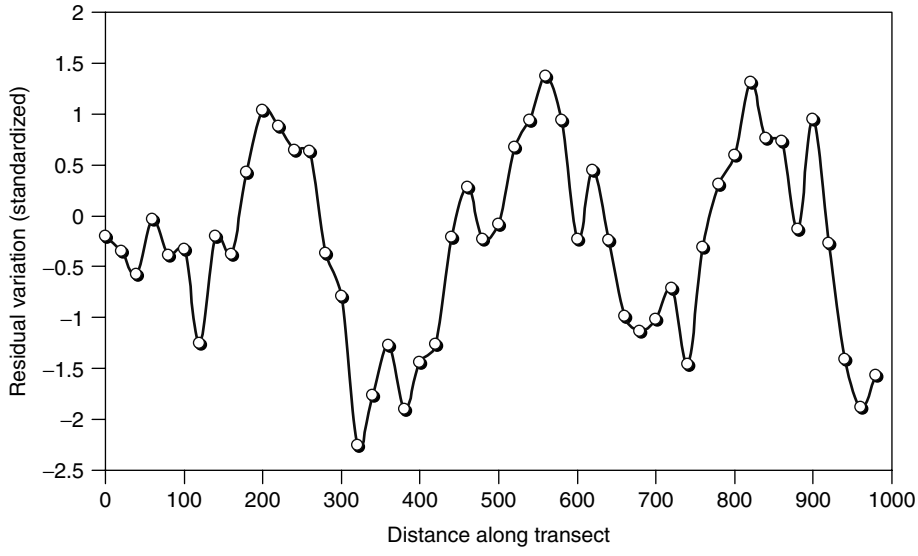


Figure 9.7 Sampling locations for the data of Figure 9.6.

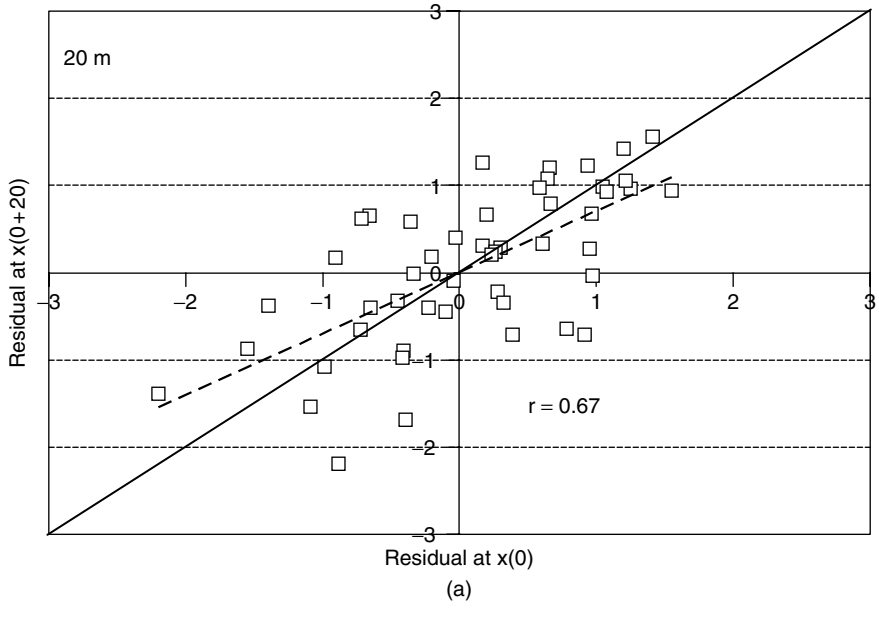


Figure 9.8a SPT data of Figure 9.7 plotted against each other for separation distance, $\delta = 20\,\text{m}$.

and their autocorrelation function are be changed. This can be seen in an analysis that DeGroot (1996) made, the results of which are shown in Figure 9.10. As the flexibility of the trend increases, the variance of the residuals goes down, and in general the extent of correlation is reduced. From a practical point of view, the selection of a trend line or curve is in effect a decision on how much of the data scatter to model as a deterministic

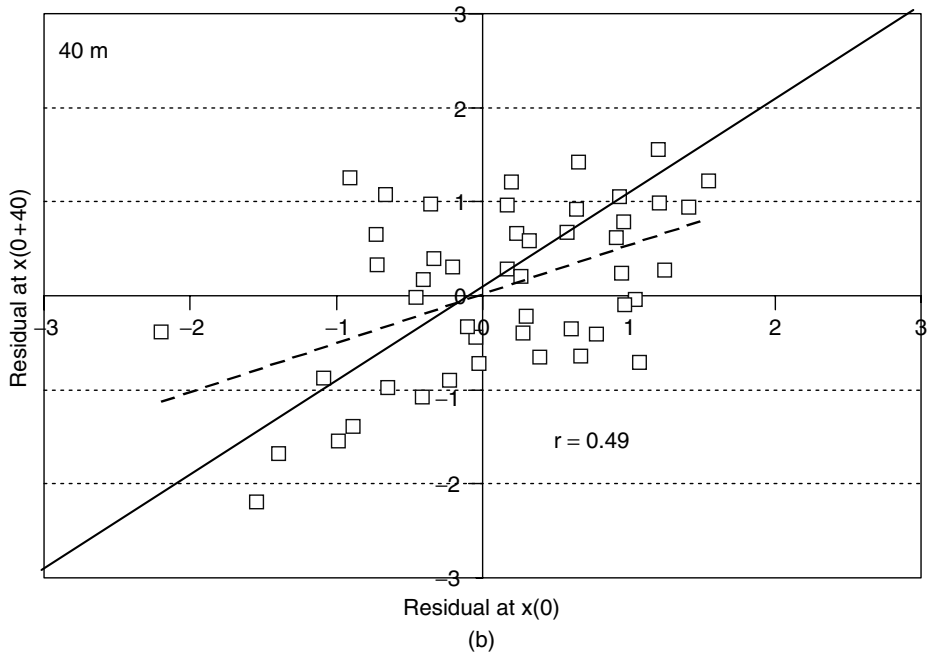


Figure 9.8b SPT data of Figure 9.7 plotted against each other for separation distance, $\delta = 40\,\text{m}$.

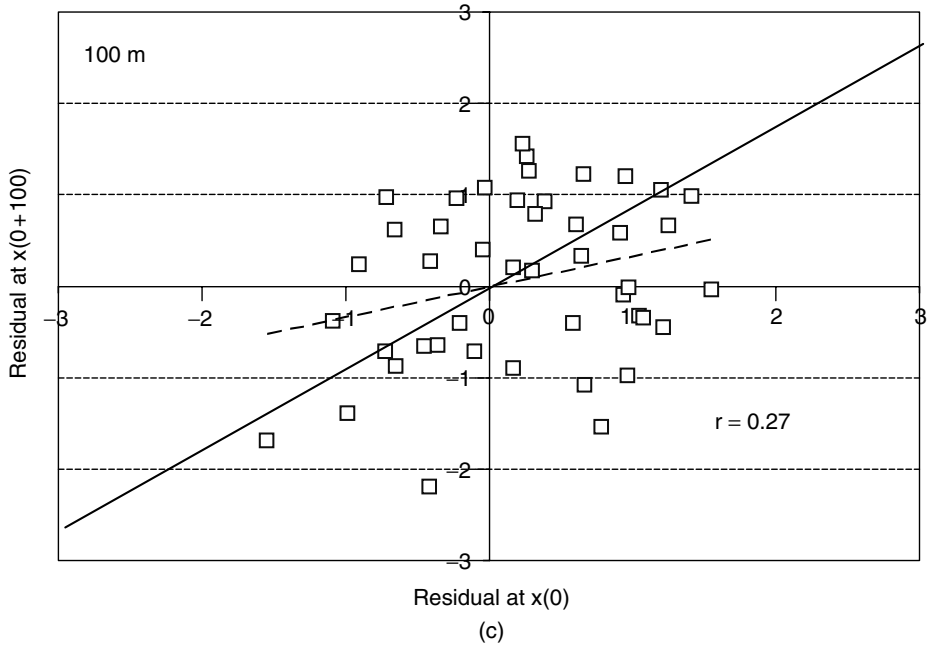
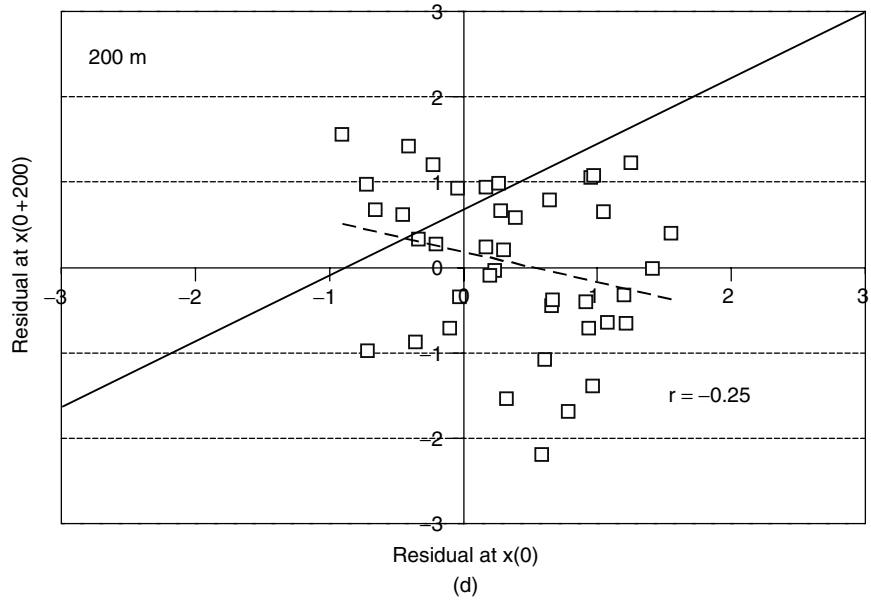
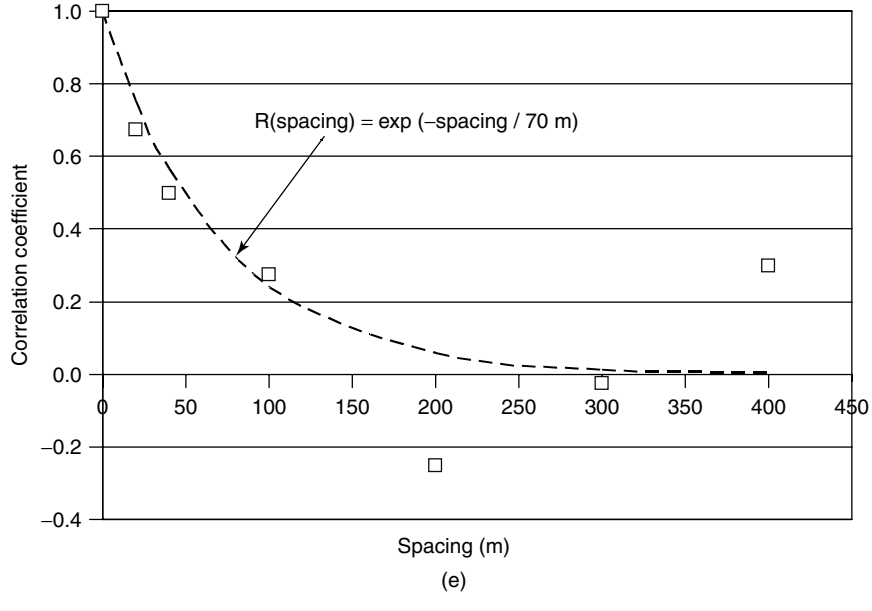


Figure 9.8c SPT data of Figure 9.7 plotted against each other for separation distance, $\delta = 100\,\text{m}$.



(d)

Figure 9.8d SPT data of Figure 9.7 plotted against each other for separation distance, $\delta = 200\,\text{m}$.



(e)

Figure 9.8e Autocorrelation function for the data of Figure 9.7.

function of space, and how much to treat statistically. Dividing spatial variability into a deterministic part and a statistical part is a matter of convenience. Prudence requires that each datum be judged for what it might say about a soil deposit, but engineering analysis requires models of soil properties for making predictions.

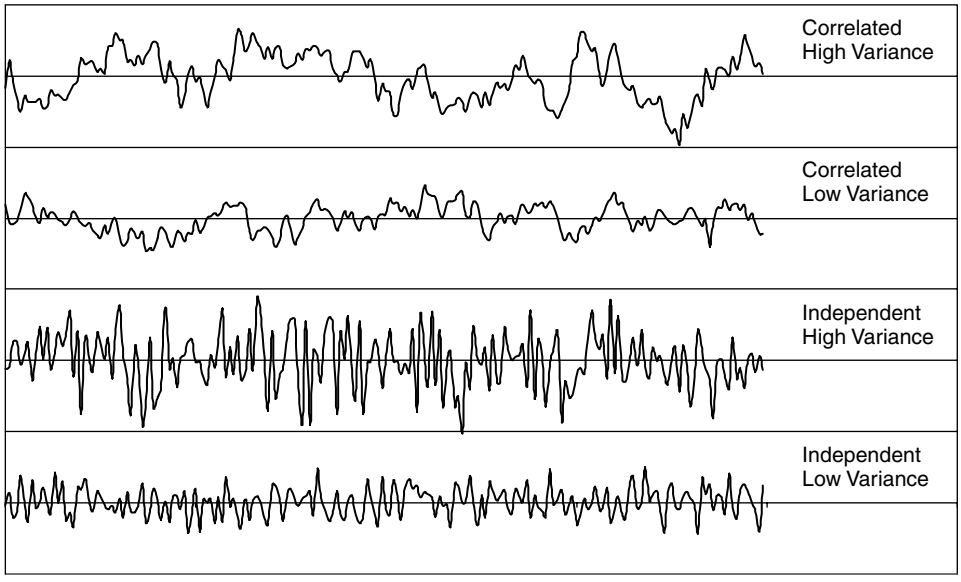


Figure 9.9 Effect of correlation structure on residual variation.

As a rule of thumb, trend surfaces should be kept as simple as possible without doing injustice to a set of data or ignoring the geologic setting. The problem with using trend surfaces that are very flexible, as for example high order polynomials, is that the number of data from which the parameters of those equations are estimated is limited. As was seen in Equation (9.9), the sampling variance of the trend coefficients is inversely proportional to the degrees of freedom, $v = (n - k - 1)$, in which n is the number of observations and k is the number of parameters in the trend. The more parameter estimates that a trend surface requires, the more uncertainty there is in the numerical values of those estimates. Uncertainty in regression coefficient estimates increases rapidly as the flexibility of the trend equation increases.

9.2.2 Autocorrelation of a deck of cards

A simple but illustrative example of autocorrelation is that of an ordered deck of cards, due to Jensen (1997). Consider a fresh deck of standard cards arranged in suites and by pip value. Let the Ace be one, and so forth, up to the King being 13. The sum of the values in the deck is 364, so the mean is $m = 364/52 = 7$. The standard deviation is 3.78.

To begin, the cards in the deck are strictly ordered. The pip values in each suit are arranged from Ace to King, and one suit follows another. The autocorrelation function, calculated by Equation (9.18), is a series of repeating curves (Figure 9.11). The correlation between every thirteenth card is perfect, and the correlation for other separations follows a regular, repeating pattern. The correlation between every seventh card is strongly negative, because if the first card in the pair has a pip value greater than seven, the mean, the pip value of the second card will have a value less than seven, and conversely. Note there is nothing random in the order of the cards, the ordering is completely known. Nonetheless, an autocorrelation function can be calculated.

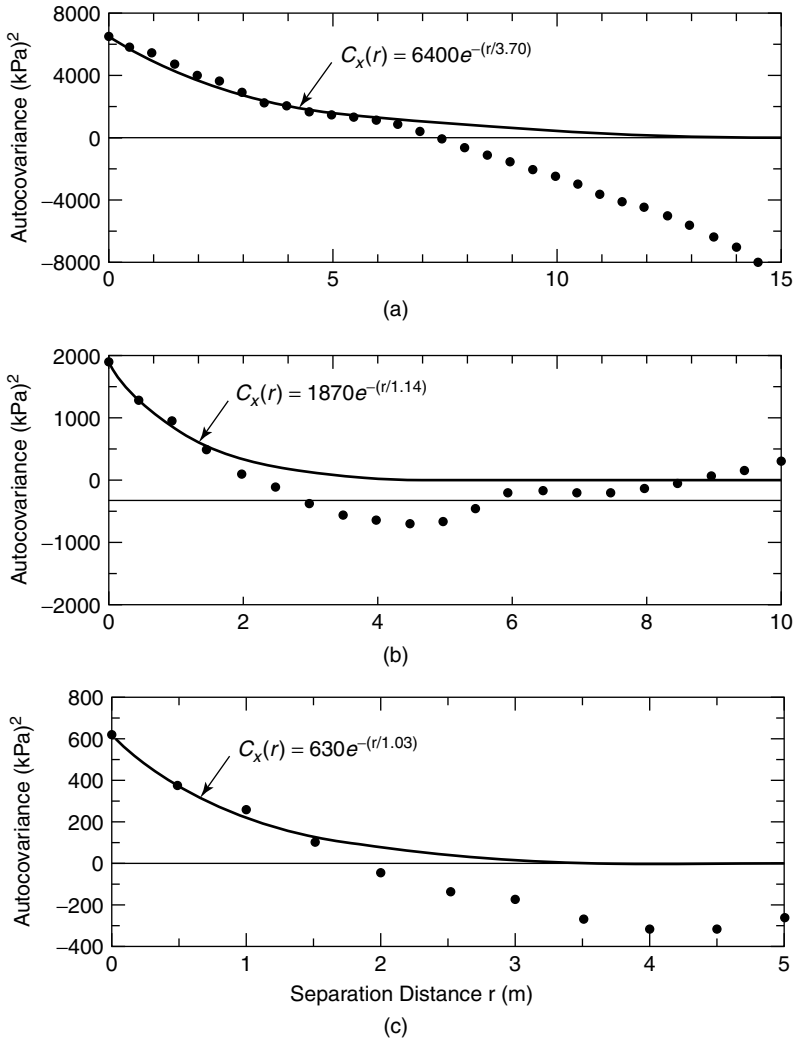


Figure 9.10 Method of moments autocovariance plots for DMT data, assuming (a) mean value trend, (b) single function linear regression trend, and (c) bilinear regression trend (note differences in x-axis scales) (DeGroot 1996).

If the cards are now shuffled, the ordering becomes less perfect. Neither the mean nor the standard deviation changes as we shuffle the deck, but the autocorrelation function does. It becomes less regular. Figure 9.11 shows the experimentally developed autocorrelation functions calculated after one, two, and three shuffles of the deck. After the first shuffle some, of the original pattern of the autocorrelation remains, but after the third shuffle the autocorrelation is essentially a spike at zero. The ordering of the cards has become almost unrelated.

Note again, even after the cards have been shuffled, there is nothing "random" in the ordering of the cards, the ordering is again completely known to us, because we looked. Nonetheless, an autocorrelation function can be calculated and used to summarize

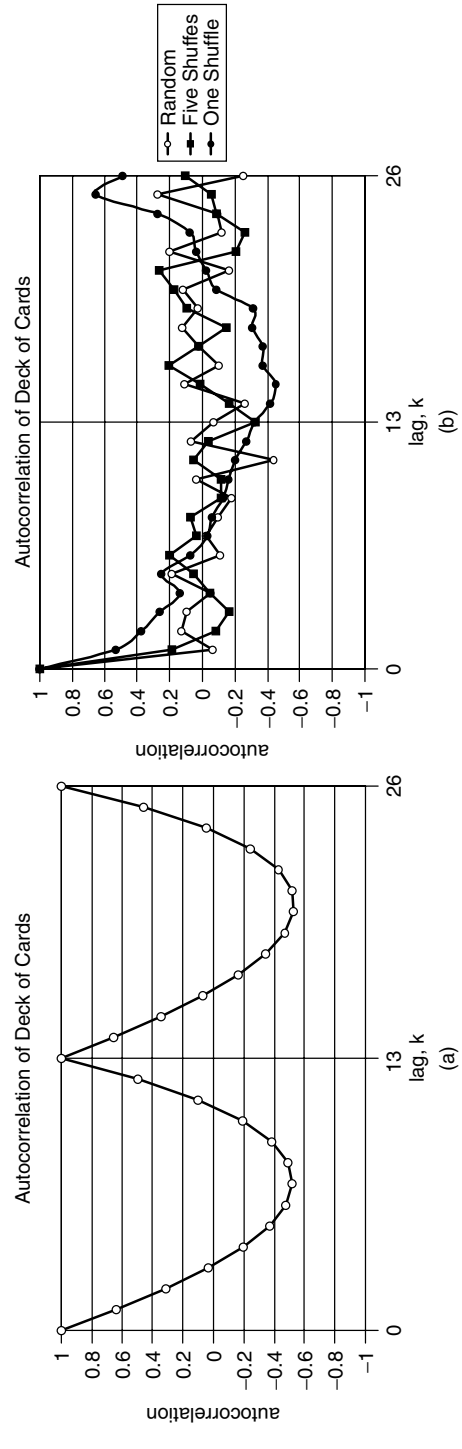


Figure 9.11 Autocorrelation function of an ordered deck of playing cards. (a) Original configuration, Originally, on left, fresh out of the package, and (b) after one and five shuffles and explicitly randomized (example reworked after Jensen 1997).

the pattern of variation in the deck. This is exactly the same situation with the spatial variation of soil or rock properties. The spatial pattern may or may not be known to us, but it is not random. We use autocorrelation to describe the pattern of spatial variation in a parsimonious way.

9.2.3 Autocovariance and autocorrelation functions

If $z(x_i) = t(x_i) + u(x_i)$ is a continuous variable and the soil deposit is zonally homogeneous, then at locations i and j, which are close together, the residuals u_i and u_j should be expected to be similar. That is, the variations reflected in $u(x_i)$ and $u(x_j)$ are associated with one another. When the locations are close together, the association is usually strong. As the locations become more widely separated, the association usually decreases. As the separation between two locations i and j approaches zero, $u(x_i)$ and $u(x_j)$ become the same, the association becomes perfect. Conversely, as the separation becomes large, $u(x_i)$ and $u(x_j)$ become independent, the association becomes zero. This is the behavior observed in Figure 9.8 for SPT data.

This spatial association of residuals off the trend $t(x_i)$ is summarized by a mathematical function describing the correlation of $u(x_i)$ and $u(x_j)$ as separation distance δ increases. This description is called the *autocorrelation function*. In concept, the autocorrelation function is a mathematical way of summarizing the correlations from the scatter plots of Figure 9.8 and shown in Figure 9.8e. Mathematically, the autocorrelation function is

$$R_z(\delta) = \frac{1}{Var[u(x)]} E[u(x_i)u(x_{i+\delta})] \tag{9.15}$$

in which $R_z(\delta)$ is the autocorrelation function, $Var[u(x)]$ is the variance of the residuals across the site, and $E[u(x_i)u(x_{i+\delta})] = Cov[u(x_i), u(x_{i+\delta})]$ is the covariance of the residuals spaced at separation distance δ. By definition, the autocorrelation at zero separation is $R_z(0) = 1.0$; and empirically, for most geotechnical data, autocorrelation decreases to zero as δ increases.

If $R_z(\delta)$ is multiplied by the variance of the residuals, $Var[u(x)]$, the autocovariance function, $C_z(\delta)$ is obtained:

$$C_z(\delta) = E[u(x_i)u(x_{i+\delta})] \tag{9.16}$$

The relationship between the autocorrelation function and the autocovariance function is the same as that between the correlation coefficient and the covariance, except that autocorrelation and autocovariance are functions of separation distance, δ.

The SPT data shown earlier in Figure 9.5 come from a site shown in Figure 9.12 which overlies hydraulic bay fill in Kawasaki Harbor (Japan). The SPT data were taken in the silty fine sand between elevations +3 and −7 m, and show little if any trend horizontally, so a constant horizontal trend at the mean of the data is assumed (Figure 9.13). Figure 9.14 shows the means and variability of the SPT data with depth. Figure 9.15 shows autocovariance functions in the horizontal direction estimated for three intervals of elevation. At short separation distances the data show distinct association, i.e. correlation. At large separation distances the data exhibit essentially no correlation.

In natural deposits, correlations in the vertical direction tend to have much shorter distances than in the horizontal direction. A ratio of about one to ten for these correlation

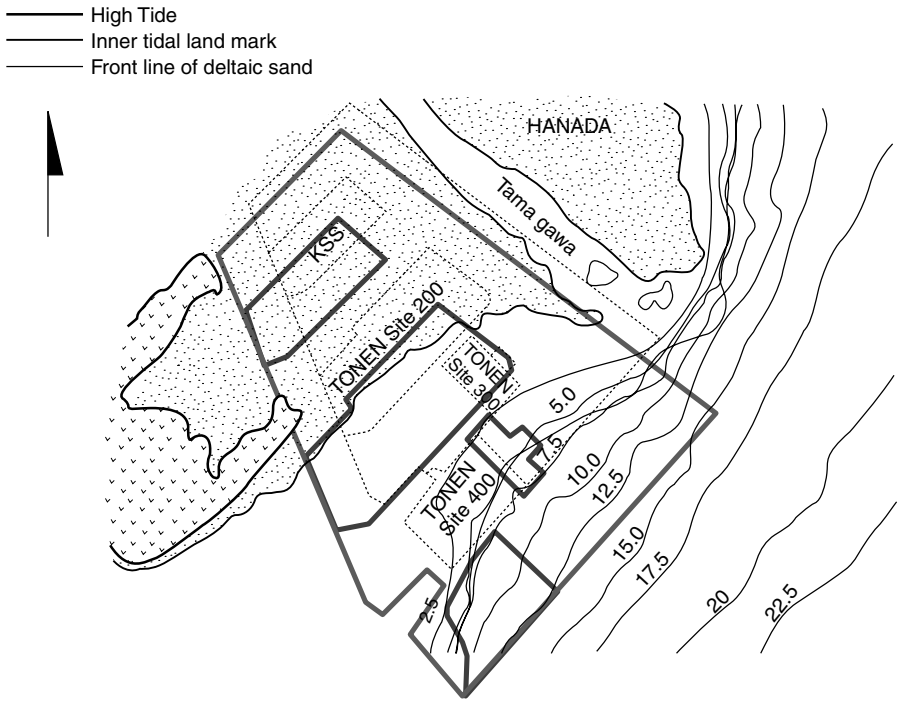


Figure 9.12 Land reclamation site showing property boundary and original water depth contours. (Lambe, T. W., and Associates, 1982. *Earthquake Risk to Palio 4 and Site 400*, Longboat Key, FL, reproduced by permission of T. W. Lambe.)

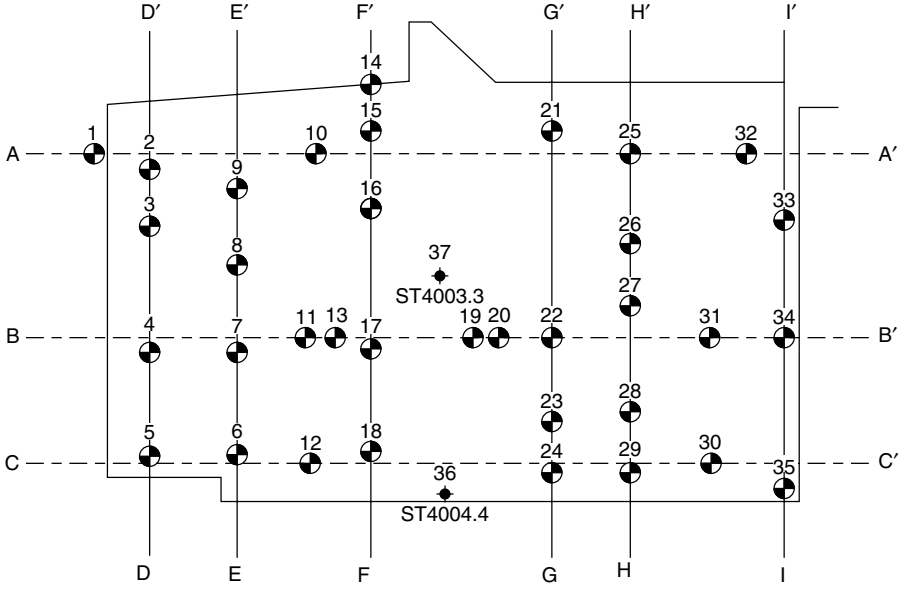


Figure 9.13 Boring locations at Site 400. (Lambe, T. W., and Associates, 1982. *Earthquake Risk to Palio 4 and Site 400*, Longboat Key, FL, reproduced by permission of T. W. Lambe.)

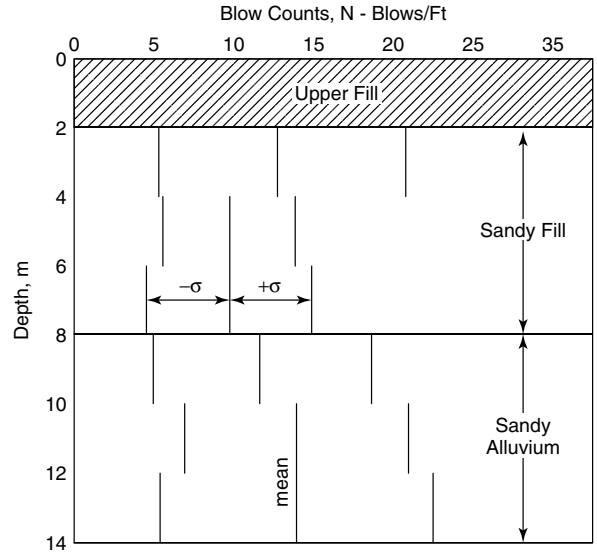


Figure 9.14 Soil model and blow count data used to describe Site 400. (Lambe, T. W., and Associates, 1982. *Earthquake Risk to Palio 4 and Site 400*, Longboat Key, FL, reproduced by permission of T. W. Lambe.).

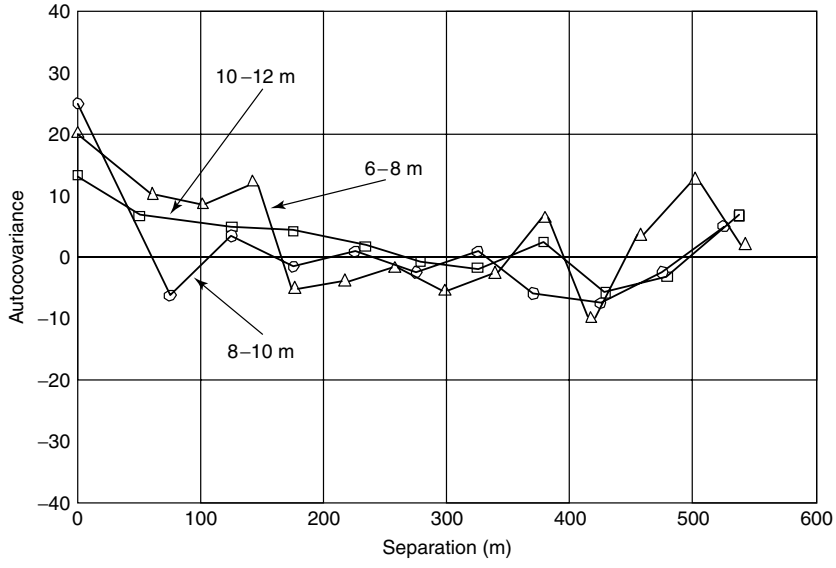


Figure 9.15 Autocorrelation function for the SPT data at Site 400, separated by depth interval. (Lambe, T. W., and Associates, 1982. *Earthquake Risk to Palio 4 and Site 400*, Longboat Key, FL, reproduced by permission of T. W. Lambe.).

distances is common. Horizontally, autocorrelation may be isotropic (i.e. $R_z(\delta)$ in the northing direction is the same as $R_x(\delta)$ in the easting direction) or anisotropic, depending on geologic history. However, in practice, isotropy is often assumed. Also, autocorrelation is typically assumed to be the same everywhere within a deposit. This assumption,

called stationarity, to which we will return, is equivalent to assuming that the deposit is statistically homogeneous.

It is important to emphasize, again, that the autocorrelation function is an artifact of the way soil variability is separated between trend and residuals. Since there is nothing innate about the chosen trend, and since changing the trend changes $R_z(\delta)$, the autocorrelation function reflects a modeling decision. The influence of changing trends on $R_z(\delta)$ is illustrated in data analyzed by Javete (1983) (Figure 9.16). Figure 9.17 shows autocorrelations of water content in San Francisco Bay Mud within an interval of 3 ft. Figure 9.18 shows the autocorrelation function when the entire site is considered. The difference comes from the fact that in Figure 9.17 the mean trend is taken locally within the 3 ft interval, and in Figure 9.18 the mean trend is taken globally across the site. The schematic drawing of Figure 9.19 suggests why the autocorrelations should differ.

Autocorrelation can be found in almost all spatial data that are analyzed using a model of the form of Equation 9.1. For example, Figure 9.20 (Baecher *et al.* 1980) shows the autocorrelation of rock fracture density in a copper porphyry deposit; Figure 9.21 (Tang 1979) shows autocorrelation of cone penetration resistance in North Sea Clay; Figure 9.22 (Baecher 1980) shows autocorrelation of water content in the compacted clay core of a rock-fill dam. An interesting aspect of the last data is that the autocorrelations they reflect are more a function of the construction process through

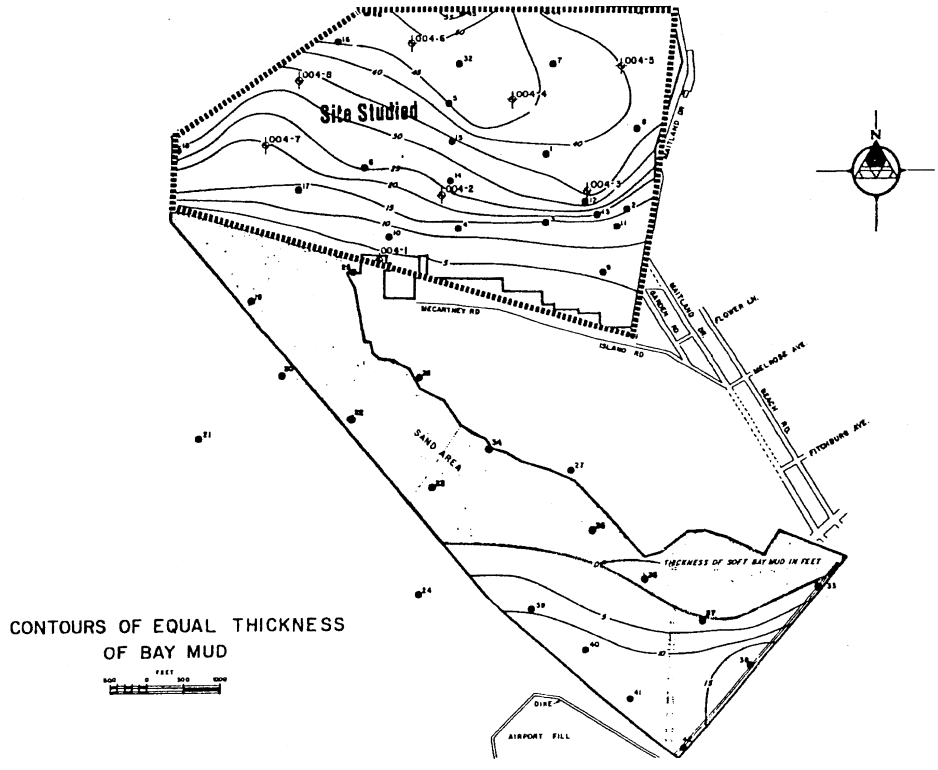


Figure 9.16 Study area for San Francisco Bay Mud measurements analyzed by Javete (1983) (reproduced with the author's permission).

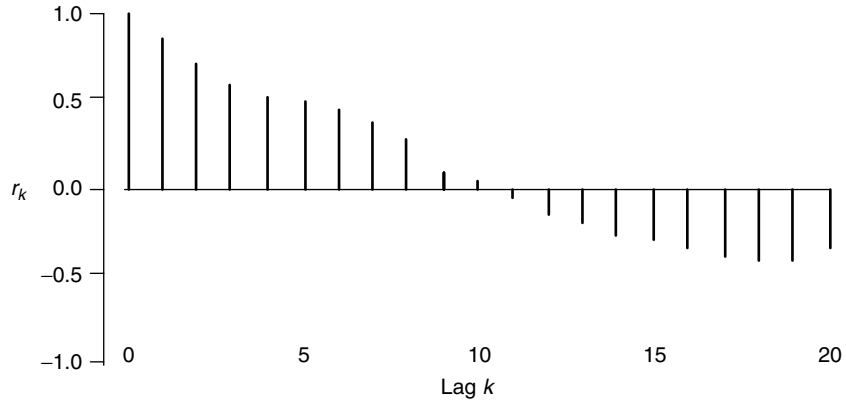


Figure 9.17 Autocorrelation function of water content over small intervals of San Francisco Bay Mud (after Javete 1983). Distance expressed in lag intervals of 0.5 inch (1.2 cm) (reproduced with the author's permission).

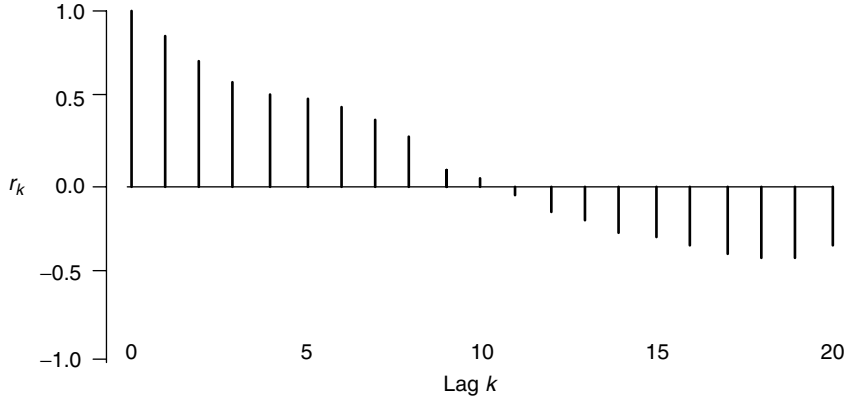


Figure 9.18 Autocorrelation function of water content over small intervals of San Francisco Bay Mud (after Javete 1983). Distance expressed in lag intervals of 25 feet (7.7 m) (reproduced with the author's permission).

which the core of the dam was placed than simply of space, *per se*. The time stream of borrow materials, weather, and working conditions at the time the core was placed led to trends in the resulting physical properties of the compacted material. In a similar way, the time stream of geological conditions – now long lost to our knowledge – that led to particular soil or rock formations, also led to the spatial correlations we see today.

For the purposes of modeling and analysis, it is usually convenient to approximate the autocorrelation structure of the residuals by a smooth function. In Figure 9.8e, the calculated autocorrelations of the residuals were approximated by the exponential function

$$\mathrm{R_z}(\delta) = \exp(-\delta/\delta_0) \tag{9.17}$$

in which δ_0 is a constant having units of length. Other functions commonly used to represent autocorrelation are shown in Table 9.1.

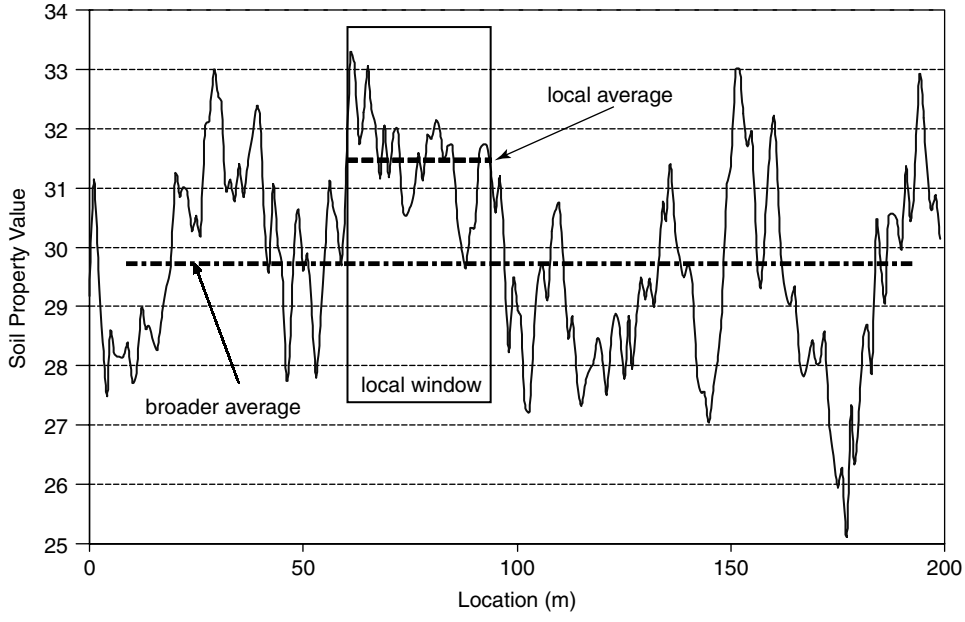


Figure 9.19 Effect of observation scale on autocorrelation function.

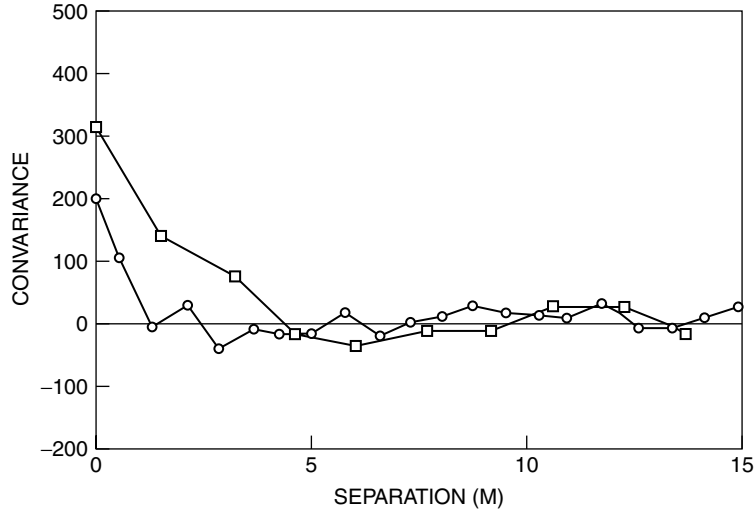


Figure 9.20 Autocorrelation function of rock joint density in a copper porphyry (Baecher *et al.* 1980).

9.3 Estimating Autocorrelation and Autocovariance

In the foregoing discussion, we have described the behavior of spatial variations in soil and rock masses, and shown that this variation might be decomposed into large-scale trends and small-scale residual fluctuations. The remaining spatial structure of the residual

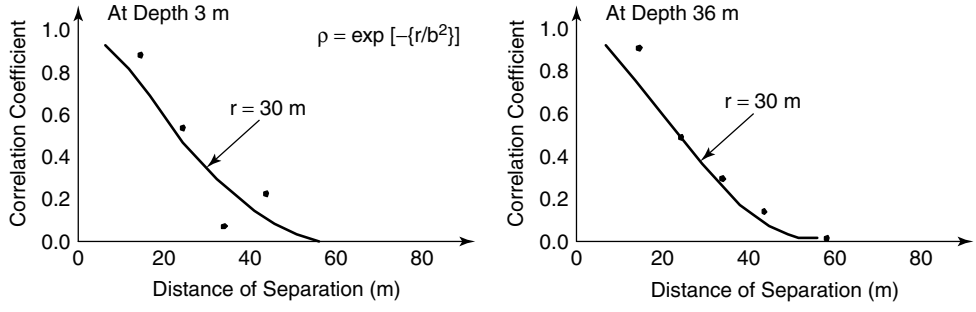


Figure 9.21 Autocorrelation functions for penetration measurements of bottom soils in the North Sea. (Tang, W. H., 1979. 'Probabilistic Evaluation of Penetration Resistance.' *Journal Geotechnical Engineering Division, ASCE* **105**(10), pp. 1173–1191, reproduced by permission of the American Society of Civil Engineers).

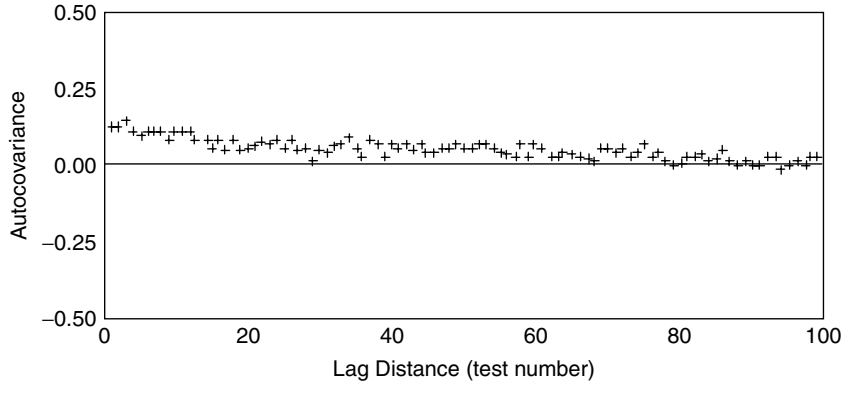


Figure 9.22 Autocorrelation estimates for soil compaction data in clay core of Carters Dam, Georgia, showing weak but significant correlation structure with sequence of soil placement in addition to spatial location within core (Baecher 1980).

Table 9.1 One-dimensional autocorrelation models

Model	Equation	Limits of validity
White noise	$R_x(\delta) = \begin{cases} 1 \; if \; \delta = 0 \\ 0 \; otherwise \end{cases}$	R^{n}
Linear	$R_x(\delta) = \begin{cases} 1 - \lvert\delta\rvert/\delta_0 \; if \; \delta \le \delta_0 \\ 0 \; otherwise \end{cases}$	R^{l}
Exponential	$R_x(\delta) = \exp(-\delta/\delta_0)$	R^{l}
Squared exponential (Gaussian)	$R_x(\delta) = \exp^2(-\delta/\delta_0)$	R^{d}
Power	$\mathrm{C}_z(\delta) = \sigma^2\{1 + (\lvert\delta\rvert^2/\delta_0^2)^{-\beta}$	$\mathrm{R}^{\mathrm{d}}, \beta > 0$

variations manifests in a waviness, or lack of independence, from one observation to the next in space. This lack of independence can be summarized by plotting the observed covariance or correlation of the residuals as a function of the latter's separation distance. The resulting graph is called an autocovariance or autocorrelation function, respectively. Thus far, we have spoken loosely of how one goes about estimating autocovariance from field measurements. In this section, attention turns to statistical issues of inferring autocovariance.

9.3.1 Conceptual methodology

Estimating autocovariance from sample data is the same as making any other statistical estimate. The sample data differ from one set of observations to another, and thus the estimates one makes of autocovariance vary from one sample to another. The important questions are, how much do these estimates vary, and, as a result, how much might one be in error in drawing inferences about autocovariances? As in Chapter 4, there are a number of ways one might approach sampling and inference for autocovariance. In this section, we consider common frequentist approaches: moment estimates, and maximum likelihood estimates. Bayesian inference for autocorrelation has not been widely used in geotechnical and geoscience applications, and is not as well developed. For further discussion of Bayesian approaches to estimating autocorrelation, see Zellner (1971) and Cressie (1991).

Moment estimates and maximum likelihood estimates both arise out of relative frequentist concepts. In each case, a mathematical function of the sample observations is used as an estimate of the true population parameters, θ, one wishes to determine

$$\hat{\theta} = \mathrm{g}\{\mathrm{z}_1, \ldots, \mathrm{z}_\mathrm{n}\} \tag{9.18}$$

in which $\{z_1, \ldots, z_n\}$ is the set of sample observations, and $\hat{\theta}$ can be a scalar, vector, or matrix. For example, the sample mean might be used as an estimator of the true population mean. Such a function is called an *estimator* or a *statistic* of the data. The realized value of $\hat{\theta}$ for a particular sample $\{z_1, \ldots, z_n\}$ is called an *estimate*.

The observations $\{z_1, \ldots, z_n\}$ are presumed to arise as observations of random variables from a joint probability distribution $f_z(\mathbf{z})$, in which $\mathbf{z}$ is a vector. Usually, the distributional form of $f_z(\mathbf{z})$ is fully specified, although in some cases, as for example when using moment estimators, only the existence of low order moments needs to be assumed. Over replicate samples, the observations $\{z_1, \ldots, z_n\}$ are random variables, and thus $\hat{\theta} = \mathrm{g}\{\mathrm{z}_1, \ldots, \mathrm{z}_\mathrm{n}\}$ is a random variable.

As the probabilistic properties of the $\{z_1, \ldots, z_n\}$ are known from the assumed model of spatial variation, the corresponding probabilistic properties of $\hat{\theta}$ can be calculated as a function of the true population parameters, $\boldsymbol{\theta}$. The probability distribution of $\hat{\theta}$ conditioned on $\boldsymbol{\theta}$, $\mathrm{f}_{\hat{\theta}}(\hat{\theta}|\theta)$, is called the *sampling distribution* of $\hat{\theta}$. The standard deviation of the sampling distribution is sometimes called, the *standard error*.

The quality of the estimate obtained in this way depends on how variable the estimator $\hat{\theta}$ is about the true value $\boldsymbol{\theta}$; that is, the sampling distribution is central to knowing how good an estimate might be. Note that this entire discussion is predicated on a frequency concept of probability. The sampling distribution, and hence the goodness of an estimate, has to do with how the estimate might have come out if another sample and therefore

another set of observations other than that actually in hand had been made. Inferences made in this way do not admit of a probability distribution directly on the true population parameter, $\boldsymbol{\theta}$ (what Bayesians would call the *state of nature*). Therefore, there is a gap between the result a traditional estimator provides and the information needed by the reliability analyst in forecasting the uncertain engineering performance of a soil or rock mass.

Various measures of estimator quality were introduced in Chapter 4. It is typically the case that an estimator that is optimal in one sense is not optimal in all other senses. For example, the sample variance for a scalar random variable is a minimum squared error estimator, but it is not unbiased.

As introduced in Chapter 4, Bayesian estimation differs from frequentist estimation in that degree-of-belief theory allows probabilities to be assigned directly to states of nature (e.g. population parameters, $\boldsymbol{\theta}$). Thus, Bayesian methods start with an *a priori* probability distribution on $\boldsymbol{\theta}$, $f(\boldsymbol{\theta})$, which is updated by the likelihood of observing the sample $\{z_1, \ldots, z_n\}$ through Bayes' Theorem,

$$f'(\theta|z_1, \ldots, z_n) \propto f^0(\theta)\mathrm{L}(\theta|z_1, \ldots, z_\mathrm{n}) \tag{9.19}$$

in which $f(\boldsymbol{\theta}|z_1, \ldots, z_n)$ is the updated or *a posteriori* pdf of $\boldsymbol{\theta}$ conditioned on the observations, and $\mathrm{L}(\theta|z_1, \ldots, z_n) = f_{z_1,\ldots,z_n}(z_1, \ldots, z_n|\theta)$ is the likelihood of $\boldsymbol{\theta}$, which equals the conditional probability of $\{z_1, \ldots, z_n\}$ as a function of $\boldsymbol{\theta}$. Thus, the Bayesian method gives a direct pdf on $\boldsymbol{\theta}$, but at the expense of needing to specify some *a priori* distribution, which may be taken as uniform, but which may be hard to estimate.

9.3.2 Moment estimation

The most common method of estimating autocorrelation or autocovariance functions for soil and rock properties is the method of moments (Chapter 4). Moment estimators use the statistical moments of the set of data (e.g. sample means, variances, and covariances) as estimators of the corresponding moments of the population being sampled. Moment estimators of autocovariance are conceptually and operationally simple, but have poor sampling properties for small sample sizes.

Consider the measurements $\{z_1, \ldots, z_n\}$ made at equally spaced locations $\mathbf{x} = \{x_1, \ldots, x_n\}$ along a line, as for example in a boring. The sample autocovariance of the measurements for separation δ is

$$\hat{C}_z(\delta) = \frac{1}{(n-\delta)} \sum_{i=1}^{n-\delta} [\{z(x_i) - t(x_i)\}\{z(x_{i+\delta}) - t(x_{i+\delta})\}] \tag{9.20}$$

in which $\hat{\mathrm{C}}_\mathrm{z}(\delta)$ is the estimator of the autocovariance function at δ, $(n - \delta)$ is the number of data pairs having separation distance δ, and $t(x_i)$ is the trend removed from the data at location x_i. Often, $t(x_i)$ is simply replaced by the spatial mean μ_z, estimated by the mean of the sample, $m_z = (1/n)\sum z(x_i)$. At $\delta = 0$, $\hat{\mathrm{C}}_\mathrm{z}(\delta)$ reduces to the sample variance. The corresponding moment estimator of the autocorrelation, $\hat{\mathrm{R}}_\mathrm{z}(\delta)$ is obtained by dividing

both sides of Equation (9.20) by the sample variance, to obtain

$$\hat{R}_z(\delta) = \frac{1}{(n-\delta)s_z^2} \sum_{i=1}^{n-\delta} (z(x_i) - t(x_i))(z(x_{i+\delta}) - t(x_{i+\delta})) \tag{9.21}$$

in which s_z is the sample standard deviation. Computationally, this simply reduces to taking all data pairs of common separation distance δ, calculating the correlation coefficient of that set, then plotting the result against δ. The results of Figure 9.23 were obtained from the James Bay data of Christian *et al.* (1994) using this moment estimator. When the samples are non-uniform in space, data pairs are usually grouped within intervals and the same procedure used.

In the general case, measurements are seldom uniformly spaced, at least in the horizontal plane and seldom lie on a line. For such situations the sample autocovariance can still be used as an estimator, but with some modification. The most common way to accommodate non-uniformly placed measurements is by dividing separation distances into bands, and then taking the averages of Equation (9.24) within those bands. This introduces some error into the estimate, but for engineering purposes it is usually of sufficient accuracy.

The James Bay Project is a large hydroelectric development in the northern region of Québec, Canada. The data used in this case study are from an investigation into the stability of dykes on a soft marine clay (Ladd *et al.* 1983). The site is interesting, among other reasons, in that the statistical properties of the clay strength have been analyzed by several authors. The marine clay at the site is approximately 8 m thick and overlies a lacustrine clay. The depth-averaged results of field vane tests conducted in 35 borings as part of the soil exploration program were selected for the correlation analysis. Nine

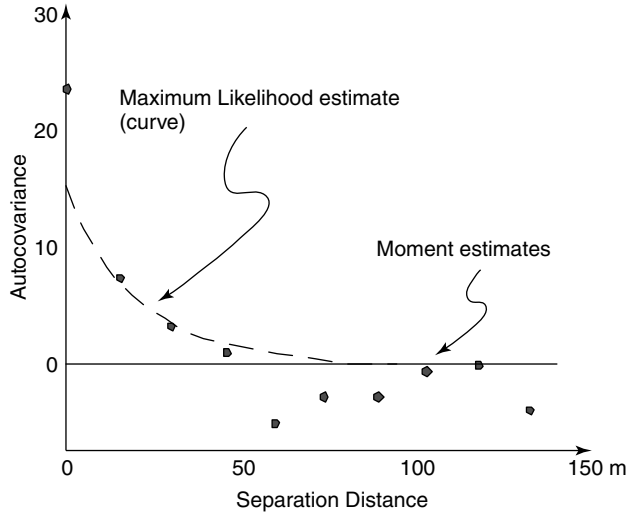


Figure 9.23 Autocovariance function for field vane strength data from marine clay at James Bay. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating autocovariance of *in situ* Soil properties. *Journal of the Geotechnical Engineering Division, ASCE* **119**(GT1): 147–166, reproduced with permission of the American Society of Civil Engineers).

of the borings were concentrated in one location. Information on the correlation structure of the soil at this site was used by Ladd *et al.* (1983) for the purpose of conducting a reliability analysis of dyke stability, which is elaborated further in Chapter 15.

To estimate the autocovariance function of the cone penetration strengths, first a constant mean was removed from the data. Then, the product of each pair of residuals was calculated, and plotted against separation distance. A moving average of these products was used to obtain the estimated points. Note the precipitous drop in covariance in the neighborhood of the origin, and also the negative sample moments in the vicinity of 50–100 m separation. Note, also, the large scatter in the sample moments at large separation distance. From these estimates. a simple exponential curve was fit by inspection, and this intersects the ordinate at about 60% of the sample variance. This yields an autocovariance function of the form

$$C_z(\delta) = \begin{cases} 22\,\text{kPa}^2, & \text{for } \delta = 0 \\ 13\exp\{-\delta/23\,\text{m}\}, & \text{for } \delta > 0 \end{cases} \tag{9.22}$$

in which variance is in kPa^2 and distance is in meters.

The moment estimator of the autocovariance function is non-parametric, in that no assumptions are needed about the mathematical shape of the autocovariance function, except that the second moments exist. This is a highly desirable property, because there is little theoretical justification in most geotechnical applications for *a priori* choosing one functional shape over another. Usually, if a functional form for the autocovariance is needed, a mathematical equation is fitted to the moment estimates by inspection.

The moment estimator is consistent, in that as the sample size becomes large, $E[(\hat{\theta} - \theta)^2] \to 0$, the sample autocovariance function becomes ever closer to the true autocovariance function in mean square error (m.s.e.) On the other hand, the moment estimator is only asymptotically unbiased. Unbiasedness means that the expected value of the estimator over all ways the sample might have been taken equals the actual value of the function being estimated. This is true only as $n \to \infty$. For finite sample sizes, the expected values of the sample autocovariances can differ significantly from the actual values, that is, the moment estimators given by Equations (7.24) and (7.25) are biased at small n. Figure 9.24, from Weinstock (1963), shows how much disparity can exist between the actual autocovariance and the expected value of the autocovariances observed in a sample. Compare Weinstock's results to Figure 9.23. The normalized length of record in those data is between 5 and 6 (i.e. about 400 m/70 m). The negative correlation observed at $\delta = 200$ m is reasonably consistent with this estimator bias, and illustrates one of the difficulties with moment estimators of autocovariance.

9.3.3 Maximum likelihood estimation

Maximum likelihood estimation, due originally to Fisher (Box 1978), sets the estimator to be that value of the parameter(s) $\boldsymbol{\theta}$ that leads to the greatest probability of observing those data, $\{z_1, \ldots, z_n\}$, that actually were observed. This is expressed through the likelihood function

$$L(\theta|z_1, \ldots, z_n) = f_{z_1,\ldots,z_n}(z_1, \ldots, z_n|\theta) \tag{9.23}$$

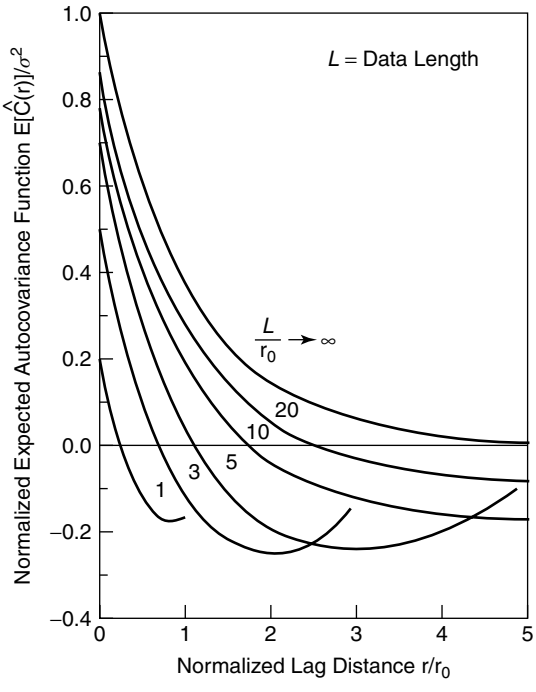


Figure 9.24 Moment estimator bias for autocorrelation as a function of the length of the window of observation. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating autocovariance of *in situ* soil properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119**(GT1): 147–166, reproduced with permission of the American Society of Civil Engineers).

which is simply the joint conditional probability density of $\{z_1, \ldots, z_n\}$, given $\boldsymbol{\theta}$. Maximum likelihood estimation is said to be a parametric method, because the distributional form $f_{z_1,\ldots,z_n}(z_1, \ldots, z_n|\theta)$ is assumed known.

In the particular case of independent $z_1, \ldots, z_n$, the likelihood becomes

$$L(\theta|z_1, \ldots, z_n) = \prod_{i=1}^{n} f_{z_i}(z_i|\theta) \tag{9.24}$$

(Obviously, this is not the case in estimating autocovariances, but we return to that below.) The estimator is then set so as to maximize the likelihood of the data, $\{z_1, \ldots, z_n\}$, over $\boldsymbol{\theta}$:

$$\hat{\theta}_{\mathrm{ML}} = \theta \text{ s.t. } \max_{\theta} L(\theta|z_1, \ldots, z_n) \tag{9.25}$$

which are obtained from

$$\frac{\partial L(\theta|z_1, \ldots, z_n)}{\partial \theta} = 0 \text{ such that } \frac{\partial^2 L(\theta|z_1, \ldots, z_n)}{\partial \theta^2} < 0 \tag{9.26}$$

These are the values of $\boldsymbol{\theta}$ which, were they to obtain, would lead to the greatest probability (density) of observing the actually observed data.

In practice, the estimate is usually found by maximizing the log-likelihood:

$$\mathrm{LL}(\theta|z_1, \ldots, z_n) = \sum_{i=1}^{n} \log f_{z_i}(z_i|\theta) dz_i \tag{9.27}$$

which, because it deals with a sum rather than a product, and because many common probability distributions involve exponential terms, is more convenient. As with any functional optimization, special cases are sometimes encountered involving multiple local maxima of $LL(\boldsymbol{\theta}|z_1, \ldots, z_n)$, discontinuities in LL or its first derivative, and so on, but these are exceptions rather than the rule.

Although linguistically, the term *likelihood* seems to imply a probability statement about the state of nature, $\boldsymbol{\theta}$, in fact the maximum likelihood is another frequentist estimator, and the probability statement is linked firmly to the observations, $\{z_1, \ldots, z_n\}$, not the parameters, $\boldsymbol{\theta}$. The appeal of the maximum likelihood estimator is that it possesses many desirable sampling properties in frequentist theory. Among others, it has minimum variance (even though not necessarily unbiased), is consistent (which implies asymptotic unbiasedness, if not unbiasedness for small n), and is asymptotically Normal regardless of the pdf of $\mathbf{z}$. Further, the maximum likelihood estimator of a function of $\boldsymbol{\theta}$, $g(\boldsymbol{\theta})$, is the function of the maximum likelihood estimator $\boldsymbol{\theta}_{\mathrm{ML}}$, or $g(\boldsymbol{\theta}_{\mathrm{ML}})$. The asymptotic variance of $\hat{\theta}_{\mathrm{ML}}$ is

$$\lim_{\mathrm{n}\to\infty} \mathrm{Var}[\hat{\theta}_{\mathrm{ML}}] = I_{\mathrm{z}}(\theta) = \mathrm{nE}[-\partial^2 \mathrm{LL}/\partial\theta^2] \tag{9.28}$$

in which $I_z(\boldsymbol{\theta})$ is Fisher's Information (Barnett 1982). This means that, asymptotically,

$$\hat{\theta}_{\mathrm{ML}} \to \mathrm{N}(\theta, \mathrm{nE}[-\partial^2 \mathrm{LL}/\partial\theta^2]) \tag{9.29}$$

Obviously, because the maximum likelihood estimator maximizes probability, it requires a parametric description of the pdf of $\boldsymbol{z}$; that is, a distributional form must be assumed, and the autocovariance function must be specified analytically.

Figure 9.25 shows the results of simulated sampling experiments in which spatial fields were generated from a multivariate Gaussian pdf with specified mean trend and autocovariance function. Samples of sizes $n = 36$, 64, and 100 were taken from these simulated fields, and maximum likelihood estimators used to obtain estimates of the parameters of the mean trend and autocovariance function. The smooth curves show the respective asymptotic sampling distributions, which in this case conform well with the actual estimates (DeGroot and Baecher 1993).

A significant advantage of the maximum likelihood estimator in dealing with spatial data is that it allows simultaneous estimation of the spatial trend and autocovariance function of the residuals. Let the trend be represented by a regression surface $T(\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$, in which $\mathbf{x}$ is a matrix of location coordinates, and $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_n\}$ is a vector of regression coefficients. Let the parameters of the autocovariance function be $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_p\}$. Assume that the random field of soil properties is isotropic Gaussian, represented by the model

$$\mathbf{z}(\mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{9.30}$$

in which $\boldsymbol{\varepsilon}$ is a vector of correlated residuals defined by a parametric autocovariance function, $C_\varepsilon(\boldsymbol{\delta})$. The autocovariance function is assumed to be twice differentiable with respect to $\boldsymbol{\theta}$, and the autocovariance matrix $\boldsymbol{\Sigma} = Cov[z_i, z_j]$, is assumed positive definite.

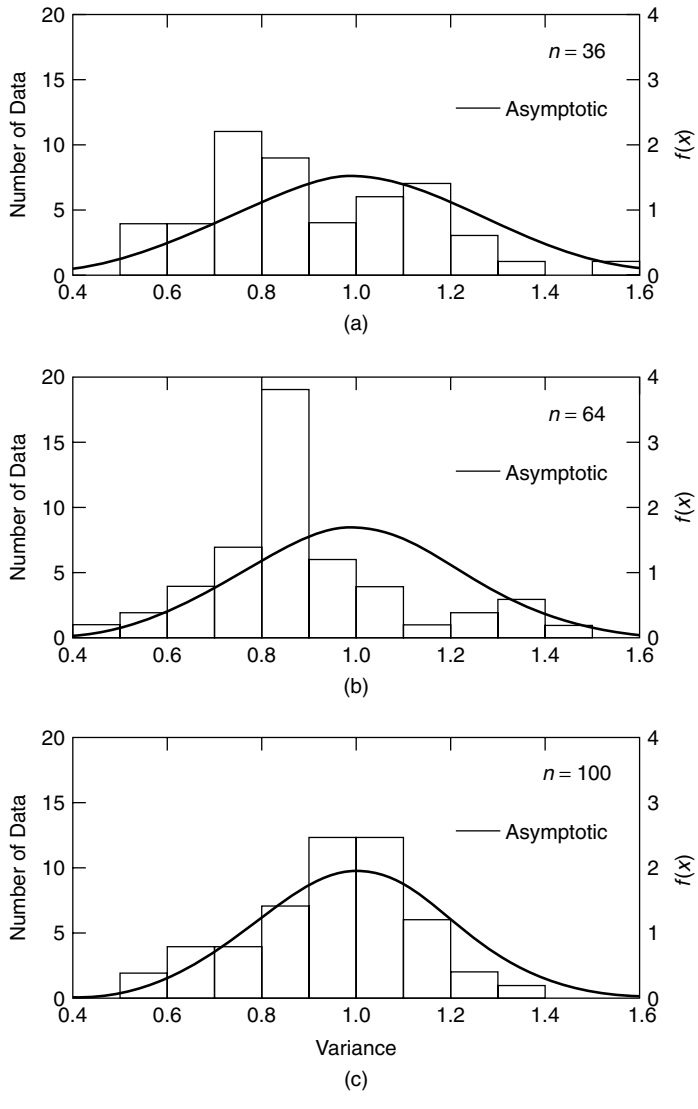


Figure 9.25 Histograms of maximum likelihood variance estimates compared to theoretical asymptotic distribution. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating Autocovariance of in Situ Soil Properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119**(GT1) pp. 147–166, reproduce with permission of the American Society of Civil Engineers.)

Given a set of measurements $\mathbf{z} = \{z(\boldsymbol{x}_1), z(\boldsymbol{x}_2), \ldots, z(\boldsymbol{x}_n)\}$, the vector of unknown parameters is defined as $\boldsymbol{\phi} = \{\boldsymbol{\beta}, \boldsymbol{\theta}\}$, and the log likelihood for $\boldsymbol{\phi}$ is

$$\mathrm{LL}(\boldsymbol{\phi}|\mathbf{Z}) = -\frac{\mathrm{n}}{2}\ell\mathrm{n}\Sigma - \frac{1}{2}(\mathbf{Z} - \boldsymbol{\beta}\mathbf{X})^t\mathbf{C}^{-1}(\mathbf{Z} - \boldsymbol{\beta}\mathbf{X}) \tag{9.31}$$

in which $\mathbf{X}$ is an $n \times q$ design matrix of locations (i.e. the matrix of coordinates of the sample locations), and $\mathbf{C}$ is the covariance matrix, with terms $Cov(z_i, z_j)$. The maximum

likelihood estimators of $\boldsymbol{\beta}$ and $\mathbf{C}_z(\boldsymbol{\delta})$ are found by maximizing $L(\boldsymbol{\phi}|\mathbf{Z})$ with respect to $\boldsymbol{\phi}$. An algorithmic procedure due to Mardia and Marshall (1984) for finding this maximum is given as an appendix to this chapter.

DeGroot and Baecher (1993) analyzed the James Bay data in two ways using maximum likelihood estimation. First, they removed a constant mean from the data, and applied Equations 9.26 and 9.31 to the residuals to obtain

$$\mathrm{C}_z(\delta) = \begin{cases} 23 & \text{for } \delta = 0 \\ 13.3\exp\{-\delta/21.4\}, & \text{for } \delta > 0 \end{cases} \tag{9.32}$$

in which variance is in kPa^2 and distance is in meters. Then, using Equation 9.31 and estimating the trend implicitly, they obtained

$$\begin{aligned} \hat{\beta}_0 &= 40.7\,\text{kPa} \\ \hat{\beta}_1 &= -2.0 \times 10^{-3}\,\text{kPa/m} \\ \hat{\beta}_2 &= -5.9 \times 10^{-3}\,\text{kPa/m} \\ \mathrm{C}_z(\delta) &= \begin{cases} 23\,\text{kPa}^2 & \text{for } \delta = 0 \\ 13.3\,\text{kPa}^2\exp\{-\delta/21.4\,\text{m}\}, & \text{for } \delta > 0 \end{cases} \end{aligned} \tag{9.33}$$

This curve is superimposed on the moment estimators of Figure 9.23.

The small values of $\hat{\beta}_1$ and $\hat{\beta}_2$ suggest that the assumption of constant mean is reasonable. Substituting a squared-exponential model for the autocovariance, results in

$$\begin{aligned} \hat{\beta}_0 &= 40.8\,\text{kPa} \\ \hat{\beta}_1 &= -2.1 \times 10^{-3}\,\text{kPa/m} \\ \hat{\beta}_2 &= -6.1 \times 10^{-3}\,\text{kPa/m} \\ \mathrm{C}_z(\delta) &= \begin{cases} 22.9\,\text{kPa}^2 & \text{for } \delta = 0 \\ 12.7\,\text{kPa}^2\exp\{-(\delta/37.3\,\text{m})^2\}, & \text{for } \delta > 0 \end{cases} \end{aligned} \tag{9.34}$$

The data presented in this case suggest that a sound approach to estimating autocovariance should involve both the method of moments and maximum likelihood. The method of moments gives a plot of autocovariance versus separation, providing an important graphical summary of the data, which can be used as a means of determining if the data suggest correlation and for selecting an autocovariance model (for example, selecting the exponential model over the squared exponential in this case). This provides valuable information for the maximum likelihood method, which then can be used to obtain estimates of both autocovariance parameters and trend coefficients.

9.3.4 Bayesian estimation

Bayesian inference for autocorrelation has not been widely used in geotechnical and geostatistical applications, and is less well developed than moment estimates. A surprising aspect of Bayesian inference of spatial trends and autocovariance functions is that for many of the non-informative prior distributions one might choose to reflect little or no prior

information about process parameters (e.g. the Jeffreys prior, the Laplace prior, truncated parameter spaces), the posterior pdf's calculated through Bayes' theorem are themselves improper, usually in the sense that they do not converge toward zero at infinity, and thus the total probability or area under the posterior pdf is infinite.

Following Berger (Berger 1993; Berger *et al.* 2001) and Kitanidis (1985; 1997), the spatial model is typically written as a multi-Normal (Gaussian) random field:

$$z(\mathbf{x}) = \sum_{i=1}^{k} f_i(\mathbf{x})\beta + \varepsilon(\mathbf{x}) \tag{9.35}$$

in which $f_i(\boldsymbol{x})$ are unknown deterministic functions of the spatial locations $\boldsymbol{x}$, and $\varepsilon(\boldsymbol{x})$ is a zero-mean spatial random function. This is a slight change from Equation (9.30) in that the functional terms $f_i(\boldsymbol{x})$ make the model more general.

The random term is the same spatially correlated component as in Equation (9.30), with isotropic autocovariance function $C_{z|\theta}(\delta) = Cov(z_i, z_j) = \sigma_z^2 R_{z|\theta}(\delta)$, in which as before, δ is vector separation distance, and σ_z^2 is the variance. The autocovariance depends on a vector of unknown parameters, $\boldsymbol{\theta}$, which describe the range and shape of the function. The autocovariance function is assumed to be non-negative and to decrease monotonically with distance from 1.0 at zero separation, to 0 at infinite separation. These assumptions fit most common autocovariance functions in geotechnical applications, including the exponential, squared-exponential, spherical, quadratic, and Matérn.

The Likelihood of a set of observations, $\mathbf{z} = \{z_1, \ldots, z_n\}$, is then

$$L(\boldsymbol{\beta}, \sigma|\mathbf{z}) = (2\pi\sigma^2)^{-n/2}|R_\theta|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^t R_\theta{}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\right\} \tag{9.36}$$

in which $\mathbf{X}$ is the $(n \times k)$ matrix defined by $X_{ij} = f_j(X_i)$, R_θ is the matrix of correlations among the observations dependent on the parameters $\boldsymbol{\theta}$, and $|R_\theta|$ indicates the trace of the correlation matrix of the observations.

In usual fashion, a prior non-informative distribution on the parameters $(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta})$ might be represented as $f(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta}) \propto (\sigma^2)^{-a} f(\boldsymbol{\theta})$, for various choices of the parameter a and of the marginal pdf $f(\boldsymbol{\theta})$. The obvious choices might be $\{a = 1, f(\boldsymbol{\theta}) = 1\}$, $\{a = 1, f(\boldsymbol{\theta}) = 1/\boldsymbol{\theta}\}$, or $\{a = 0, f(\boldsymbol{\theta}) = 1\}$; but each of these leads to an improper posterior pdf, as does the well-known Jeffreys prior, which is proportional to the root of the determinant of Fisher's Information matrix, as discussed in Chapter 4. A proper, informative prior does not share this difficulty, but it is correspondingly hard to assess from usually subjective opinion. Given this problem, Berger *et al.* (2001) suggest the reference non-informative prior:

$$f(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta}) \propto \frac{1}{\sigma^2}\left(|W_\theta^2| - \frac{|W_\theta|^2}{(n-k)}\right)^{1/2} \tag{9.37}$$

in which

$$\mathbf{W}_\theta^2 = \frac{\partial \mathbf{R}_\theta}{\partial \boldsymbol{\theta}}\mathbf{R}_\theta^{-1}\{\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{R}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{R}_\theta^{-1}\} \tag{9.38}$$

This does lead to a proper posterior. The posterior pdf is usually evaluated numerically, although, depending on the choice of autocovariance function model and on the extent to

which certain of the parameters $\boldsymbol{\theta}$ of that model are known, closed form solutions can be obtained (Kitanidis 1985). Berger *et al.* (2001) present a numerically calculated example using terrain data from Davis (1986).

9.4 Variograms and Geostatistics

In mining, the importance of autocorrelation for estimating ore reserves has been recognized for many years. In mining *geostatistics* a function related to the autocovariance, called the *variogram* (Matheron 1971), is more commonly used to express the spatial structure of data. The variogram requires a less restrictive statistical assumption on stationarity than the autocovariance function does, and it is therefore sometimes preferred for inference problems. On the other hand, the variogram is more difficult to use in spatial interpolation and engineering analysis, and thus for geotechnical purposes autocovariance is more commonly used. In practice, the two ways of characterizing spatial structure are closely related.

Whereas the autocovariance is defined as the expected value of the product of two observations (Equation (9.21)), the variogram 2γ is usually defined as the expected value of the squared difference

$$2\gamma = \mathrm{E}[\{z(x_i) - z(x_j)\}^2] = \mathrm{Var}[z(x_i) - z(x_j)] \tag{9.39}$$

which is a function of only the increments of the spatial properties, not their absolute values. Cressie (1991) points out that, in fact, the common definition of the variogram as the mean squared difference – rather than as the variance of the difference – limits applicability to a more restrictive class of processes than necessary, and thus the latter definition is to be preferred. Nonetheless, one finds the former definition more commonly referred to in the literature. The term γ is usually referred to as the *semivariogram*, although caution must be exercised because different authors interchange the terms. The concept of average mean-square difference has been used in many applications, including turbulence (Kolmogorov 1941) and time series analysis (Jowett 1952), and is alluded to in the often cited work of Matérn (1960).

As noted, the advantage of the variogram over autocovariance is that it makes less restrictive assumptions on the stationarity of the spatial properties being sampled, specifically, only that their increment and not their mean is stationary. Furthermore, the use of geostatistical techniques has expanded broadly, so that a great deal of experience has been accumulated with variogram analysis, not only in mining applications, but also in environmental monitoring, hydrology, and even geotechnical engineering (Chiasson *et al.* 1995; Soulie and Favre 1983; Soulie *et al.* 1990).

For spatial variables with stationary means and autocovariances (i.e. second-order stationary processes, as discussed in Chapter 8), the variogram and autocovariance function are directly related. Expanding the definition of the variogram

$$2\gamma = \mathrm{Var}[z(x_i) - z(x_j)] = \mathrm{Var}[z(x_i)] - \mathrm{Var}[z(x_j)] - 2\mathrm{Cov}[z(x_i), z(x_j)] \tag{9.40}$$

thus

$$\begin{aligned} 2\gamma(\delta) &= 2\{\mathrm{C_z}(0) - \mathrm{C_z}(\delta)\} \\ \gamma(\delta) &= \mathrm{C_z}(0) - \mathrm{C_z}(\delta) \end{aligned} \tag{9.41}$$

So, the variogram and autocovariance function, if the latter exists, are directly related. For the data of Figure 9.23, with their respective autocovariance function, the corresponding variogram is shown in Figure 9.26. Common analytical forms for one-dimensional variograms are given in Table 9.2.

For a stationary process, as $|\delta| \to \infty$, $C_z(\boldsymbol{\delta}) \to 0$; thus, $\gamma(\boldsymbol{\delta}) \to C_z(0) = Var[z(\mathbf{x})]$. This value at which the variogram levels off, $2C_z(\boldsymbol{\delta})$, is called the *sill* value. The distance at which the variogram approaches the sill is called, the *range*. More mathematical properties of the variogram and of incrementally stationary process are considered in Chapter 8. The sampling properties of $2\gamma(\boldsymbol{\delta})$ are summarized in some detail by Cressie (1991). Unfortunately, the geostatistical literature itself is reasonably quiet on the topic of statistical sampling properties.

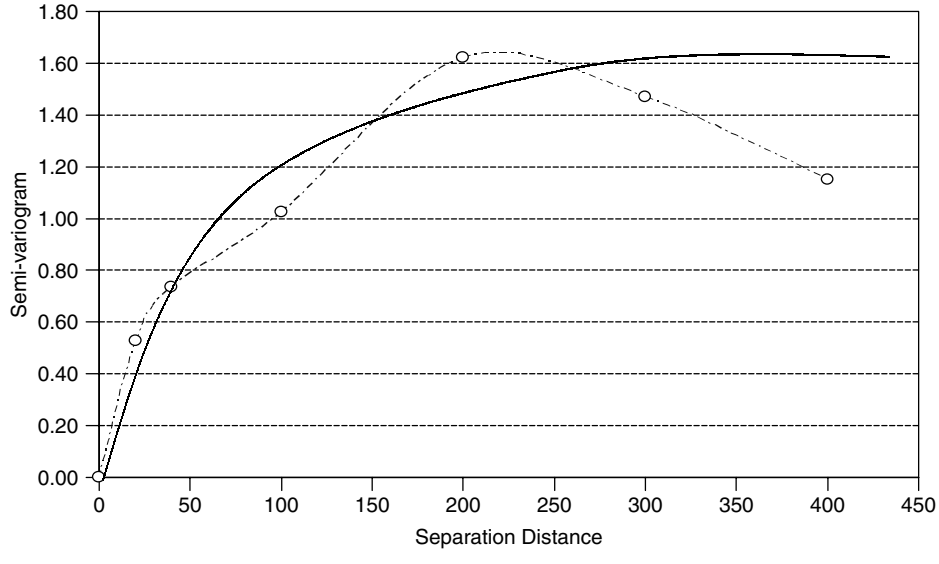


Figure 9.26 Semi-variogram for field vane data of Figure 9.23.

Table 9.2 One-dimensional variogram models

Model	Equation	Limits of validity
Nugget	$\gamma(\delta) = \begin{cases} 0 \text{ if } \delta = 0 \\ 1 \text{ otherwise} \end{cases}$	R^n
Linear	$\gamma(\delta) = \begin{cases} 0 \text{ if } \delta = 0 \\ c_0 + b\|\delta\| \text{ otherwise} \end{cases}$	R^1
Spherical	$\gamma(\delta) = \begin{cases} (1.5)(\delta/a) - (1/2)(\delta/a)^3 \text{ if} \\ 1 \text{ otherwise} \end{cases}$	R^n
Exponential	$\gamma(\delta) = 1 - \exp(-3\delta/a)$	R^1
Gaussian	$\gamma(\delta) = 1 - \exp(-3\delta^2/a^2)$	R^n
Power	$\gamma(\delta) = h^\omega$	R^n, $0 < \omega < 2$

Appendix: Algorithm for Maximizing Log-likelihood of Autocovariance

To maximize $L(\boldsymbol{\phi}|\mathbf{z})$, we set the first derivative of the log likelihood $LL(\boldsymbol{\phi}|\mathbf{z})$ with respect to $\boldsymbol{\phi}$ equal to zero. Then, we use a scoring procedure developed by Mardia and Marshall (1984). This method uses an updating expression of the form

$$\boldsymbol{\phi}_{i+1} = \hat{\boldsymbol{\phi}}_i + [\mathbf{B}^{-1}\mathbf{L}'] \tag{9.42}$$

where $\mathbf{B}^{-1}$ is the inverse of Fisher's information matrix, and $\mathbf{L}'$ is the first derivative vector of the log-likelihood function.

Starting with an initial estimate vector of the autocovariance parameters, $\hat{\boldsymbol{\theta}}_0$, the regression coefficients are computed as

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x})^{-1}(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{z}) \tag{9.43}$$

and $\hat{\boldsymbol{\theta}}$ is updated using

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_{\mathrm{i}} + [\mathbf{B}_{\theta}{}^{-1}\mathbf{L}_{\theta}{}']_{\mathrm{i}} \tag{9.44}$$

The information matrix, $\mathbf{B}$ is given by

$$\mathbf{B} = -\mathrm{E}[\mathbf{L}''] = \mathrm{diag}(\mathbf{B}_{\beta}, \mathbf{B}_{\theta}) \tag{9.45}$$

where $\mathbf{L}''$ is the second derivative of the log-likelihood function. The components $\mathbf{B}_{\beta}$ and $\mathbf{B}_0$ are computed as

$$\begin{aligned} \mathbf{B}_{\beta} &= (\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}) \\ \mathbf{B}_{\theta} &= \frac{1}{2}\mathrm{t}_{\mathrm{ij}} \end{aligned} \tag{9.46}$$

where $(1/2)t_{ij}$ is the (i, j)th element of $\mathbf{B}_{\theta}$.

$$\mathrm{t}_{\mathrm{ij}} = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_{\mathrm{i}}}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_{\mathrm{j}}}\right) \tag{9.47}$$

The first-derivative vectors of the log-likelihood function

$$\mathbf{L}' = (\mathbf{L}_{\beta}^{\mathrm{T}}, \mathbf{L}_{\theta}^{\mathrm{T}})^{\mathrm{T}} \tag{9.48}$$

are equal to

$$\begin{aligned} \mathbf{L}_{\beta} &= -(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}\boldsymbol{\beta}) + \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{z} \\ \mathbf{L}_{\theta} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\theta}_{\mathrm{k}}}\right) - \frac{1}{2}\mathbf{w}^{\mathbf{T}}\left(-\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\theta}_{\mathrm{k}}}\boldsymbol{\Sigma}^{-1}\right)\mathbf{w} \end{aligned} \tag{9.49}$$

where

$$\mathbf{w} = \mathbf{z} - \mathbf{x}\hat{\boldsymbol{\beta}} \tag{9.50}$$

In some instances, in may be necessary to pre-multiply the updating term $[\mathbf{B}_\theta^{-1}\mathbf{L}_\theta']$ by a scaling factor $\alpha(0 \leq \alpha \leq 1)$ to ensure stability of the algorithm during iterations.

A simple application of the solution algorithm is solving for the parameters required to describe the structure of a soil property measured in one dimension (e.g. blow counts versus depth). The trend model is taken as

$$\mathrm{z(x)} = \beta_0 + \beta_1 \mathrm{x} \tag{9.51}$$

with an autocovariance function of the form, $\mathrm{C}(\delta|\sigma^2, \delta_0)$. Thus, given the measurements $\mathbf{z}(\mathbf{x})$ at locations $x_1, \ldots, x_n$, the maximum likelihood estimates of β_0, β_1, σ^2, and r_0 fully describe the model. This example can be readily extended to incorporate two-dimensional trends and autocovariance models with added noise.

10 Random field theory

Thus far we have considered descriptive summaries of spatial variability in soil and rock formations. We have seen that spatial variability can be described by a mean trend or trend surface, about which individual measurements exhibit residual variation. This might also be thought of as separating the spatial variability into large-scale and small-scale increments, and then analyzing the two separately. We have seen that residual variations themselves usually exhibit some spatial structure, which can be described by an autocorrelation function. In this section, we consider more mathematical approaches to modeling spatial variation, specifically random field theory. Random field theory is important for two reasons: first, it provides powerful statistical results which can be used to draw inferences from field observations and plan spatial sampling strategies; secondly, it provides a vehicle for incorporating spatial variation in engineering and reliability models. Random field theory is part of the larger subject of stochastic processes, of which we will only touch a small part. For more detailed treatment, see Adler (1981), Christakos (1992; 2000), Christakos and Hristopoulos (1998), Parzen (1964), or Vanmarcke (1983).

10.1 Stationary Processes

The application of random field theory to geotechnical issues is based on the assumption that the spatial variable of concern, $z(\mathbf{x})$, is the realization of a random process. When this process is defined over the space $\mathbf{x} \in \mathbf{S}$, the variable $z(\mathbf{x})$ is said to be a *stochastic process*. In this book, when $\mathbf{S}$ has dimension greater than one, and especially when $\mathbf{S}$ is a spatial domain, $z(\mathbf{x})$ is said to be a *random field*. This usage is more or less consistent across civil engineering, although the geostatistics literature uses a vocabulary all of its own, to which we will occasionally refer.

A random field is defined as the joint probability distribution

$$F_{x_1,\ldots,x_n}(z_1,\ldots,z_n) = P\{z(x_1) \le z_1,\ldots,z(x_n) \le z_n\} \tag{10.1}$$

This joint probability distribution describes the simultaneous variation of the variables $\mathbf{z}$ within the space $\mathbf{x} \in \mathbf{S}_x$. Let

$$E[z(\mathbf{x})] = \mu(\mathbf{x}) \tag{10.2}$$

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

for all $\mathbf{x} \in \mathbf{S}$, be the mean or trend of $z(\mathbf{x})$, presumed to exist for all $\mathbf{x}$; and let

$$Var[z(\mathbf{x})] = \sigma^2(\mathbf{x}) \tag{10.3}$$

be the variance, also assumed to exist for all $\mathbf{x}$. The covariances of $z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n)$ are defined as

$$Cov[z(\mathbf{x}_i), z(\mathbf{x}_j)] = E[(z(\mathbf{x}_i) - \mu(\mathbf{x}_i)) \cdot (z(\mathbf{x}_j) - \mu(\mathbf{x}_j))] \tag{10.4}$$

10.1.1 Stationarity

The random field is said to be *second-order stationary* (sometimes referred to as weak or wide-sense stationary) if $E[z(\mathbf{x})] = \mu$ for all $\mathbf{x}$, and $Cov[z(\mathbf{x}_i), z(\mathbf{x}_j)]$ depends only upon vector separation of $\mathbf{x}_i$, and $\mathbf{x}_j$, not on absolute location:

$$Cov[z(\mathbf{x}_i), z(\mathbf{x}_j)] = C_z(\mathbf{x}_i - \mathbf{x}_j) \tag{10.5}$$

in which $C_z(\mathbf{x}_i - \mathbf{x}_j)$ is the autocovariance function. The random field is said to be *stationary* (sometimes referred to as strong or strict stationarity) if the complete probability distribution, $F_{x_1,\ldots,x_n}(z_1, \ldots, z_n)$, is independent of absolute location, depending only on vector separations among $\mathbf{x}_i \ldots \mathbf{x}_n$ (Table 10.1). Obviously, strong stationarity implies second-order stationarity. In the *geotechnical* literature, stationarity is sometimes referred to loosely as *statistical homogeneity*. If the autocovariance function depends only on the absolute separation distance of $\mathbf{x}_i$, and $\mathbf{x}_j$, not vector separation (i.e. direction), the random field is said to be *isotropic*.

10.1.2 Ergodicity

A final, but important and often poorly understood property is *ergodicity*. Ergodicity is a concept originating in the study of time series, in which one observes individual time series of data, and wishes to infer properties of an ensemble of all possible time series. The meaning and practical importance of ergodicity to the inference of unique realizations of spatial random fields is less well defined and is much debated in the literature.

Simply, but inelegantly, stated, ergodicity means that the probabilistic properties of a random process (field) can be completely estimated from observing one realization of that

Table 10.1 Summary properties of random fields

Property	Meaning
Homogeneous, stationary	Joint probability distribution functions are invariant to translation; joint PDF depends on relative, not absolute, locations.
isotropic	Joint probability distribution functions are invariant to rotations
ergodic	All information on joint pdf's can be obtained from a single realization of the random field.
second-order stationary	$E[z(x)] = \mu$, for all $x \in S$ $Cov[z(x_1), z(x_2)] = C_z(x_1, x_2)$, for all $x_1, x_2 \in S$

process. For example, the stochastic time series $z(\xi) = v + \varepsilon(\xi)$, in which v is a discrete random variable and $\varepsilon(\xi)$ is an autocorrelated random process of time ξ, is non-ergotic. In one realization of the process there is but one value of v, and thus the probability distribution of v cannot be estimated. One would need to observe many realizations of $z(\xi)$ to have sufficiently many observations of v, to estimate $F_v(v)$. Another non-ergotic process of more relevance to the spatial processes in geotechnical engineering is $z(x) = \mu x + \varepsilon(x)$, in which the mean of $z(x)$ varies linearly with location x. In this case, $z(x)$ is non-stationary; $Var[z(x)]$ increases without limit as the window within which $z(x)$ is observed increases, and the mean m_z of the sample of $z(x)$ is a function of the location of the window.

It is important to note that, ergodicity of a stochastic process (random field) implies strong stationarity. Thus, if a assumed second-order stationary field is also assumed to be ergotic, the latter assumption dominates, and the second-order restriction can be relaxed (Cressie, 1991).

The meaning of ergodicity for spatial fields of the sort encountered in geotechnical engineering is less clear, and has not been widely discussed in the literature.[1] An assumption weaker than full ergodicity, which nonetheless should apply for spatial fields, is that the observed sample mean m_z and sample autocovariance function $\hat{C}_z(\delta)$ converge in mean squared error to the respective random field mean and autocovariance function as the volume of space within which they are sampled increases (Cressie, 1991; Christakos, 1992). This means that, as the volume of space increases, $E[(m_z - \mu)^2] \to 0$, and $E[\{(\hat{C}_z(\delta) - C_z(\delta)\}^2] \to 0$. In Adler (1981), the notion of a continuous process in space, observed continuously within intervals and then compared to the limiting case of the interval tending to infinity, is introduced.

When the joint probability distribution $F_{x_1,\ldots,x_n}(z_1, \ldots, z_n)$ is multivariate Normal (Gaussian), the process $\mathbf{z}(\mathbf{x})$ is said to be a Gaussian random field. A sufficient condition for ergodicity of a Gaussian random field is that $\lim_{\|\delta\| \to \infty} C_z(\boldsymbol{\delta}) = 0$. This can be checked empirically by inspecting the sample moments of the autocovariance function to ensure they converge to 0. Cressie (1991) notes philosophical limitations of this procedure. Essentially all the analytical autocovariance functions common in the geotechnical literature obey this condition, and few practitioners – or even most theoreticians – are concerned about verifying ergodicity. Christakos (1992) suggests that, in practical situations, it is difficulty or impossible to verify ergodicity for spatial fields and, thus, ergodicity must be considered only a falsifiable hypothesis, in the sense that it is judged by the successful application of the random field model.

10.1.3 Nonstationarity

A random field that does not meet the conditions of Section 10.1.1 is said to be non-stationary. Loosely speaking, a non-stationary field is statistically heterogeneous. It can be heterogeneous in a number of ways. In the simplest case, the mean may be a function of location, for example, if there is a spatial trend that has not been removed. In a more complex case, the variance or autocovariance function may vary in space. Depending on the way in which the random field is non-stationary, sometimes a transformation of

[1] Cressie cites Adler (1981) and Rosenblatt (1985) with respect to ergodicity of spatial fields.

variables can convert a non-stationary field to a stationary or nearly stationary field. For example, if the mean varies with location, perhaps a trend can be removed.

In the field of geostatistics, a weaker assumption is made on stationarity than that described in Section 10.1.1. Geostatisticians usually assume only that increments of a spatial process are stationary (i.e. differences $|z_1 - z_2|$) and then operate on the probabilistic properties of those increments. This leads to the use of the variogram (Chapter 9) rather than the autocovariance function as a vehicle for organizing empirical data. The variogram describes the expected value of squared differences of the random field, whereas the autocovariance describes the expected values of products. Stationarity of the latter implies stationarity of the former, but not the reverse.

Like most things in the natural sciences, stationarity is an assumption of the model, and may only be approximately true in the world. Also, stationarity usually depends upon scale. Within a small region, such as a construction site, soil properties may behave as if drawn from a stationary process; whereas, the same properties over a larger region may not be so well behaved.

10.2 Mathematical Properties of Autocovariance Functions

By definition, the autocovariance and autocorrelation functions are symmetric, meaning

$$C_z(\delta) = C_z(-\delta) \text{ and } R_z(\delta) = R_z(-\delta) \tag{10.6}$$

and they are bounded, meaning

$$C_z(\delta) \leq C_z(0) = \sigma^2 \text{ and } |R_x(\delta)| \leq 1 \tag{10.7}$$

In the limit, as absolute distance $|\boldsymbol{\delta}|$ becomes large:

$$\lim_{|\delta|\to\infty} \frac{C_z(\boldsymbol{\delta})}{|\boldsymbol{\delta}|^{-(n-1)/2}} = 0 \tag{10.8}$$

10.2.1 Valid autocovariance functions

In general, for $C_z(\boldsymbol{\delta})$ to be a permissible autocovariance function, it is necessary and sufficient that a continuous mathematical expression be non-negative-definite, that is

$$\sum_{i=1}^{m}\sum_{j=1}^{m} k_i k_j C_z(\boldsymbol{\delta}) \geq 0 \tag{10.9}$$

for all integers m, scalar coefficients $k_1, \ldots k_m$, and vectors $\boldsymbol{\delta} = |\mathbf{x}_i - \mathbf{x}_j|$. This condition follows from the requirement that variances of linear combinations of the $z(\mathbf{x}_i)$, of the form

$$Var\left[\sum_{i=1}^{m} k_i z(x_i)\right] = \sum_{i=1}^{m}\sum_{j=1}^{m} k_i k_j C_z(\boldsymbol{\delta}) \geq 0 \tag{10.10}$$

be non-negative (Cressie 1991). The argument for this condition can be based on spectral representations of $C_z(\boldsymbol{\delta})$, following from Bochner (1955), but beyond the present scope (see also Yaglom 1962). Christakos (1992) discusses the mathematical implications of this condition on selecting permissible forms for the autocovariance. Suffice it to say that analytical models of autocovariance common in the geotechnical literature usually satisfy the condition.

Autocovariance functions valid in R^d, the space of dimension d, are valid in spaces of lower dimension, but the reverse is not necessarily true. That is, a valid autocovariance function in 1-D is not necessarily valid in 2-D or 3-D. Christakos (1992) gives the example of the linearly declining autocovariance:

$$C_z(\delta) = \begin{cases} \sigma^2(1 - \delta/\delta_0), & \text{for } 0 \le \delta \le \delta_0 \\ 0, & \text{for } \delta > \delta_0 \end{cases} \tag{10.11}$$

which is valid in 1-D, but not in higher dimensions. This is easily demonstrated by considering a 2×2 square grid of spacing $\delta_0/\sqrt{2}$, and constants $a_{i,1} = a_{i,2} = a_{j,1} = 1$, and $a_{j,2} = -1.1$, which yields a negative value of the variance.

Linear sums of valid autocovariance functions are also valid. This means that if $C_{z1}(\boldsymbol{\delta})$ and $C_{z2}(\boldsymbol{\delta})$ are valid, then the sum $C_{z1}(\boldsymbol{\delta}) + C_{z2}(\boldsymbol{\delta})$ is also a valid autocovariance function. Similarly, if $C_z(\boldsymbol{\delta})$ is valid, then the product with a scalar, $\alpha C_z(\boldsymbol{\delta})$, is also valid.

10.2.2 Separable autocovariance functions

An autocovariance function in d-dimensional space R^d is *separable* if

$$C_z(\boldsymbol{\delta}) = \prod_{i=1}^{d} C_{zi}(\delta_i) \tag{10.12}$$

in which $\boldsymbol{\delta}$ is the d-dimensioned vector of orthogonal separation distances $\{\delta_1, \ldots, \delta_d\}$, and $C_i(\delta_i)$ is the one-dimensional autocovariance function in direction i. For example, the autocovariance function in R^d,

$$\begin{aligned} C_z(\boldsymbol{\delta}) &= \sigma^2 \exp\{-a^2|\boldsymbol{\delta}|^2\} \\ &= \sigma^2 \exp\{-a^2(\boldsymbol{\delta}_1^2 + \cdots + \boldsymbol{\delta}_d^2)\} \\ &= \sigma^2 \prod_{i=1}^{d} \exp\{-a^2\boldsymbol{\delta}_i^2\} \end{aligned} \tag{10.13}$$

is separable into its d, one-dimensional components.

The function is *partially separable* if

$$C_z(\boldsymbol{\delta}) = C_z(\delta_i)C_z(\boldsymbol{\delta}_{j\neq i}) \tag{10.14}$$

in which $C_z(\boldsymbol{\delta}_{j\neq i})$ is a $(d-1)$ dimension autocovariance function, implying that the function can be expressed as a product of autocovariance functions of lower dimension fields. The importance of partial separability to geotechnical applications, as noted

by Vanmarcke (1983), is the 3-D case of separating autocorrelation in the horizontal plane from that with depth

$$C_z(\delta_1, \delta_2, \delta_3) = C_z(\delta_1, \delta_2)C_z(\delta_3) \tag{10.15}$$

in which δ_1, δ_2, are horizontal distances, and δ_3 is depth.

10.3 Multivariate (vector) random fields

Thus far, we have considered random fields of the scalar variable $z(\mathbf{x})$, in which $\mathbf{x}$ is a vector of location coordinates. By direct analogy, we can define a random field of the vector variable $\mathbf{z}(\mathbf{x})$, in which $\mathbf{z}$ is an ordered n-tuple of variables. For example, water content in a clay stratum can be modeled as a scalar random field. The joint pair of variables water content and shear strength might be modeled as a vector (bi-variate) random field. Note, that in this case each of the variables, water content and shear strength, would have its own autocorrelated properties, but the pair of variables would also be cross correlated one to the other, and as a function of spatial separation.

In analogy to Equation (10.1), a vector random field is defined by the joint probability distribution

$$F_{x_1,\ldots,x_n}(\mathbf{z}_1, \ldots, \mathbf{z}_n) = P\{\mathbf{z}(\mathbf{x}_1) \leq \mathbf{z}_1, \ldots, \mathbf{z}(\mathbf{x}_n) \leq \mathbf{z}_n\} \tag{10.16}$$

in which $\mathbf{z}_i$ is a vector of properties, and $\mathbf{x}$ is spatial location, $\mathbf{x} \in \mathbf{S}$. Let

$$E[\mathbf{z}(\mathbf{x})] = \boldsymbol{\mu}(\mathbf{x}) \tag{10.17}$$

for all $\mathbf{x} \in \mathbf{S}$, be the mean or trend of $\mathbf{z}(\mathbf{x})$, presumed to exist for all $\mathbf{x}$; and let

$$Var[\mathbf{z}(\mathbf{x})] = \Sigma_{\mathbf{z}}(\mathbf{x}) \tag{10.18}$$

be the covariance matrix of the z_i, z_j, such that

$$\Sigma_z(\mathbf{x}) = \{Cov[z_i(\mathbf{x}), z_j(\mathbf{x})]\} \tag{10.19}$$

is the matrix comprising the variances and covariances of the z_i, z_j, also assumed to exist for all $\mathbf{x}$. The cross-covariances of $\mathbf{z}(\mathbf{x}_1), \ldots, \mathbf{z}(\mathbf{x}_n)$ as functions of space are defined as

$$\mathbf{C}_z(\boldsymbol{\delta}) = Cov[z_h(\mathbf{x}_i), z_k(\mathbf{x}_j)] = E[(z_h(\mathbf{x}_i) - \mu_h(\mathbf{x}_i)), (z_k(\mathbf{x}_j) - \mu_k(\mathbf{x}_j))] \tag{10.20}$$

in which $Cov[z_h(\mathbf{x}_i), z_k(\mathbf{x}_j)]$ is the covariance of the (h, k)th components of $\mathbf{z}$, at locations $\mathbf{x}_i$ and $\mathbf{x}_j$. Again, by analogy, the spatial cross-correlation function is

$$\mathbf{R}_{\mathbf{z}}(\delta) = \frac{Cov[z_h(\mathbf{x}_i), z_k(\mathbf{x}_j)]]}{\sigma_{z_h}(\mathbf{x}_i)\sigma_{z_k}(\mathbf{x}_j)} \tag{10.21}$$

The matrix $\mathbf{C}_z(\boldsymbol{\delta})$ is non-negative definite, and symmetric, as a covariance matrix is.

10.4 Gaussian random fields

The Gaussian random field is an important special case, because it is widely applicable due to the Central Limit Theorem (Chapter 2), has mathematically convenient properties, and is widely used in practice. The probability density distribution of the Gaussian or Normal variable is

$$f_z(z) = -\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \tag{10.22}$$

for $-\infty \leq z \leq \infty$. The mean is $E[z] = \mu$, and variance $Var[z] = \sigma^2$. For the multivariate case of vector $\mathbf{z}$, of dimension n, the corresponding pdf is

$$f_z(\mathbf{z}) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-\tfrac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\} \tag{10.23}$$

in which $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{\Sigma}$ the covariance matrix,

$$\Sigma_{ij} = \{Cov[z_i(\mathbf{x}), z_j(\mathbf{x})]\} \tag{10.24}$$

Gaussian random fields have the following convenient properties (Adler 1981): (1) They are completely characterized by the first and second-order moments: the mean and autocovariance function for the univariate case, and mean vector and autocovariance matrix (function) for the multivariate case; (2) any subset of variables of the vector is also jointly Gaussian; (3) the conditional probability distributions of any two variables or vectors are also Gaussian distributed; (4) if two variables, z_1 and z_2 are bi-variate Gaussian, and if their covariance $Cov[z_1, z_2]$ is zero, then the variables are independent.

10.5 Functions of random fields

Thus far, we have considered the properties of random fields themselves. In this section, we consider the extension to properties of functions of random fields.

10.5.1 Stochastic integration and averaging

Spatial averaging of random fields is among the most important considerations for geotechnical engineering. Limiting equilibrium failures of slopes depend upon the average strength across the failure surface. Settlements beneath foundations depend of the average compressibility of subsurface soils. Indeed, many modes of geotechnical performance of interest to the engineer involve spatial averages – or differences among spatial averages – of soil and rock properties. Spatial averages also play a significant role in mining geostatistics, where average ore grades within blocks of rock have important implications for planning. As a result, there is a rich literature on the subject of averages of random fields, only a small part of which can be reviewed here.

Consider the one-dimensional case of a continuous, scalar stochastic process (1D random field), $z(x)$, in which x is location, and $z(x)$ is a stochastic variable with mean μ_z, assumed to be constant, and autocovariance function $C_z(r)$, in which r is separation

distance, $r = (x_1 - x_2)$. The spatial average or mean of the process within the interval [0,X] is

$$M_X\{z(x)\} = \frac{1}{X}\int_0^X z(x)dx \tag{10.25}$$

The integral is defined in the common way, as a limiting sum of $z(x)$ values within infinitesimal intervals of x, as the number of intervals increases. We assume that $z(x)$ converges in a mean square sense, implying the existence of the first two moments of $z(x)$. The weaker assumption of convergence in probability, which does not imply existence of the moments, could be made, if necessary (see Parzen 1964, 1992) for more detailed discussion.

If we think of $M_X\{z(\mathbf{x})\}$ as a sample observation within one interval of the process $z(x)$, then over the set of possible intervals that we might observe, $M_X\{z(\mathbf{x})\}$ becomes a random variable with mean, variance, and possibly other moments. Consider first, the integral of $z(x)$ within intervals of length X. Parzen (1964) shows that the first two moments of $\int_0^X z(x)dx$ are

$$E\left[\int_0^X z(x)dx\right] = \int_0^X \mu(x)dx = \mu X \tag{10.26}$$

$$Var\left[\int_0^X z(x)dx\right] = \int_0^X \int_0^X C_z(x_i - x_j)dx_i dx_j = 2\int_0^X (X - r)C_z(r)dr \tag{10.27}$$

and that the autocovariance function of the integral $\int_0^X z(x)dx$ as the interval [0,X] is allowed to translate along dimension x, is (Vanmarcke 1983),

$$C_{\int_0^X z(x)dx}(r) = Cov\left[\int_0^X z(x)dx, \int_r^{r+X} z(x)dx\right] = \int_0^X \int_0^X C_z(r + x_i - x_j)dx_i dx_j \tag{10.28}$$

The corresponding moments of the spatial mean $M_X\{z(\mathbf{x})\}$ are

$$E[M_X\{z(\mathbf{x})\}] = E\left[\frac{1}{X}\int_0^X z(x)dx\right] = \int_0^X \frac{1}{X}\mu(x)dx = \mu \tag{10.29}$$

$$Var[M_X\{z(\mathbf{x})\}] = Var\left[\frac{1}{X}\int_0^X z(x)dx\right] = \frac{2}{X^2}\int_0^X (X - r)C_z(r)dr \tag{10.30}$$

$$\begin{aligned} C_{M_X\{z(\mathbf{x})\}}(r) &= Cov\left[\frac{1}{X}\int_0^X z(x)dx, \frac{1}{X}\int_r^{r+X} z(x)dx\right] \\ &= \frac{1}{X^2}\int_0^X \int_0^X C_z(r + x_i - x_j)dx_i dx_j \end{aligned} \tag{10.31}$$

The effect of spatial averaging is to smooth the process. The variance of the averaged process is smaller than the original process $z(x)$, and the autocorrelation of the averaged process is wider. Indeed, averaging is sometimes referred to as, smoothing (Gelb *et al*. 1974).

The reduction in variance from $z(x)$ to the averaged process $M_X\{z(x)\}$ can be represented in a *variance reduction function*, $\gamma(X)$:

$$\gamma(X) = \frac{Var[M_X\{z(\mathbf{x})\}]}{Var[z(x)]} \tag{10.32}$$

The variance reduction function is 1.0 for $X = 0$, and decays to zero as X becomes large. Using Equation (10.32), $\gamma(X)$, can be calculated from the autocovariance function of $z(x)$ as

$$\gamma(X) = \frac{\frac{2}{X^2}\int_0^X (X-r)C_z(r)dr}{Var[z(x)]} = \frac{2}{X}\int_0^X \left(1 - \frac{r}{X}\right) R_z(r)dr \tag{10.33}$$

in which $R_z(r)$ is the autocorrelation function of $z(x)$. Note, the square root of $\gamma(X)$ gives the corresponding reduction of the standard deviation of $z(x)$. Table 10.2 gives 1-D variance reduction functions for common autocovariance functions. It is interesting to note that each of these functions is asymptotically proportional to $1/X$. Based on this observation, Vanmarcke (1983) proposed a *scale of fluctuation*, θ, such that

$$\theta = \lim_{X\to\infty} X\gamma(X) \tag{10.34}$$

or $\gamma(X) = \theta/X$, as $X \to \infty$; that is θ/X is the asymptote of the variance reduction function as the averaging window expands. The function $\gamma(X)$ converges rapidly to this asymptote as X increases. For θ to exist, it is necessary that $R_z(r) \to 0$ as $r \to \infty$, that is, that the autocorrelation function decreases faster than $1/r$. In this case, θ can be found from the integral of autocorrelation function (the moment of $R_z(r)$ about the origin),

$$\theta = 2\int_0^\infty R_z(r)dr = \int_{-\infty}^\infty R_z(r)dr \tag{10.35}$$

Table 10.2 Variance reduction functions for common 1D autocovariance (after Vanmarcke, 1983)

Model	Autocorrelation	Variance reduction function	Scale of fluctuation, θ
White noise	$R_x(\delta) = \begin{cases} 1 & \text{if } \delta = 0 \\ 0 & \text{otherwise} \end{cases}$	$\gamma(X) = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{otherwise} \end{cases}$	0
Linear	$R_x(\delta) = \begin{cases} 1 - \lvert\delta\rvert/\delta_0 & \text{if } \delta \le \delta_0 \\ 0 & \text{otherwise} \end{cases}$	$\gamma(X) = \begin{cases} 1 - X/3\delta_0 & \text{if } X \le \delta_0 \\ (\delta_0/X)[1 - \delta_0/3X] & \text{otherwise} \end{cases}$	δ_0
Exponential	$R_x(\delta) = \exp(-\delta/\delta_0)$	$\gamma(X) = 2(\delta_0/X)^2\left(\frac{X}{\delta_0} - 1 + \exp^2(-X/\delta_0)\right)$	$4\delta_0$
Squared exponential (Gaussian)	$R_x(\delta) = \exp^2(-\lvert\delta\rvert/\delta_0)$	$\gamma(X) = (\delta_0/X)^2\left[\sqrt{\pi}\frac{X}{\delta_0}\Phi(-X/\delta_0) + \exp^2(-X/\delta_0) - 1\right]$ in which Φ is the error function	$\sqrt{\pi}\delta_0$

(It should be noted that this variance reduction function is not to be confused with the variance reduction schemes described in Chapter 17.)

This concept of summarizing the spatial or temporal scale of autocorrelation in a single number, typically the first moment of $R_z(r)$, is used by a variety of other workers, and in many fields. Taylor (1921) in hydrodynamics called it the *diffusion constant* (Papoulis and Pillai 2002), Christakos (1992) in geoscience calls $\theta/2$ the *correlation radius*, and Gelhar (1993) in groundwater hydrology calls θ the *integral scale*.

In two dimensions, the equivalent expressions to Equations (10.25) and (10.27) for the mean and variance of the planar integral, $\int_0^X \int_0^X z(x)dx$, are

$$E\left[\int_0^X z(x)dx\right] = \int_0^X \mu(x)dx = \mu X \tag{10.36}$$

$$Var\left[\int_0^X z(x)dx\right] = \int_0^X \int_0^X C_z(x_i - x_j)dx_i dx_j = 2\int_0^X (X - r)C_z(r)dr \tag{10.37}$$

Papoulis (2002) discusses averaging in higher dimensions, as do Elishakoff (1999) and Vanmarcke (1983).

10.5.2 Stochastic differentiation

The issue of continuity and differentiability of a random field depends on the convergence of sequences of random variables $\{\mathbf{z}(\mathbf{x}_a), \mathbf{z}(\mathbf{x}_b)\}$, in which $\mathbf{x}_a, \mathbf{x}_b$ are two locations, with (vector) separation $\mathbf{r} = |\mathbf{x}_a - \mathbf{x}_b|$. The random field is said to be *continuous in mean square* at $\mathbf{x}_a$, if for every sequence $\{\mathbf{z}(\mathbf{x}_a), \mathbf{z}(\mathbf{x}_b)\}$, $E^2[\mathbf{z}(\mathbf{x}_a) - \mathbf{z}(\mathbf{x}_b)] \to 0$, as $\mathbf{r} \to 0$.

The random field is said to be *continuous in mean square* throughout, if it is continuous in mean square at every $\mathbf{x}_a$. Given this condition, the random field $\mathbf{z}(\mathbf{x})$ is *mean square differentiable*, with partial derivative

$$\frac{\partial \mathbf{z}(\mathbf{x})}{\partial \mathbf{x}_i} = \lim_{|r|\to 0} \frac{\mathbf{z}(\mathbf{x} + r\boldsymbol{\delta}_i) - \mathbf{z}(\mathbf{x})}{r} \tag{10.38}$$

in which the delta function $\boldsymbol{\delta}_i$ is a vector of all zeros, except the ith term, which is unity.

While stronger, or at least different, convergence properties could be invoked, mean square convergence is often the most natural form in practice, because we usually wish to use a second-moment representation of the autocovariance function as the vehicle for determining differentiability.[2] A random field is mean square continuous if and only if its autocovariance function, $C_z(\mathbf{r})$, is continuous at $|\mathbf{r}| = 0$. For this to be true, the first derivatives of the autocovariance function at $|\mathbf{r}| = 0$ must vanish

$$\frac{\partial C_{\mathbf{z}}(\mathbf{r})}{\partial \mathbf{x}_i} = 0, \text{ for all } i \tag{10.39}$$

[2] Other definitions of convergence sometimes applied to continuity and differentiability include *convergence with probability one, convergence in probability,* and *weak* or *in-distribution convergence*. For further discussion and definitions of these convergence criteria, see Adler (1981).

If the second derivative of the autocovariance function exists and is finite at $|\mathbf{r}| = 0$, then the field is mean square differentiable, and the autocovariance function of the derivative field is

$$C_{\partial \mathbf{z}/\partial \mathbf{x}_i}(\mathbf{r}) = \partial^2 C_{\mathbf{z}}(\mathbf{r})/\partial \mathbf{x}_i^2 \tag{10.40}$$

and the variance of the derivative field can be found by evaluating the autocovariance $C_{\partial \mathbf{z}/\partial \mathbf{x}_i}(\mathbf{r})$ at $|\mathbf{r}| = 0$. Similarly, the autocovariance of the second derivative field $\partial^2 z(x)/\partial x_i \partial x_j$ is

$$C_{\partial^2 \mathbf{z}/\partial \mathbf{x}_i \partial \mathbf{x}_j}(\mathbf{r}) = \partial^4 C_{\mathbf{z}}(\mathbf{r})/\partial \mathbf{x}_i^2 \partial \mathbf{x}_j^2 \tag{10.41}$$

The cross covariance function of the derivatives $\partial z(x)/\partial x_i$ and $\partial z(x)/\partial x_j$ in separate directions is

$$C_{\partial \mathbf{z}/\partial \mathbf{x}_i, \partial \mathbf{z}/\partial \mathbf{x}_j}(\mathbf{r}) = -\partial^2 C_{\mathbf{z}}(\mathbf{r})/\partial \mathbf{x}_i \partial \mathbf{x}_j \tag{10.42}$$

Importantly, for the case of homogeneous random fields, the field itself, $z(x)$, and its derivative field, $\partial z(x)/\partial x_i$, are uncorrelated (Vanmarcke 1983).

So, the behavior of the autocovariance function in the neighborhood of the origin is the determining factor for mean-square local properties of the field, such as continuity and differentiability (Cramér and Leadbetter 1967). Unfortunately, the properties of the derivative fields are sensitive to this behavior of $C_z(\mathbf{r})$ near the origin, which in turn is sensitive to the choice of autocovariance model. Empirical verification of the behavior of $C_z(\mathbf{r})$ near the origin is exceptionally difficult.

10.5.3 Linear functions of random fields

Assume that the random field $z(x)$, is transformed by a deterministic function $g(.)$, such that

$$y(x) = g[z(x)] \tag{10.43}$$

$g[z(x_0)]$ is a function of z alone, that is, not of x_0, and not of the value of $z(x)$ at any x other than x_0. Also, we assume that the transformation does not depend on the value of x; that is, the transformation is space- or time-invariant, $y(z + \delta) = g[z(x + \delta)]$. Thus, the random variable $y(x)$ is a deterministic transformation of the random variable $z(x)$, and its probability distribution can be obtained from the derived distribution methods of Chapter 3. Similarly, the joint distribution of the sequence of random variables $\{y(x_1), \ldots, y(x_n)\}$ can be determined from the joint distribution of the sequence of random variables $\{x(x_1), \ldots, x(x_n)\}$. The mean of $y(x)$ is then

$$E[y(x)] = \int_{-\infty}^{\infty} g(z) f_z(z(x)) dz \tag{10.44}$$

and the autocorrelation function is

$$R_y(y_1, y_2) = E[y(x_1)y(x_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(z_1)g(z_2) f_z(z(x_1)z(x_2)) dz_1 dz_2 \tag{10.45}$$

Papoulis (2002) shows that the process $y(x)$ is (strictly) stationary if $z(x)$ is (strictly) stationary.

The solution to Equations (10.44) and (10.45) for non-linear transformations may be difficult, but for linear functions general results are available. The mean of $y(x)$ for linear $g(z)$ is found by transforming the expected value of z(x) through the function

$$E[y(x)] = g(E[z(x)]) \tag{10.46}$$

The autocorrelation of $y(x)$ is found in a two-step process

$$R_{yy}(x_1, x_2) = L_{x_1}[L_{x_2}[R_{zz}(x_1, x_2)]] \tag{10.47}$$

in which L_{x_1} is the transformation applied with respect to the first variable $z(x_1)$ with the second variable treated as a parameter, and L_{x_2} is the transformation applied with respect to the second variable $z(x_2)$ with the first variable treated as a parameter.

10.5.4 Excursions (level crossings)

A number of applications arise in geotechnical practice for which one is interested not in the integrals (averages) or differentials of a stochastic process, but in the probability that the process exceeds some threshold, either positive or negative. For example, we might be interested in the probability that a stochastically varying water inflow into a reservoir exceeds some rate or in the weakest interval or seam in a spatially varying soil mass. Such problems are said to involve *excursions* or *level-crossings* of a stochastic process. The following discussion follows Cramér and Leadbetter (1967), Parzen (1964), and Papoulis (2002).

To begin, consider the *zero-crossings* of a random process: the points x_i at which $z(x_i) = 0$. For the general case, this turns out to be a surprisingly difficult problem. Yet, for the continuous Normal case, a number of statements or approximations are possible. Consider a process $z(x)$ with zero mean and variance σ^2. For the interval $[x, x + \delta]$, if the product

$$z(x)z(x + \delta) < 0 \tag{10.48}$$

then there must be an odd number of zero-crossings within the interval, for if this product is negative, one of the values must lie above zero and the other beneath. Papoulis (2002) demonstrates that, if the two (zero-mean) variables $z(x)$ and $z(x + \delta)$ are jointly normal with correlation coefficient

$$r = \frac{E[z(x)z(x + \delta)]}{\sigma_x \sigma_{x+\delta}} \tag{10.49}$$

then

$$\begin{aligned} P(z(x)z(x + \delta) < 0) &= \frac{1}{2} - \frac{\arcsin(r)}{\pi} = \frac{\arccos(r)}{\pi} \\ P(z(x)z(x + \delta) > 0) &= \frac{1}{2} + \frac{\arcsin(r)}{\pi} = \frac{\pi - \arccos(r)}{\pi} \end{aligned} \tag{10.50}$$

The correlation coefficient, of course, can be taken from the autocorrelation function, $R_z(\delta)$; thus

$$\cos[\pi P(z(x)z(x + \delta) < 0)] = \frac{R_z(\delta)}{R_z(0)} \tag{10.51}$$

and the probability that the number of zero-crossings is positive is just the complement of this.

For simplicity, let the probability of an odd number of zero-crossings within the interval $[x, x+\delta]$ be $p_0(\delta)$. If δ is small, then this probability is itself small compared to 1.0. Thus, the probability of two or more zero-crossings is negligible. Therefore, the probability of exactly one zero-crossing, $p_1(\delta)$, is approximately, $p_1(\delta) \approx p_0(\delta)$, and thus, from Equation (10.50), expanding the cosine in a Fourier series and truncating to two terms:

$$1 - \frac{\pi^2 p_1^2(\delta)}{2} = \frac{R_z(\delta)}{R_z(0)} \tag{10.52}$$

or

$$p_1(\delta) \approx \frac{1}{\pi}\sqrt{\frac{2[R_z(0) - R_z(\delta)]}{R_z(0)}} \tag{10.53}$$

In the case of a regular autocorrelation function, for which the derivative $dR_z(0)/d\delta$ exists and is zero at the origin, the probability of a zero-crossing is approximately

$$p_1(\delta) \approx \frac{\delta}{\pi}\sqrt{-\frac{d^2R_z(0)/d\delta^2}{R_z(0)}} \tag{10.54}$$

The non-regular case, for which the derivative at the origin is not zero (e.g. $dR_z(\delta) = exp(\delta/\delta_0)$), is discussed by Parzen (1964). Elishakoff (1999) and Vanmarcke (1983) treat higher dimensional results.

The related probability of the process crossing an arbitrary level, z^*, can be approximated by noting that, for small δ and thus $r \to 1$, from Equation (10.50),

$$P[\{z(x) - z^*\}\{z(x+\delta) - z^*\} < 0] \approx P[\{z(x)\}\{z(x+\delta)\} < 0]e^{\frac{-\arcsin^2(r)}{2\sigma^2}} \tag{10.55}$$

For small δ, the correlation coefficient $R_z(\delta)$ is approximately 1, and the variances of $z(x)$ and $z(x+\delta)$ are approximately $R_z(0)$, thus

$$p_{1,z^*}(\delta) \approx p_1(\delta)e^{\frac{-\arcsin^2(r)}{2R_z(0)}} \tag{10.56}$$

and for the regular case

$$p_1(\delta) \approx \frac{\delta}{\pi}\sqrt{\frac{-d^2R_z(0)d\delta^2}{R_z(0)}}e^{\frac{-\arcsin^2(r)}{2R_z(0)}} \tag{10.57}$$

Many other results can be found for continuous Normal processes, e.g. the average density of the number of crossings within an interval, the probability of no crossings (i.e. drought) within an interval, and so on. A rich literature is available of these and related results (Adler 1981; Christakos 1992, 2000; Christakos and Hristopulos 1998; Cliff and Ord 1981; Cramér and Leadbetter 1967; Cressie 1991; Gelb *et al.* 1974; Papoulis and Pillai 2002; Parzen 1964; Yaglom 1962).

11 Spatial Sampling

Sampling theory for spatial processes has developed principally from a frequentist point of view, involving estimators, sampling distributions, and confidence limits, and this is the approach reflected in this chapter. There is no reason that parallel results cannot be developed from a Bayesian view, and Chapter 4 presents Bayesian sampling results for simple independent identically distributed (IID) situations. IID means that sample observations are assumed to be drawn independently of one another and all from the same sampled population. This chapter briefly summarizes a large literature on frequentist sampling theory as applied to spatial variables, and discusses how that theory applies to the spatial sampling problems encountered in site characterization. Many excellent books are available on sampling theory, and more complete discussions are available in Cochran (1977), Thompson (2002), and Levy and Lameshow (1999), among others. More detailed treatment of spatial sampling can be found in Matérn (1960), Cliff and Ord (1981), Ripley (1981), Cressie (1991), Berger (1993), or Stein (1999).

11.1 Concepts of Sampling

The purpose of sampling is to obtain estimates of population parameters (e.g. means, variances, covariances) or possibly to characterize the entire population distribution without observing and measuring every element in the sampled population. Spatial sampling differs from IID sampling in that the observations may display spatial structure, for example, autocorrelation.

An *estimator*, in the frequentist sense, is a sample statistic that can be used to estimate true population parameters. It can be chosen on several bases (Table 4.1), and individual estimators seldom satisfy all the desired criteria simultaneously. An *estimate* is the realization of a particular estimator for a specific set of sample observations.

Estimates are not exact. Uncertainty is reflected in the variance of their distribution about the true parameter value they estimate. This variance is, in turn, a function of both the sampling plan and the sampled population. By knowing this variance and making assumptions about the distribution shape, confidence limits on true population parameters can be set.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

A *sampling plan* is a program of action for collecting data from a sampled population. Common plans are grouped into many types: for example, simple random, systematic, stratified random, cluster, traverse, line intersects, and so on. In deciding among plans, or in designing a specific program once the type plan has been chosen, one attempts to obtain the highest precision for a fixed sampling cost or the lowest sampling cost for a fixed precision. The contrast with Bayesian methods is that the latter attempt to maximize the overall expected utility of sampling and engineering design.

11.2 Common Spatial Sampling Plans

Statistical sampling is a common activity in many human enterprises, from the national census, to market research, to scientific research. As a result, common situations are encountered in many different endeavors, and a family of sampling plans has grown up to handle these situations. In this section, we consider four such sampling plans: simple random sampling, systematic sampling, stratified random sampling, and cluster sampling. We revisit some of these again in Chapter 17 in discussing Monte Carlo methods.

11.2.1 Simple random sampling

The characteristic property of simple random sampling is that individual observations are chosen at random from the sampled population, and each element of the sampled population has an equal probability of being observed.

An unbiased estimator of the population mean from a simple random sample, $\mathbf{x} = \{x_1, \ldots, x_n\}$, is the sample mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{11.1}$$

This estimator has sampling variance,

$$Var(\overline{x}) = \frac{\sigma^2}{n}\frac{N-n}{N} \tag{11.2}$$

where σ^2 is the (true) variance of the sampled population, and N is the total sampled population size. The term, $(N-n)/N$, is called the *finite population factor*, which for n less than about 10% of N, can safety be ignored. However, since σ^2 is usually unknown, it is estimated by the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{11.3}$$

in which the denominator is taken as $n-1$ rather than n, reflecting the loss of a degree-of-freedom due to estimating the mean from the same data. The estimator is unbiased, but does not have minimum variance (Chapter 4).

The only choice (i.e. allocation) to be made in simple random sampling is the sample size n. Since the sampling variance of the mean is inversely proportional to sample size,

$Var(\overline{x}) \propto n^{-1}$, a given estimator precision can be obtained by adjusting the sample size, if σ is known or assumed. A sampling plan can be optimized for total cost by assuming some relationship between $Var(\overline{x})$ and cost in construction or design. A common assumption is that this cost is proportional to the square root of the variance, usually called the *standard error* of the mean, $\sigma_{\overline{x}} = Var^{1/2}(\overline{x})$.

Consider a set of direct shear tests performed on silty sand specimens collected using a simple random sampling plan. Under a normal stress of 60 kPa, the tests yield the results: $\boldsymbol{x} = \{38, 51, 43, 39, 48, 45, 42, 45, 49\,\text{kPa}\}$. The mean strength is estimated from the sample mean, $\overline{x} = 44.4\,\text{kPa}$, and the standard deviation from the sample standard deviation, $s = 4.2\,\text{kPa}$. The standard error in the estimator for the mean, from Equation (11.2) is, $\sqrt{Var(\overline{x})} = 4.2/\sqrt{9} = 1.4\,\text{kPa}$. This implicitly assumes that the total volume of the specimens is small compared to the volume of the soil *in situ*, as the finite population factor is ignored.

11.2.2 Systematic sampling

In systematic sampling the first observation is chosen at random, and subsequent observations are chosen periodically throughout the population. For example, grid sampling is a systematic plan. An unbiased estimate of the mean from a systematic sample is the same as Equation (11.1). The sampling variance of this estimate is

$$Var(\overline{x}) = \left(\frac{N-1}{N}\right)\sigma_w^2 + \left(\frac{k(n-1)}{N}\right) \tag{11.4}$$

where k is the interval between samples ($k = N/n$), and σ_w^2 is the variance of elements within the same systematic sample

$$s_w^2 = \frac{1}{k}\sum_{i=1}^{k}\frac{\sum_{j=1}^{n}(x_{ij} - \overline{x})^2}{n-1} \tag{11.5}$$

in which x_{ij} is the jth member of the ith interval of the sample. When only one systematic sample has been taken (i.e. one set of n observations at spacing k) the variance of the mean cannot be evaluated unless an assumption is made about the nature of the sampled population. The conventional assumption is that the population can be modeled by a linear expression of the form, $x_i = \mu + e_i$, in which μ is the mean and e_i is a zero-mean random perturbation. For constant mean, this leads to (Cochran 1977)

$$Var(\overline{x}) = \frac{1}{n}\left(\frac{N-n}{N}\right)\left(\frac{\Sigma(x_i - \overline{x})^2}{n-1}\right) \tag{11.6}$$

and for linearly trending mean, $\mu = \mu_0 + b_i$,

$$Var(\overline{x}) = \frac{1}{n}\left(\frac{N-n}{N}\right)\left(\frac{\Sigma(x_i - 2x_{i+k} + x_{i+2k})}{6(n-2)}\right) \tag{11.7}$$

This estimate depends upon the population being adequately described by a linear trend. A general way of estimating $Var(\overline{x})$ from one systematic sample does not exist. Matérn

(1986) among others deals with this problem. Standard convention is to assume the linear model, although in geotechnical practice different choices are sometimes called for.

The parameter over which allocation is optimized for systematic sampling is k, the spacing between observations, although, in more than one dimension the pattern of systematic sampling can also be optimized (e.g. triangular vs. rectangular grids, etc.). From Equation (11.4), if s^2 and s_w^2 are approximately the same, $Var(\overline{x}) \approx s_w^2(k/N)$, so precision is controlled by changing the ratio of sample spacing to population size (e.g. borehole spacing to formation or site dimension). Using the cost model, $c(n) = c_0 + c_1 n$, where c_0 is a setup cost and c_1 is unit cost, an optimal k can determined, as before, by assuming a risk cost proportional to $\sqrt{Var(\overline{x})}$.

11.2.3 Stratified random sampling

A heterogeneous population can sometimes be divided into subpopulations that are internally homogeneous. For each homogeneous subpopulation, usually called a *stratum* – somewhat confusing in the geological context, but common statistical usage – precise estimates of stratum characteristics can be obtained by random sampling. Estimates of the total population characteristics can then be made by combining the individual stratum estimates. For certain populations, stratifying before sampling is more efficient than taking samples directly from the total population. Sampling plans that specify a simple random sample in each stratum are called *stratified random sampling* plans.

An unbiased estimator of the mean of the total sampled population is (Thompson 2002)

$$\overline{\overline{x}} = \frac{1}{N}\sum_{h=1}^{m} N_h \overline{x}_h \tag{11.8}$$

where $\overline{\overline{x}}$ is the population mean, m is the number of strata, and h denotes the stratum (i.e. N_h is the size of the hth stratum, and $\overline{x}_h$ is the corresponding mean). The variance of this estimate is

$$Var(\overline{x}) = \sum_h w_h^2 \frac{s_h^2}{n_h}(1 - f_h) \tag{11.9}$$

where $w_h = N_h/N$ and $f_h = n_h/N_h$. Since the sample from each stratum is simple random, the estimate of the variance within each can be taken from Equation (11.2). Then, an estimate of the variance of the total population is

$$Var(\overline{x}) = \frac{1}{N}\sum_h s_h^2 + s^2_{among\ means} \tag{11.10}$$

The data of Table 11.1 are standard penetration test blow counts measured in a sequence of sand and silty sand strata that are part of a harbor reclamation project. The data have been divided among the five apparent geological strata within which they were measured. The sizes of the respective strata are measured in meters of thickness, and are assumed large compared to the volume of soil tested. The strata means and standard deviations are estimated assuming simple random sampling within each by Equations (11.1) and (11.3), respectively. The overall formation mean is estimated by taking the sum-product of the

Table 11.1 Standard penetration test blow counts in silty sand reclaimed land by stratum

Stratum	N(h)	n(h)	Measurements 1	2	3	4	5	6	Stratum Mean	Stratum SD
1	1.67	5	7.80	11.41	6.88	10.65	5.36		8.42	2.55
2	3.33	4	12.19	13.27	13.41	11.96			12.71	0.74
3	6.67	6	10.47	12.50	6.81	9.27	10.72	12.35	10.35	2.12
4	4.00	6	11.09	20.49	15.04	16.41	14.46	15.00	15.42	3.05
5	4.67	5	8.17	11.57	10.01	10.65	12.03		10.49	1.51
						Sum Product over N(h) =			11.61	2.00

strata means over the relative sizes of the strata, according to Equation (11.8), to obtain 11.61 bpf. The sampling variance is calculated by Equation (11.9) to obtain $Var(\overline{x}) = 0.18\ \text{bpf}^2 = 0.18\,\text{bpf}^2$.

Assuming that the strata are obviously separable and that the relative sizes of the sampled strata are known, how many elements do we sample from each stratum? If the cost function is of the form $c(n) = c_0 + \Sigma c_h n_h$, then $Var(\overline{x})$ is minimized when $n_h \propto N_h s_n / \sqrt{c_h}$ (Cochran 1977). In words, the sample size within a stratum should be larger when (1) the stratum is larger, (2) the variance of the stratum is larger, and (3) the cost of sampling in the stratum is lower. For the common case when the unit cost is the same in each stratum, $Var(\overline{x})$ is minimized when $n_h = n(N_h s_n / \Sigma N_h s_n)$ (Neyman 1934). This is sometimes called, the *Neyman allocation*. A second common plan is *proportional allocation*, $n_h = n(N_h/N)$.

The sampling of some populations of interest to geotechnical practice, for example the orientation of rock joints (Einstein and Baecher 1983), presents questions of sampling one population for two or more independent variables, e.g. strike and dip. The difficulty with such plans is that the optimum allocation on one variable may not be the optimal allocation on the other.

For Neyman allocation the easiest approach is to determine the optimum allocation on each variable and average the results. Since the precision of a stratified sampling plan is insensitive to minor deviations from optimal allocation, the averaging procedure often offers almost as much precision as the optimum allocation for each variable. In the case of positive correlation between the two variables, averaging works even better. For cases where the optimum allocations for each variable are widely different, other allocating techniques have been proposed by Yates (1948).

11.2.4 Cluster sampling

In cluster sampling, aggregates or *clusters* of elements are selected from the sampled population as units rather than as individual elements, and properties of the clusters are determined. From the properties of the clusters, inferences can be made on the total sampled population. Plans that specify to measure every element within clusters are called *single-stage cluster plans*, since they specify only one level of sampling; plans that specify that cluster properties be estimated by simple random sampling are called *two-stage cluster plans*, since they specify two levels of sampling. Higher order cluster plans are sometimes used.

We consider the simplest case first: of M possible clusters, m are selected; the ratio $f_1 = m/M$ is called, as before, the *sampling fraction.* Each cluster contains the same number of elements, N, some number n of which are selected for measurement ($f_2 = n/N$). An unbiased estimate of the average of each cluster is

$$\overline{x}_i = \frac{1}{n_i} \sum_j x_{ij} \tag{11.11}$$

where x_{ij} is the jth element of the ith cluster. An unbiased estimate of the average of the total population is

$$\overline{\overline{x}}_i = \frac{1}{m} \sum_i \overline{x}_j = \frac{1}{mn} \sum_i \sum_j x_{ij} \tag{11.12}$$

and the variance of this estimator is

$$Var(\overline{\overline{x}}) = \frac{(1 - f_1)}{m} s_1^2 + \frac{f_1(1 - f_2)}{mn} s_2^2 \tag{11.13}$$

in which s_1^2 is the estimated variance among cluster means

$$s_1^2 = \frac{\Sigma(\overline{x}_i - \overline{\overline{x}})^2}{n - 1} \tag{11.14}$$

and s_2^2 the estimated variance within clusters

$$s_2^2 = \frac{\Sigma\Sigma(x_{ij} - \overline{x}_i)^2}{m(n - 1)} \tag{11.15}$$

In the more general case, not all of the clusters are of equal size. For example, the numbers of joints appearing in different outcrops are different. With unequal sized clusters the selection plan for clusters is not as obvious as it was previously. The relative probability, z_i, of selecting different sized clusters is now a parameter of the plan. Commonly, the z_i are either taken all equal (simple random sampling of the clusters) or proportional to size. The precisions of these two plans are different. For selection with equal probability an unbiased estimate of the true total population mean is

$$\overline{\overline{x}} = \frac{\sum_i N_i \overline{x}_i}{\sum_i N_i} \tag{11.16}$$

and the variance of this estimator is

$$Var(\overline{\overline{x}}) = \frac{(1 - f_1)}{m\overline{N}^2} \frac{\Sigma(N_i\overline{x}_i - \overline{x}_m)^2}{(m - 1)} + \frac{f_1}{m^2\overline{N}^2} \sum \frac{N_i^2(1 - f_{2i})s_{2i}^2}{n_i} \tag{11.17}$$

in which $\overline{N}$ is the average value of N_i, and $\overline{x}_m = \sum_i N_i x_i / m$. If the assumption is made that $n_i \propto N_i$, then the plan is self-weighting and this simplifies to

$$Var(\overline{\overline{x}}) = \frac{(1 - f_1)}{m\overline{N}_2} \frac{\Sigma(N_i\overline{x}_i - \overline{x}_m)^2}{(m - 1)} + \frac{f_1(1 - f_2)}{m\overline{n}\overline{N}} \sum N_i^2 s_{2i}^2 \tag{11.18}$$

This assumption (i.e. $n_i \propto N_i$) is frequently valid; for example, proportionally more joints are typically sampled from larger outcrops than from smaller outcrops.

For selection with probability proportional to size, an unbiased estimate of the total population mean is

$$\overline{\overline{x}} = \frac{1}{n} \sum_i \overline{x}_i \tag{11.19}$$

and the variance is

$$Var(\overline{\overline{x}}) = \frac{\Sigma(\overline{x}_i - \overline{\overline{x}})^2}{m(n-1)} \tag{11.20}$$

In all cases, the variance of the total population can be estimated from the variances between elements within clusters and the variance between the means of the clusters:

$$\sigma^2 = \sigma^2_{among\ means} + \sigma^2_{within\ clusters} \tag{11.21}$$

Joint surveys and many other types of geotechnical sampling may be based on cluster plans because the cost of sampling many individual joints on one outcrop (i.e. a cluster) is less than the cost of traveling between outcrops.

The variance of geological populations is often a function of spatial extent. Indeed, this is the principal argument in the geostatistical literature for favoring variograms over autocovariance functions. If we consider the strength of soil specimens taken close together, the variance among specimens is usually smaller than the variance among specimens taken from many locations in one area of the site, which, in turn, is smaller than the variance among specimens taken from all across the site. Cluster techniques allow us to evaluate variance as a function of the "extent" of the distribution in space by nesting the variances. Examples of the analysis of nested levels of variance in geology are found in Pettijohn (1987) and Potter and Siever (1953), who analyze variance in cross-bedding as a function of scale.

To analyze nested levels of variance we first assume a model for the total population. Consider Figure 11.1, in which an abstract map of a population is shown. The population is divided into elements. The variance within each smallest element is s_4^2, since the

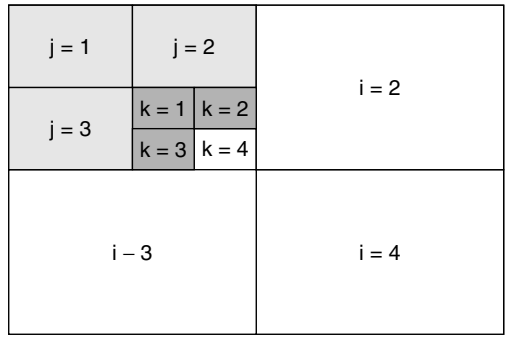


Figure 11.1 Nested model of a sampled population, showing 54 smallest level elements structured within four levels of nesting.

example has four levels of nesting. The means of smallest level elements within each second-smallest element are themselves distributed with some variance, s_3^2, and so forth.

The whole population, therefore, can be modeled by some 'grand' mean, μ, plus random components accounting for the variance at all four levels. Mathematically,

$$x_{ijkl} = \mu + a_i + b_{ij} + c_{ijk} + d_{ijkl} \tag{11.22}$$

where a_i, is a zero-mean error term associated with variability among the largest elements, and d_{ijkl} is the error term associated with variability among the smallest elements within the next largest element. If the variance among means at level a is, s_a^2, then the total variance is $s_i^2 + s_j^2 + s_k^2 + s_l^2$; and the variance in the estimate of the grand mean, μ, is

$$Var(\overline{\overline{x}}) = \frac{(1-f_1)}{n_1}s_1^2 + \frac{f_1(1-f_2)}{n_1 n_2}s_2^2 + \frac{f_1 f(1-f_3)}{n_1 n_2 n_5}s_3^2 + \cdots \tag{11.23}$$

where f_i is the sampling fraction at level a, and n_a is the size of the sample at that level, respectively (Cochran 1977).

As an example, a large site is divided into nested clusters according to the logical scheme of Figure 11.1, but with three rather than four levels of nesting. At the smallest cell level, five specimens are tested in each cell. The average of all the tests ($N = 5 \times 4 \times 4 \times 4 = 320$) is 22.56. The standard deviations among the nested means at each level, respectively, are $s_1 = 5.2$, $s_2 = 1.4$, and $s_3 = 0.4$. Thus, the standard error in the estimate of the site mean, by Equation (11.23), ignoring the finite population correction, is $\sqrt{Var(\overline{\overline{x}})} = \sqrt{\{(5.2)^2/4 + (1.4)^2/16 + (0.4)^2/64\}} = 5.4$.

11.3 Interpolating Random Fields

A problem common in site characterization is interpolating among spatial observations to estimate soil or rock properties at specific locations where they have not been observed. The sample observations themselves may have been taken under any number of sampling plans: random, systematic, cluster, or so forth. What differentiates this spatial estimation question from the sampling theory estimates in preceding sections of this chapter is that the observations display spatial correlation. Thus, the assumption of IID observations underlying the estimator results is violated in an important way. This question of spatial interpolation is also a problem common to the natural resources industries such as forestry (Matérn 1986) and mining (Matheron 1971), but also to geohydrology (Kitanidis 1997), and to environmental monitoring (Switzer 1995).

Consider the case for which the observations are sampled from a spatial population with constant mean, μ, and autocovariance function $C_z(\delta) = \mathrm{E}[z(x_i)z(x_{i+\delta})]$. The set of observations $\mathbf{z} = \{z_i, \ldots, z_n\}$ therefore has mean vector $\boldsymbol{\mu}$ in which all the terms are equal, and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} Var(z_1) & \cdots & Cov(z_1, z_n) \\ \vdots & \ddots & \vdots \\ Cov(z_n, z_1) & \cdots & Var(z_n) \end{bmatrix} \tag{11.24}$$

These terms are found from the autocovariance function as $Cov(z(x_i)z(x_j)) = C_z(\delta_{ij})$, in which δ_{ij} is the (vector) separation between locations x_i and x_j.

In principle, we would like to estimate the full distribution of $z(x_0)$ at an unobserved location x_0, but in general, this is computationally intensive if a large grid of points is to be interpolated. Instead, the most common approach is to construct a simple linear unbiased estimator based on the observations

$$\hat{z}(x_0) = \sum_{i=1}^{n} w_i z(x_i) \tag{11.25}$$

in which the weights $\mathbf{w} = \{w_1, \ldots, w_n\}$ are scalar values chosen to make the estimate in some way optimal. Usually, the criteria of optimality are unbiasedness and minimum variance. This is sometimes called the Best Linear Unbiased Estimator (BLUE).

The BLUE estimator weights are found by expressing the variance of the estimate $\hat{z}(x_0)$ using a first-order second-moment formulation (Chapter 14), and minimizing the variance over $\mathbf{w}$ using a Lagrange multiplier approach subject to the condition that the sum of the weights equals one (Hildebrand 1976). The solution in matrix form is

$$\mathbf{w} = \mathbf{G}^{-1}\mathbf{h} \tag{11.26}$$

in which $\mathbf{w}$ is the vector of optimal weights. The matrices $\mathbf{G}$ and $\mathbf{h}$ are the covariance matrix of the observations and the vector of covariances of the observations related to the value of the spatial variable at the interpolated location, x_0, respectively,

$$\mathbf{G} = \begin{bmatrix} Var(z_1) & \cdots & Cov(z_1, z_n) & 1 \\ \vdots & \ddots & \vdots & 1 \\ Cov(z_n, z_1) & \cdots & Var(z_n) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad \mathbf{h} = \begin{bmatrix} Cov(z_1, z_0) \\ \vdots \\ Cov(z_n, z_0) \\ 1 \end{bmatrix} \tag{11.27}$$

in which the terms $z(x_i)$ are replaced by z_i for convenience. The resulting estimator variance is

$$\begin{aligned} Var(\hat{z}_0) &= E[(z_0 - \hat{z}_0)^2] \\ &= Var(z_0) - \sum_{i=1}^{n} w_i \, Cov(z_0, z_i) - \lambda \end{aligned} \tag{11.28}$$

in which λ is the Lagrange multiplier resulting from the optimization. This is a surprisingly simple and convenient result, and forms the basis of the increasingly vast literature on the subject of so-called *Kriging* in the field of geostatistics (below). For regular grids of observations, like a grid of borings, an algorithm based on Equations (11.25) and (11.26) can be established for the points within an individual grid cell, and then replicated for all cells to form an interpolated map of the larger site or region (Cressie 1991; David 1977; Isaaks and Srivastava 1989; Journel and Huijbregts 1978).

In the mining industry (Matheron 1971), and increasingly in other applications (Kitanidis 1997), it has become common to replace the autocovariance function as a measure of spatial association with the *variogram* (Chapter 9). In place of the expected value of the

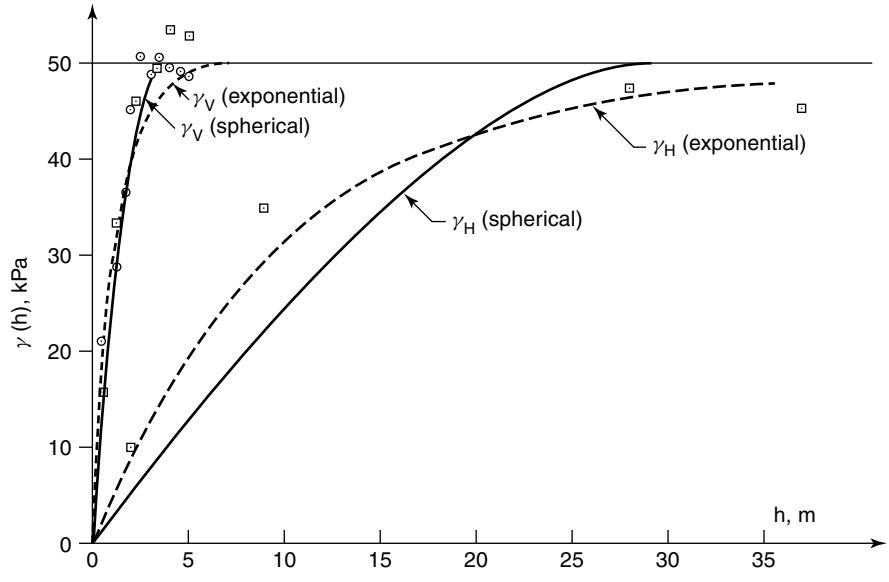


Figure 11.2 Models of horizontal and vertical variograms. (Soulie, M., Montes, P. and Silvestri, V., 1990. 'Modeling spatial variability of soil parameters.' *Canadian Geotechnical Journal* **27**, pp. 617–630, reproduced with permission of the National Research Council of Canada.)

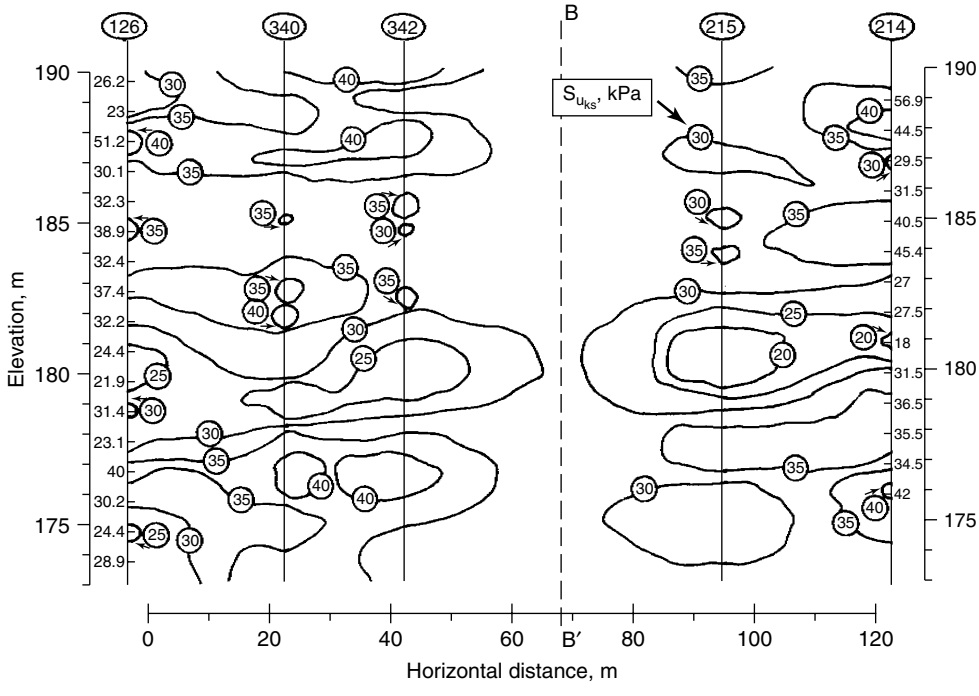


Figure 11.3 Contours of Kriged shear vane strength values. (Soulie, M., Montes, P. and Silvestri, V., 1990. 'Modeling spatial variability of soil parameters.' *Canadian Geotechnical Journal* **27**, pp. 617–630, reproduced with permission of the National Research Council of Canada.)

product of two variables as used in the autocovariance function, the variogram, $2\gamma(\delta)$, uses the expected value of the squared difference:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & 1 \\ \gamma_{n1} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad \boldsymbol{\lambda} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n1} \\ 1 \end{bmatrix} \tag{11.29}$$

In the geostatistics literature, this interpolation procedure is called *Kriging*, named by Matheron (1971) after Krige (1951).

Soulie and his associates (Soulie and Favre 1983; Soulie *et al*. 1990) have used Kriging to interpolate shear vane strengths of a marine clay both horizontally and vertically among

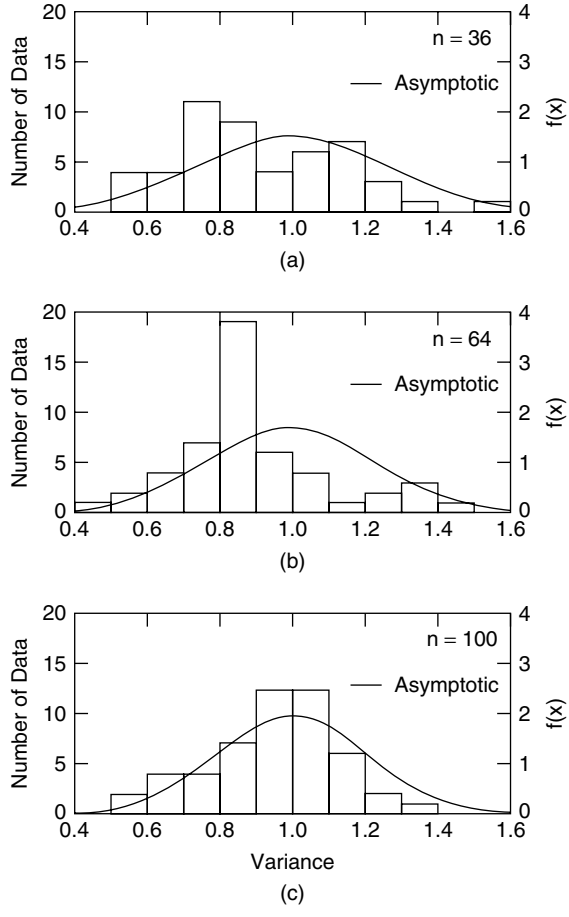


Figure 11.4 Histograms of maximum likelihood estimates of spatial variance, $\sigma^2 = Var(z(x))$, compared to the theoretical asymptotic distribution for the pdf of the estimates. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating autocovariance of in situ soil properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119** (GT1), pp. 147–166, reproduced with permission of the American Society of Civil Engineers.)

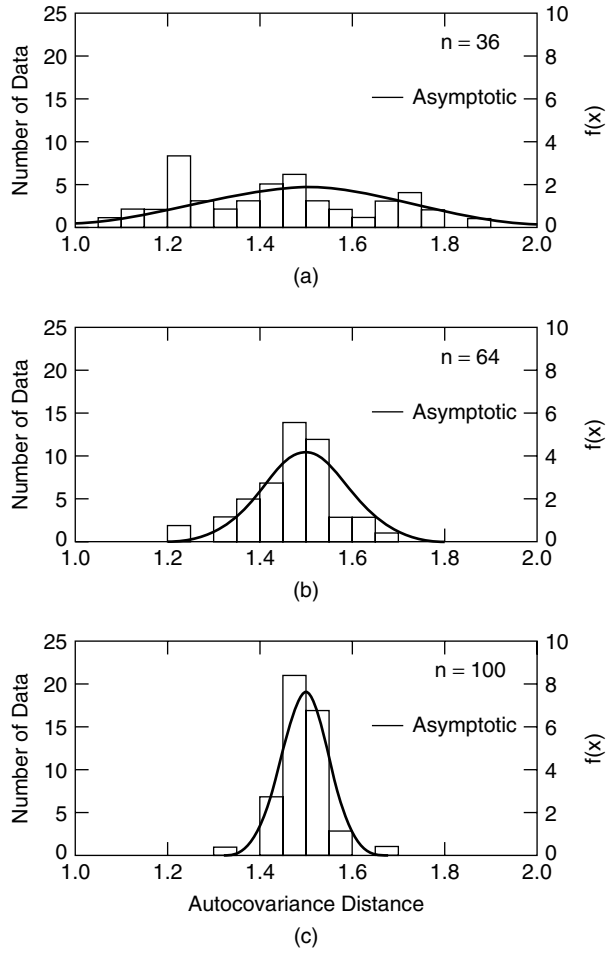


Figure 11.5 Histograms of maximum likelihood estimates of autocovariance distance, r, compared to the theoretical asymptotic distribution for the pdf of the estimates. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating autocovariance of in situ soil properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119** (GT1), pp. 147–166, reproduced with permission of the American Society of Civil Engineers.)

borings. To do so, they estimated semi-variograms based on spherical and exponential models (Table 9.2) separately for the horizontal and vertical directions (Figure 11.2), and then used Equations (11.25) and (11.29) to interpolate the observations (Figure 11.3).

11.4 Sampling for Autocorrelation

The literature addressing statistical aspects of estimating autocovariance functions or variograms from data is not extensive. Most texts on geostatistics quickly pass over the estimation of variograms based on intuitive plotting of sample moments, and quickly move on to Kriging and other applications of the variogram. Cressie (1991) provides an

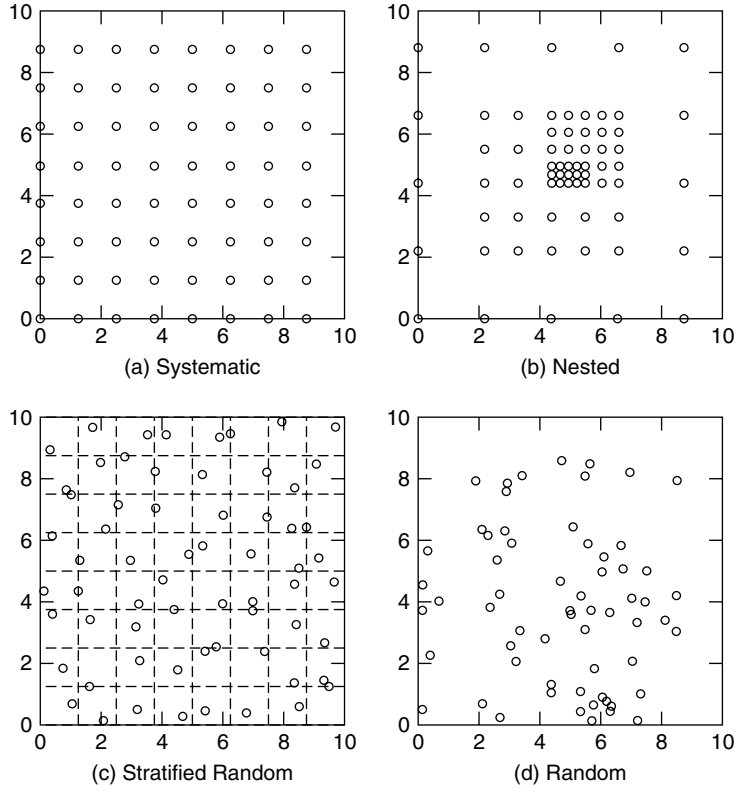


Figure 11.6 Layout of sampling plans for $n = 64$ points. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating autocovariance of in situ soil properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119** (GT1), pp. 147–166, reproduced with permission of the American Society of Civil Engineers.)

overview and summary of various issues and statistical questions surrounding the estimation of variograms. The parallel literature of time series analysis, albeit in one-dimension rather than three or more, is much more extensive on statistical questions, and is another starting point for working with spatial data (Kendall and Stuart 1976). Mardia and Marshall (1984) in a widely-cited work, present a maximum likelihood approach to estimating autocovariance functions, which is discussed in Chapter 9 of this book. The method is, of course, closely related to Bayesian approaches. The issue of estimating these functions was treated in Chapter 9.

The question here is, how to design a sampling plan for estimating the autocovariance or variogram function, and the answer follows the results of DeGroot (DeGroot 1996; DeGroot and Baecher 1993). DeGroot carried out a series of simulation experiments to investigate the properties of maximum likelihood estimators of autocorrelation along the lines of Mardia and Marshall. Simulations were performed by generating pseudo-random standard Normal variates $z \sim N(0, 1)$, and then taking the transformation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{AZ}$, where $\mathbf{I}$ is the identify matrix and $\mathbf{A}$ is the Cholesky decomposition (Chapter 17) of the autocovariance matrix $\mathbf{V} = \mathbf{AA}^{\mathrm{t}}$. The mean was taken as constant, $\boldsymbol{\mu} = 0$. Maximum likelihood estimates of trend and autocovariance parameters were obtained using Mardia and

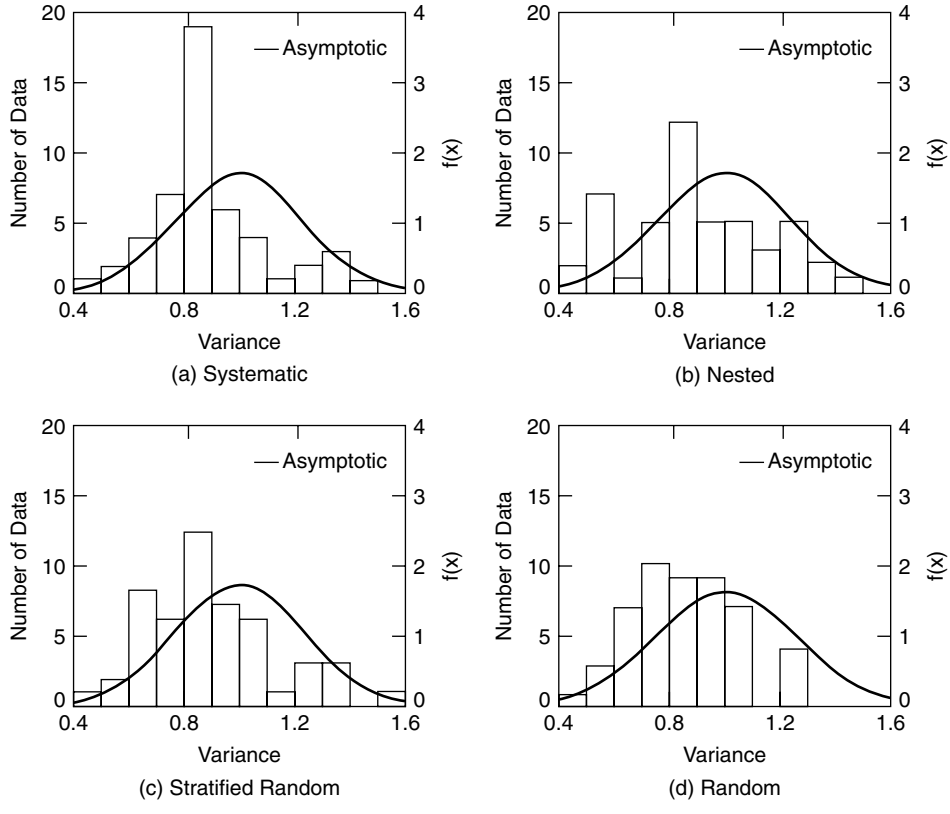


Figure 11.7 Histograms of maximum likelihood estimates of spatial variance, $\sigma^2 = Var(z(x))$, for various sampling plans, for $n = 64$ points. (DeGroot, D. J. and Baecher, G. B., 1993. 'Estimating autocovariance of in situ soil properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119** (GT1), pp. 147–166, reproduced with permission of the American Society of Civil Engineers).

Marshall's algorithm for maximum likelihood estimation (Chapter 9). Since the principal constraint facing most geotechnical site characterization is limited numbers of observations, the analyses focused on small sample sizes, n, and results were expressed as a function of n.

Initially, square grids of, respectively, $n = 36$, 64, and 100 observations were simulated; corresponding to 6×6, 8×8, and 10×10 grid patterns, uniform over the same site area. Thus, the respective proportional grid spacings were, 1.7, 1.25, and 1.0. Data were generated using a squared exponential autocovariance function

$$C_z(\delta) = \sigma^2 \exp(-\delta/r) \tag{11.30}$$

in which $\sigma^2 = Var(z(x))$ is the spatial variance and r is the so-called autocorrelation distance (i.e. the distance at which the autocorrelation diminishes to $1/e$). The resulting asymptotic distribution of the maximum likelihood estimates of σ^2 and r along with the simulated results from 100 trials are shown in Figures 11.4 and 11.5. The fit, even for a 6×6 grid, is reasonable in most cases.

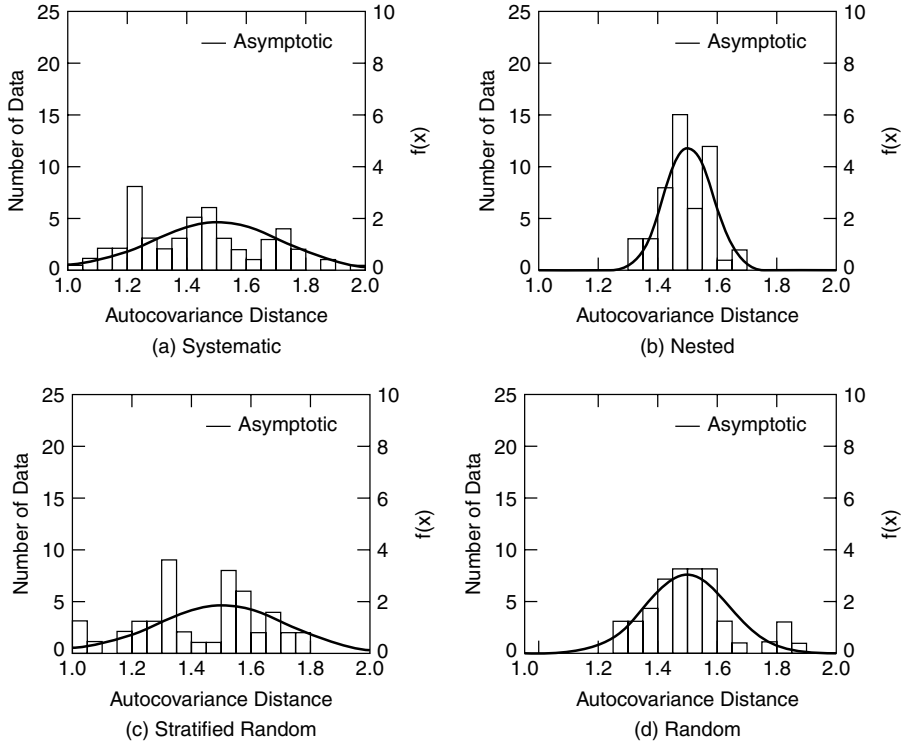


Figure 11.8 Histograms of maximum likelihood estimates of autocovariance distance, r, for various sampling plans, for $n = 64$ points. (DeGroot, D. J. and Baecher, G. B., 1993. "Estimating autocovariance of in situ soil properties.' *Journal of the Geotechnical Engineering Division, ASCE* **119** (GT1), pp. 147–166, reproduced with permission of the American Society of Civil Engineers.)

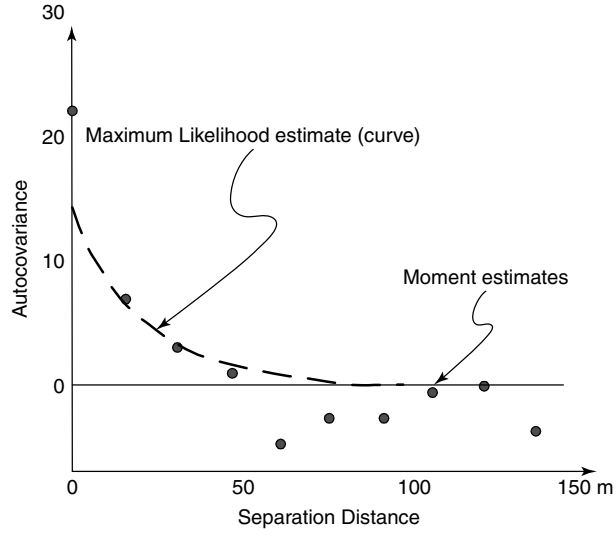


Figure 11.9 Comparison of maximum likelihood estimate of the autocovariance function, and the moment estimate, for data from James Bay (after Christian *et al*. 1994).

In comparing different sampling plans, at least two criteria of optimality might be considered: the quality of the estimates of trend and autocovariance function, and the quality of interpolations using Kriging or equivalent methods. Olea (1984) studied the latter, and concluded that for spatial interpolation with known variograms, systematic spatial samples were superior to other sampling plans in the sense of minimizing the variance of interpolations, and that cluster plans fared the worst. DeGroot studied the former (Figure 11.6), and concluded that for estimating the autocovariance function either nested or cluster plans fared better, and gridded plans fared less well (Figures 11.7 and 11.8). A comparison of the maximum likelihood estimate of autocovariance for *in situ* clay strength data from James Bay (Christian *et al*. 1994) with traditional moment estimates is shown in Figure 11.9.

12 Search Theory

Faced with designing a structure on limestone, one would naturally be concerned about karst. We might have a general idea of the geometry of solution features in the region, and also of the size and depths of cavities that would cause trouble for the structure being planned. How much effort should be invested trying to find such cavities – which, after all, might not exist – and how should that effort be distributed over the site?

This is typical of a class of site characterization problems called, *search*. The problem of search is to locate a geological feature of certain description – usually referred to as an anomaly – in an efficient way, subject to a prior probability distribution on its location, and in the presence of measurement noise. In site characterization, the problem is to maximize the probability of finding anomalies at a cost commensurate with the risk cost of not finding them. These anomalies might be solution features, faults, weak clay layers, high permeability sand lenses, serpentinized joints, or any other unusual geological feature with the potential to affect engineering performance adversely.

The suspicion that adverse geological anomalies exist, and what those anomalies might look like, rests on geological knowledge, as does the prior probability distribution of anomaly locations across the site, that is, their spatial distribution. These suspicions can be expressed as subjective probabilities and evaluated by the methods of Chapter 21 or they might be based on statistical analysis of observations already made. In this chapter, we assume that these evaluations can and have been made.

Throughout, we refer to the anomalies for which we search as *targets*, the distribution of effort in space as *allocation*, and the distribution of effort in time as *strategy*. We consider both *single-stage search*, in which all the available effort is expended in one allocation, and *sequential search*, in which the effort is expended in sequential allocations after each of which the observations are evaluated and probabilities updated. We consider throughout the discussion that the geometries and dimensions of targets are either known, assumed, or described parametrically by a probability distribution.

12.1 Brief History of Search Theory

Search problems are important to geotechnical site characterization, but not surprisingly, they arise in a broad variety of human enterprises, from finding sailors lost at sea (NSRC

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

2000), to finding enemy submarines in war (Stone 1975), to finding mineral resources (Wignall and De Geoffroy 1985), to finding data in databases (Hand *et al.* 2001), to finding consumer products (Barron and Peterson 1976), to finding information in libraries or on the internet (Morse 1970). The literature on search theory is large, and one may even need the rudiments of that theory to find things within its own literature.

The initial major theoretical contribution to rationalizing search procedures came as a result of antisubmarine warfare efforts by the United States military during World War II. The product of many people's work, it was first published by Koopman (1946) as a classified report to the Department of the Navy. The report was declassified in 1958, but selected portions of it had been earlier revised and published in the professional literature (Koopman 1956a, 1956b, 1957). An updated version of the original wartime report was published some years later in the form of an introductory textbook (Koopman 1980). Such military applications of search theory have continued to be developed up to the present (Baston and Garnaev 2000). Stone (1975) published a comprehensive review of the mathematical theory of optimal search developed in the quarter century following World War II, mostly under funding of the US Department of the Navy, and has continued to publish on the topic (Stone 1983; 1989). Other surveys of search theory have been published by Washburn (1989), Dobbie (1968), and Benkoski *et al.* (1991).

12.1.1 Civilian applications of search theory

After World War II, search theory was seen to have application to a host of non-military problems. Gluss (1959) applied the theory to searching for defective components in electronic circuitry; Posner (1963) applied the theory to searching for satellites lost in space; Morse (1970) applied the theory to searching for books in libraries and to establishing policies for new-book shelving practices. Dobbie (1963) and de Guenin (1961) extended search theory to exploration techniques more general than Koopman's, while Charnes and Cooper (1958) modified the basic search problem of Koopman for solution by linear-programming techniques. In recent years, search theory has formed the basis for international land-sea rescue operations (NSRC 2000). A recent primer on optimal search theory from the perspective of search and rescue has been published online by Frost (2000).

The most common optimizing criterion is maximum probability of a find, although a minor amount of work has been based on changes in uncertainty as measured by Shannon's information or entropy (Danskin 1962; Mela 1961). Various reviews and annotated bibliographies of search theory articles up to the mid-1980s have been published (Chudnovsky and Chudnovsky 1989; Enslow 1966; Haley and Stone 1980; Nunn 1981). Benkoski *et al.* (1991) have presented a survey through 1990.

Interest in sequential procedures of search – procedures in which each unit of effort is allocated on the basis of all preceding allocations – began appearing in the literature during the middle 1960s (Black 1962; Charnes and Cooper 1958; Dobbie 1968), and were treated from an entirely different approach than were single-stage allocations. Sequential search is among the sequential decision problems for which analytical solutions can be found.

Starting with the work of the French economist and Nobel laureate Allais (1957), varieties of search theory have been applied to a broad class of problems in mineral and in oil and gas exploration. Applications to mining are many (Brown 1960; Drew 1966;

Engle 1957; Slichter 1955; Wignall and De Geoffroy 1985; Wignall and De Geoffroy 1987). Applications to oil and gas exploration are discussed by Harbaugh *et al*. (1995, 1977). One of the more intriguing applications is to the search for seafloor mineral deposits (Wagner 1979).

12.1.2 Search theory in geotechnology

In the geotechnical literature, Baecher (1972) presented an overview of search theory, and suggested applications to specific issues in site characterization (Baecher 1979a, 1979b). Wu and his students have discussed the nature of geological information in formulating search problems (Wu *et al*. 1996; Wu and Wong 1981), and Tang and his students have worked on geometric models of anomalies and on Bayesian updating of probability distributions of shape, size, and location of anomalies (Tang 1979, 1987; Tang and Halim 1988; Tang *et al*. 1988; Tang and Quek 1986; Tang and Saadeghvaziri 1983), as well as on the influence of anomalies on engineering performance (Tang 1990; Tang *et al*. 1988). Halim (1991) has written a PhD dissertation on the effect of anomalies on geotechnical performance, including the use of search theory to characterize the distribution of anomalies at specific sites. More recently, McGrath, Gilbert, and Tang (Gilbert and McGrath 1997a, 1997b, 1999; Gilbert and Tang 1989; McGrath and Gilbert 1999) used first-order second-moment approaches to search in a Bayesian context.

12.2 Logic of a Search Process

From an analytical perspective, the logic of a search process in the earth sciences is divided into two parts: a geological model describing the character, number, and possible location of anomalies; and an observation model describing the probability of detecting existing anomalies with particular distributions and intensities of effort expended in searching (Baecher 1979b). These combine to yield a likelihood function describing the conditional probability or probability density of the observations actually made, given parameter values of the two models. The likelihood function is subsequently combined with prior probabilities through Bayes' Theorem to draw inferences (Figure 12.1)

$$f_{\mathbf{x}}(\mathbf{x}|data) \propto f_{\mathbf{x}}(\mathbf{x})L(data|\mathbf{x}) \tag{12.1}$$

in which $\mathbf{x}$ is a parameter or set of parameters describing the distribution of anomaly shapes, sizes, orientations, or other attributes.

The most common geological models simplify target features as regular geometric shapes with random size, shape (e.g. obliquity), and location distributions. For example, solution features might be idealized as independently positioned ellipsoids, with spatially variable but preferred orientations, and having some probability distribution over horizontal and vertical location. Clearly, this is a highly idealized view of solution cavities, which, in fact, usually have tortuous shapes, and must be interconnected. Nonetheless, important conclusions are possible even with simple models, as has been demonstrated in various applications of search theory.

Tang and Quek (1986) developed a simple model for use in characterizing the distribution of boulders in sedimentary deposits in Singapore on the basis of boring logs. The

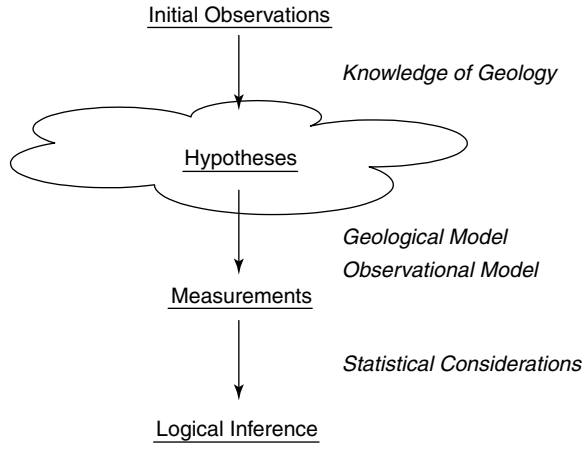


Figure 12.1 Schematic diagram of the logic of search models and inference.

problem is that rock inclusions are encountered in borings which otherwise appear to penetrate a deep stratum of clay (Figure 12.2). These are thought to reflect the presence of large boulders in the formation, but the question is, on the basis of these intersections, can quantitative conclusions be drawn about the size, number, or volume fraction of these boulders?

To address this question, the boulders, which presumably might have quite irregular shapes, were idealized as spheres of unknown diameter. The intersected lengths observed in the boring logs are then random chords through the spheres (Figure 12.3).[1] If the boring may strike the boulder anywhere within the horizontal plane, the likelihood of observing a chord of length l given that the diameter of a boulder is x, is shown to be

$$f_l(l|x) = \frac{2l}{x^2} \quad \text{for } 0 \le l \le x \tag{12.2}$$

Thus, for a given prior probability distribution on boulder diameter, Bayes' Theorem can be used in conjunction with the likelihood of Equation (12.2) to infer a posterior distribution of boulder diameter

$$\begin{aligned} f_X(x|l_1, l_2, \ldots, l_n) &\propto f_X(x) L(l_1, l_2, \ldots, l_n|x) \\ &\propto f_X(x) \prod_{i=1}^{n} \frac{2l_i}{x^2} \end{aligned} \tag{12.3}$$

making the presumption that the intersections are independent of one another, which is not precisely so unless the boulders are widely spaced, but it is perhaps a reasonable approximation.

[1] The exact definition of randomness in generating 'random chords' can cause non-invariance to coordinate transformations. See, for example, discussion of what has become known as the *Bertrand paradox* (Bertrand 1907). As a general rule, the realm of geometric probability is complicated, and such things as coordinate systems can introduce differences in outcomes. For more in depth discussion see Kendall and Moran (1963).

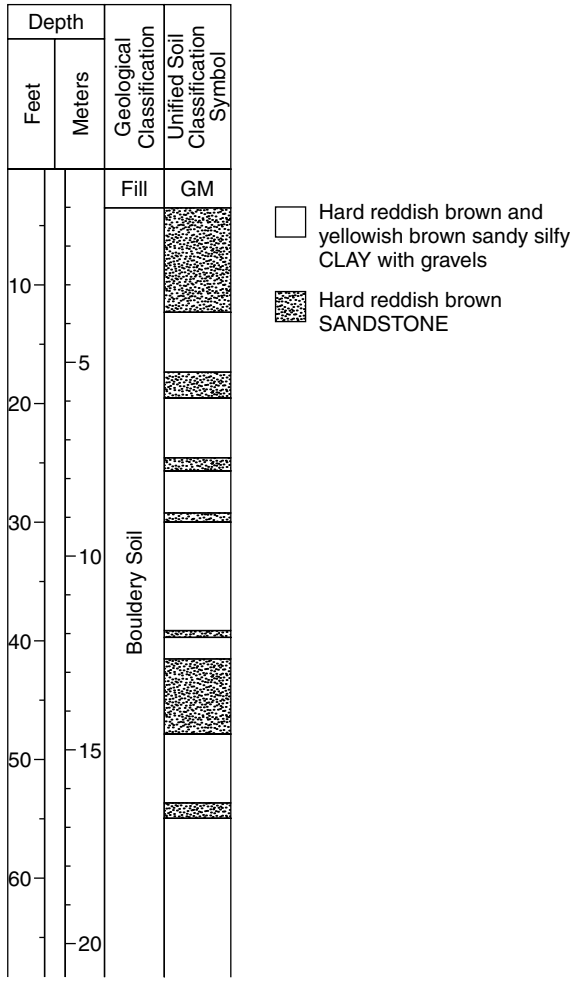


Figure 12.2 Boring log showing intersections with boulders buried in a sandy-silty clay. (Tang, W. H. and Quek, S. T., 1986. 'Statistical model of boulder size and fraction.' *Journal of Geotechnical Engineering, ASCE*, **112** (1), pp. 79–90, reproduced with permission of the American Society of Civil Engineers.)

In a similar way, a number of workers in the minerals industry have used simple two-dimensional shapes such as circles, ellipses, and rectangles to model ore deposits or oil pools for the purposes of search (Drew 1966; Savinskii 1965; Slichter 1955). For example, Figure 12.4 shows clay lenses in a sandy stratum modeled as 2D ellipses. The probability of one of these lenses being intersected by one or more borings is a function of its size, obliquity, and orientation, as has been derived and summarized in tables by Savinskii (1965). A number of more complex but statistically more tractable models based on Poisson flats have been developed by Lippman (1974), Veneziano (1979), Dershowitz *et al*. (1991), and others (see Figure 12.5).

The most common observational models are simple probabilities of detection. For example, Cone Penetration Testing (CPT) might be used in characterizing a site, and

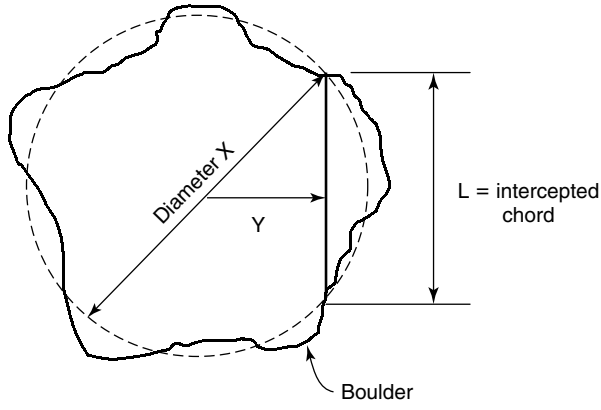


Figure 12.3 Geometry (geological) model of boulder (Tang, W. H. and Quek, S. T., 1986. 'Statistical model of boulder size and fraction.' *Journal of Geotechnical Engineering, ASCE*, **112** (1), pp. 79–90, reproduced with permission of the American Society of Civil Engineers).

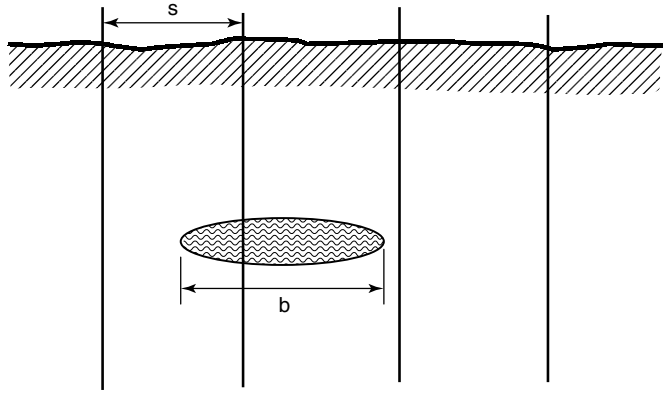


Figure 12.4 Geometric (geological) model of clay lenses as 2D ellipses in a sandy stratum (Baecher 1972).

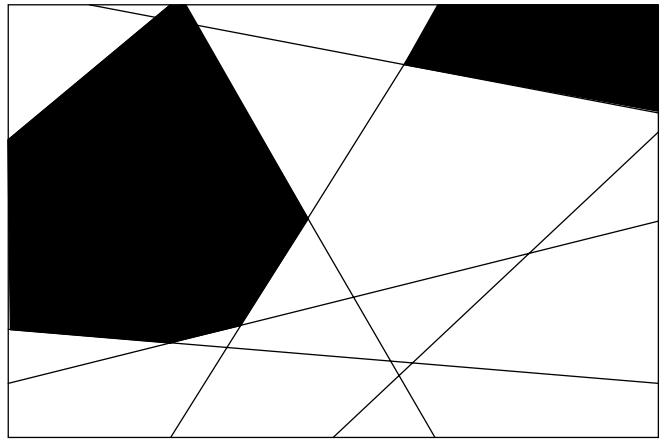


Figure 12.5 Poisson flat model of rock joints (Veneziano 1979).

the question arises whether soft layers are being found with adequate probability. The observation model for this type of search might be a simple conditional probability of detecting a soft layer given that it is intersected by a specific penetration. This probability might be further conditioned on the thickness of the target layer, and on how differentially soft that layer is compared to the rest of the soil formation. As we shall see in the next section, such simple detection models can lead to mathematically interesting and tractable models.

12.3 Single Stage Search

Single-stage search refers to the commitment of the total available search effort in one stage, the outcome of any one unit of effort not being reviewed until the entire effort has been expended. This is a common case in site characterization. A program of exploration is planned, a contract let to perform the exploration, and the results submitted to the engineer in the form of a report.

We shall consider several allocation schemes for the search effort, starting with the analytically simplest and proceeding through allocations that are in some sense optimal. We begin with random search, then discuss the use of grid patterns, and end with allocations that maximize the probability of finding the target when the prior pdf of target location is non-uniform.

Random search is not a technique often used in practice, but it serves as a starting point for considering more complex allocations, and it also provides a bounding case with which to compare other, presumably more efficient, allocations (Tang 1987). *Random search* refers to a procedure that allocates each unit of effort independently of all others. The conditions for random search are: (i) each small element of area within the site has an equal probability of being searched by a unit of effort; (ii) each unit of effort has an equal probability at being located at any elemental area within the site; and (iii) the selection of the location of any one unit of effort in no way influences the selection of the location of any other unit.

The probability of finding a target using random search follows the Binomial distribution since each unit of effort is independent and has the same probability of success. Let the probability of success of any one unit of effort (e.g. a boring) be p. The probability of not detecting the target with any one unit is, therefore, $(1 - p)$. Thus, the probability of not detecting the target with any of n units is

$$\mathrm{P}(\textit{no finds}|n) = (1 - p)^n \tag{12.4}$$

The probability of at least one detection is the complement, or

$$\mathrm{P}(\textit{one or more finds}|n) = 1 - (1 - p)^n \tag{12.5}$$

which for small $p (p \leq 0.1)$ is approximately

$$\mathrm{P}(\textit{one or more finds}|n) \tilde{=} 1 - e^{-np} \tag{12.6}$$

This is compared to a simple linear model in Figure 12.6.

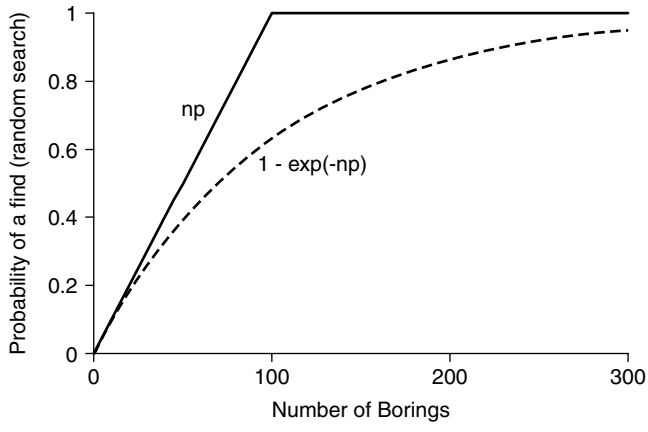


Figure 12.6 Probability of detecting one or more target with random search, $p = 0.01$.

For small values of n, the probability of finding at least one target is approximately linear in n, but as n grows larger an increasing deviation from linearity reflects the growing chance that more than one boring will strike a target, perhaps even the same target. For small exploration effort or for low probabilities of success per unit, random search provides approximately the same probability of a find as do systematic plans. However, as the overall probability of a find rises, the probability of a find with random plans becomes smaller than that with systematic plans.

12.4 Grid Search

Locating exploration effort in a grid pattern is the most common allocation in geotechnical practice, the primary reasons being that grid allocations are easy to specify, and coverage of the site is ensured. Nearly every textbook in applied soil mechanics and engineering geology discusses the appropriate grid spacings for particular structures or projects (Table 12.1).

In search problems, grid allocations are most appropriate for targets lacking preferred locations at a site because they distribute effort uniformly, but they can be used whether or not preferred locations are suspected. In the latter case, however, grid allocations are

Table 12.1 Recommended boring spacings in site characterization for various classes of project (Sowers and Sowers 1979)

	Boring spacing	
Structure or project	ft	m
Highway	1000–2000	300–600
Earth dam	100–200	30–60
Borrow pits	100–400	30–120
Multistory building	50–100	15–30
Manufacturing plants	100–300	30–90

usually sub-optimal (that is, other allocations of the same effort may return a higher probability of finding targets).

12.4.1 Point Grids

Analyzing grid allocations for the probability of finding an object of given geometry and size is generally straightforward. Grids are repeating geometries, and only one cell need be considered in an analysis. Consider first the one dimensional problem of finding an object whose location can be described by the single variable, x.

Let the size of the object in the x-direction be b (Figure 12.7). If the center of the object is located at x, than a find will be made if one point of the exploration grid falls within the interval $[x - \mathrm{b}/2,\ x + \mathrm{b}/2]$. Letting the grid spacing be s, the probability of a find, neglecting boundary conditions, is[2]

$$\mathrm{P}\{find|b, s\} = b/s \tag{12.7}$$

When we expand the problem to two dimensions, the probability of a find becomes dependent on geometry as well as on target size. Consider a circular target of diameter b whose center is randomly located in the (x, y) plane (i.e. each small element of area $\delta x \delta y$ has an equal probability of containing the center at the circle). The exploration grid can be broken down into repeating elements, each containing a boring at its center. If the center of the target falls within a circle of diameter b about a boring, a find will be made (Figure 12.8). If it falls outside of this circle, no find will be made. Since the elementary element is repeating over the zone of search, the probability of finding a target becomes

$$\mathrm{P}(find|b, s) = \frac{\pi^2}{4}\left(\frac{b}{s}\right)^2 \tag{12.8}$$

When $b > s$ the circles within which a find is recorded overlap. This complicates the problem only in that the area within which a find is recorded in each element is more complicated to specify analytically. The solution for square targets follows the same algorithm, except that expectations are now taken over, say, a presumed uniform probability distribution of target orientation (Baecher 1972). Figure 12.9 shows results for circles and arbitrarily oriented squares.

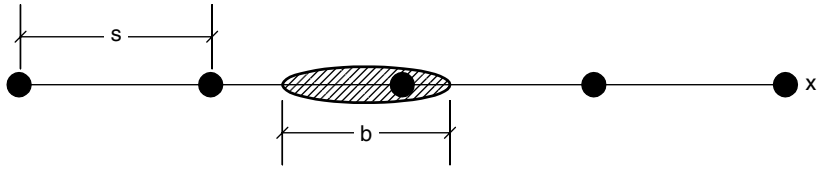


Figure 12.7 Geometry model for one-dimensional search.

[2] We have implicitly assumed the target to be continuous in setting up the problem statement. However, $p(find|b, s)$ depends only upon the length of the intersection of the target with the x-axis and not target continuity.

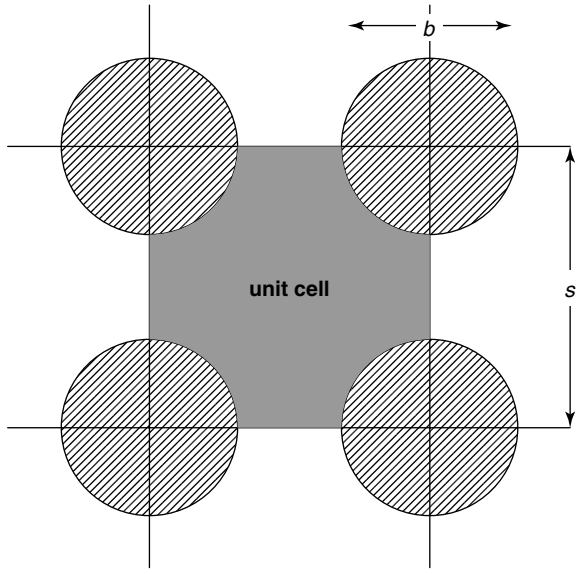


Figure 12.8 Geometry model for two-dimensional search with a circular target of diameter b.

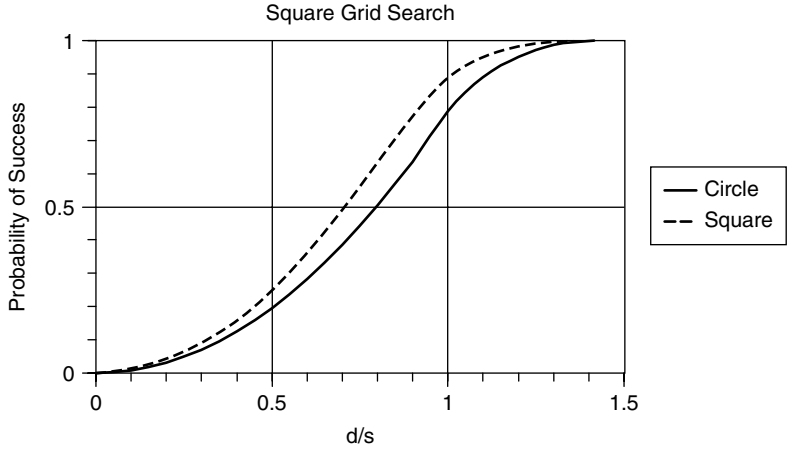


Figure 12.9 Probability of finding circular and square targets with a square grid of borings. Solid lines are analytical solutions; dashed lines are interpolations.

Other target and grid shapes may be treated in a similar fashion, first by determining the elemental circle around each boring for which a find occurs as a function of target orientation, then by taking the expectation over the probability density function of target orientation. This approach applies whether or not the pdf of target orientation is uniform. For example, Slichter (1955) presents results for rectangular targets using line grids (e.g. as used in geophysical searches), and Savinskii (1965) presents results for elliptical targets using rectangular point grids. Similar results may be straightforwardly obtained for other grid shapes, for example circular targets and triangular girds (Figure 12.10), or for oblique

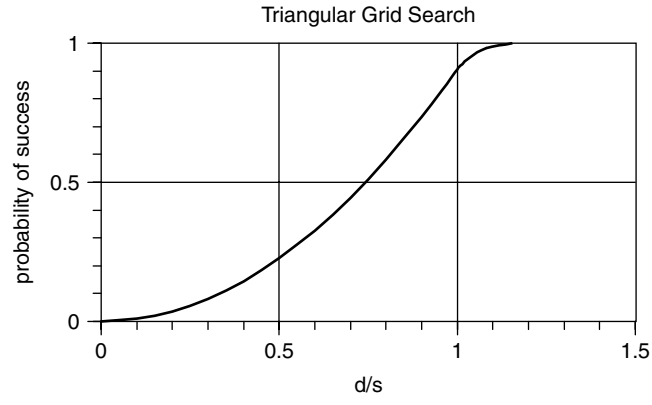


Figure 12.10 Probability of finding circular targets with triangular grids (after Slichter 1955).

target geometries, or any combination of grid and target geometry. For complex target or grid geometries, the Monte Carlo procedures of Chapter 17 provide a powerful and inexpensive way of calculating probabilities of a find.

12.4.2 Line Grids

We have thus far considered only point grids: grids in which the effort is allocated at the intersections of grid lines. In geophysical work, field reconnaissance, and certain other search techniques the effort is allocated along the grid lines themselves. Thus, the allocation is continuous along lines rather than localized at points. The probability of a find using line grids equals the probability that a line intersects the target. We consider only rectangular grids and circular targets to illustrate how the probability of a find is determined.

The probability of a find with a rectangular grid is the probability that the center of the target falls in hatched area of the elemental unit shown in Figure 12.11

$$P\{find|a_1, a_2, b) = \begin{cases} 1 - \left[\dfrac{(a_1 - b)(a_2 - b)}{a_1 a_2}\right] & \text{for } b \le a_2 \le a_1 \\ 1.0 & \text{otherwise} \end{cases} \tag{12.9}$$

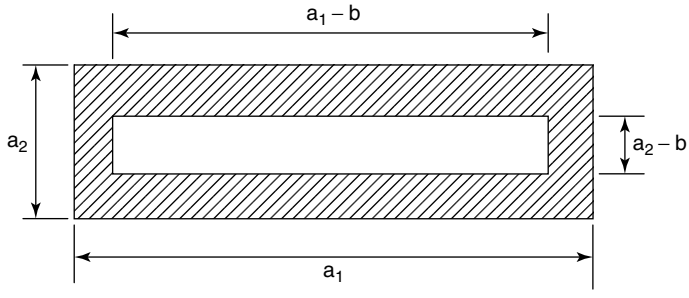


Figure 12.11 Elemental areas for line search with a rectangular line grid and circular target.

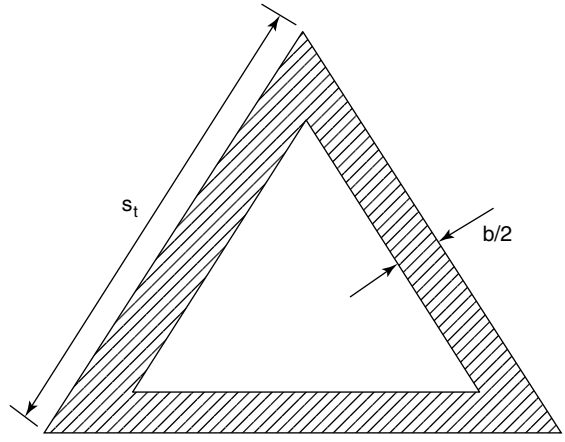


Figure 12.12 Elemental areas for line search with a triangular line grid and circular target obliquity.

The probability of a find with a triangular grid equals the probability that the target center falls into the cross-hatched area of the elemental unit shown in Figure 12.12:

$$P\{find|b, s_t\} = \begin{cases} 1 - \left(\dfrac{s_t - b(Csc\dfrac{\pi}{6} + Sin\dfrac{\pi}{6})}{s_t} \right)^2 & \text{for } b \leq 0.575 s_t \\ 1.0 & \text{otherwise} \end{cases} \tag{12.10}$$

12.4.3 Target Obliquity

Up to this point, we have considered equi-dimensional planar targets (i.e. circles and squares). We now consider targets that have one dimension larger than the other, specifically, ellipses and rectangles. Once the target ceases to be equi-dimensional, target orientation acquires importance, and prior probabilities of target orientation need to be considered.

Let the ratio of the maximum dimension of the target, a, to the minimum dimension, b, be the target obliquity, $\lambda = a/b$. We first concern ourselves with uniform prior probabilities of target orientation, and evaluate the sensitivity of the probability of a find to changes in target obliquity; following this we consider rectangular grids in searching for oblique targets for which some orientation information exists prior to searching. Finally, we consider grids that maximize the probability of a find for a given target and prior pdf of orientation.

The graphs in Figure 12.13 and Figure 12.14 show the relationships between probability of a find and target obliquity for elliptical and rectangular targets with uniform pdf of orientation. The important thing to note here is that once the grid spacing exceeds about 1.5 times the square root of the target area, the probability of a find is insensitive to changes in target obliquity, and the error for spacings greater than 1.5 times the square root of the target area is less than 5%. This means that in most practical applications the assumption of a circular target causes little error. As the grid spacing becomes smaller,

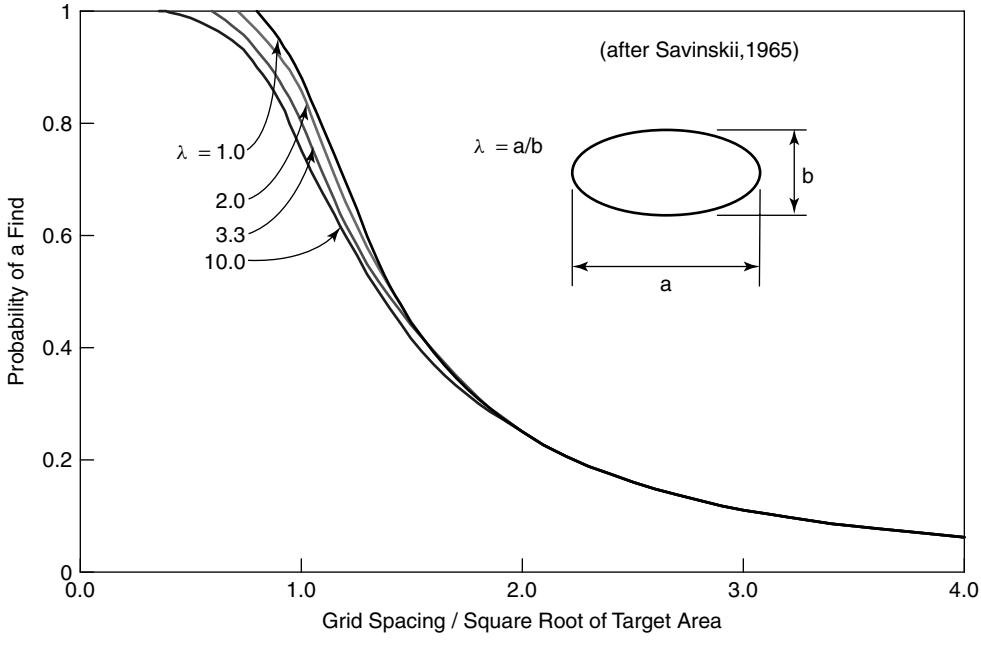


Figure 12.13 Probability of finding elliptical targets with square grids (after Savinskii 1965).

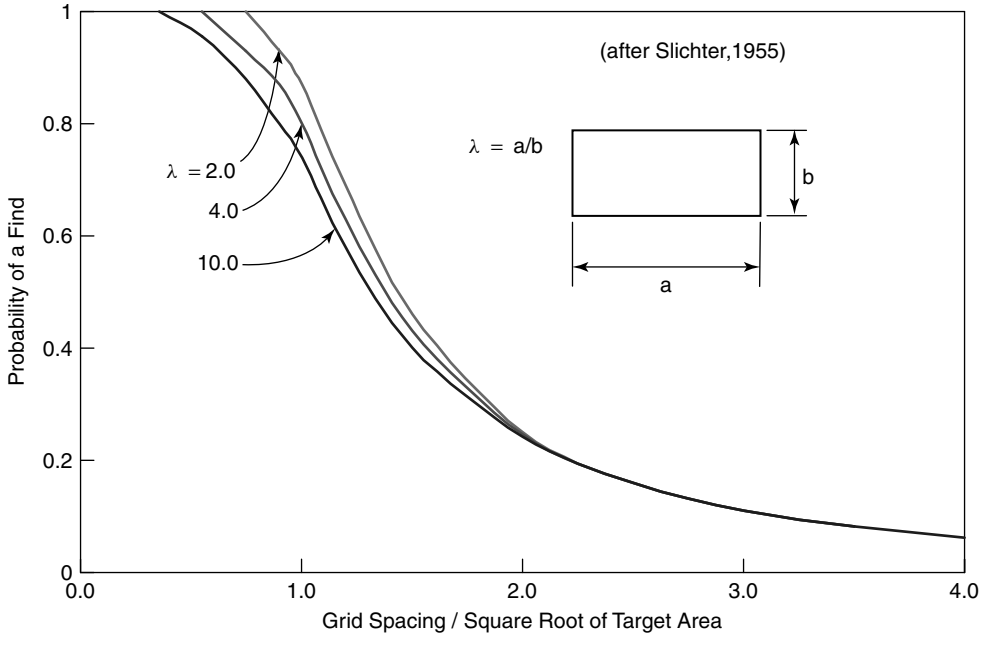


Figure 12.14 Probability of finding rectangular targets with square grids (after Slichter 1955).

however, the actual obliquity of the target is important and must be considered. Errors of 20–25% can occur by neglecting target obliquity in the region of spacings on the order of the size of the target. In geological problems, such situations seldom occur.

In the case that oblique targets have preferred orientation, the probability of a find can be augmented by sizing and orienting the boring grid to take account of this information. As seen in the last two figures, the probability of finding an elliptical target and the probability of finding a rectangular target are essentially the same when the obliquities of the two are the same, so we may restrict attention to the elliptical case with little error.

Consider an elliptical target of dimensions a by b and a rectangular grid, s_1 by s_2. Savinskii (1965) has provided tables of the probability that rectangular grids intersect elliptical targets as a function of grid spacings, target geometry, and target orientation relative to the grid. Baecher (1972) built upon these to calculate the probability of intersecting elliptical targets with orientations described by a probability distribution. While the results of these analyses are beyond the scope of this book, and in any event only results for relatively simple cases have been tabulated, nonetheless useful observations suggest themselves. First, when the preferred target orientation is known, the optimum grid obliquity is approximately proportional to the target obliquity, Secondly, the optimal orientation of the long axis of the grid is parallel to the long axis of the target.

12.5 Inferring Target Characteristics

The numbers and sizes of targets found with a given search effort offers evidence that may be used to refine our assessment of the total number and sizes of targets existing at a site. We first consider the evidence provided by the total number of targets found.

Let the actual number of targets at a site be n, and the number found in a given search with effort Ψ be m. If the prior pmf of n is $p(n)$, the posterior pmf can be found by Bayes' Theorem

$$p(n|m) = \frac{p(n)L(m|n, \Psi)}{\sum_n p(n)L(m|n, \Psi)} \tag{12.11}$$

The probability that one or more targets exist somewhere at the site when no find is recorded is simply $p(\textit{one or more}|m) = 1 - p(0|m)$.

Consider the problem of trying to estimate the number of soft clay lenses interbedded at a site from the records of a boring program. On a regional basis, the prevalence of lens of the type in question is guessed to have an average rate of about λ per some measure of unit area. As a simplification, assume the lenses are randomly and independently distributed in the horizontal plane. Under these assumptions, the number of lenses could be modeled as a Poisson distribution

$$p(n) = \frac{(\lambda A)^n e^{-\lambda A}}{n!} \tag{12.12}$$

in which A is the site area. What inferences can we draw about the total number of lenses if a search has encountered m of them?

To avoid geometric complexities, assume that the vertical dimensions of the lenses are small compared to the deposit depth, and hence the probability of two lenses being at the same elevation and intersecting one another is negligible. We further assume that systematic borings of spacing s, completely penetrating the deposit, have been used. The

likelihood of m finds, from the binomial theorem, is

$$p(m|n) = C_m^n p(find)^m (1 - p(find))^{n-m} \tag{12.13}$$

in which C_m^n is the combinations of n things taken m at a time, and $p(find)$ is found from Figure 12.13. Consider a site area 5000×2000 m searched by a square grid of borings 100 m on center. If the lenses are approximately circular with 50 m diameters, and the prior pmf of the number of lenses is Poisson with $\lambda A = 10$ (i.e. about 10 lenses would be expected at a site this large), then the posterior pmf would be

$$\begin{aligned} p(n|m) &\propto p(n)L(n|m) \\ &\propto \left[\frac{\lambda A^n e^{-\lambda A}}{n!}\right]\left[\frac{n!}{m!(n-m)!} p(find)^m (1 - p(find))^{n-m}\right] \end{aligned} \tag{12.14}$$

for which the prior and posterior pmf's of n are shown in Figure 12.15, for the case of $m = 3$ finds of 50 m lenses with a 100 m boring spacing. Note, this problem could also be formulated by assigning a prior pdf to the spatial density of lenses, λ, and then using the sample observations to draw inferences on the posterior pdf of λ, then forming a predictive pmf on n by integrating the Poisson function over the uncertainty in λ (Chapter 4).

An interesting result of general usefulness is the posterior probability that a target exists undetected. That is, if a certain level of effort has been expended to find a target, but fails to do so, what is the probability that the target yet exists? This is straightforward application of Bayes' Theorem

$$p(\text{target exists}|0) \propto p(\text{target exists})L(0|\text{target exists}, \Psi) \tag{12.15}$$

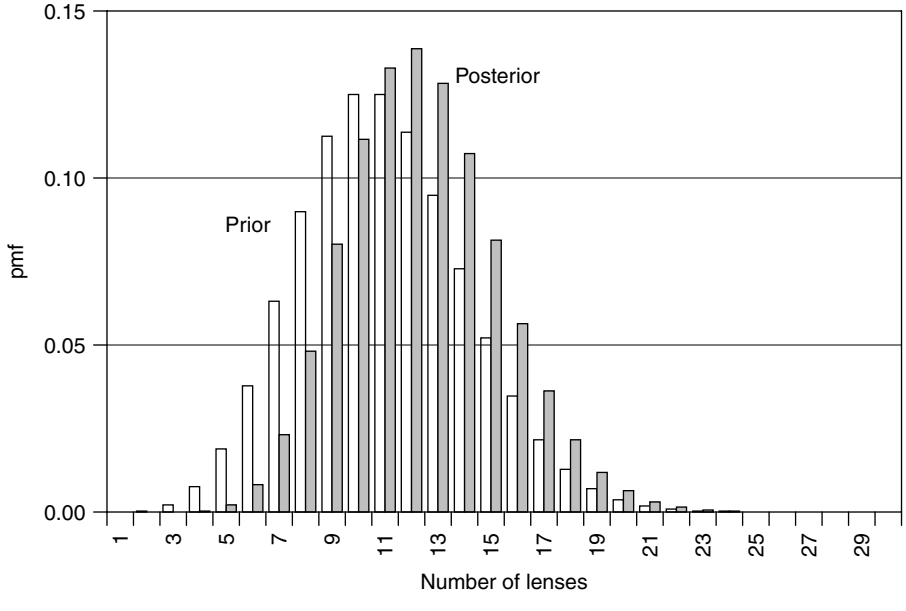


Figure 12.15 Prior and posterior pmf's for total number of 50 foot lenses, given that $m = 3$ have been found with a grid spacing of 100 feet.

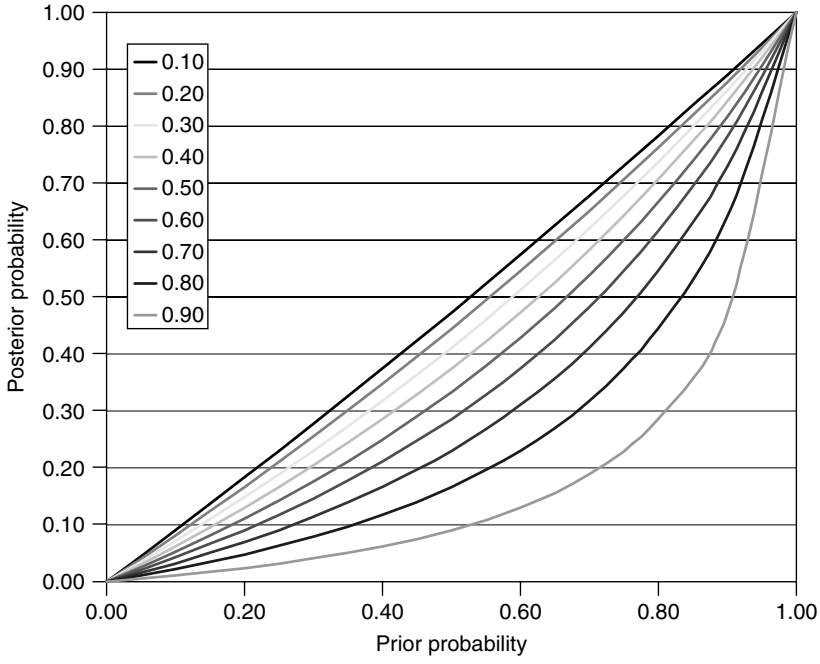


Figure 12.16 Posterior probability of a target existing undetected as a function of the prior probability and the conditional probability of finding an existing target with the search program.

as shown in Figure 12.16 as a function of the prior probability that the target exists. This is an interesting chart. Note that, if the conditional probability of finding a target is zero, the posterior probabilities are the same as the priors, since the search provides no information on the existence of targets. As the probability of finding an existing target increases, that is, the search efficacy increases, the posterior probability of a target existing becomes less and less dependent on the prior probability, at least up to a prior of about 0.6 or 0.7. That is, the search provides more and more information, to the posteriors depend more and more on the search outcomes. Finally, the search efficacy has to be reasonably high, say greater than about 0.7 before the posterior probability is strongly affected by the result of not finding a target.

Were target dimensions not deterministic but themselves represented by some prior pdf, we could compute a posterior pdf of dimension in much the way as we have just treated total number of targets. We do this by first assuming the distribution of b to belong to a family of distributions (Normal, Gamma, etc.), and then record the prior uncertainty about b in the form of a joint pdf on the parameters of its distribution. This uncertainty can be updated in the same way that the pmf of number of targets was updated, and it can be integrated to get the likelihood of the sample results for number of targets.

Tang and his students (Halim and Tang 1993; Halim 1991; Halim and Tang 1990; Halim and Tang 1991; Halim *et al*. 1991; Tang 1979, 1987, 1995; Tang 1990; Tang and Halim 1988; Tang *et al*. 1988a, 1988b; Tang and Quek 1986; Tang and Saadeghvaziri 1983) have used a Bayesian approach to draw inferences of target or anomaly size based on systematic drilling. Suppose that n borings are randomly located within a site of overall area, A_0. The likelihood that exactly m of these borings would intersect a target of area a can be

modeled as a Binomial process as in Equation (12.13), in which $p(find) = a/A_0$. Then, the posterior pdf of target size, $f(a|m, n)$, can be determined from Bayes' Theorem to be

$$\begin{aligned} f(a|m,n) &\propto f(a)L(a|m,n) \\ &\propto k\frac{n!}{m!(n-m)!}(a/A_0)^m(1-a/A_0)^{n-m} \end{aligned} \tag{12.16}$$

in which the prior pdf is assumed uniform, $f(a) = k$. Note that one might also use the non-informative prior $f(a) \propto 1/a$, following Jeffreys (1983), in which case the posterior would differ slightly. For the important special case that $m = 0$, that is, there are no finds

$$f(a|m=0,n) = \frac{(n+1)}{A_0}\left(1-\frac{a}{A_0}\right)^n \tag{12.17}$$

in which $a \leq A_0$, and for which $E[a] = A_0/(n+2)$ and $Var[a] = A_0^2(n+1)/(n+2)^2(n+3)$. For relative size, a/A_0, the corresponding posterior pdf is (Figure 12.17)

$$f\left(\frac{a}{A_0}|m=0,n\right) = (n+1)\left(1-\frac{a}{A_0}\right)^n \tag{12.18}$$

This suggests that should a large number of borings fail to detect a target, any target that is still present must be small. Tang (1987) also presents results for gridded borings and for informative prior pdf's on target size.

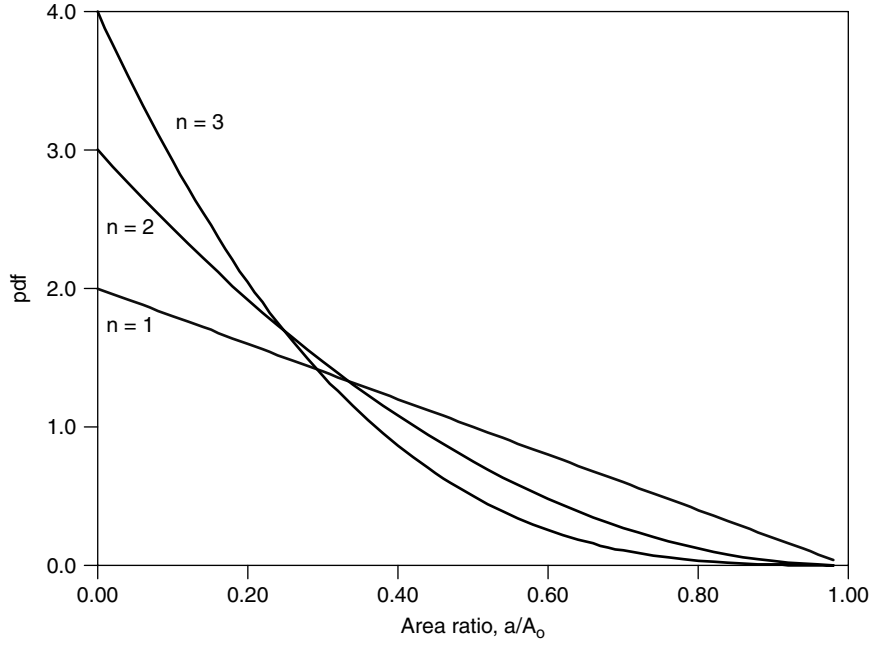


Figure 12.17 Probability distribution over the normalized area of a target, when no finds have been recorded with $n = 1,2,3$ borings.

12.6 Optimal Search

Grid and random search patterns allocate effort uniformly over the region of search. When no prior information exists to indicate likely areas for targets (i.e. the prior probabilities of target location are uniform), there is no reason to allocate effort other than uniformly: A uniform allocation produces the highest probability a find. However, when prior probabilities are not uniform – when some areas are more likely to contain a target than others – non-uniform allocations may increase the probability of a find. These allocations place more effort where a target is more likely to be.

The probability of finding a target in a particular elemental area, A, is the product of the probability that it exists there, and the conditional probability of finding it, given that it exists there,

$$p(\textit{find in } A) = p(\textit{find target exists in } A)\, p(\textit{target exists in } A) \tag{12.19}$$

Allocating more effort to those areas with a higher probability of containing a target – and thus less effort to those areas with a lower probability – increases the term $p(\textit{find}|\textit{target exists in } A)$ in those areas where *p(target exists in A)* is greatest.

12.6.1 Detection functions

The *detection function* is the conditional probability of finding a target as a function of the amount of effort expended

$$\textit{detection function} \equiv p(\textit{find}|\textit{target exists in A, search effort expended}) \tag{12.20}$$

For grid allocations, the detection function was assumed to be constant (i.e. if a target existed at a grid point it was assumed to be detected with some fixed probability, p). When the prior probability distribution of location is uniform, the exact value of p has no effect on the optimal allocation of effort.

Borings are discrete quantities of effort, and their density of allocation as a function of area or length is a series of points. Correspondingly, the conditional probability of finding a target, given that it exists at point x, is a series of spikes (Figure 12.18). For search segments or areas that are large relative to the boring spacing,[3] the allocation of effort over x (where X is the space of the search) can be expressed as

$$\phi\{x\} \propto s^{-k} \tag{12.21}$$

in which $\phi\{x\}$ is the allocation of effort at x, and k is the dimension of the search (i.e. $k = 1$ for search over a line, $k = 2$ for search over an area, etc.). Because the conditional probability of a find with systematic observations is proportional to s^{-k}, the detection function becomes

$$p(\phi\{x\}) \propto \phi\{x\} \tag{12.22}$$

[3] The requirement of large elemental area relative to boring spacing avoids complications of boundary effects.

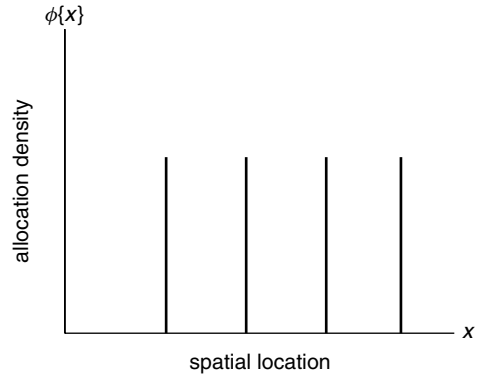


Figure 12.18 Allocation of borings over a line, and conditional probability of finding a target at location x.

Thus, systematic observations such as boring grids obey a *linear detection function*. From Equation (12.20), the probability of finding a target is

$$\begin{aligned} f(\textit{find}) &= \int f_X(x) N \varphi\{x\} dx \\ &= N \int f_X(x) \varphi\{x\} dx \end{aligned} \tag{12.23}$$

in which N, a normalizing constant, is in turn maximized by

$$\varphi\{x\} = \begin{cases} \Psi & x \text{ such that } f_X(x) \text{ is maximal} \\ 0 & \text{otherwise} \end{cases} \tag{12.24}$$

and Ψ, the total effort, is a constant. In other words, by placing all the effort in those regions having the highest probability of containing a target – and no effort in other regions – the overall probability of a find is maximized.

12.6.2 Exponential saturation detection functions

When the prior probability distribution of location is not uniform, imperfect detection assumes greater importance. Most detection functions encountered in practice exhibit diminishing returns. That is, allocations of increasingly large amounts of exploration effort do not produce proportionally large increases in the conditional probability of a find. In this case, the optimum amount of effort allocated in a particular area is a function both of the probability of the target existing there and of the rate of diminishing returns in detecting a target.

Let the prior probability distribution of target location be $f_X(x)$. We assume that the target exists somewhere, that is, the integral of $f_X(x)$ over the complete space is 1.0, $\int_X f_X(x)\, dx = 1$. The only restriction on $f_X(x)$ is that it be continuous except for finite jumps. As above, the amount of search effort allocated at x is represented by the function

$\phi\{x\}$, such that $\phi\{x\} \geq 0$ for all x. The integral of $\phi\{x\}$ over all x is the total search effort, $\int_X \phi\{x\}dx = \Psi$.

Consider the one-dimensional case of search along a line, X. Let the probability of detecting a target – known to exist within some interval $(x, x + dx)$ – with one quantum of search effort, be p. If n quanta are used, the probability of detection increases according to

$$p(\textit{find}|\textit{target exists}) = 1 - (1 - p)^n \tag{12.25}$$

which, for p smaller than about 0.1, is approximately

$$p(\textit{find}|\textit{target exists}) \cong 1 - e^{-np} = 1 - e^{-\varphi\{x\}} \tag{12.26}$$

Koopman (1946) called this the *exponential saturation detection function*.

12.6.3 Optimal one-dimensional search

If we combine the detection function of Equation (12.26) with the, possibly non-uniform, prior probability distribution of target location, $f_X(x)$, the probability of detecting a target within the interval $(x, x + dx)$ becomes

$$p(\textit{detecting target in}(x, x + dx)) = f_X(x)[1 - e^{-\varphi\{x\}}] \tag{12.27}$$

and the probability of a find anywhere is the integral over the search space, X

$$p(\textit{find}) = \int_X f_X(x)[1 - e^{-\varphi\{x\}}]\,dx \tag{12.28}$$

The optimal allocation of search effort $\phi^*\{x\}$ maximizes $p(\textit{find})$ while being everywhere non-negative and integrating to the total effort, $\int_X \phi\{x\}dx = \Psi$. Koopman shows that the function $\phi^*\{x\}$ maximizing Equation (12.28) satisfies the condition

$$f_X(x)e^{-\varphi\{x\}} = \lambda \tag{12.29}$$

in which λ is a constant. Taking the logarithm of both sides gives

$$\phi\{x\} = \ln f_X(x) - \ln \lambda \tag{12.30}$$

and integrating over x gives

$$\int \phi\{x\}dx = \Psi = \int [\ln f_X(x) - \ln \lambda]dx \tag{12.31}$$

in which Ψ is the total search effort.

Koopman noted that Equation (12.31) suggests a simple graphical procedure by which to determine $\phi^*\{x\}$. This is carried out by graphing the logarithm of $f_X(x)$, and drawing in a horizontal line of height ln (λ) (Figure 12.19). The difference between these two under optimal allocation, from Equation (12.31), equals the total search effort Ψ. The line ln (λ)

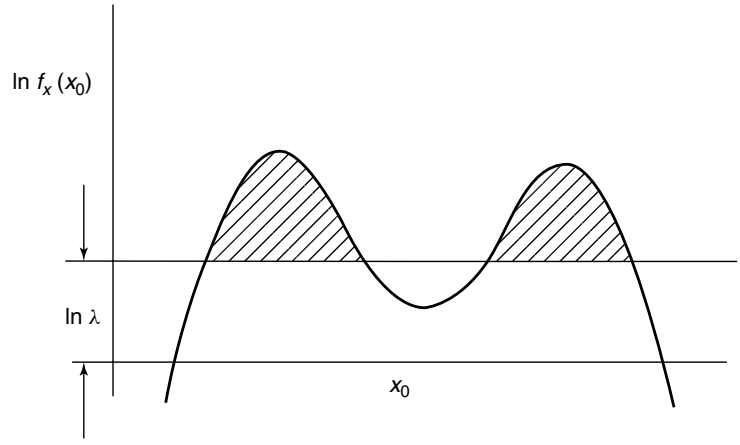


Figure 12.19 Graphical procedure for determining the optimal allocation $\varphi * \{x\}$.

is moved up or down until the area between the curves satisfies this criterion. Transforming the logarithmic ordinate to arithmetic yields the optimal allocation of effort, $\phi^*\{x\}$.

The probability of finding a target given that it exists can be found by substituting terms into Equation (12.28) to obtain

$$p(\textit{find}) = \int [f_X(x) - \lambda] dx \tag{12.32}$$

This shows that the probability of a find is the area between the curves $f_X(x)$ and λ (Figure 12.19).

In the event that the target is not found by the optimal allocation of effort, the posterior distribution of target location can be determined from Bayes' Theorem

$$f_X(x|\textit{no find}) \propto f_X(x) L(\textit{no find}|\textit{targest exists at} x) \tag{12.33}$$

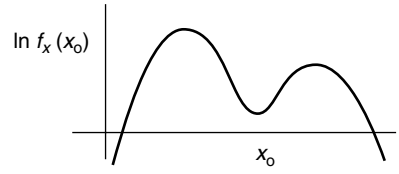
in which the probability of *no find* given that the target exists at x is the likelihood. After some manipulation, this posterior pdf can be shown to be (Baecher 1972)

$$f_X(x|\textit{no find}) = \begin{cases} \dfrac{f_X(x)}{(1-P)} & \text{for } x \; s.t. \; \varphi\{x\} > 0 \\ \dfrac{\lambda}{(1-P)} & \text{for } x \; s.t. \; \varphi\{x\} = 0 \end{cases} \tag{12.34}$$

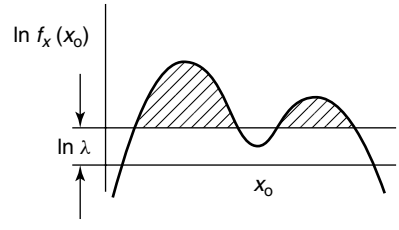
in which P is the probability of finding the target with the optimal allocation. Thus, the posterior pdf of location for any x is proportional to the distance between the x-axis and either the curve $f_X(x)$ or the curve λ, whichever is smaller, by the factor $(1-p)^{-1}$. Successive allocations of effort without a find yield posterior pdf's of location that are found by successively lowering the horizontal line representing λ (Figure 12.20).

Consider a large reconnaissance effort in which exploration is undertaken to route a gas pipeline across a 9 km-long by 1 km-wide section of terrain, and in which the objective of the program is to identify potentially troublesome geological conditions that might hinder

Step 1:Graph the natural logarithm of the prior pdf, $\ln f_X(x)$

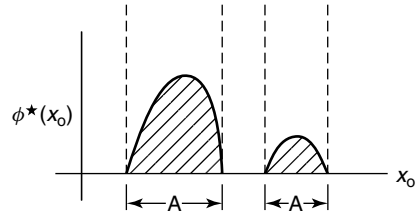


Step II: Draw a line parallel to the x-axis truncating the graph of $\ln f_X(x)$. This line is a plot of $\ln\psi$, where ψ is the total search effort



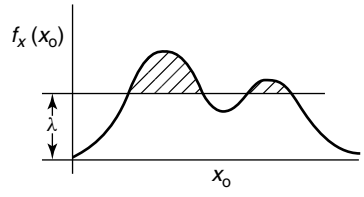
(a)

Step III: Determine the area under the curve in $\ln f_X(x)$ and above the line $\ln\lambda$. This area is the RHS of Equation Equation 12-31. Under the optimal allocation this should equal the total search effort, ψ. By moving the horizontal line up and down (i.e. by varying λ), this area can be made to equal ψ) and the value of λ thus determined satisfies the criterion for optimal allocation of equation

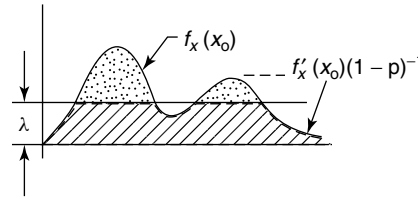


(b)

Step IV: The probability of finding the target given that it exist somewhere at the site is just the area between the curves $f_X(x)$ and λ.



Step V: In the event the target is not found, the posterior probability distribution of target location is proportional to the distance between the *x*-axis and either $f_X(x)$ or λ, whichever is less.



(c)

Figure 12.20 Step-by-step graphical procedure for determining the optimal allocation $\varphi * \{x\}$.

the project. About 500 person-days of effort have been allocated to the program (i.e. two person-years), and as a rough estimate, the field teams can investigate an area of about 100 m by 100 m per person per day. Thus, the total available effort (denominated in units of site area) is

$$\Psi = 500 \text{ days} \times (100 \times 100 \text{ m}^2 \text{ day}) = 5 \text{ km}^2 \tag{12.35}$$

We presume the probability of finding troublesome conditions follows an exponential saturation relationship with respect to level of effort. How should the 500 days – equaling 5 km^2 of effort – best be allocated over the 9 km^2 of potential right-of-way?

To begin, the right-of-way is divided into nine, one-km long by one-km wide sections. Subjective probabilities are assessed (Chapter 21) to establish a prior probability distribution (pmf) of where difficult geological conditions might be expected to occur. That is, some sections are likely to be more problematic than others – based on regional geology and perhaps other experience in the region – and it would make sense to allocate more effort in those subsections than in others. This prior pmf is shown in Figure 12.21, and the natural logarithms of the pmf in Figure 12.22. Trial values of $\ln(\lambda)$ are chosen, beginning at -1.0 and moving down to -4.0, and the answers interpolated to find the optimal $\lambda \approx -2.85$ (Figure 12.23). Transforming from logarithms to arithmetic scale, the optimal allocation of effort is shown in Figure 12.24. The allocation that maximizes the chance of finding unsuitable conditions is that which devotes effort to subsections 2,3,4,7, and 8; while ignoring sub-sections 1,5,6, and 9. The reasons for ignoring the latter sub-sections in the search are that the prior probability of problematic conditions existing in them is simply too low and that sufficiently diminished returns are not reached in the more likely subsections with the total amount of available effort.

12.6.4 Optimal search in higher dimensions

Koopman's theory of optimal search is readily extended to higher dimensions by substituting a multivariate pdf of location for $f_X(x)$. The construction to determine $\phi^*\{\boldsymbol{x}\}$, in which

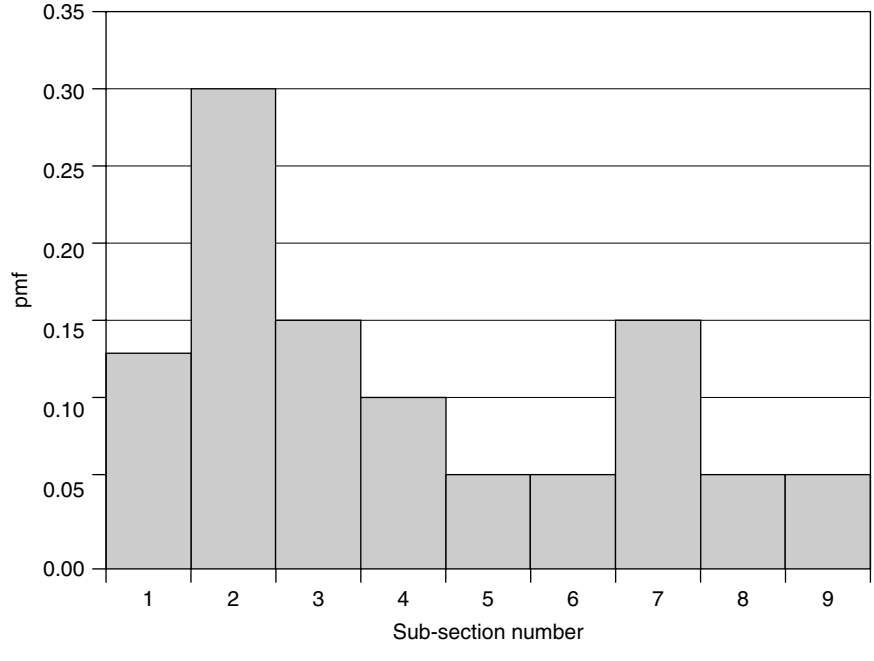


Figure 12.21 Prior pmf of potentially difficult geological conditions along the proposed right-of-way of a new gas pipeline.

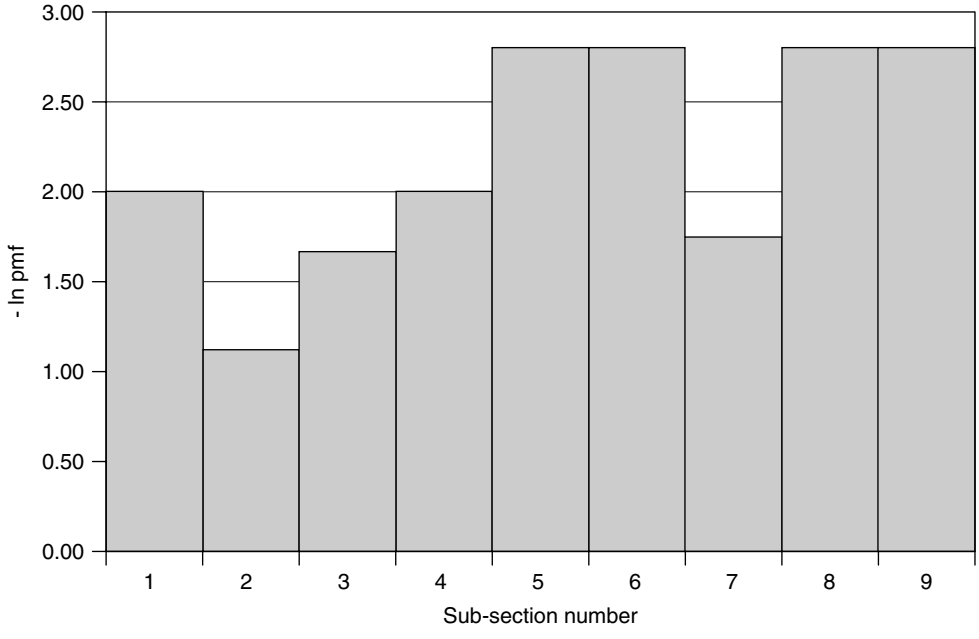


Figure 12.22 Natural logarithms of the prior pmf of potentially difficult geological conditions along the proposed right-of-way of a new gas pipeline.

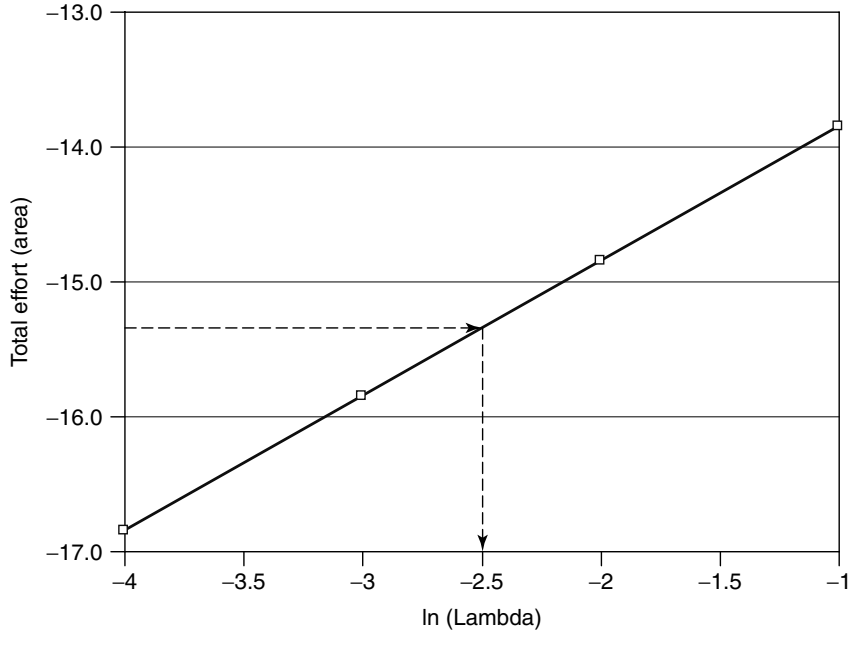


Figure 12.23 Natural logarithms of the prior pmf of potentially difficult geological conditions along the proposed right-of-way of a new gas pipeline.

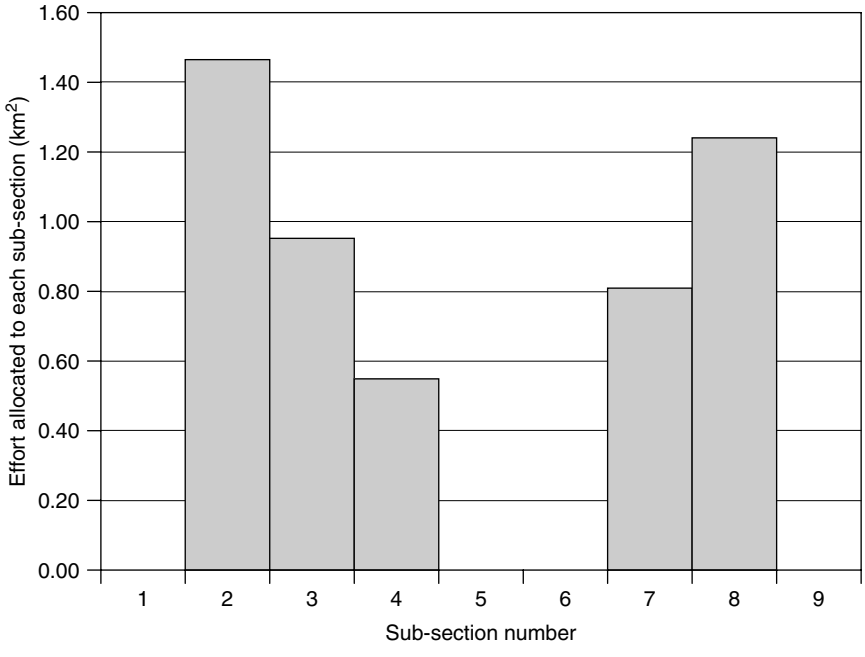


Figure 12.24 Optimal allocation of reconnaissance effort.

$\boldsymbol{x}$ is a vector, is analogous to the construction already considered. In the two-dimensional case, the volume between the surfaces $f_{X_1X_2}(x_1, x_2)$ and $\ln(\lambda)$ is varied by changing the value of the constant λ until this volume equals the total search effort. The optimum is then to allocate effort only to those elemental areas for which $f_{X_1X_2}(x_1, x_2) > \ln\lambda$, and to expend effort at each in proportion to

$$\varphi^*\{x_1, x_2\} \propto |\ln f_{X_1X_2}(x_1, x_2) - \ln\lambda| \tag{12.36}$$

Thus, the construction in multi-dimensions is strictly analogous to the one-dimensional case.

Koopman's theory is limited by the assumption of exponential saturation of the detection function, which only applies to search techniques with small conditional probability of finding a target. This assumption is violated in many geological exploration problems. de Guenin (1961) extended Koopman's theory to cover any detection function displaying diminishing returns, that is, all detection functions for which $dP(\phi\{x\})/d\phi\{x\}$ is a decreasing function of $\phi\{x\}$.

De Guenin shows that a more general statement of Equation (12.29) for optimal allocation is

$$f_X(x)\frac{dP(\varphi^*\{x\})}{d\varphi\{x\}} = \lambda \tag{12.37}$$

which in the exponential case reduces to Equation (12.29). Based on this modification, de Guenin proposes an iterative graphical procedure, somewhat more involved than Koopman's, by which the optimal allocation can be determined.

In some cases more than one target type may be of interest. For example, at a dam site both faulting and weathered zones unrelated to faulting may be of concern. Optimal search theory can be applied to such problems by restating the objective not as maximizing probability of a find, but maximizing expected utility (Chapter 5). There are utilities associated with finding each type target and these utilities are usually not the same. Furthermore, the detection function for each target typically differs. In such cases one seeks an optimal allocation that maximizes the expected total utility of search. Baecher (1972) considers this problem further.

12.7 Sequential Search

A sequential search procedure is one in which the pdf of location is made current and a new optimal strategy determined after each increment of effort is allocated, as opposed to single stage procedures in which all of the effort is expended before updating. Since each allocation is made on the basis of all past information, sequential procedures frequently produce greater probabilities of finds than do single-stage procedures. However, the increased probability of a find may not be sufficient to compensate for discontinuities in field work and for the increased effort required by constant updating.

One may formulate the problem of sequential search in a manner similar to the single-stage problem already considered. We are searching for a target that is thought to exist somewhere along a line (extensions to higher dimensions are straightforward), and the prior pdf of the target's location will be denoted by $f_X(x)$. We break from the single-stage formulation in that we specify that discrete quantities of effort are expended in searching. These discrete quantities may be located at any point along the continuous x-axis within the region of search. We once again assume that the search technique is not perfectly efficient, that there exists a positive conditional probability of the target being missed in a search of location x_i, even though the target exists at x_i. This conditional probability of a miss is denoted as before

$$\alpha_i = \Pr(\text{miss at } x_i | \text{target at } x_i) \tag{12.38}$$

We seek a strategy of search, that is, a rule telling us where to look at each stage that is in some sense optimal. To judge the goodness of each strategy we consider four criteria commonly of interest in site exploration: (1) minimizing the expected amount of effort required to find the target; (2) maximizing the probability of finding the target with a given amount of effort; (3) minimizing the conditional probability that the target exists somewhere at the site when no find is made; and (4) minimizing the expected amount of effort required to lower the conditional probability that the target exists to a specified level. It can be demonstrated that each of these criteria leads to the same optimal strategy (Baecher 1972). The importance of this property is that a unique optimum strategy exists.

The subject of optimal sequential decisions forms a major component of the theory of statistical decisions. The basic functional equations for the problem of sequential search are difficult to solve directly (De Groot 1970; Stone 1975), but for several cases the optimal strategy can be found by other analytical means (Chew 1967). Without loss of

generality, since each of the criteria above leads to the same optimal strategy, we consider optimality in the sense of the first criterion, minimizing the expected effort to find a target.

For any strategy δ, let the cost of the search at the nth stage be c_n and the probability that the target is found at the nth stage be λ_n. Assume that the cost of searching location x_i is c_i. The probability of finding the target at location x_i is

$$\Pr(\text{find at } x_i) = f_X(x_i)(1 - \alpha_i)dx \tag{12.39}$$

The expected cost of finding the target using strategy δ is

$$\begin{aligned} E[C(\delta)] &= \sum_{n=1}^{\infty} c_n \Pr(N \geq n) \\ &= c_1 + c_2(1 - \lambda_1) + c_3(1 - \lambda_1 - \lambda_2) + \ldots . \end{aligned} \tag{12.40}$$

where N is a random variable equal to the number of searches until a find is made. For the present we only consider prior distributions for which the integral of $f_X(x_i)$ is 1.0 (i.e. the target exists somewhere), since the expected value of cost is infinite when there is a positive probability of continuing the search indefinitely.

One can show that the strategy δ^* that minimizes $E[C(\delta)]$ must satisfy the relation that $\lambda_{n+1}/c_{n+1} < \lambda_n/c_n$ (De Groot 1970). The proof of this assertion is easily demonstrated. Suppose δ is a strategy for which $E[c(\delta)] < \infty$, but for which $(\lambda_n/c_n) < (\lambda_{n+1}/c_{n+1})$ for some value of n. Let δ^* be the strategy that is similar to δ at every stage, except that stages n and $(n + 1)$ are reversed. Then

$$\begin{aligned} E[C(\delta)] - E[C(\delta^*)] &= c_n(1 - \lambda_1 - \cdots - \lambda_{n-1}) \\ &\quad + c_{n+1}(1 - \lambda_1 - \cdots - \lambda_n) \\ &\quad - c_{n+1}(1 - \lambda_1 - \cdots - \lambda_{n-1}) \\ &\quad - c_n(1 - \lambda_1 - \cdots - \lambda_{n-1} - \lambda_{n+1}) \\ &= c_n\lambda_{n+1} - c_{n+1}\lambda_n > 0 \end{aligned} \tag{12.41}$$

Thus, the expected total cost for δ^* is lower than for δ.

The optimal strategy, therefore, specifies that at any stage in the search, the next location to be searched is the one that maximizes the probability of a find per unit cost. A sequential strategy like this, which at each stage specifies the short-term best decision is said to be *myopic*. In the case for which all the c_i's are equal, which is not uncommon in site investigation, the optimal strategy specifies searching the location with the highest probability of a find $f_X^{(n)}(x_i)(1 - \alpha_i)$, where at any given stage, n, $f_X^{(n)}(x)$ is the current posterior pdf of target location.

Part III

13 Reliability Analysis and Error Propagation

Reliability analysis deals with the relation between the loads a system must carry and its ability to carry those loads. Both the loads and the resistance may be uncertain, so the result of their interaction is also uncertain. Today, it is common to express reliability in the form of a *reliability index*, which can be related to a *probability of failure*. It should be understood in this context that 'failure' includes not only catastrophic failure – as in the case of a landslide – but also, in Leonards' (1975) phrase, any "unacceptable difference between expected and observed performance."

13.1 Loads, Resistances and Reliability

The loading to which an engineering system is exposed is Q. The available resistance is R. In this context, 'loading' and 'resistance' must be taken in their broadest sense; that is, they include not only forces and stresses, but also seepage, settlement, and any other phenomena that might become design considerations. The values of both R and Q are uncertain, so these variables have mean or expected values, variances, and covariances, as well as other statistical descriptors that will be ignored for the present. The notation is as follows:

μ_R, μ_Q:	mean values of R and Q, respectively.
$E[R], E[Q]$:	expected values of R and Q, respectively, $= \mu_R, \mu_Q$.
σ_R, σ_Q:	standard deviations of R and Q, respectively.
Ω:	coefficient of variation $= \sigma/\mu$.
σ_R^2, σ_Q^2:	variances of R and Q, respectively, also denoted $Var[Q], Var[R]$
ρ_{RQ}:	correlation coefficient between R and Q.
$Cov[R, Q]$:	covariance of R and $Q = \rho_{RQ}\sigma_R\sigma_Q$.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

The *margin of safety*, M, is the difference between the resistance and the load:

$$M = R - Q \tag{13.1}$$

From the elementary definitions of mean and variance, it follows that *regardless of the probability distributions of R and Q* the mean value of M is

$$\mu_M = \mu_R - \mu_Q \tag{13.2}$$

and the variance of M is

$$\sigma_M^2 = \sigma_R^2 + \sigma_Q^2 - 2\rho_{RQ}\sigma_R\sigma_Q. \tag{13.3}$$

A *reliability index*, β, is defined as

$$\beta = \frac{\mu_M}{\sigma_M} = \frac{\mu_R - \mu_Q}{\sqrt{\sigma_R^2 + \sigma_Q^2 - 2\rho_{RQ}\sigma_R\sigma_Q}} \tag{13.4}$$

which expresses the distance of the mean margin of safety from its critical value ($M = 0$) in units of standard deviation. If the load and resistance are uncorrelated, the correlation coefficient is zero, and

$$\beta = \frac{\mu_M}{\sigma_M} = \frac{\mu_R - \mu_Q}{\sqrt{\sigma_R^2 + \sigma_Q^2}} \tag{13.5}$$

Figure 13.1 shows plots of typical probability distributions of R and Q. Figure 13.2 (a) shows the resulting probability distribution of M. The probability of failure must be the probability that M is less than 0.0, which is the shaded area in the figure.

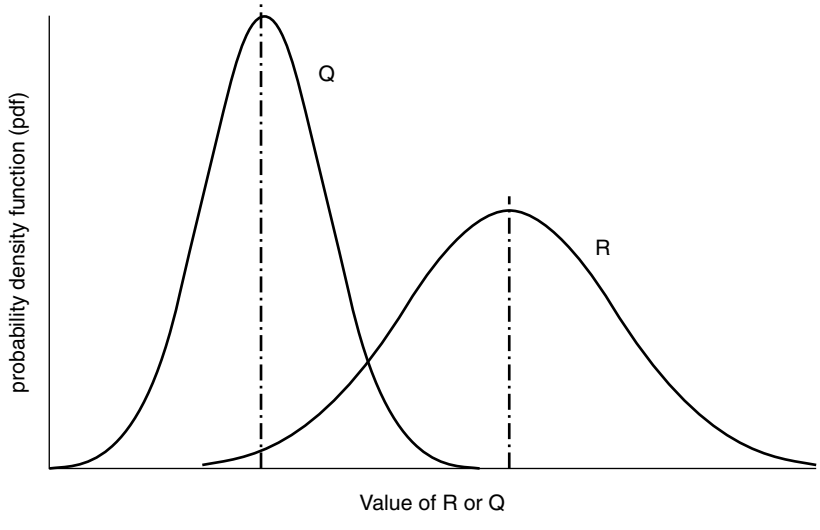


Figure 13.1 Probability densities for typical resistance (R) and load (Q).

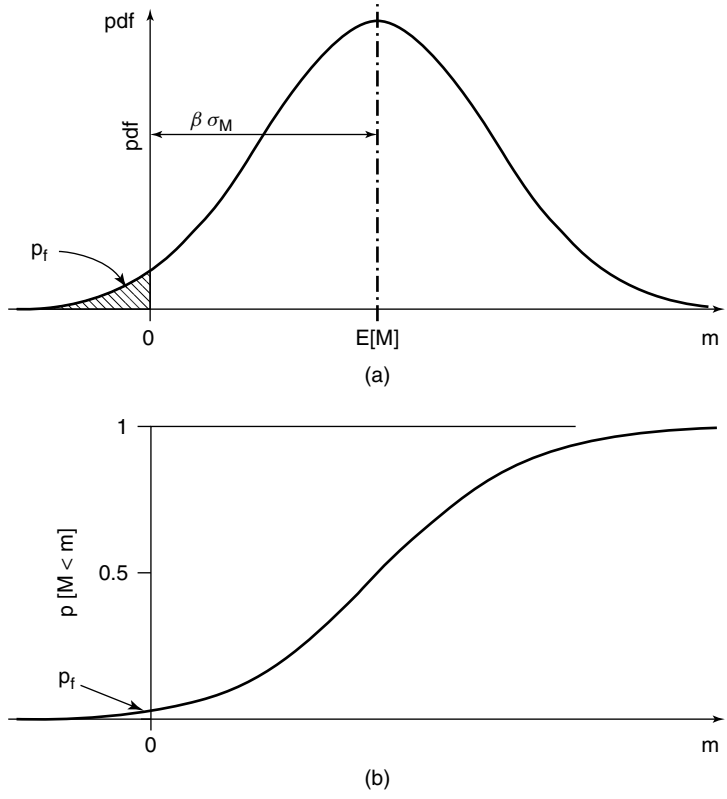


Figure 13.2 Probability density (a) and cumulative probability (b) for margin (M). Note that the area in (a) under the curve and to the left of the axis is the probability of failure identified in (b).

Figure 13.2 (b) shows the cumulative distribution corresponding to the distribution function of Figure 13.2 (a). The probability of failure is the intercept of the cumulative distribution function with the vertical axis at $M = 0$.

In the special case that R and Q are Normally distributed, M is Normally distributed as well. Thus, the reliability index, β, which normalizes M with respect to its standard deviation, is a standard Normal variate, usually designated Z. The standardized Normal distribution, with zero mean and unit standard deviation, is widely tabulated (Abramowitz and Stegun 1964; Burington and May 1970), and modern spreadsheets include it as a function. Usually, the tabulation expresses the integral Φ of the standardized Normal distribution between $-\infty$ and positive values of the parameter Z. We want the probability of failure p_f, which is the integral between $-\infty$ and values of the parameter Z located below the mean value, that is, negative values of Z. Due to the symmetry of the Normal distribution, p_f is simply $1 - \Phi(\beta) = \Phi(-\beta)$.

Geotechnical engineers are more accustomed to working with the factor of safety, F, defined by

$$F = \frac{R}{Q} \tag{13.6}$$

Failure occurs when $F = 1$, and a reliability index is defined by

$$\beta = \frac{E[F] - 1}{\sigma_F} \tag{13.7}$$

Calculations of the reliability index are more difficult when it is expressed in terms of the factor of safety because F is the ratio of two uncertain quantities while M is their difference. To avoid this problem, some researchers have assumed that R and Q are logNormally distributed (that is, the logarithms of R and Q are Normally distributed) so that the logarithm of their ratio becomes the difference between their logarithms. Then, the formulation is identical to Equations (13.1)–(13.5), but numerical results for a given problem will be different. Also, in estimating the statistical parameters of R and Q from experimental and field data, calculations must be made on the logarithms of the data rather than the arithmetic values.

13.2 Results for Different Distributions of the Performance Function

Either M or F can describe the performance of a geotechnical structure, so either will be called the *performance function*. Table 13.1 gives values of the probability of failure for several distributions of the performance function and for a range of values of the reliability index. The same information is plotted in Figure 13.3. The 'Normal' case applies when M is Normally distributed. It also applies when the logarithm of F is Normally distributed and the standard deviation of the logarithm of F can be calculated. The 'Triangular' case applies when M has a symmetric triangular distribution. The 'logNormal' cases arise when β is defined by Equation (13.7) in terms of F, but F is logNormally distributed. The relation between β and the probability of failure is then not unique but depends on

Table 13.1 Probability of failure for various distributions of performance function

Reliability index	Probability of failure				
	Normal distribution	Triangular distribution	LogNormal distribution		
			$\Omega = 0.05$	$\Omega = 0.10$	$\Omega = 0.15$
0.0	5.000×10^{-1}	5.000×10^{-1}	5.100×10^{-1}	5.199×10^{-1}	5.297×10^{-1}
0.5	3.085×10^{-1}	3.167×10^{-1}	3.150×10^{-1}	3.212×10^{-1}	3.271×10^{-1}
1.0	1.586×10^{-1}	1.751×10^{-1}	1.583×10^{-1}	1.571×10^{-1}	1.551×10^{-1}
1.5	6.681×10^{-2}	7.513×10^{-2}	6.236×10^{-2}	5.713×10^{-2}	5.111×10^{-2}
2.0	2.275×10^{-2}	1.684×10^{-2}	1.860×10^{-2}	1.437×10^{-2}	1.026×10^{-2}
2.5	6.210×10^{-3}	0.0	4.057×10^{-3}	2.298×10^{-3}	1.048×10^{-3}
3.0	1.350×10^{-3}	0.0	6.246×10^{-4}	2.111×10^{-4}	4.190×10^{-5}
3.5	2.326×10^{-4}	0.0	6.542×10^{-5}	9.831×10^{-6}	4.415×10^{-7}
4.0	3.167×10^{-5}	0.0	4.484×10^{-6}	1.977×10^{-7}	6.469×10^{-10}
4.5	3.398×10^{-6}	0.0	1.927×10^{-7}	1.396×10^{-9}	4.319×10^{-14}
5.0	2.867×10^{-7}	0.0	4.955×10^{-9}	2.621×10^{-12}	

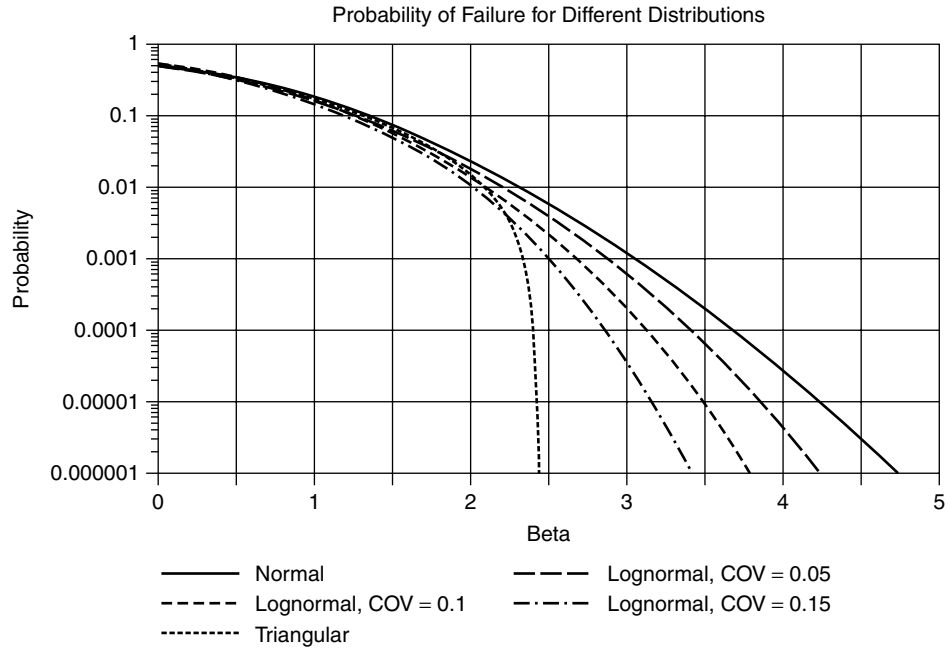


Figure 13.3 Probability of failure versus reliability index for various distributions.

the coefficient of variation (Ω) of F:

$$\Omega[F] = \frac{\sigma_F}{E[F]} \tag{13.8}$$

Results are shown for three values of Ω: 0.05, 0.10, and 0.15.

Table 13.1 and Figure 13.3 show that for most of the range of the reliability index the assumption of a Normal distribution is conservative and that, for values of β less than about 2, there is little difference between the results. For very small values of β the probability of failure is actually slightly larger for the Normal distribution than for the others. These results suggest that it is reasonable to assume a Normal distribution in the absence of further information and that the assumption will probably overestimate the probability of failure.

A further argument in favor of using the Normal distribution to relate β to the probability of failure is the C*entral Limit Theorem*, a fundamental result of probability theory. It states that, for a wide variety of conditions, the distribution of the sum of a large number of random variables converges to a Normal distribution. The theorem is not proven here, but it is a basic probabilistic concept first developed by Laplace (Stigler 1986). As long as there is a reasonably large number of variables, no one variable dominates, and the variables are not highly dependent on each other, the theorem applies with only weak conditions, regardless of the distributions of the individual variables. However, see footnote 3 of Chapter 3. An extension of this argument is that geological and geotechnical phenomena are determined by the combined contributions of a large number of small effects and, therefore, that the distribution of the overall effect

ought to be Normal. A different argument is that these combinations are not the *sum* of individual effects but their *product*, in which case the resulting distribution should reflect the sum of their logarithms and be logNormally distributed. Most workers would probably agree that it is reasonable to assume that M is Normally distributed. Other distributions apply to some geotechnical variables. For example, probability of exceeding a particular seismic acceleration is often represented by an exponential function. Gamma and Beta distributions are appropriate for other parameters. There is more controversy over the appropriate distribution for F. However, the authors believe that it is reasonable, simple, and conservative to assume that F is Normally distributed unless demonstrated otherwise.

One argument against using a Normal distribution is that it allows negative values, which are obviously impossible for F, while the logNormal distribution does not admit of negative values. The trouble with this argument is that the error introduced by the portion of the Normal distribution that has negative values of F is almost always so small that it has virtually no effect on the results. For example, in practice 1.25 is a relatively low value of $E[F]$, and 0.25 is a relatively large value of σ_F. These correspond to $\beta = 1.0$ and $p_F = 1.59 \times 10^{-1}$. A table of the standardized Normal distribution shows that the probability of $F < 0.0$ is 2.9×10^{-7}, which is insignificant compared to the computed probability of failure. Therefore, in practice, this argument has little merit.

Example 13.1 – Vertical Cut in Cohesive Soil

The example describes a case in which the calculations can be done exactly. Figure 13.4 shows a vertical cut in a purely cohesive soil. The critical sliding surface is inclined 45 degrees from the horizontal. The shear strength that resists sliding on this surface is c. The shearing force that tends to cause sliding is $\gamma H/4$. Therefore, the margin of safety is

$$M = c - \gamma H/4 \tag{13.9}$$

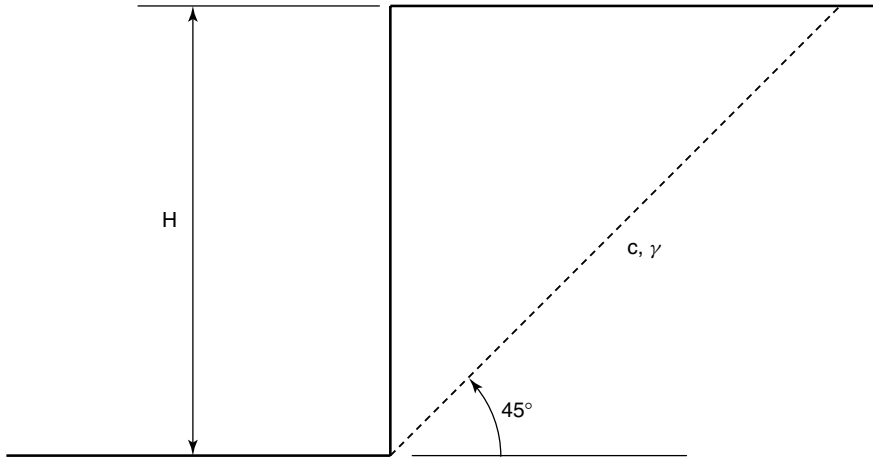


Figure 13.4 Geometry of vertical cut in cohesive soil.

(It should be noted that, if curved failure surfaces are allowed, the critical failure mode has a factor 3.77 instead of 4 in this equation.)

The cut is given to be 10 m deep. Then, the uncertainty in M is contributed by c and γ, which are the basic uncertainties in the analysis. Test data for the soil suggest the following means, standard deviations, and correlation coefficient for the variables c and γ:

$$\mu_c = 100\,\text{kPa} \qquad \sigma_c = 30\,\text{kPa}$$
$$\mu_\gamma = 20\,\text{kN/m}^3 \quad \sigma_\gamma = 2\,\text{kN/m}^3 \quad \rho_{c\gamma} = 0.5$$

The expected value of the margin of safety from Equation (13.2) is

$$\mu_M = 100 - (20)(10)/4 = 50\,\text{kPa}$$

and the variance of the margin of safety is

$$\sigma_M^2 = (30)^2 + (2)^2(10/4)^2 - 2(0.5)(30)(10/4)(2) = 900 + 25 - 240 = 775$$

The reliability index is then

$$\beta = 50/\sqrt{775} = 1.80$$

If the strength and the unit weight are Normally distributed, so is the margin of safety, and the probability of failure is

$$p_f = 3.62 \times 10^{-2}$$

If the shear strength and the unit weight were uncorrelated, the expected value of the margin would be unchanged, but the variance would be

$$\sigma_M^2 = (30)^2 + (2)^2(10/4)^2 = 925$$

and

$$\sigma_M^2 = (30)^2 + (2)^2(10/4)^2 = 900 + 25 = 925$$
$$\beta = 50/\sqrt{925} = 1.64$$
$$p_f = 5.01 \times 10^{-2}$$

Thus, the fact that c and γ are positively correlated – that is, that they tend to vary together – reduces the uncertainty about the stability of the cut. High values of c are associated with high values of γ, and vice versa. The chance that the difference $(c - \gamma H/4)$ is less than zero is smaller than it would be if c and γ were not associated with each other. This is reasonable, for we have more confidence in an estimate of the stability of a cut if we know that strength and unit weight are correlated than if we know nothing about their correlation.

13.3 Steps and Approximations in Reliability Analysis

In the above description of reliability theory and in the simple example just presented, each portion of the reliability analysis could be carried out exactly, and the steps flowed seamlessly into each other. Real problems are not so simple. The analyses may not be tractable, and approximations may be necessary in any or all of the steps. The approximations define the differences between the procedures for performing practical reliability analyses.

To explain what is involved in the various approaches to reliability calculations, it is useful to set out the steps explicitly. The goal of the analysis is to estimate the probability of failure p_f, with the understanding that 'failure' may involve any unacceptable performance. The steps are:

1. *Establish an analytical model.* There must be some way to compute the margin of safety, factor of safety, or other measure of performance. It can be as simple as Equation (13.9), or it can be an elaborate computational procedure. There may be error, uncertainty, or bias in the analytical model, which can be accounted for in the reliability analysis. For example, the well-known bearing capacity formulas are notoriously conservative and thus systematically biased.
2. *Estimate statistical descriptions of the parameters.* The parameters include not only the properties of the geotechnical materials but also the loads and geometry. Usually, the parameters are described by their means, variances, and covariances, but other information such as spatial correlation parameters or skewness may be included as well. The forms of the distributions of the parameters may be important.
3. *Calculate statistical moments of the performance function.* Usually this means calculating the mean and variance of the performance function. In the example of the vertical cut in cohesive soil the calculation can be done exactly, but most practical cases require approximation.
4. *Calculate the reliability index.* Often this involves the simple application of Equation (13.5). Sometimes, as in the Hasofer–Lind (1974) method, the computational procedure combines this step with step 3.
5. *Compute the probability of failure.* If the performance function has a well defined probabilistic description, such as the Normal distribution, this is a simple calculation. In many cases the distribution is not known or the intersection of the performance function with the probabilistic description of the parameters is not simple. In these cases the calculation of the probability of failure is likely to involve further approximations.

The engineer who is carrying out a reliability analysis should understand that most practical methods of reliability analysis involve approximations, even if one or more steps are exact. One should expect that different methods will give different answers. In particular, the calculation of the probability of failure from the reliability index usually implies the assumption that the margin or factor of safety is Normally distributed, and this is seldom the case. Therefore, it is often a good idea to compare results from two or more approaches to gain an appreciation of the errors involved in the computational procedures.

13.4 Error Propagation – Statistical Moments of the Performance Function

The evaluation of the uncertainty in the computed value of the margin of safety and the resulting reliability index is a special case of *error propagation*. The basic idea is that uncertainties in the values of parameters propagate through the rest of the calculation and affect the final result. For example, an engineer might measure properties of a soil in the laboratory, use these in some constitutive relation, and employ the results to compute the stability of an embankment using a simplified method of slices. Each of these steps involves some error and uncertainty on its own, and uncertainties in the original measurements of soil properties will affect the numbers calculated at each subsequent step. The study of error propagation is aimed at dealing rationally with this problem. There is a large literature on this subject, but Taylor's (1997) book is an excellent practical introduction.

The treatment of error propagation starts by recognizing that the result of the calculations can be considered a function g of the several input parameters and variables, $(X_1, X_2, \ldots, X_n)$ evaluated at some point $(x_1, x_2, \ldots, x_n)$:

$$g = g(x_1, x_2, \ldots, x_n) \tag{13.10}$$

If there is only one independent variable X and the value of g is known for some value of X, say $\overline{X}$, then the value of g can be found for any other value x by using the well known Taylor series

$$g(x) = g(\overline{X}) + \frac{1}{1!}(x - \overline{X}) \cdot \frac{dg}{dx} + \frac{1}{2!}(x - \overline{X})^2 \frac{d^2g}{dx^2} + \frac{1}{3!}(x - \overline{X})^3 \frac{d^3g}{dx^3} + \cdots \tag{13.11}$$

This equation is exact, provided all terms out to infinity are used. In practical applications x is chosen to be near $\overline{X}$, so higher order terms become small and the series can be truncated after only a few terms. However, in some reliability applications $(x - \overline{X})$ may not be small, and complications ensue.

In the study of error propagation more than one independent variable usually appears, so a generalization of Equation (13.11) is needed. Also, the initial value of each independent variable X_i is usually taken to be its mean value, μ_{X_i}. There are several equivalent forms of the Taylor series for multiple variables, but a convenient one is

$$\begin{aligned} g(x_1, x_2, \ldots, x_n) = g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}) &+ \frac{1}{1!} \sum_{i=1}^{n} (x_i - \mu_{X_1}) \frac{\partial g}{\partial x_i} \\ &+ \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - \mu_{X_i})(x_j - \mu_{X_j}) \frac{\partial^2 g}{\partial x_i \partial x_j} \\ &+ \frac{1}{3!} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (x_i - \mu_{X_i})(x_j - \mu_{X_j})(x_k - \mu_{X_k}) \frac{\partial^3 g}{\partial x_i \partial x_j \partial x_k} + \cdots \end{aligned} \tag{13.12}$$

The partial derivatives are taken at $\mu_{X_1}, \mu_{X_2}, \ldots$, etc., but notation to this effect is left out of Equation (13.12) to reduce clutter.

13.4.1 First order methods

The simplest and most widely used methods start by assuming that all the $(x_i - \mu_{X_i})$ terms are small, so their squares, cubes, and higher powers will be even smaller, and can be ignored. Since only the first order terms are included, methods based on this assumption are called *First Order Reliability Methods* (FORM), although this term is sometimes limited to geometrical reliability methods based on the Hasofer–Lind (1974) approach that is described in Chapter 16. Then, the first order terms give

$$g(x_1, x_2, \ldots, x_n) \approx g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}) + \sum_{i=1}^{n} (x_i - \mu_{X_1}) \frac{\partial g}{\partial x_i} \tag{13.13}$$

The approximation sign replaces the equal sign because this is now an approximation once the higher order terms have been removed. To find the expected value of g, it is necessary to integrate g multiplied by the joint probability density function of the variables X_1 through X_n from $-\infty$ to $+\infty$. Equation (13.13) is the sum of $n+1$ terms, so each term can be integrated in turn and the results added. The term $g(\mu_{X_1}, \mu_{X_2}, \ldots \mu_{X_n})$ is a constant. So are all the partial derivatives, for they have been evaluated at the means of the individual variables. Also, the integral from $-\infty$ to $+\infty$ of a probability density function multiplied by a constant is simply that constant. This all leads to

$$\mu_g \approx g(\mu_{X_1}, \mu_{X_2}, \ldots \mu_{X_n}) + \sum_{i=1}^{n} \int_{-\infty}^{\infty} (x_i - \mu_{X_i}) f_{X_i}(x_i) dx_i \tag{13.14}$$

but each of the terms after the summation sign must be identically zero, so

$$\mu_g \approx g(\mu_{X_1}, \mu_{X_2}, \ldots \mu_{X_n}) \tag{13.15}$$

This is the intuitively reasonable result that the expected value of a function of several variables is approximately equal to the value of the function calculated with the mean values of all the variables. However, it remains an approximation. For example, a well known result from elementary probability theory is that, if $g(X) = X^2$, then

$$E[g] = \mu_g = E[X^2] = (\mu_X)^2 + \sigma_X^2 \neq (\mu_X)^2 \tag{13.16}$$

The variance of a function g is

$$Var[g] = \sigma_g^2 = E[(g - \mu_g)^2] \tag{13.17}$$

and from Equations (13.13) and (13.15),

$$\sigma_g^2 \approx E\left[\left(\sum_{i=1}^{n} (x_i - \mu_{X_1}) \frac{\partial g}{\partial x_i}\right)^2\right] \tag{13.18}$$

Again, multiplying the expression in brackets by the probability density function and integrating over the complete range of probabilities leads to an expression for the variance. However, in this case care must be taken to multiply out the terms before interchanging the order of integration and summation. The result is

$$\begin{aligned}\sigma_g^2 &\approx \sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{X_iX_j}\sigma_{X_i}\sigma_{X_j}\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j} \\ &= \sum_{i=1}^{n}\sigma_{X_i}^2\left(\frac{\partial g}{\partial x_i}\right)^2 + \sum_{i=1}^{n}\sum_{j\neq i}^{n} Cov(X_i, X_j)\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\end{aligned} \tag{13.19}$$

Some comments are in order. First, Equation (13.19) requires the values of some partial derivatives. Sometimes it is possible to differentiate the function g and evaluate the corresponding terms exactly. The more common situation is that g is not so tractable and the partial derivatives must be found numerically. The easiest way to do this is to use central differences. The function g is evaluated with each of the variables set at its mean value; this yields the estimate of μ_g in Equation (13.15). Then the partial derivative for each variable in turn is found by increasing and decreasing the variable by a small amount, finding the difference between the two resulting values of g, and dividing the difference by twice the small increment. In mathematical notation, for variable X_i:

$$\frac{\partial g}{\partial x_i} \approx \frac{1}{2\varepsilon_i}\{g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_i} + \varepsilon_i, \ldots, \mu_{X_n}) - g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_i} - \varepsilon_i, \ldots \mu_{X_n})\} \tag{13.20}$$

in which ε_i is the small increment in variable X_i.

As an example, consider the vertical cut of example 13.1 and assume that the performance is expressed by the factor of safety F. Then,

$$F = \frac{4c}{\gamma H}, \quad \frac{\partial F}{\partial c} = \frac{4}{\gamma H}, \quad \frac{\partial F}{\partial \gamma} = -\frac{4c}{\gamma^2 H} \tag{13.21}$$

When these are evaluated at the mean values of the variables, the results are 2.0, 0.02, and -0.1, respectively. F is linear in c, so the central difference approximation is exact, but F is not linear in γ. If the increment and decrement in γ are 1.0, then F calculated with the incremented value of γ is 1.90476, and with the decremented value is 2.11526. Then

$$\frac{\partial F}{\partial \gamma} \approx \frac{(1.90476 - 2.10526)}{2 \cdot 1} = -0.10025$$

which is an accurate estimate. As larger increments in γ are used, the accuracy of the estimate deteriorates. For example, if the increment and decrement in γ were each 5, the resulting estimate of the partial derivative would be -0.1167.

Secondly, as can be seen in the example, the increments and decrements in the variables should be small. In the limit the derivative is defined as the result of taking the increment to zero, but that is computationally impossible. The derivative is estimated from the difference between two numbers, and, if the increment is too small, rounding errors could overwhelm the calculation. Therefore, the optimal choice is an increment as small

as possible without unacceptable rounding error. The easiest way to verify this is to compare the change in the magnitude of the estimate of the partial derivative for different increments and decrements of the independent variable. For example, if the increment and decrement in γ for the above example is 0.5, the estimate of $\partial F/\partial\gamma$ is -0.100063. Comparing this value with the result for an increment and decrement of 1.0 shows that the estimate of the derivative is robust and the contribution of the rounding error is insignificant.

On the other hand, the increment should be no larger than necessary to obtain an accurate result. The example shows that, the larger the increment, the larger the error in the estimate of the derivative. It has been suggested that the increment in each variable should be equal to the standard deviation of that variable. This is not a good idea. The standard deviation, or, equivalently, the variance, enters directly into Equation (13.20), and there is no need to introduce it into the calculation of the partial derivatives. One reason for the recommendation to use the standard deviation may be confusion between the Taylor series approach and Rosenblueth's point-estimate method, which is discussed in Chapter 15. These approaches are not the same.

Thirdly, if g is a linear function of the variables x_i, the first order approximations become exact, so

$$g = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i \tag{13.22}$$

$$\mu_g = \sum_{i=1}^{n} a_i\mu_{X_i} \tag{13.23}$$

$$\sigma_g^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\rho_{X_iX_j}\sigma_{X_i}\sigma_{X_j} \tag{13.24}$$

When there are two variables, R and Q, these equations reduce to Equations (13.1)–(13.3).

Fourthly, none of the results described in this section depends on the forms of the distributions of the variables or the functions, provided the distributions and functions are regular and smooth.

13.4.2 Second order methods

One possible extension of the first order approximation is to include terms in Equation (13.12) up to the second derivatives. Two principal justifications for this are: (1) the terms $(x_i - \overline{X}_i)$are not small enough to vanish when squared but do become small when cubed; and (2) the third and higher derivatives of g vanish or become small. Furthermore, it can be argued that the addition of second order terms improves the approximations enough to warrant the additional trouble. When second order terms are included, the techniques are called *Second Order Reliability Methods* (SORM), although this term too has sometimes been limited to second order geometrical reliability methods.

Removing all terms with third and higher derivatives from Equation (13.12) and performing the same integration as was done for the first order approximation yields the

following estimate for the mean of g:

$$\mu_g \approx g(\mu_{X_1}, \mu_{X_2}, \ldots \mu_{X_n}) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{X_i}\sigma_{X_j}\rho_{X_iX_j}\frac{\partial^2 g}{\partial x_i \partial x_j} \tag{13.25}$$

The expression for the variance when second-order terms are included is more complicated. Ang and Tang (1975) give the following equation for the variance when g is a function of one variable x:

$$\sigma_g^2 \approx \sigma_X^2 \cdot \left(\frac{dg}{dx}\right)^2 - \frac{1}{4}(\sigma_X^2)^2 \cdot \left(\frac{d^2g}{dx^2}\right)^2 + E[(x-\mu_X)^3] \cdot \frac{dg}{dx} \cdot \frac{d^2g}{dx^2} + \frac{1}{4}E[(x-\mu_X)^4] \cdot \left(\frac{d^2g}{dx^2}\right)^2 \tag{13.26}$$

This expression involves the square of the second derivatives and moments higher than second order. If these higher order terms are removed, Equation (13.26) reduces to

$$\sigma_g^2 \approx \sigma_X^2 \cdot \left(\frac{dg}{dx}\right)^2 \tag{13.27}$$

The corresponding expressions for multiple variables are much more complicated, but, if the squares of the second derivatives and the moments higher than second order are removed, the result is the same as Equation (13.19). In other words, the simplified form of the second order expression for the variance is the same as the first order expression.

13.4.3 Some useful results

Probability theorists have developed many relations among random variables; civil engineers will find some of these explained in standard references, such as Benjamin and Cornell (1970), Ang and Tang (1975), and Ayyub and McCuen (1997). Equations (13.22), (13.23), and (13.24) give the mean and variance of the sum of a set of random variables. Similar derivations lead to the following results for the product of a set of random variables when they are mutually independent:

$$g = a_1x_1 \cdot a_2x_2 \cdot \cdots \cdot a_nx_n = \prod_{i=1}^{n} a_ix_i \tag{13.28}$$

$$\mu_g = \prod_{i=1}^{n} a_i\mu_{X_i} \tag{13.29}$$

$$\sigma_g^2 \approx \sum_{i=1}^{n}\left(\prod_{\substack{j=1\\ j\neq i}}^{n}\mu_{X_i}^2\right)\sigma_{X_i}^2 \tag{13.30}$$

Note that Equation (13.29) is exact, but expression (13.30) is approximate. When there are only two mutually independent variables, Equation (13.30) reduces to

$$\sigma_g^2 \approx \mu_{X_1}^2 \sigma_{X_2}^2 + \mu_{X_2}^2 \sigma_{X_1}^2 \tag{13.31}$$

and the coefficients of variation (Ω) are approximately related by

$$\Omega_g^2 = \frac{\sigma_g^2}{\mu_{X_1}^2 \mu_{X_2}^2} \approx \Omega_{X_1}^2 + \Omega_{X_2}^2 \tag{13.32}$$

If the standard deviations are very small compared to the magnitudes of the variables themselves, and the variables are mutually independent, a further elimination of higher order terms leads to

$$\Omega_g^2 \approx \Omega_{X_1}^2 + \Omega_{X_2}^2 + \cdots + \Omega_{X_n}^2 = \sum_{i=1}^{n} \Omega_{X_i}^2 \tag{13.33}$$

This expression is the basis for calculating the uncertainty in laboratory measurements when the various contributing factors combine by multiplication.

In the same case that the variables are mutually independent, the exact expression for the relations among the coefficients of variation is

$$1 + \Omega_g^2 = (1 + \Omega_{X_1}^2)(1 + \Omega_{X_2}^2) \cdots (1 + \Omega_{X_n}^2) = \prod_{i=1}^{n} (1 + \Omega_{X_i}^2) \tag{13.34}$$

Equation (13.34) reduces to the approximate Equation (13.33) when all the variances are small. If there are only two variables, Equation (13.34) becomes

$$\sigma_g^2 = \mu_{X_1}^2 \sigma_{X_2}^2 + \mu_{X_2}^2 \sigma_{X_1}^2 + \sigma_{X_1}^2 \sigma_{X_2}^2 \tag{13.35}$$

Finally, when there are two correlated variables, the mean value of their product is

$$\mu_g = \mu_{X_1} \mu_{X_2} + \rho_{X_1 X_2} \sigma_{X_1} \sigma_{X_2} \tag{13.36}$$

The exact equations for the variance are more complicated.

Example 13.2 – Error Propagation in Staged Construction

Noiray (1982) estimates the uncertainties in the values of the shear strength for use in the analysis of the stability of a limestone storage area built over soft clay. Many factors must be considered, such as how the parametric uncertainty is divided between systematic and spatial uncertainty, the uncertainty in the loads, and the bias and uncertainty in the stability analysis, but the present discussion concentrates on the shear strength at a particular depth. Because the failures of concern would be rapid, the analysis uses undrained strength and

bases its calculations on the SHANSEP model (Ladd and Foott 1974), which describes the undrained shear strength of a soft clay by

$$s_u = \overline{\sigma}_{v0} \cdot s \cdot \left(\frac{\overline{\sigma}_{vm}}{\overline{\sigma}_{v0}} \right)^m \tag{13.37}$$

where s_u is the undrained shear strength, $\overline{\sigma}_{v0}$ is the *in situ* effective vertical stress, $\overline{\sigma}_{vm}$ is the maximum past effective vertical stress, and s and m are empirical parameters.

The first step in evaluating error propagation is to find the means and standard deviations for the parameters. Based on field and laboratory tests, profiles of existing and maximum past effective vertical stress are developed and regression lines drawn through them. At a depth of 20 ft., $\overline{\sigma}_{v0}$ is 2 KSF, and its standard deviation is vanishingly small. At the same depth the mean value of $\overline{\sigma}_{vm}$is 4.18 KSF, and the standard deviation in the estimate of the mean is 0.97 KSF. From laboratory tests on soils from the site and from comparisons to similar soils, the mean value of m is taken to be 0.86, with a standard deviation of 0.03; the corresponding values for s are taken to be 0.213 and 0.023. Both these parameters are dimensionless. It is reasonable to assume that the parameters are mutually independent.

The second step is to propagate these errors or uncertainties through Equation (13.37) using the first order approximation technique. This states that the best first order estimate of the mean of the strength is the value calculated using the mean values of the parameters. Substituting the mean values of the parameters into Equation (13.37) gives $\mu_{s_u} \approx 0.803 KSF$. Equation (13.19) is then used to estimate the variance or standard deviation in the strength, and this requires the partial derivatives of s_u. One way to find these is to differentiate Equation (13.37) analytically and perform the necessary algebra to arrive at

$$\Omega^2[s_u] \approx \Omega^2[s] + m^2 \cdot \Omega^2[\overline{\sigma}_{vm}] + \ln^2(\overline{\sigma}_{vm}/\overline{\sigma}_{v0}) \cdot \sigma_m^2 + (1-m) \cdot \Omega^2[\overline{\sigma}_{v0}]m \tag{13.38}$$

where $\Omega[X]$ represents the coefficient of variation of X. Substitution and simplification gives $\sigma_{s_u} \approx 0.183\, KSF$. Another approach is to avoid the tedium of differentiation by using the central difference technique. If the increments and decrements in m and s are 0.01 and those in $\overline{\sigma}_{vm}$ are 0.1, the resulting estimate is again $\sigma_{s_u} \approx 0.183\, KSF$. Actually, the calculations agree to six decimal places, so the central difference approach is quite accurate. One should note, however, that both these calculations are first order approximations.

The next step would be to combine these estimates of uncertainty with estimates of other factors contributing to the analysis to obtain an estimate of the overall uncertainty in the stability of the structure. Similar error propagation studies would be performed for the other factors. It would, of course, be possible to perform the entire estimate in one step by doing a central difference estimate for each original parameter, carrying its effect through the entire calculation. In addition to being tedious and unwieldy, such an approach would not reveal the relative contributions of the individual components of the problem. Breaking the problem into steps and performing the error propagation at each step is not only easier, it is less prone to error, and it gives a deeper insight into the behavior of the physical system.

13.5 Solution Techniques for Practical Cases

The example of the vertical cut is simple enough that it is possible to perform the calculations exactly. In more typical cases, the analyst must usually employ a technique that yields an approximation to the true value of the reliability index and the probability of failure. Several methods are available, each having advantages and disadvantages. Among the most widely used are

- The *First Order Second Moment* (FOSM) method. This method uses the first terms of a Taylor series expansion of the performance function to estimate the expected value and variance of the performance function. It is called a second moment method because the variance is a form of the second moment and is the highest order statistical result used in the analysis. If the number of uncertain variables is N, this method requires either evaluating N partial derivatives of the performance function or performing a numerical approximation using evaluations at $2N + 1$ points.
- The *Second Order Second Moment* (SOSM) method. This technique uses the terms in the Taylor series up to the second order. The computational difficult is greater, and the improvement in accuracy is not always worth the extra computational effort. SOSM methods have not found wide use in geotechnical applications.
- The *Point Estimate* method. Rosenblueth (1975) proposed a simple and elegant method of obtaining the moments of the performance function by evaluating the performance function at a set of specifically chosen discrete points. One of the disadvantages of the original method is that it requires that the performance function be evaluated 2^N times, and this can become a very large number when the number of uncertain parameters is large. Recent modifications reduce the number of evaluations to the order of $2N$ but introduce their own complications.
- The *Hasofer–Lind* method. Hasofer and Lind (1974) proposed an improvement on the FOSM method based on a geometric interpretation of the reliability index as a measure of the distance in dimensionless space between the peak of the multivariate distribution of the uncertain parameters and a function defining the failure condition. This method usually requires iteration in addition to the evaluations at $2N$ points. The acronym FORM often refers to this method in particular.
- *Monte Carlo simulation*. In this approach the analyst creates a large number of sets of randomly generated values for the uncertain parameters and computes the performance function for each set. The statistics of the resulting set of values of the function can be computed and β or p_F calculated directly. The method has the advantage of conceptual simplicity, but it can require a large set of values of the performance function to obtain adequate accuracy. Furthermore, the method does not give insight into the relative contributions of the uncertain parameters that is obtained from other methods. The computational effort can be reduced significantly by using statistical techniques known as variance reduction schemes, and these should be employed whenever possible.

The subsequent chapters discuss these methods in detail. It should be noted that selecting a method involves several tradeoffs. Each method involves different computational effort, provides a different level of accuracy, and yields a different insight into the effects of the individual parameters. In addition, some of the methods provide values of the reliability

index, and a further assumption about the distribution of the performance function is necessary to obtain the probability of failure.

13.6 A Simple Conceptual Model of Practical Significance

Lambe and his colleagues (Lambe 1985; Lambe *et al.* 1988; Marr 2000) developed an interesting perspective on the reliability of slopes. They observed that the factor of safety is meaningful only in the context of the methodology used for analysis, design, construction, and maintenance of the facility. That is, a factor of safety computed for a slope constructed of materials whose properties have been carefully measured and whose behavior has been continuously monitored means something different from a number computed on the basis of little field work, index properties, and no monitoring.

Four levels of engineering effort are described in Table 13.2. It should be borne in mind that the choices of specific items in the table are affected by the concentration on slope stability and influenced by the state of the art in the 1980s. Lambe and his colleagues reckoned that a slope designed and built to a factor of safety of 1.5 with level II engineering had a probability of failure of 10^{-4}. The corresponding sets of

Table 13.2 Description of different levels of engineering (after Lambe 1985)

	Design	Construction	Performance	FS Determination
I	By qualified engineer	Full-time supervision by qualified engineer	Complete performance program	Effective stress analyses based on measured data (strength, geometry, pore pressure...) for site
	Based on measured data for site	Field control measurements	No malfunctions (slides, cracks, artesian heads...)	Complete flow net
	Has key components (filters, cut-off, etc.) No error, omissions Complete assessment of geologic conditions	No errors, omissions	Continuous maintenance by trained crews	
II	By qualified engineer	Part-time supervision by qualified engineer	Periodic inspection by qualified engineer	Effective stress analyses
	Has key components No errors, omissions Geology considered	No errors, omissions	No malfunctions Few field measurements Routine maintenance	
III	Approximate design using parameters inferred from index tests	Informal supervision	Annual inspection by qualified engineer	Rational analyses using parameters inferred from index tests
			No malfunctions Maintenance limited to emergency repairs	
IV	No rational design	No supervision	Occasional inspection by non-qualified person	Approximate analyses using assumed parameters

Table 13.3 Relations between factor of safety and probability of failure for different levels of engineering (after Lambe 1985)

Level of engineering	Factor of safety	Probability of failure
I	1.3	10^{-4}
II	1.5	10^{-4}
III	1.5	10^{-2}
IV	1.5	10^{-1}

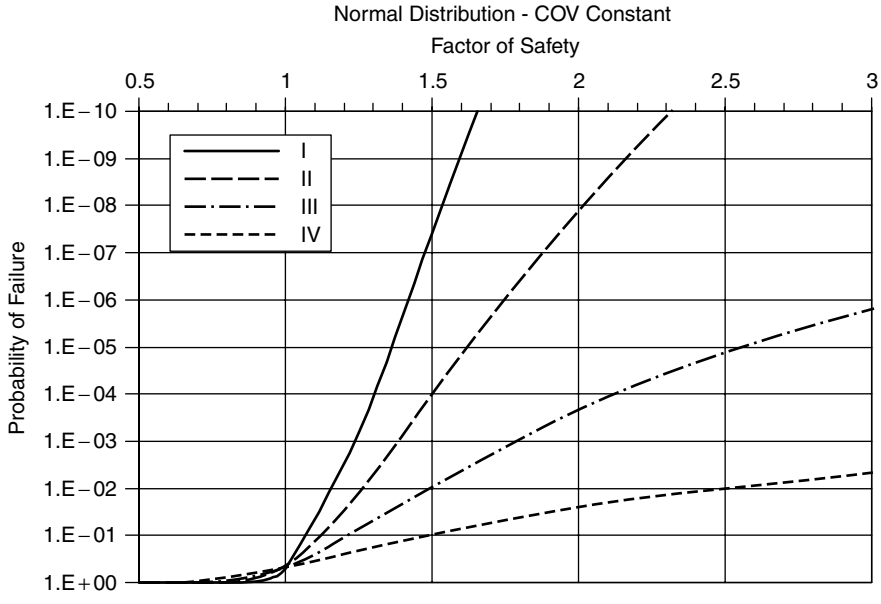


Figure 13.5 Conceptual relation between design factor of safety and probability of failure when design factor of safety is the expected value and *FS* is Normally distributed.

probabilities of failure and design factors of safety for the other levels of engineering are listed in Table 13.3. Having identified one relation between factor of safety and probability of failure for each of the levels, they sketched curves similar to those in Figure 13.5 to extrapolate to other factors of safety. These were clearly intended to represent a concept rather than a set of hard relations; as Lambe (1985) wrote, "We prepared Figure [13.5] as a numerical expression of our judgment – we had little data to prepare these plots."

Some reasonable assumptions yield a more precise extrapolation. Let it be assumed that (a) the factor of safety is Normally distributed, and (b) the coefficient of variation $\Omega(= \sigma/mu)$ is constant for all values of factor of safety. From assumption (a), the reliability index β can be computed from the probability of failure p_f:

$$\beta = -\Phi^{-1}[p_f] \tag{13.39}$$

If μ and σ are the expected value and the standard deviation of the factor of safety,

$$\beta = \frac{\mu - 1}{\sigma} = \frac{1 - 1/\mu}{\Omega} = \frac{1 - \mu}{\Omega \cdot \mu} \tag{13.40}$$

At each of the four known points, p_f and μ are known, so we can solve for Ω and compute the probability at any other mean factor of safety:

$$p_f = \Phi\left[\frac{1 - FS}{FS \cdot \Omega}\right] \tag{13.41}$$

The resulting curves are plotted in Figure 13.5.

An additional assumption might be made to the effect that the engineer actually unintentionally introduces conservatism is calculating factors of safety so that the computed factor of safety (FS_0) is actually one σ less than the mean μ. This leads to the relations

$$\Omega = \frac{FS_0 - 1}{FS_0 \cdot \beta - 1}$$
$$p_f = \Phi\left[\frac{1 - FS - \Omega}{FS \cdot \Omega}\right] \tag{13.42}$$

where FS is the factor of safety at any other point. This set of relations is plotted in Figure 13.6.

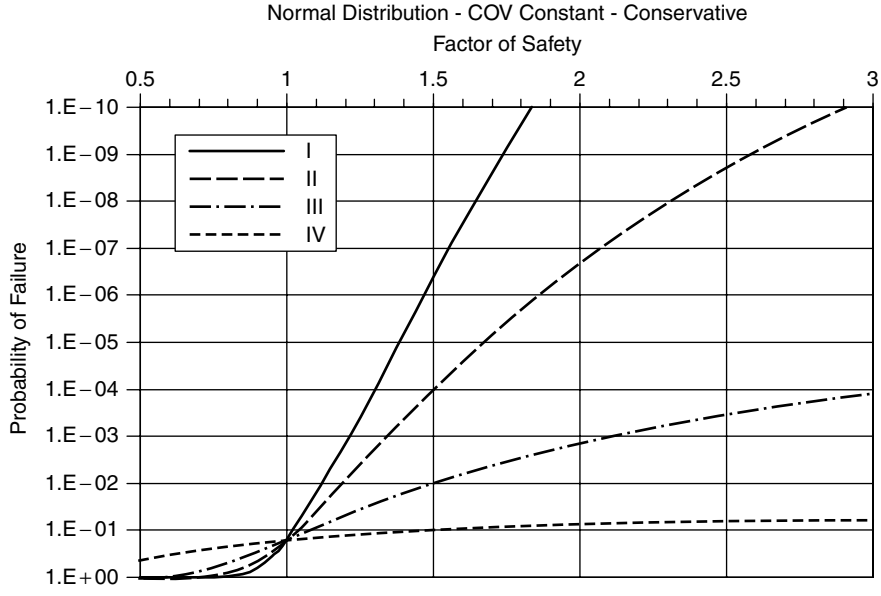


Figure 13.6 Conceptual relation between design factor of safety and probability of failure when design factor of safety is one standard deviation below the expected value and *FS* is Normally distributed.

These figures are not rigorous representations of the relations between levels of engineering and probabilities of failure because the necessary verification studies have not been done. Nevertheless, they do present a conceptual model. The owner or regulator of a constructed facility does not, in the last analysis, care about the factor of safety. What he or she should be concerned with is the probability of failure or the amortized costs associated with the facility. Figures 13.5 and 13.6 show that an owner can achieve reduced probability of failure (and therefore reduced costs) by increasing the design factor of safety or by improving the level of engineering and construction. In most cases, improved engineering reduces the probability of failure even when the factor of safety is smaller.

14 First Order Second Moment (FOSM) Methods

Chapter 13 described the first order approximation to the mean, variance, and standard deviation of a function F, based on the first terms of a Taylor series expansion of F. When the variables are uncorrelated the expressions become

$$E[F] = \mu_F \approx F(X_1, X_2, \ldots, X_n) \tag{14.1}$$

and

$$\sigma_F^2 \approx \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial F}{\partial x_i} \frac{\partial F}{\partial x_j} \rho_{X_i X_j} \sigma_{X_i} \sigma_{X_j} \tag{14.2}$$

where the x's are the values of the variables X that enter into the calculation of F. As is shown in the following example, a modeling error can be added to these expressions. The function F could be any relevant function, but for present purposes it is the factor of safety calculated by some appropriate numerical method. Once the mean and variance have been calculated it is a simple matter to compute the reliability index β:

$$\beta = \frac{E[F] - 1}{\sigma_F} \tag{14.3}$$

The usual practice is then to assume that the factor of safety is Normally distributed and to compute the probability of failure from β. The First Order Second Moment (FOSM) procedures that follow are easier to demonstrate than to describe, so the bulk of this chapter is a detailed presentation of the application of FOSM methods to a specific problem of the reliability of the dikes for the James Bay Project (Christian *et al*. 1994).

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

14.1 The James Bay Dikes

The James Bay hydroelectric project required the construction of approximately fifty kilometers of dikes on soft, sensitive clays (Ladd *et al.* 1983; Ladd 1991). When the stability of an embankment in such circumstances is evaluated, professional opinions often diverge regarding the strengthening of foundation soil during construction, selecting appropriate factors of safety, and evaluating strength properties. The owners formed an international committee to address and resolve the issues of the stability of the slopes. In particular, the engineers had to determine appropriate factors of safety to be used in the design of dikes for different parts of the project ranging in height from 6–23 m.

Ladd *et al.* (1983) provide details of the use of the mean first-order reliability method to evaluate the single- or multi-stage construction of a typical dike, whose cross-section is shown in Figure 14.1. The goals of the analysis included understanding the relative safety of different designs, obtaining insight about the influence of different parameters, and establishing consistent criteria for preliminary designs.

The first design cases to be considered were the construction of embankments in a single stage, either to a height of 6 m without berms or 12 m with one berm. Undrained shear strength values were obtained from field vane tests, and the stability analyses were done using the simplified Bishop circular arc method of slices. The second design case called for multi-stage construction. Following the installation of vertical drains, the first stage would be built to a height of 12 m. Then the zone with the drains would be allowed to consolidate to 80% of primary consolidation, and a second stage would be built to a height of 23 m. In this case the undrained shear strength would be obtained from consolidated-undrained shear tests combined with knowledge of the stress histories (Ladd

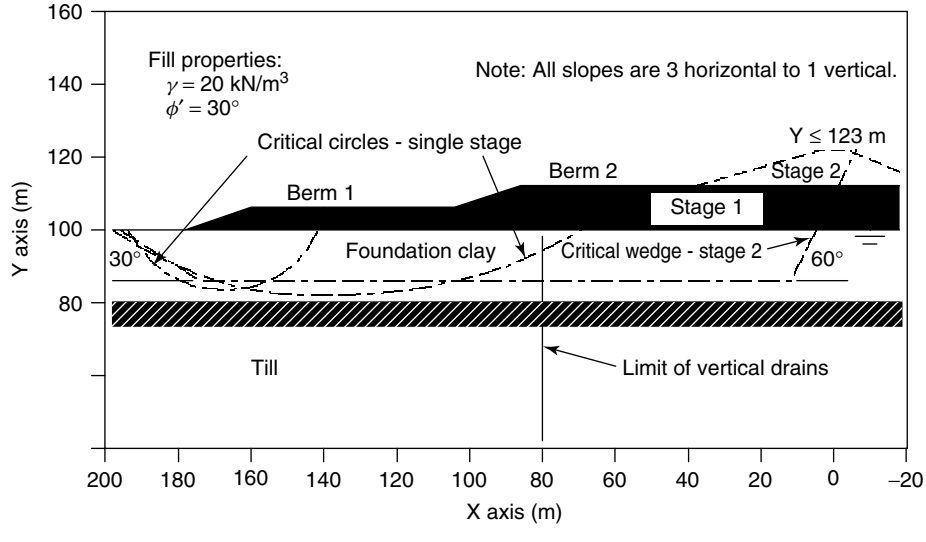


Figure 14.1 Cross-section of typical dike. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

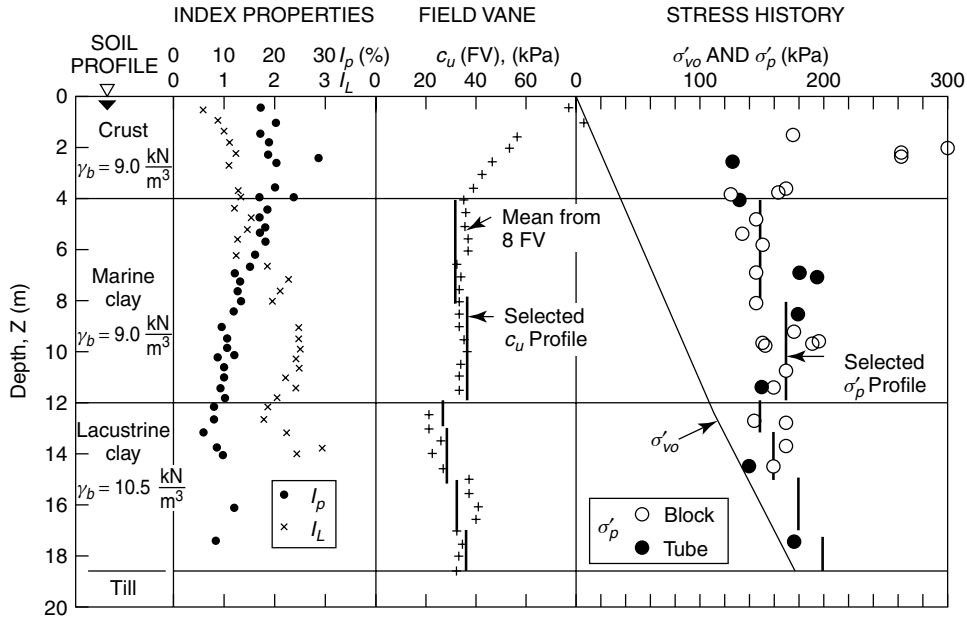


Figure 14.2 Simplified soil profile. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

1991). The stability analysis was done by the Morgenstern–Price method of slices for a wedge-shaped failure pattern.

Figure 14.2 is a simplified soil profile, showing that the foundation is composed of 4 m of crust underlain by 8 m of a highly sensitive marine clay underlain in turn by 6.5 m of lacustrine clay, for a total thickness of clay of 18.5 m. It also shows the distribution of vane shear strength and the effective vertical stress history.

14.2 Uncertainty in Geotechnical Parameters

The uncertainty in the stability of a slope is the result of many factors. Some, such as ignorance of geological details missed in the exploration program, are difficult to treat formally; others, such as the estimates of soil properties are more amenable to statistical analysis. The first part of this book discusses the statistical description of geological profiles and geotechnical properties, upon which this chapter builds.

Uncertainties in soil properties arise from two sources: scatter and systematic error in estimates of properties. The former consists of inherent spatial variability in the properties and random testing errors in measurements. The latter consists of systematic statistical errors due to the sampling process and bias in the methods of measurement. The following sections describe how this conceptual view was applied to the properties of the soils in the James Bay Dikes.

14.2.1 Data Scatter: Spatial Variability and Noise

Field vane tests conducted with the Nilcon device on the James Bay clays exhibited an unusually large degree of scatter. If α is Bjerrum's (1972) empirical factor for correcting the bias in field vane data, then a common model of measurement error is

$$\alpha c_u(FV) = c_u + c_e \tag{14.4}$$

where c_u is the actual undrained strength, $c_u(FV)$ indicates the value measured by the field vane test, and c_e is a random experimental error with a zero mean. This equation assumes that α is known precisely. The engineer needs to know the actual *in situ* shear strength, c_u, including its spatial variability, but the random experimental variations represented by c_e must be eliminated. One way to do this is to use the autocovariance function, provided it does not change with position (i.e. the spatial variation of c_u is stationary).

Chapter 9 described how the autocovariance could be used to separate the noise from the spatial variability. Figure 9.23 shows the application of maximum likelihood estimators to the horizontal variation of vane shear data at James Bay. Because the autocovariance functions for soil properties are different in the vertical and horizontal directions, the vertical autocorrelations should be made using values from the same boring, and the horizontal autocorrelations using data from different borings at the same elevation. Then each pair of data is used to compute the autocovariance as a function of the separation distance r. The results are averaged or smoothed within each interval of r. DeGroot and Baecher (1993) and Soulié *et al.* (1990) have published more rigorous analyses of these data.

It must be emphasized that the purpose of the exercise is to separate the observed variation in measured soil properties into two categories. The first is the spatial variance in the field values, which should represent a real effect that occurs in the field and needs to be taken into account in the slope stability analysis. The second consists of random scatter due to error in measurements and small-scale fluctuations in soil properties. While these appear in the observed data, they do not affect the behavior of the soil in the field and should not be included in the stability analysis.

14.2.2 Systematic Error

In addition to the data scatter and spatial variability, there are systematic errors in the estimation of the soil properties. The systematic error contributes to the analytical evaluation of slope stability differently from data scatter because spatial variability averages over the volume of soil while systematic error is constant. Therefore, systematic error propagates almost unchanged through an analysis and can have a greater influence on predictive uncertainty than does spatial variability.

In the present example two factors contribute to systematic error. First, a limited number of tests is used to determine the mean or trend of shear strength, and different set of measurements would yield a different estimate. To the extent that the statistical estimate of mean or trend is wrong, it is so everywhere. Standard statistical methods, such as those described by Benjamin and Cornell (1970), provide variances of estimates obtained by

regression analysis. In the present case of a mean value of a parameter X, the variance in the estimate from n measurements is approximately

$$V\{E[x]\} \approx V[x]/n \tag{14.5}$$

The second component of systematic error is bias. The experimental technique may not measure directly the quantity of interest or may measure it in some systematically erroneous way. The need to correct data from field vane tests is well known. The factor α in Equation (14.4) is a function of the plasticity index, I_p. However, there is scatter in the data defining this relationship (Ladd *et al.* 1977, 1983; Azzouz *et al.* 1983). The scatter, which has a coefficient of variation of about 25%, represents the uncertainty in using this correction factor. On the basis of other observations on this marine clay, the mean of α was taken to be 1.0, but, of course, this would not be true for most clays.

In separating spatial variability and systematic error it is easiest to think of spatial variation as scatter of the soil property about the mean trend and of systematic error as uncertainty in the location of the mean trend itself. The only component of the former is inherent spatial variability after random measurement error has been removed, while the latter consists of measurement bias and statistical estimation error due to a finite number of measurements. The contribution of spatial variability depends on the size of the region or failure surface over which it is effective in relation to the autocorrelation distance, r_0.

14.3 FOSM Calculations

As explained in Chapter 13, there are two direct ways to evaluate Equation (14.2). First, the function can be differentiated formally to give a closed form expression for the variance of F. Second, the function can be differentiated numerically by divided differences to give approximations to the partial derivatives.

14.3.1 Computing Uncertainty in Factor of Safety

The computation of factor of safety using the method of slices for an actual geometry is an example of a case in which direct differentiation may not be feasible and numerical methods of evaluating the variance becomes attractive. The approach is to evaluate F for the values of the parameters at the condition of interest and then to change each of the parameters in turn by a small amount and re-evaluate F. The differences, divided by the increments in the parameters, provide estimates of the derivatives. As in all numerical work, the user must exercise care not to extrapolate a linear analysis too far beyond the region in which actual results exist.

14.3.2 Uncertainty in Soil Profile and Fill Properties

The single stage cases involved many kilometers of embankment with a modest exploration program at any particular site along the embankment. The analysis had to include uncertainty about the thickness of the crust and the depth to the till as well as the unit weight and friction angle of the fill. The multi-stage case involved a limited length of embankment with an extensive exploration and instrumentation program. The geometry

Table 14.1 Soil profile and fill property uncertainties for reliability analysis. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

Variable	Expected value	Variance		
		Spatial	Systematic	Total
Depth of crust	4 m	0.96	0.04	1.0
Depth to till	18.5 m	0.0	1.0	1.0
Fill density	20 kN/m^3	1.0	1.0	2.0
Fill friction ϕ'	30°	1.0	3.0	4.0

and the soil properties were better known, and the major uncertainty was in the properties of the fill and the partially consolidated clay.

The depth of the crust and the depth to the till were both treated as uncertain variables. From the results of field exploration, the uncertainties listed in Table 14.1 were established. The separation of the uncertainty into spatial and systematic components is important because averaging over the failure surface affects the contribution of spatial variability. The depth of the crust affects the undrained strength profile, and its spatial variance was estimated from the field vane data across the entire axis of the dike. The systematic statistical uncertainty for the vane soundings was estimated from Equation (14.5) with $n = 27$.

Uncertainty in the depth to the till affects the depth of the failure surfaces in the analysis. Since the critical surfaces are nearly tangent to the top of the till for $H = 12$ m, this uncertainty does not affect the average across the profile but is a systematic uncertainty even though the estimate of variance reflects spatial variance along the cross-section. That is, there is only one critical failure surface at any location. The location of this surface and its associated F are dictated by the uncertain depth of the till at that location, but the uncertainty in the depth of the till comes about by the variation in depth over the whole project site. The estimate of systematic error reflects the effects of a finite number of borings.

Uncertainty in the density of the fill was taken to be composed equally of spatial and systematic components. This assumption and estimated magnitudes of the variances were based on judgment. The assumption that $E[\phi'] = 30°$ reflected Canadian practice.

14.3.3 Uncertainty in Shear Strength of Foundation Clay

Table 14.2 presents the uncertainty in c_u used for the single stage analyses, based on the field vane data for the marine and lacustrine clays. The relatively small coefficient of variation selected for the bias in α for the marine clay ($\Omega = 0.075$) reflected detailed comparison with results from laboratory CK_oU shear tests with different modes of failure, described by Lefebvre *et al*. (1988). This bias was doubled for the lacustrine clay, for which a similarly detailed comparison had not been made. The other components of spatial and systematic uncertainty were developed from the techniques described above. For example, the statistical coefficient of variation for the marine clay ($[(0.24)^2/62]^{1/2} = 0.03$) follows from Equation (14.5). As the total variance due to several independent

Table 14.2 Undrained shear strength uncertainty based on vane shear data. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

	Clay	
Field vane statistics	Marine	Lacustrine
Number of tests	62	37
Mean, kPa	34.5	31.2
Data scatter, COV	0.236	0.272
Spatial variability, COV	0.183	0.272
Systematic error, COV:		
Statistical	0.030	0.045
Correction factor:	0.075	0.15
Sub-total	0.08	0.16
Total COV	0.20	0.32

variates is the sum of the individual variances, coefficients of variation and standard deviations are combined by taking the square root of the sum of the squares of the individual components.

For the multi-stage case, the shear strengths were established from a combination of undrained strength ratios and the *in situ* stress history. For the overconsolidated intact clay beyond the limits of the vertical drains, no consolidation was assumed, and c_u/σ'_p ratios were applied to the σ'_p profile shown in Figure 14.2. For the clay within the limits of the vertical drains, 80% consolidation was assumed under berm number 2, and the c_u/σ'_{vc} ratios were applied to $\sigma'_{vc} = \sigma'_{v0} + 192$ kPa ($= 0.80 \times 12\,\text{m} \times 20\,\text{kN/m}^3$). Therefore, it is necessary to deal with uncertainty in both c_u/σ'_p or c_u/σ'_{vc} and σ'_p or σ'_{vc}. Morgenstern–Price wedge stability analyses used anisotropic undrained strength profiles for compression, direct simple shear, and extension modes of failure.

The undrained shear strength ratios were obtained from CK_oU triaxial compression and extension and Geonor direct simple shear tests run using the recompression technique and treated for strain compatibility to account for the effects of progressive failure (Ladd *et al*. 1983; Ladd 1991). Table 14.3 shows best estimates of these ratios. These were not intended to be conservative. In this table the subscript C indicates compression failure, D direct simple shear failure on a horizontal plane, and E extension failure.

Table 14.4 presents the uncertainties in the undrained strength ratios, the vertical preconsolidation and consolidation stresses, and the resulting values of undrained shear strength, τ. Some uncertainties were developed by combining engineering judgment, experience, and experimental data. As in Table 14.3, these are not conservative values of parameters, but best estimates.

14.3.3.1 Undrained strength ratios

The coefficient of variation for the intact and consolidated clay were treated collectively for convenience. The spatial coefficient of variation was assumed equal to the average

Table 14.3 Undrained strength ratios from laboratory CK_0U tests. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

Deposit	State	τ_C/σ'	τ_D/σ'	τ_E/σ'
B-6 Marine	OC	0.26	0.225	0.16
	NC	0.26	0.225	0.16
B-6 Lacustrine	OC	0.225	0.19	0.14
	NC	0.25	0.215	0.12

Note: $\sigma' = \sigma'_p$ for OC clay, and $\sigma' = \sigma'_{vc}$ for NC clay.

Table 14.4 Uncertainty in undrained shear strength estimates for multi-stage analysis. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

		Coefficient of Variation		
Parameter	Clay	Spatial	Systematic	Total
Intact clay				
τ/σ'_p	Marine	0.067	0.076	0.10
	Lacustrine	0.084	0.092	0.12
σ'_p	Marine	0.18	0.10	0.21
	Lacustrine	0.27	0.15	0.31
τ	Marine	0.19	0.13	0.23
	Lacustrine	0.28	0.18	0.33
Consolidated clay				
τ/σ'_{vc}	Marine	0.067	0.076	0.10
	Lacustrine	0.084	0.092	0.12
σ'_{vc}	Marine	0.09	0.08	0.12
	Lacustrine	0.09	0.08	0.12
τ	Marine	0.11	0.11	0.16
	Lacustrine	0.12	0.12	0.17

total scatter measured in τ/σ' for each of the three modes of failure. The systematic coefficient of variation had two components: one due to the statistical error calculated from Equation (14.5), and the other the estimated bias. The latter assumed that the triaxial tests underestimated plane strain shear strengths by $10 \pm 5\%$ ($\Omega = 0.05$) and that the strain rate used for the tests overestimated the *in situ* strengths by $10 \pm 5\%$ ($\Omega = 0.05$). Thus, these two errors cancel in the mean but cause a bias $\Omega = \sqrt{2} \times 0.05 = 0.07$.

14.3.3.2 Preconsolidation pressures and consolidation stresses

Because very few oedometer tests were run (see Figure 14.2), the spatial coefficient of variation of σ'_p was assumed equal to that obtained from the field vane data, that is $\Omega = 0.18$ and 0.27 from Table 14.2 for the marine and lacustrine clays, respectively. The systematic coefficients of variation reflected a combination of statistical error and judgment regarding the likely bias and were higher for the lacustrine clay because of the lack of data within the bottom half of the deposit (again, see Figure 14.2). The vertical consolidation stress, σ'_{vc}, assumed no uncertainty in σ'_{v0} and therefore reflected uncertainty in the average degree of consolidation, U, which would be estimated from field piezometer data and the load imposed by berm 2. The values in Table 14.4 were all based on judgment and assumed the following: for spatial uncertainty, $\Omega[U] = 0.075$ and $\Omega[\text{load}] = 0.05$ to give $\Omega[\sigma'_{vc}] = 0.09$; for systematic uncertainty, $\Omega[U] = 0.075$ and $\Omega[\text{load}] = 0.0$ to give $\Omega[\sigma'_{vc}] = 0.075$, which was rounded to 0.08.

14.3.3.3 Undrained strength

The coefficient of variation for shear strength τ in Table 14.4 is obtained from

$$\Omega^2[\tau] = \Omega^2[\tau/\sigma'] + \Omega^2[\sigma'] \tag{14.6}$$

The variance in τ actually used in the analysis then becomes

$$Var[\tau] = \Omega^2[\tau] \cdot (E[\tau])^2 \tag{14.7}$$

where $E[\tau]$ is the mean undrained strength within the intact or consolidated layer of clay.

14.3.4 Calculating the Reliability Index

There are six steps to the calculation:

1. Identifying all the variables that affect the stability of the embankment; examples are the geometry, the weight and strength of fill materials, and the relevant properties of the foundation soils.
2. Determining the best estimate (usually the mean value) of each variable, $E[X_i]$, and using these to calculate the best estimate of the factor of safety, $E[F]$ by the method of slices.
3. Estimating the uncertainty in each variable and, in particular, its variance, $Var[X_i]$, based on the uncertainties in soil properties.
4. Performing sensitivity analyses by calculating the partial derivatives of F with respect to each of the uncertain variables or by approximating each derivative by the divided difference $\Delta F/\Delta X_i$.
5. Using Equation (14.2) to obtain $Var[F]$.
6. Calculating the reliability index β from Equation (14.3).

Table 14.5 Variance composition, single-stage dike, $H = 12$ m. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

		Variance		$(\Delta F/\Delta x_i)^2 \cdot Var(x_i)$	
Parameter	$\Delta F/\Delta x_i$	Systematic	Total	Systematic	Total
ϕ'	0.01	3.0	4.0	0.0003	0.0004
γ_{fill}	0.06	1.0	2.0	0.0036	0.0072
D_{crust}	0.008	0.04	1.0	nil	0.0001
D_{till}	0.056	1.0	1.0	0.0031	0.0031
$c_u(M)$	0.0137	7.6	47.6	0.0014	0.0089
$c_u(L)$	0.0215	24.9	99.7	0.0115	0.0461
			$V[F] =$	0.0199	0.0658
			Spatial $V[F] =$	0.0459	
			$V_{ave}[F]$ at $R = 0.2 =$	$0.0199 + 0.0092 = 0.0291$	
			$\beta =$	$(1.453 - 1.0)/\sqrt{0.0291}$	
			$=$	2.66	

The numbers in Tables 14.1–14.4 are the result of the first three steps. Tables 14.5 and 14.6 show the results of the last three steps for two specific cases. In these tables column 3 contains the systematic variances, and column 4 the total variances. The differences are the spatial variances. Similarly, column 5 contains the contribution of the systematic variances to the variance in F, and column 6 the total variances. Again the differences are the spatial contributions. Columns 5 and 6 are summed to obtain the systematic variance in F and the total variance in F. Their difference is the spatial variance, which is reduced as explained below.

First, however, it should be noted that the procedure minimizes F over the range of failure surfaces and then calculates β. It does not minimize β directly. The minimum F surface and the minimum β surface may not be identical (Lacasse and Nadim 1996). Oka and Wu (1990) observed that, in a situation with two different but likely modes of failure, reliability calculations based on one mode only can be seriously in error. This turned out not to be the situation here.

14.3.4.1 Single stage embankments

The results for the single stage dike built with one berm to a height of 12 m, without the model error, are displayed in Table 14.5. The circular arc analyses gave, for the critical failure circle, $E[F] = 1.453$. The spatial variations tend to average out over the failure surface. By integrating the statistically varying shear strength over the failure surface it can be shown (see Chapter 9) that the effective spatial variance of the average shear strength is

$$\frac{V[c_{u,ave}]}{V[c_u]} \approx \frac{2r_o}{L} \tag{14.8}$$

Table 14.6 Variance composition, multi-stage dike $H = 23$ m. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

		Variance		$(\Delta F/\Delta x_i)^2 \cdot Var(x_i)$	
Parameter	$\Delta F/\Delta x_i$	Systematic	Total	Systematic	Total
ϕ'	0.0088	3.0	4.0	0.0002	0.0003
γ_{fill}	0.055	1.0	2.0	0.0030	0.0061
Intact clay					
$\tau(M)$	0.0018	13.3	41.5	nil	0.0001
$\tau(L)$	0.012	26.3	88.5	0.0038	0.0127
Consolidated clay					
$\tau(M)$	0.0021	52.7	111.5	0.0002	0.0005
$\tau(L)$	0.009	62.0	124.4	0.0050	0.0101
			$V[F] =$	0.0122	0.0298
			Spatial $V[F] =$	0.0176	
			$V_{ave}[F]$ at $R = 0.07 =$	$0.0122 + 0.0012 = 0.0134$	
			$\beta =$	$(1.427 - 1.0)/\sqrt{0.0134}$	
			$=$	3.69	

where r_o is the autocorrelation distance and L is the length of the curve of the intersection of the failure surface with the plane of analysis. For the geometry of this case the ratio becomes 0.2. This is applied to the spatial $V[F]$ of 0.046.

For the case with a single stage to 6 m, $E[F] = 1.50$, the critical failure circle is much smaller, and the reduction ratio for the spatial variation in Equation (14.8) becomes 0.7. This leads to significantly larger variation in the computed factor of safety and a much lower value of β (Table 14.7).

Figure 14.3 shows graphically how the different parameters contribute to the variance in the factor of safety for the two single stage cases. The bars represent the sum of the systematic uncertainty and the spatial uncertainty reduced due to averaging over the failure surface. The importance of the uncertainty in the strength of the lacustrine clay is obvious. The density of the fill and the strength of the marine clay are less important. The 6 m dike has a much smaller β than does the 12 m dike (1.84 versus 2.66), even though its factor of safety is slightly higher (1.50 versus 1.45). The reduced reliability index reflects a somewhat larger systematic uncertainty and a much larger spatial variability.

14.3.4.2 Multi-stage embankment

Table 14.6 shows the results for the multi-stage dike. The failure surface in this case is large (see Figure 14.2) and mostly horizontal, so the appropriate autocorrelation distance for the shear strength is also large. The reduction ratio for spatial variation of shear strength

Table 14.7 Summary of reliability results and effects of model error. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers)

Case (No. Stages)	$E[F]$[1]	$Var[F]$ Reduced spatial	$Var[F]$ Prior systematic	$Var[F]$ Model error[2]	$Var[F]$ Total	$\beta = \frac{E[F]-1}{\sqrt{Var[F]}}$
Sgl., $H = 6$ m						
w/o mod. err.	1.500	0.050	0.024		0.074	1.84
with mod. err.	1.575	0.050	0.024	0.012	0.086	1.96
Sgl., $H = 12$ m						
w/o mod. err.	1.453	0.009	0.020		0.029	2.66
with mod. err.	1.526	0.009	0.020	0.011	0.041	2.60
Multi, $H = 23$ m						
w/o mod. err.	1.427	0.0012	0.0122		0.0134	3.69
with mod. err.	1.498	0.0012	0.0122	0.0110	0.0244	3.19

[1] $E[F]$ with model error = $E[F]$ w/o model error × 1.05
[2] $V[e] = (0.07 \times \text{increased } E[F])^2$

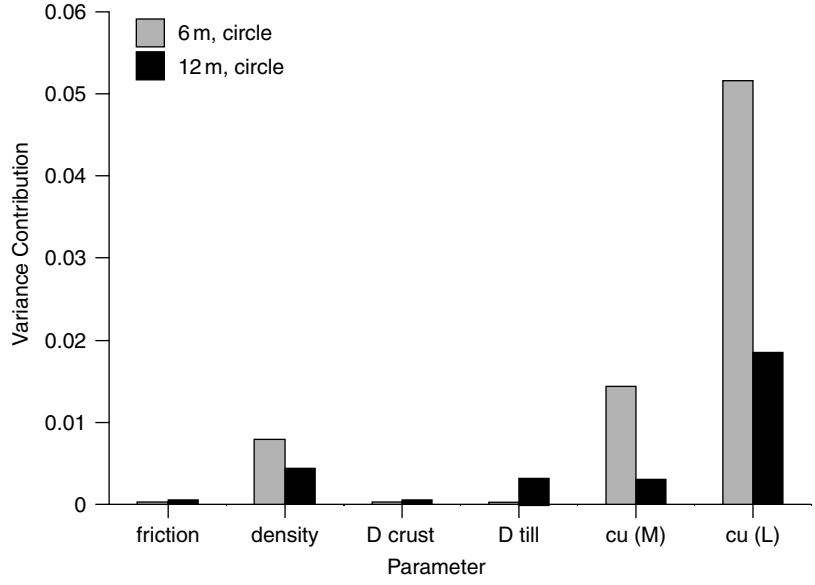


Figure 14.3 Variance components for single-stage dikes. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

becomes 0.07. Because the overall uncertainty is much lower in this case, the reliability index, at 3.69, is the largest of all the cases. As in the single-stage cases, the strength of the lacustrine clay and the density are the most important contributors to the overall

uncertainty. Since very little of the failure surface passes through the marine clay, its uncertainty contributes little. The failure surface passes about equally through the intact and consolidated lacustrine clays, and each contributes about equally to the uncertainty. Although the coefficients of variation are larger for the intact clay, the mean strength is smaller. The contributions of the different parameters are illustrated in Figure 14.4, which again shows the importance of the lacustrine clay. This figure shows the breakdown of the uncertainty between spatial and systematic contributions and also shows the important effect of averaging the spatial uncertainty.

14.3.4.3 Model error

Even if the soil parameters have been estimated correctly, the predictions of an engineering model can be expected to deviate from reality. Estimating the error introduced by this factor can be difficult; the most effective approach is to rely on empirical observations when they exist. An alternative is to use formal analysis. In the present case three sources of model error were considered: three-dimensional failure compared to two-dimensional analysis (i.e. end effects), failure to find the most critical failure surface in the analysis, and numerical and rounding error. Azzouz *et al.* (1983) examined the contributions of end effects in eighteen case histories. They found that the ratio between the three-dimensional and the plane strain F was 1.11 with a standard deviation of 0.06. These numbers were rounded to 1.1 and 0.05. Failure to find the lowest factor of safety is considered to overestimate F by about 5% with a standard deviation of about 5%. The third factor was incorporated by assuming that it contributed 2% to the variability of the results. The combined effect of these three sources is to increase $E[F]$ by about 5% and to create an

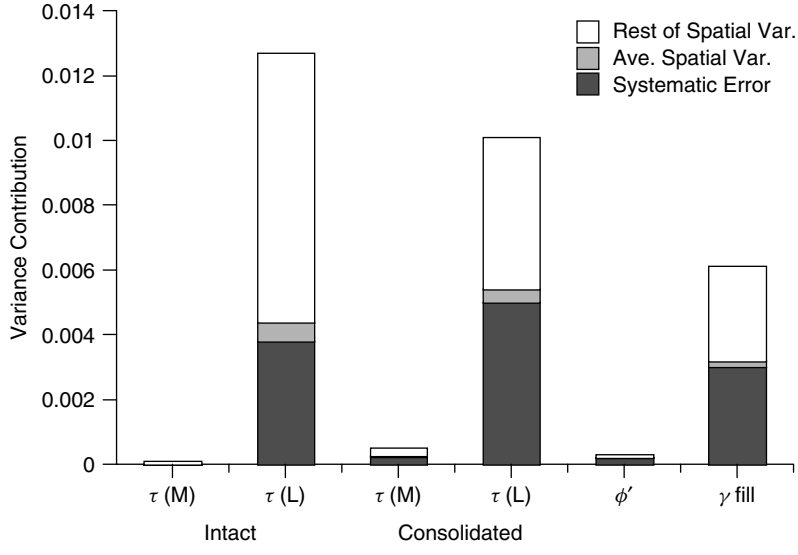


Figure 14.4 Variance components for multi-stage dike. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

added systematic uncertainty described by

$$\Omega = (0.05^2 + 0.05^2 + 0.02^2)^{1/2} = 0.073 \approx 0.07 \tag{14.9}$$

14.3.4.4 Summary of results for three cases

The results for all three cases are summarized in Table 14.7. Although the factor of safety in the multi-stage case is slightly less than for the single stage cases, the reliability index is significantly larger. Figure 14.5 summarizes the overall contributions of systematic and spatial uncertainty. It shows the effect of the reduction in spatial uncertainty from averaging the soil parameters over the failure surface. Spatial variability becomes almost insignificant for the 23 m high dike constructed in two stages. It should be noted that, in order to simplify calculations, the values of R used for spatial averaging were based on analyses of the field vane data and therefore should apply only to the undrained strength for the single stage cases with H of 6 and 12 m and the part of the multi-stage case involving the intact clay. Different values of R would exist for the fill density, friction angle, and normally consolidated undrained shear strength.

Table 14.7 shows that introducing model error can have unexpected results. For the single stage dike with $H = 6$ m, β increases from 1.84 to 1.96. There is little effect for the single stage dike with $H = 12$ m; β changes from 2.66 to 2.60. For the multi-stage dike β decreases from 3.69 to 3.19. Thus, the computed reliability index can increase or decrease when the effects of model uncertainty are introduced. For a particular configuration and set of properties an increase in $E[F]$ might be offset by an increase in $\sigma[F]$.

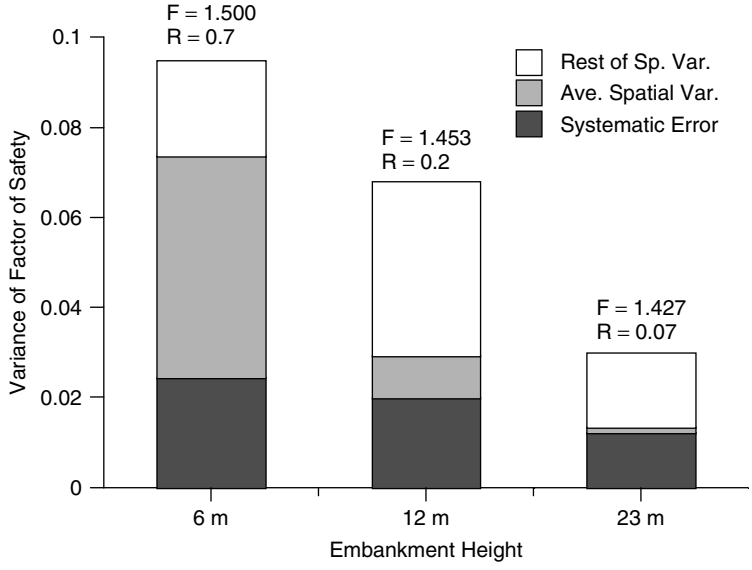


Figure 14.5 Separation of spatial and systematic uncertainties. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

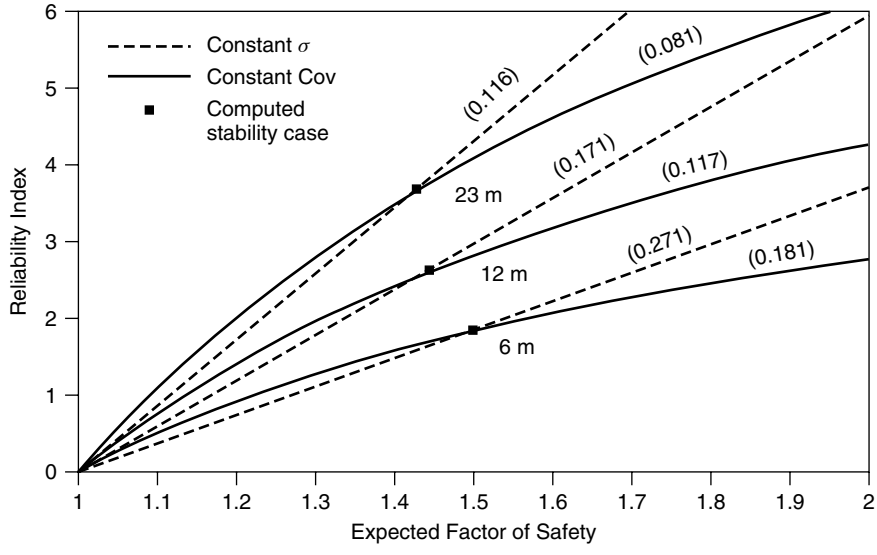


Figure 14.6 Projection of reliability index versus factor of safety (without model error). (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

14.4 Extrapolations and Consequences

The above results follow from detailed consideration of a limited set of geometries and parameters. In order to develop consistent criteria for design, it is necessary to extrapolate them to other conditions. Also, a consideration of the results of the analyses reveals much about the interactions between the parameters of the design problem. As the first page of Hamming's (1962) book on numerical methods states, "The purpose of computing is insight, not numbers."

14.4.1 Extrapolation to Other Values of Parameters

There are two simple ways to extrapolate the results. On the one hand, one could assume that for a basic geometry the variance or standard deviation of F is constant and apply Equation (14.3). On the other hand, one could assume the coefficient of variation is constant and apply the modified form of the equation:

$$\beta = \frac{E[F] - 1}{E[F] \cdot \Omega[F]} \tag{14.10}$$

The consequences of the two assumptions are shown in Figure 14.6. Constant variance means that the absolute uncertainty in F does not change with the target value of $E[F]$. Constant coefficient of variation means that the uncertainty in F changes in proportion to $E[F]$. Since the derivatives (or in this case the divided differences $\Delta F/\Delta X_i$) are for the most part nearly constant for different values of F, a projection based on constant variance seems the better choice. It is also computationally simpler.

Figure 14.6 shows the projected nominal probabilities of failure as functions of the computed factors of safety for the example stability problems. The probabilities are computed on the assumption that F is Normally distributed. Until now the form of the distribution of F has not entered into these calculations. Curves are plotted for results both with and without the effects of model uncertainty. From this figure the designer can determine the computed factors of safety that correspond to a desired target nominal probability of failure.

14.4.2 Effects of types of failure

The uncertainties associated with the three design cases differ from one another because each reflects a different mixture of spatial variability and systematic uncertainty. This has an important implication. A 1% probability of failure deriving from spatial variability alone implies that, on average, one per cent of the length of a long dike would fail. Conversely, a one per cent probability of failure due to systematic uncertainty alone would imply that, on average, one out of one hundred of a large number of similar dikes would fail along its entire length. In any specific case, such as this one, the probability of failure derives from both spatial variability and systematic uncertainty.

Figure 14.5 shows the contributions of the two components for the three cases. In the single stage case with $H = 6\,\text{m}$ the distribution of F for different sections along the dike is broad, but the uncertainty on the location of the mean is small. In the multi-stage case the distribution of F along the dike is narrow, but there is proportionally more uncertainty on the location of the mean. In other words, the uncertainty in the case of a single stage dike with $H = 6\,\text{m}$ deals principally with the fraction of the dike that will fail, but the uncertainty in the case of a multi-stage dike deals principally with the failure of all or a large portion of the dike.

14.4.3 Target Probability of Failure

Baecher *et al.* (1980) determined that the historical rate of failure of modern, well-engineered dams is about 10^{-4} per dam-year. The details of failures are often imprecise in the catalogues from which this result was computed, but it appears that about one third of the failures were due to overtopping and another one third to internal erosion, piping, or seepage. These numbers are close to what would be expected from professional judgment and experience. The remaining one third are due to slides and other mechanisms, which include the subject of the present analysis.

The relative contribution of different modes of failure should be kept in mind when a target probability of failure is being selected. It makes little sense to reduce the computed probability of failure due to slope stability problems if other causes of failure are not addressed at the same time. The historical data would suggest that, for a dam with a life of 100 years, the overall probability of failure is 0.01, in the absence of specific attention during design and construction to mitigate the hazards. Since well less than one third of the failures are due to slope instability *per se*, a target probability of failure of slope stability failure of about 0.001 seems reasonable for design purposes and consistent with past experience. Figure 14.7 indicates that to achieve this target probability the stability analysis should give factors of safety of 1.84, 1.53, and 1.36 for the 6 m, 12 m, and 23 m cases, respectively. If the effects of model uncertainty were included, the values of

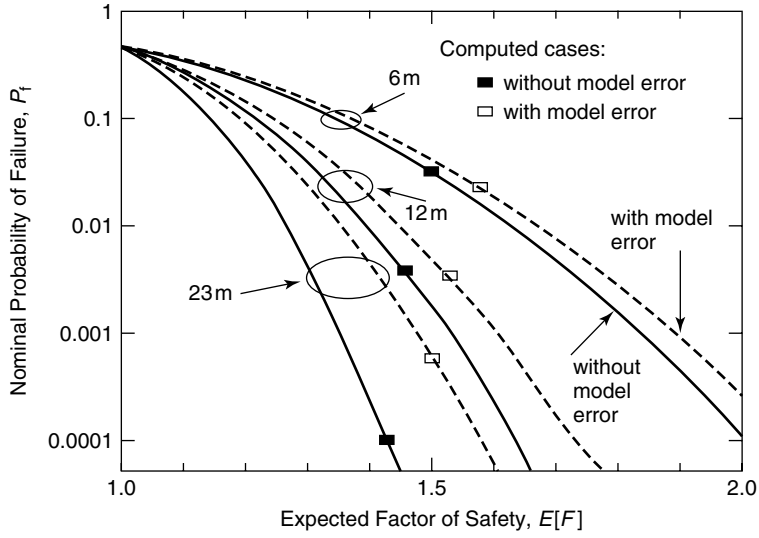


Figure 14.7 Nominal probability of failure versus computed factor of safety. (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

desired factor of safety would increase to 1.91, 1.63, and 1.48. As the computation of the effects of model uncertainty was probably somewhat too conservative and the effects on the desired factors of safety relatively small, this effect was ignored in further results.

14.4.4 Consistent Factors of Safety

The present analysis was concerned with stability during construction due to failure of the foundation clay. The consequences of such a failure can be expected to consist primarily of the costs of reconstruction. Failure of the embankment after the facility is in operation would involve much larger economic costs as well as other public consequences. This means that different target reliability indices or probabilities of failure are appropriate for different stages in the life of the project. The selection of such values is a difficult task for owners, designers, and public officials. Even for the present problem, the consequences of failure are quite different in the different cases. Because experience with staged construction is less fully evaluated and the consequences of failure of a major portion of the completed multi-stage embankment are expected to be greater than those of the failure of the 12 m high embankment, the target probability for the multi-stage case might be reduced to 0.0001, implying a calculated factor of safety of about 1.43. Conversely, the consequences of failure of a portion of the smallest dike might be only one tenth as great. The target probability then becomes 0.01 and the desired factor of safety about 1.63. Based on the revised target probabilities, one obtains the consistent, desired factors of safety shown in Table 14.8. The range in factors of safety for the different cases is now much smaller than for the same $p_f = 0.001$. These results led to the recommendation that $F = 1.50$ be adopted for feasibility studies and preliminary cost estimates for both single

Table 14.8 Consistent factors of safety (without model error). (Christian, J. T., Ladd, C. C. and Baecher, G. B., 1994, 'Reliability Applied to Slope Stability Analysis,' *Journal of Geotechnical Engineering, ASCE*, Vol. 120, No. 12, pp. 2180–2207, reproduced by permission of the American Society of Civil Engineers.)

Case	Target p_f	Consistent factor of safety
Single stage, $H = 6$ m, circular failure surface	0.01	1.63
Single stage, $H = 12$ m, circular failure surface	0.001	1.53
Multi-stage, $H = 23$ m, wedge failure surface	0.0001	1.43

and multi-stage dikes. It is worth noting that a previous committee had recommended using $F = 1.4$ for minor dikes and $F = 1.7$ for major (i.e. multistage) dikes.

14.5 Conclusions from the James Bay Study

The review of reliability analysis and its application to the problems of the embankments at the James Bay project illustrate several major points that apply to reliability analyses in general. The relative importance of the different observations will vary from case to case.

14.5.1 General Conclusions

The reliability index β provides a more meaningful measure of stability than the factor of safety F. Geotechnical engineers have long recognized that F has little physical meaning and that the choice of a satisfactory value is fraught with difficulty. In contrast, the reliability index describes safety by the number of standard deviations (i.e. the amount of uncertainty in the calculated value of F) separating the best estimate of F from its defined failure value of 1.0. It is a more explicit approach.

The uncertainty in the values of the soil properties is a major contributor to the uncertainty in the stability of slopes and embankments. It consists of two portions: scatter and systematic error. The first comprises real spatial variability and noise or measurement error. These represent variations about the trend of the parameters. The systematic error includes both a statistical uncertainty in the expected values and the effects of bias. The latter is much the more difficult to evaluate.

The structure of the spatial variation can be used to estimate the level of random noise in soil property data and to eliminate it from the calculation of reliability index. The effects of the spatial variability on the computed reliability index are further reduced because the variability is averaged over a region or failure surface, and it is only its average contribution to the uncertainty that is of interest.

Several difficulties arise in practical assessment of scatter and systematic error. There are often not enough data to evaluate the autocorrelation distance statistically. Evaluating the effects of bias introduces many additional concerns. Bias is often very important, and

the associated uncertainty can be a significant contributor to the overall uncertainty. Since bias is often ignored in theoretical treatments of analytical procedures and it is difficult to quantify, the engineer must often rely on judgment to establish its contribution. The strength of the reliability analysis is not that one can get a better estimate of each of these uncertainties but that one can deal with them explicitly and coherently.

Uncertainties in soil properties yield a lower bound estimate of the probability of failure, not the absolute probability of failure. That requires a more elaborate probabilistic risk analysis involving fault trees or other methods of evaluating risk due to all contingencies (e.g. Whitman 1984). However, for most practical applications the calculation of relative probability of failure is sufficient for parametric analysis.

A 1% chance that the whole slope will fail is different from a near certainty that one per cent of a long embankment will fail, and reliability analysis allows such a distinction.

14.5.2 Analysis of the James Bay Embankments

The analysis of the James Bay embankments illustrates a multi-step procedure for evaluating the reliability index. Once the uncertainties in the soil properties, geometry, loads, and other contributing factors have been established, it is necessary to propagate their effects through the stability analysis. Equation (14.2) provides one way to do so. The FOSM approach is a powerful approximate method to deal with the calculations. The Taylor series combined with a divided difference approximation to the derivatives has proven effective.

In the analysis, the effects of both spatial and systematic uncertainty are considered. The averaging effect tends to reduce greatly the contribution of the spatial uncertainty. In the case of the multi-stage construction spatial uncertainty contributes so little that it could almost be ignored without affecting the results. The small embankment with a short failure surface has a large uncertainty in F due to spatial variation in undrained strength. This means that several small failures along the length of the embankment are likely. The multi-stage embankment with a long failure surface has almost all the uncertainty in systematic error, which leads to a small probability of failure along a long section of the embankment. When the autocorrelation distance is not known, the analysis can at least give point estimates of the reliability index without providing information on the length of embankment likely to fail. The example also illustrates the contribution of model error, and the difficulty in determining the magnitude of such bias effects.

However the solution is carried out, it is important to deal rationally with the spatial variability of the parameters. In particular, the averaging effect on variations that are due to inherent spatial scatter and are not systematic must be accounted for, or the resulting probabilities of failure will be too high. The technique used here is one appropriate method.

14.5.3 Improvements in the James Bay Analysis

The preceding sections describe the reliability analysis of the James Bay dikes as it was actually carried out. Several factors were not included in the analysis, and some of the procedures could have been improved upon. This is typical of most engineering calculations; they are done under the pressures of the job and with the tools available

at the time. Nevertheless, we should mention some areas where the procedures could be modified.

First, the analysis ignored the possible correlation between the parameters and variables that entered into the calculation. Little additional effort is needed to incorporate correlation in an FOSM analysis. Equation (14.2) includes all the necessary terms, and more rows can be added to Tables 14.5 and 14.6 to incorporate the correlated terms. In the case of the James Bay dikes there was not enough information available to include correlation effects. Furthermore, the effect of most reasonable correlations would be to increase the reliability and decrease the probability of failure. Therefore, ignoring the correlation effects was conservative in this case, but, if data were available to establish the correlation coefficients, their inclusion would certainly be a meaningful extension of the analysis.

The analysis of the multi-stage dike relied on a Morgenstern-Price analysis of a failure that incorporated a long, flat middle section. More recent work indicates that a failure mechanism involving a curved or wavy middle section can give lower factors of safety in some cases (Ladd 1996).

The analysis found the minimum factor of safety for the various modes of failure. Further iterations could have been introduced to find the failure mode that gave the lowest value of β instead of the lowest value of F. It does not seem that the overall results would have changed very much, but it would be a desirable improvement in methodology, especially if the results were very near the margin for a decision.

The procedure for reducing the contribution of the spatial variations by averaging their effects over the length of the failure surface assumed that the spatial correlation for the strength parameters applied to all the variables. This is probably not the case. A better procedure would be to apply the averaging correction to each variable separately.

14.6 Final Comments

In summary the FOSM method consists of the following steps:

1. Identify the significant variables contributing to the uncertainty.
2. Find the mean values, variances, correlation coefficients, and autocorrelation distances for the variables.
3. Determine how the variances are distributed between spatial and systematic uncertainty and remove measurement errors, if possible.
4. Find the expected value of the performance function, in this case F.
5. Find the partial derivatives of the performance function with respect to the variables, usually employing some form of numerical differencing.
6. Using a format like Tables 14.5 or 14.6, find the contributions of each variable to the systematic and spatial variance of the performance function. Modify the spatial contribution to account for averaging effects.
7. Compute the variance in the performance function, as in Tables 14.5 and 14.6.
8. Compute β and the probability of failure.
9. Examine the results to determine the relative contributions of various parameters and to extrapolate the results into meaningful conclusions.

One of the great advantages of the FOSM method is that it reveals the relative contribution of each variable to the overall uncertainty in a clear and easily tabulated manner.

This is very useful in deciding what factors need more investigation and even in revealing some factor whose contribution cannot be reduced by any realistic procedure. Many of the other methods do not provide this useful information.

Evaluation of the partial derivatives or their numerical approximations is critical to the use of the method. When numerical differentiation is used, the increment in the independent variable should be as small as possible until rounding error affects the result. It is customary to take the derivatives at a point defined by the expected values of the independent variables. In some cases the point at which the derivatives are taken can have an effect on the result. This question is central to the geometrical definition of the reliability index, which is the subject of Chapter 16.

15 Point Estimate Methods

In a paper slightly longer than two pages appearing in the October 1975 Transactions of the National Academy of Science (USA), Emilio Rosenblueth of the National Autonomous University of Mexico published a method for numerically approximating the moments of functions of random variables (Rosenblueth 1975). Starting from the low-order moments of the independent variable X, the method provides approximations for the low-order moments for the dependent variable Y. The procedure, which Rosenblueth called the *point-estimate method*, has become a staple of geotechnical reliability analyses. The method is straightforward, is easy to use, and requires little knowledge of probability theory. It is widely employed in practice, often to good effect (Harr 1987; Wolff 1996; Duncan 1999) and has been commented on further by several workers in reliability analysis (Lind 1983; Wong 1985; Harr 1989; Li 1992).

While the point-estimate method is popular in practice, it has many detractors. Rosenblueth's original paper contains many unfortunate typographical errors that make the equations difficult to interpret, and the very simplicity of the method suggests to some that the answers are overly approximate. A deeper look at the point-estimate method brings to light interesting relationships with well-known numerical methods of approximate quadrature and identifies categories of problems for which it is well suited, as well as some for which it is not. Rosenblueth was well acquainted with both the theory and practice of numerical analysis and knew the appropriate and inappropriate applications of the point-estimate method to engineering problems. This chapter attempts to explicate these relations and applications in a way that is more accessible to the practicing geotechnical profession and to place the point-estimate within the broader category of numerical methods, such as the finite element method. The method is a powerful tool and more rigorously based than many workers in geotechnical engineering and reliability analysis give it credit for being.

15.1 Mathematical Background

The basic mathematical problem is that of a random variable or variables X with probability distribution function (pdf) $f_X(x)$ and another variable Y, which is a deterministic

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

function of X, $Y = g(X)$. X could be soil properties, geometrical parameters, seismic loadings, and so on; Y could be a factor of safety, settlement, quantity of flow, and so on. Assume that $g(X)$ is well behaved and that the mth order moments of $f_X(x)$ exist. The question then is how to approximate the low-order moments of $f_Y(y)$ using only the low-order moments of $f_X(x)$ and the function $g(X)$.

The approaches most commonly used in geotechnical reliability analysis (e.g. FOSM) start by approximating $g(X)$ by a Taylor series, truncating it to low order terms, and computing the approximate moments from the result. This requires calculating or approximating the derivatives of $g(X)$, which can involve significant algebraic or numerical effort, as well as iteration when $g(X)$ is non-linear. Rosenblueth approached the problem by replacing the continuous random variable X with a discrete random variable whose probability mass function (pmf) $p_X(x)$ has the same moments of order m as does $f_X(x)$. He transformed $p_X(x)$ through $Y = g(X)$ to obtain another discrete function with a pmf denoted $p_Y(y)$. He then used this pmf to calculate the moments, which were assumed to approximate the moments of Y in the continuous case.

The first moment of $f_X(x)$ about the origin is the mean, μ_X:

$$\mu_X = \int x \cdot f_X(x) \cdot dx \tag{15.1}$$

The higher-order central moments of $f_X(x)$ of order m are

$$\mu_{Xm} = \int (x - \mu_X)^m \cdot f_X(x) \cdot dx \tag{15.2}$$

We use μ_{Xm} for the mth central moment to distinguish it from the mth power of μ_X. The second central moment, μ_{X2}, is the variance, and its square root is the standard deviation, σ_X. The discrete pmf has non-zero probabilities at a limited number of discrete points only. The corresponding moments of the discrete pmf $p_X(x)$ are

$$\mu_{Xm} = \sum (x - \mu_X)^m \cdot p_X(x) \tag{15.3}$$

Then, equating the moments of $f_X(x)$ and $p_X(x)$ yields

$$\mu_{Xm} = \int (x - \mu_X)^m \cdot f_X(x) \cdot dx = \sum (x - \mu_{Xm}) \cdot p_X(x) \tag{15.4}$$

It is well known that any probability density function can be represented to any desired degree of accuracy by taking moments of high enough order. In principle, there are infinitely many pmf's of X satisfying the low-order moments of Equation (15.4). To limit these to a unique representation, Rosenblueth considered only pmf's with two, three, or some other small number of discrete masses, the number chosen depending on the order of the moments in Equation (15.4).

This approach falls within a long-standing practice of approximating complicated functions by a series of simpler functions. As Equations (15.1) and (15.2) show, the moments of a pdf are integrals, and the large body of experience with numerical quadrature as an approximation to integration is relevant to the present problem. In particular, Gaussian quadrature procedures are concerned with choosing the optimal values of the coordinates at

which to evaluate the integrand (called the Gauss points in the finite element literature) and the corresponding weights. Rosenblueth's method is in fact an application of the Gaussian quadrature procedures to the problem of finding the moments of a random function.

15.2 Rosenblueth's Cases and Notation

Rosenblueth (1975) deals with three cases: (1) when Y is a function of one variable X, whose mean, variance, and skewness are known; (2) when Y is a function of one variable X whose distribution is symmetrical and approximately Gaussian; and (3) when Y is a function of n variables $X_1, X_2, \ldots, X_n$, whose distributions are symmetric and which may be correlated. In most cases the calculations are made at two points, and Rosenblueth uses the following notation:

$$E[Y^m] \approx P_+ y_+^m + P_- y_-^m \tag{15.5}$$

In this equation,

Y is a deterministic function of X, $Y = g(X)$,
$E[Y^m]$ is the expected value of Y raised to the power m,
y_+ is the value of Y evaluated at a point x_+, which is greater than the mean, μ_x,
y_- is the value of Y evaluated at a point x_-, which is less than μ_x, and
P_+, P_- are weights.

The problem is then to find appropriate values of x_+, x_-, P_+, and P_-.

15.2.1 Case 1

In the first case – Y a function of one variable X whose mean, standard deviation, and skewness are known – Rosenblueth gives four conditions that must be satisfied for the low-order moments of X to be modeled accurately:

$$P_+ + P_- = 1 \tag{15.6}$$

$$P_+ x_+ + P_- x_- = \mu_x \tag{15.7}$$

$$P_+ (x_+ - \mu_x)^2 + P_- (x_- - \mu_x)^2 = \sigma_x^2 \tag{15.8}$$

$$P_+ (x_+ - \mu_x)^3 + P_- (x_- - \mu_x)^3 = \nu_x \sigma_x^3 \tag{15.9}$$

In these equations σ_x is the standard deviation of X, and ν_x is the skewness (i.e. $\nu_X \equiv \mu_X^3/\sigma_X^3$). A simple solution to these equations (Rosenblueth 1981) is

$$x_+ = \mu_x + \left[\frac{\nu_x}{2} + \sqrt{1 + \left(\frac{\nu_x}{2}\right)^2} \right] \sigma_x \tag{15.10}$$

$$x_- = \mu_x + \left[\frac{\nu_x}{2} - \sqrt{1 + \left(\frac{\nu_x}{2}\right)^2} \right] \sigma_x \tag{15.11}$$

$$P_+ = \frac{1}{2}\left[1 - \frac{\nu_x}{2} \cdot \frac{1}{\sqrt{1 + (\nu_x/2)^2}}\right] \tag{15.12}$$

$$P_- = 1 - P_+ \tag{15.13}$$

A great simplification occurs when the skewness is zero or negligible. The distribution of X is then symmetric, and

$$P_+ = P_- = \tfrac{1}{2} \quad x_+ = \mu_x + \sigma_x \quad x_- = \mu_x - \sigma_x \tag{15.14}$$

15.2.2 Case 2

In the case that X is symmetric and approximately Gaussian, Rosenblueth proposes that x can be estimated at more than two points. For example, a three-point estimate would use a central point at $x = \mu_x$ and two points x_+ and x_- symmetrically distributed about the mean. The weight for the central point is designated P, and the other notation remains the same. Then

$$2P_+ + P = 1 \tag{15.15}$$

$$2P_+(x_+ - \mu_x)^2 = \sigma_x^2 \tag{15.16}$$

$$2P_+(x_+ - \mu_x)^4 = 3\sigma_x^4 \tag{15.17}$$

The last equation follows from the fact that the expected value of the fourth central moment of a normal distribution is $3\sigma_x^4$. The solution of these equations gives the following results, analogous to those in Equations (15.10)–(15.13):

$$P = \frac{2}{3},\ P_+ = P_- = \frac{1}{6} \tag{15.18}$$

$$x_\pm = \mu_x \pm \sqrt{3}\sigma_x \tag{15.19}$$

By analogy to case 1,

$$E[Y^m] \approx P_-(y_-)^m + P(y_\mu)^m + P_+(y_+)^m \tag{15.20}$$

where y_μ is the value of Y evaluated at μ_x.

Rosenblueth observes that higher order approximations can be obtained by using more points and suggests in passing that the points can be found from the zeros of Hermite polynomials. We will return to this issue later.

15.2.3 Case 3

The most widely used application of Rosenblueth's method follows from the third case – when Y is a function of n variables whose skewness is zero but which may be correlated. The procedure is a generalization of the procedure in case 1 when the skewness is ignored. The procedure chooses 2^n points selected so that the value of each variable is one standard deviation above or below its mean. Thus, if there were two variables X_1 and

X_2, the four points would be $(\mu_{X_1} + \sigma_{X_1}, \mu_{X_2} + \sigma_{X_2})$, $(\mu_{X_1} + \sigma_{X_1}, \mu_{X_2} - \sigma_{X_2})$, $(\mu_{X_1} - \sigma_{X_1}, \mu_{X_2} + \sigma_{X_2})$, and $(\mu_{X_1} - \sigma_{X_1}, \mu_{X_2} - \sigma_{X_2})$. If the variables are not correlated, the function Y is evaluated at each of the four points, and the weight for each point is 0.25.

If the correlation coefficient between X_1 and X_2 is ρ, the points are still located one standard deviation above or below the mean of each variable, but the weights are changed. At the two points $(\mu_{X_1} + \sigma_{X_1}, \mu_{X_2} + \sigma_{X_2})$ and $(\mu_{X_1} - \sigma_{X_1}, \mu_{X_2} - \sigma_{X_2})$ the weights become $(1 + \rho)/4$, and at the other two points they are $(1 - \rho)/4$. The locations of the points and the weights are illustrated in Figure 15.1.

When Y is a function of three variables, X_1, X_2, and X_3, there are eight points, which are located at each combination one standard deviation above or below the mean of all the variables. In defining the weights Rosenblueth uses a set of $+$ and $-$ signs as subscripts on the weight P. The convention is that the first sign refers to X_1, the second to X_2, and the third to X_3. If the point is at $\mu_{x_i} + \sigma_{x_i}$, the sign is positive; otherwise it is negative. For example, P_{-+-} refers to the point $(\mu_{x_1} - \sigma_{x_1}, \mu_{x_2} + \sigma_{x_2}, \mu_{x_3} - \sigma_{x_3})$. The convention is illustrated in Figure 15.2. Also ρ_{12} is the correlation coefficient between X_1 and X_2, and so on. Then

$$
\begin{aligned}
P_{+++} &= P_{---} = \tfrac{1}{8}(1 + \rho_{12} + \rho_{23} + \rho_{31}) \\
P_{++-} &= P_{--+} = \tfrac{1}{8}(1 + \rho_{12} - \rho_{23} - \rho_{31}) \\
P_{+-+} &= P_{-+-} = \tfrac{1}{8}(1 - \rho_{12} - \rho_{23} + \rho_{31}) \\
P_{+--} &= P_{-++} = \tfrac{1}{8}(1 - \rho_{12} + \rho_{23} - \rho_{31})
\end{aligned}
\tag{15.21}
$$

For any two variables, when the point coordinates both have the same sense of the standard deviation, the sign of the corresponding correlation coefficient is positive, and when the senses are opposite, the sign is negative.

The generalization to more than three variables is obvious. If there are n variables, then 2^n points are chosen to include all possible combinations with each variable one standard deviation above or below its mean. If s_i is $+1$ when the value of the ith variable is

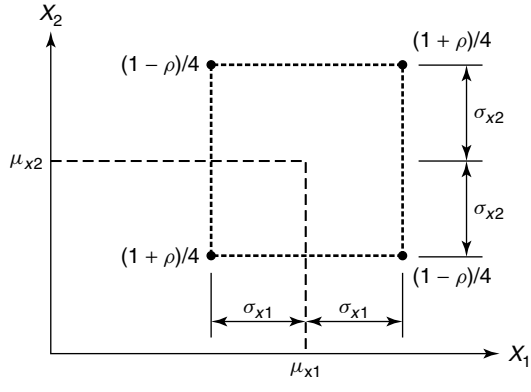


Figure 15.1 Rosenblueth's points and weights for two variables, correlated or uncorrelated. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers).

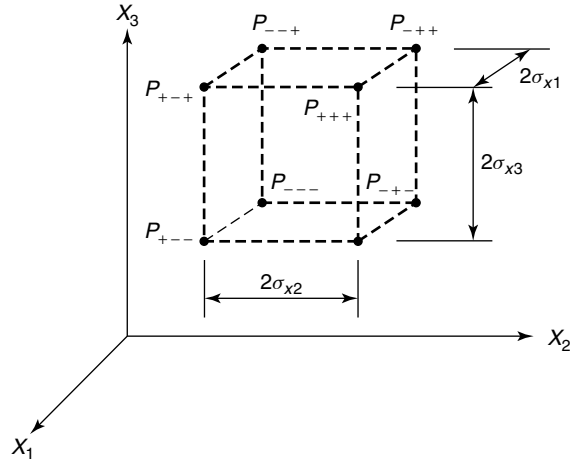


Figure 15.2 Rosenblueth's points and weights for three variables, correlated or uncorrelated. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

one standard deviation above the mean and -1 when the value is one standard deviation below the mean, then the weights are

$$P_{(s_1 s_2 \ldots s_n)} = \frac{1}{2^n}\left[1 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(s_i)(s_j)\rho_{ij}\right] \tag{15.22}$$

Again,

$$E[Y^m] \approx \sum P_i (y_i)^m \tag{15.23}$$

where y_i is the value of Y evaluated at x_i and i is an appropriate combination of $+$ and $-$ signs indicating the location of x_i which are the corners of a hypercube.

When the variables are uncorrelated, Equation (15.21) reduces to $P = 1/2^n$. This still requires evaluating Y at 2^n points, which can be onerous. Rosenblueth proposed that, when the variables are uncorrelated, the following approximations (in his notation) are also a valid generalization of Equations (15.14):

$$\frac{\overline{Y}}{y} = \frac{\overline{Y}_1}{y}\frac{\overline{Y}_2}{y}\cdots\frac{\overline{Y}_n}{y} \text{ and} \tag{15.24}$$

$$1 + V_Y^2 = (1 + V_{Y_1}^2)(1 + V_{Y_2}^2)\ldots(1 + V_{Y_n}^2) \tag{15.25}$$

In these equations, $\overline{Y}$ is the estimate of the expected value of Y, y is the value of Y evaluated at the mean values of the variables (i.e. $y = y(\mu_{x_1}, \mu_{x_2}, \ldots \mu_{x_n})$), $\overline{Y}_i$ is the mean value of Y calculated with all variables but the ith variable held constant at their mean values, V_Y is the coefficient of variation of Y that is to be found ($COV = \sigma_Y/\mu_Y$), and V_{Y_i} is the coefficient of variation of Y calculated as if the ith variable were the only random variable

and the others were all held constant at their mean values. The $\overline{Y}_i$ and V_{Y_i} can be calculated using the procedures of Equations (15.5) and (15.14). This method requires only $2n + 1$ evaluations of Y – two evaluations for each variable and one for the estimate of y.

15.3 Numerical Results for Simple Cases

The multivariate form of Rosenblueth's method for uncorrelated variables has been applied widely in geotechnical reliability studies (Wolff 1996; Duncan 1999), and will be discussed further below. However, it is worthwhile to examine the accuracy of the point-estimate method by looking at the results for relatively simple problems for which the exact answers are known.

15.3.1 Example – Vertical Cut in Cohesive Soil

The first simple case is the vertical cut in a cohesive soil that was used in Chapter 13. The margin of safety M is

$$M = c - \gamma H/4 \tag{15.26}$$

and when the height of the cut H is fixed, the exact values for the mean and variance of M are

$$\begin{aligned} \mu_M &= \mu_c - \mu_\gamma H/4 \\ \sigma_M^2 &= \sigma_c^2 + (H/4)^2\sigma_\gamma^2 - 2\rho_{c\gamma}\sigma_c(H/4)\sigma_\gamma \end{aligned} \tag{15.27}$$

When $H = 10\,\text{m}$, $\mu_c = 100\,\text{kPa}$, $\mu_\gamma = 20\,\text{kN/m}^3$, $\sigma_c = 30\,\text{kPa}$, $\sigma_\gamma = 2\,\text{kN/m}^3$, and $\rho_{c\gamma} = 0.5$, the exact results are

$$\begin{aligned} \mu_M &= 50\,\text{kPa} \\ \sigma_M^2 &= 775\,(\text{kPa})^2 \\ \sigma_M &= 27.84\,\text{kPa} \end{aligned} \tag{15.28}$$

Table 15.1 shows the calculations using Rosenblueth's multivariate method with M evaluated at four points. The first column gives the number of each point. The second gives the value of c at each point, and the third gives the value of γ at the point. It can be seen that the values of c alternate between one σ above and below the mean while the first two entries for γ are below the mean and the last two are above. This arrangement gives all four possible combinations of points above and below the mean for the two variables. The fourth column contains the value of M calculated from the values for each point. The fifth column lists the corresponding values of the weight assigned to that point. Note that this weight is $(1 \pm \rho)/4$ and that the sign is positive when the values in columns 2 and 3 are both on the same side of the means and negative when they are on opposite sides. Multiplying the values in columns 4 and 5 gives the values in the sixth column, which are the weighted values of M. These are added to give the estimate of μ_M, which

Table 15.1 Results from point-estimate method for vertical cut

Point	c (kPa)	γ (kN/m^3)	M (kPa)	p	pM	pM^2
1	70	18	25	0.375	9.375	234.375
2	130	18	85	0.125	10.625	903.125
3	70	22	15	0.125	1.875	28.125
4	130	22	75	0.375	28.125	2109.375
				Sum =	50.000	3275.000
				$Var = 3275 - (50)^2 = 775$		$\sigma = 27.84$
$H = 10\,\text{m}$, $\mu_c = 100\,\text{kPa}$, $\sigma_c = 30\,\text{kPa}$, $\mu_c = 20\,\text{kN/m}^3$, $\sigma_c = 2\,\text{kN/m}^3$						

is 50. Similarly, the products of the entries in column 5 and the square of the entries in column 4 give the entries in column 7, and the sum of these gives the estimate of $E[M^2]$. The difference between this and the square of the estimate of the mean gives the estimate of the variance. The square root of the variance is the standard deviation. The calculations in the next-to-last row show that this gives 27.84. The values of the mean and the standard deviation of M agree exactly with the theoretical values. These give a reliability index $\beta = 1.796$, and, if M is Normally distributed, a probability of failure $p_f = 3.6 \times 10^{-2}$. When the correlation coefficient is set to zero, μ_M is unchanged, and σ_M in both the exact and Rosenblueth calculations becomes 30.41 kPa. Then $\beta = 1.644$, and $p_f = 5.0 \times 10^{-2}$.

15.3.2 Example – Culmann Failure Mechanism along a Plane

A more complicated example is the margin of safety of a slope against sliding on a single plane, otherwise known as the Culmann analysis (Taylor 1948). The margin of safety is

$$M = c + \left[\frac{1}{2}\frac{H}{\sin\psi}\sin(\psi - \theta)\cos\theta\right][\tan\phi - \tan\theta]\gamma \tag{15.29}$$

where θ is the inclination of the failure plane, ψ is the inclination of the front of the slope, H is the vertical distance from the crest to the toe of the slope, and the other variables have the conventional soil mechanics meanings. If the only uncertain variables are the unit weight, the cohesion, and the tangent of the friction angle (γ, c, and $\tan\phi$), and if these variables are mutually independent, the mean and variance of M can be expressed exactly. For the following values of the parameters:

$$\begin{aligned} H &= 10\,\text{m}, \theta = 20^\circ, \psi = 26^\circ \\ \mu_\gamma &= 22\,\text{kN/m}^3, \sigma_\gamma = 2.2\,\text{kN/m}^3 \\ \mu_c &= 5\,\text{kPa}, \sigma_c = 0.5\,\text{kPa} \\ \mu_{\tan\phi} &= 0.267949(\phi = 15^\circ), \sigma_{\tan\phi} = 0.018756(\Omega = 0.07) \end{aligned} \tag{15.30}$$

the exact analytical results are

$$\mu_M = 2.633\,\text{kPa} \quad \sigma_M = 0.722\,\text{kPa} \tag{15.31}$$

Table 15.2 Results from point-estimate method for Culmann slope

Point	γ(kN/m^3)	c (kPa)	tan(ϕ)	M (kPa)	p	pM	pM^2
1	24.2	5.5	0.286706	3.405	0.125	0.4256	1.4493
2	24.2	5.5	0.249193	2.388	0.125	0.2985	0.7128
3	24.2	4.5	0.286706	2.405	0.125	0.3006	0.7230
4	24.2	4.5	0.249193	1.388	0.125	0.1735	0.2408
5	19.8	5.5	0.286706	3.786	0.125	0.4733	1.7917
6	19.8	5.5	0.249193	2.954	0.125	0.3693	1.0908
7	19.8	4.5	0.286706	2.786	0.125	0.3483	0.9702
8	19.8	4.5	0.249193	1.954	0.125	0.2443	0.4773
					Sum =	2.6333	7.4559
					$Var = 7.4559 - (2.6333)^2 = 0.52160$		$\sigma = 0.7222$
				$H = 10\,\text{m}$, $\theta = 20\,\text{deg}$, $\psi = 26\,\text{deg}$			

Table 15.2 shows the calculations for Rosenblueth's method with three independent variables. The arrangement is similar to that of Table 15.1. In this case all the weights are the same (0.125) because the variables are uncorrelated. The results are the same as the exact values to the limit of computational precision. In both cases $\beta = 3.647$, and $p_f = 1.3 \times 10^{-4}$.

It is much easier to carry out the Rosenblueth point estimates than to derive the exact expressions for the mean and variance of M. Further, in the cases not considered here that the variables were correlated or that some of the other fixed variables were uncertain, an analytical expression would be very difficult or impossible to derive, and an approximate approach would be necessary. These results inspire some confidence in the point-estimate method.

15.4 Relation to Orthogonal Polynomial Quadrature

Case 2 is actually a direct application of a well known technique for numerical integration known as Gauss-Hermite quadrature, which provides an approximate value for the integral

$$I = \int_{-\infty}^{+\infty} g(z) \exp(-z^2) dz \tag{15.32}$$

where $g(z)$ is some arbitrary function of z. The approximate solution is

$$I \approx \sum_{i=1}^{n} H_i \cdot g(z_i) \tag{15.33}$$

in which n is the number of points used in the approximation, z_i is a point at which $f(z)$ is to be evaluated, and H_i is the corresponding weight. Salzer *et al.* (1952) provide a table of the values of z_i and H_i for $n \leq 20$, but a more accessible reference is Abramowitz and Stegun (1964). The first three columns of Table 15.3 give the values for $n \leq 5$. The obvious example of a function containing $\exp(-z^2)$ and integrated from minus to

Table 15.3 Coordinates and Weights, Gauss–Hermite and Rosenblueth Quadrature

Order	Gauss–Hermite Quadrature		Equivalent Rosenblueth Values	
n	Abscissas z_i	Weights H_i	Coordinates x_i	Weights w_i
2	± 0.707107	0.886227	$\pm 1.0\sigma$	0.5
3	0.0	1.181636	$0.0\ \sigma$	0.666667
	± 1.224745	0.295409	$\pm 1.732051\ \sigma$	0.166667
4	± 0.524648	0.804914	$\pm 0.741964\ \sigma$	0.454124
	± 1.650680	0.081313	$\pm 2.334414\ \sigma$	0.045876
5	0.0	0.945309	$0.0\ \sigma$	0.533333
	± 0.958572	0.393619	$\pm 1.355626\ \sigma$	0.222076
	± 2.020183	0.019953	$\pm 2.856970\ \sigma$	0.011257

plus infinity is the Gaussian or Normal probability distribution. Hildebrand (1956), for example, points out that Hermite polynomials are 'of particular importance in the theory of statistics.' The probability density function for the normal distribution with zero mean is

$$pdf(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{15.34}$$

Let x and z be related by

$$x = \sqrt{2}\sigma z \tag{15.35}$$

Then

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} f(x) \cdot pdf(x) \cdot dx = \int_{-\infty}^{+\infty} f(\sqrt{2}\sigma z) \cdot \frac{1}{\sqrt{\pi}} \cdot \exp(-z^2) \cdot dz \\ &\approx \sum_{i=1}^{n} (\sqrt{2}\sigma z_i) \frac{H_i}{\sqrt{\pi}} = \sum_{i=1}^{n} x_i w_i \end{aligned} \tag{15.36}$$

The values of x_i and w_i are in the last two columns of Table 15.3. For $n = 3$, the values are the same as those given by Rosenblueth in Equations (15.18) and (15.19). In effect, case 2 is a restatement of Gauss–Hermite quadrature with a change of variable.

In case 1, when the skewness is zero, the values of the variable x and the weights ($\pm\sigma_x$ and 0.5) are the same as the values in Table 15.3 for $n = 2$. However, Rosenblueth's method in case 1 is not simply an application of Gauss–Hermite quadrature with two points. It is actually based on a more radical approach, namely, the assumption that the continuously distributed variable x can be replaced by a discrete variable that has values only at the discrete points and that the locations of the points and their weights can be chosen so that the moments of x are recovered. Then the integrals of functions of the discrete variable should approximate the integrals of functions of the continuous variable. The approach is then extended to the case of a skewed variable in case 1 and multiple variables in case 3.

To illustrate that the approach is more than simply an application of a well-known numerical technique for integrating functions of a normally distributed variable, assume

that x is uniformly distributed between a and b. Equations (15.10)–(15.13) then give

$$P_- = P_+ = \tfrac{1}{2}$$
$$x_\pm = \mu_x \pm \sigma_x = \frac{(b-a)}{2} \pm \frac{(b-a)}{2\sqrt{3}} \tag{15.37}$$

These are identically equal to the weights and integration points called for by two-point Gauss-Legendre quadrature, which is the form of Gaussian quadrature that applies in the case of a uniform weighting function.

Now consider a function that is exponentially distributed. The skewness coefficient is 2, and, if the parameter (hazard rate) is λ, the mean and standard deviation are both $1/\lambda$ (Evans *et al.* 1993). Equations (15.10)–(15.13) give

$$P_\pm = \frac{2 \mp \sqrt{2}}{4}$$
$$x_\pm = \frac{1}{\lambda}(2 \pm \sqrt{2}) \tag{15.38}$$

Again, these are identical to the weights and coordinates called for by Gauss-Laguerre quadrature, which is the technique appropriate for an exponential weighting function.

In these three cases, the Rosenblueth procedure yields values of the coordinates and weights that are identical to the optimum values for numerical integration using two points. In fact, for any distribution of x, the Rosenblueth procedure in case 1 gives the values of the two coordinates and two weights that recover exactly the first three moments of x and the condition that the sum of all probabilities must be 1.0. This is the optimal choice for a two-point integration scheme. Rosenblueth's method is thus analogous to the generalized method of numerical integration with orthogonal polynomials (Press *et al.* 1992) and applies for any probabilistic distribution of X. Miller and Rice (1983) developed a similar point-estimate approach based on generalized numerical quadrature and described how to obtain the coordinates and weights for multiple concentration points.

15.5 Relation with 'Gauss Points' in the Finite Element Method

Figures similar to Figure 15.3 are to be found in most textbooks on the finite element method. They illustrate the most common technique for numerical integration of various components, such as the stiffness matrix, over an irregularly shaped element. First, the element is mapped onto a square with normalized coordinates s and t. The points for Gaussian quadrature are identified along each coordinate direction and a grid is formed locating the points throughout the element. The weights are the products of the weights for the individual directions. Then the entire process is mapped back onto the actual element. Extensions to more than two dimensions or to curved isoparametric elements are obvious.

Rosenblueth's third case is a similar extension of the one-dimensional Gaussian quadrature points to a grid of higher dimensions. Whereas the finite element method works almost exclusively with polynomial definitions of the functions to be integrated, Rosenblueth's method incorporates the effects of different probabilistic distributions for the underlying

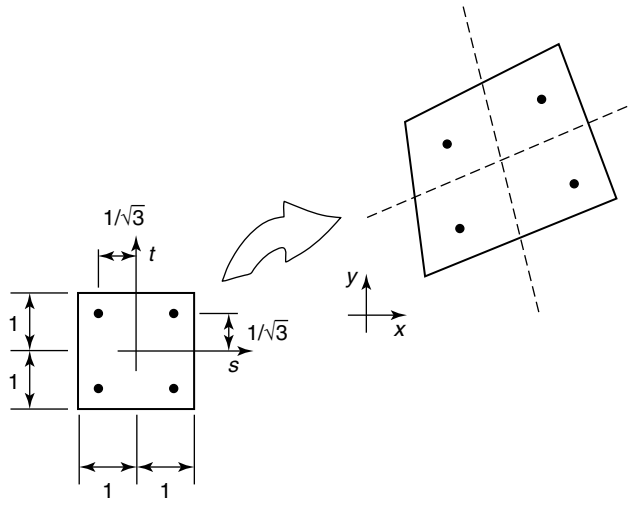


Figure 15.3 Typical four-point gaussian quadrature and mapping pattern employed. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

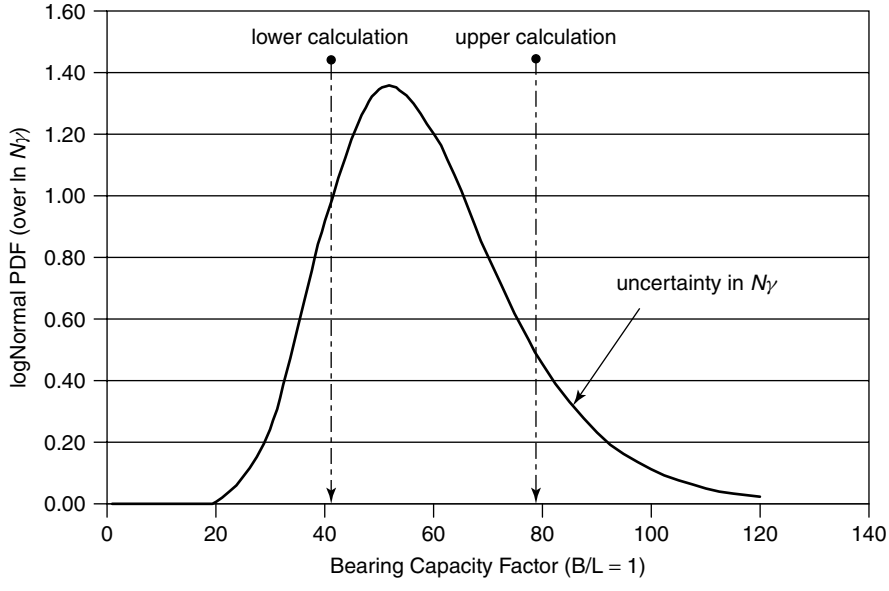


Figure 15.4 Points at which the bearing capacity is calculated using the point-estimate method with one uncertain variable, the bearing capacity factor N_γ using empirical results from Ingra and Baecher (1983). (Christian, J. T. and Baecher, G. B., 2003, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons).

variables. In fact, the method implicitly allows each of these variables to have a different distribution, so long as the mean and variance of the variable are known.

Ingra and Baecher (1983) performed logarithmic regression analyses on 145 model and prototype footing tests to estimate the expected value and variance of N_γ for footing length to width ratios of 1.0. The uncertainty in N_γ is logNormally distributed with skewness coefficient of 0.928. Using the point-estimate method to obtain the mean and variance of the bearing capacity when the only uncertainty is in the evaluation of N_γ, the calculations would be made at two values of the bearing capacity coefficient. For $\phi = 35^\circ$, Figure 15.4 shows the distribution of N_γ and the points at which the evaluations should be made. The corresponding weights are 0.710 and 0.290.

15.6 Limitations of Orthogonal Polynomial Quadrature

Although Rosenblueth's method is a simple way of obtaining the coordinates and weights for evaluating the mean and variance of a function of an arbitrarily distributed variable x, it does have limitations. First, two points may not be adequate to obtain accurate estimates of the moments for the function Y in a particular case. Rosenblueth gives some examples where the procedure has to be modified for some difficult functions, but in most practical cases the functions are well behaved.

A more serious deficiency appears when the method is used to evaluate moments of Y higher than the second moment. By evaluating the skewness coefficient of an arbitrary function Y from the case 1 procedure and working through the algebra, one discovers that, *regardless of the function Y, Rosenblueth's two-point procedure will always return the skewness of X instead of the skewness of Y*. The procedure should not be used for moments higher than the second, but most practical reliability calculations do not use the higher moments anyway. Harr (1987) observes that the point-estimate method cannot be used for moments higher than those used to define the points and weights. The limitation is actually stronger: because the two-point method returns the skewness of X not Y, it should not be used for moments higher than the second order *even when the third moments are used to find the points and weights.*

Rosenblueth's method is most widely used in case 3 to estimate the mean and variance of a function of several variables. It is worth noting that, in most applications to geotechnical engineering of which the authors are aware, the technique has been used for uncorrelated variables. The additional effort required to include the correlations is insignificant, and one of the strengths of Rosenblueth's method is the ease with which it can be applied to multivariate problems.

15.6.1 Example – Normal to Log-Normal Transformation

Although the two previous examples show excellent results from the point-estimate method, this is not always the case. Consider the case when X is a standard normal variable; that is, its mean is 0 and its standard deviation is 1. Let $Y = e^X$. Then Y is logNormally distributed with $\mu_Y = 1.649, \sigma_Y = 2.161$, and the skewness coefficient $\nu_Y = 6.185$. Now, Gauss–Hermite quadrature applies for the case where the probability distribution of the underlying variable (X) is normal, so successively higher orders of Gauss–Hermite quadrature can be used to estimate the moments of Y. When $n = 2$, this is the Rosenblueth case 1 method, and, when $n = 3$, this is the Rosenblueth case 2 method.

Table 15.4 Estimates of moments for lognormal distribution

	$X = \exp(Y); Y = \mathrm{N}[0,1]$		
	μ_Y	σ_Y	ν_Y
Exact Values:	1.648721	2.161197	6.184877
Estimate, n = 2	1.543081	1.175201	0.0
Estimate, n = 3	1.638192	1.822602	1.822157
Estimate, n = 4	1.647969	2.071355	3.425399
Estimate, n = 5	1.648678	2.143514	4.79062

The results are in Table 15.4. It can be seen that, while $n = 3$ gives a reasonable estimate of μ_Y, it is necessary to go at least to $n = 4$ to get a reasonable estimate of σ_Y and the estimate of the skewness ν_Y is still quite far off when $n = 5$. These results indicate that the point-estimate methods, and, for that matter, Gauss–Hermite quadrature, should be used with caution when the function of $g(X)$ changes the probability distribution severely.

15.7 Accuracy, or When to Use the Point-Estimate Method

There is an unfortunate tendency for persons using numerical tools to employ them without first considering their accuracy. It is almost as though the numerical procedure were some sort of machinery enclosed in a black box whose workings were not to be investigated. Once the procedure is described in books and manuals, the results are seldom questioned. This attitude can lead to some very inaccurate work.

When numerical procedures are involved, it is often difficult to determine the accuracy of the technique, for one of the principal reasons for employing numerical methods is that the problem is not tractable analytically. Nevertheless, this section offers some guidance on when the Rosenblueth method with two points may or may not be expected to give accurate results.

15.7.1 Accuracy for $Y = X^2$

Gaussian quadrature is a numerical approximation to the integral

$$I = \int g(z) \cdot f(z) \cdot dz \tag{15.39}$$

in which the limits of integration have been left out because they can vary depending on the form or the functions to be integrated. In Equation (15.39), $g(z)$ is the function to be evaluated at the integration points, and $f(z)$ is the so-called weighting function. The approximate integral is

$$I \approx \sum H_i \cdot g(z_i) \tag{15.40}$$

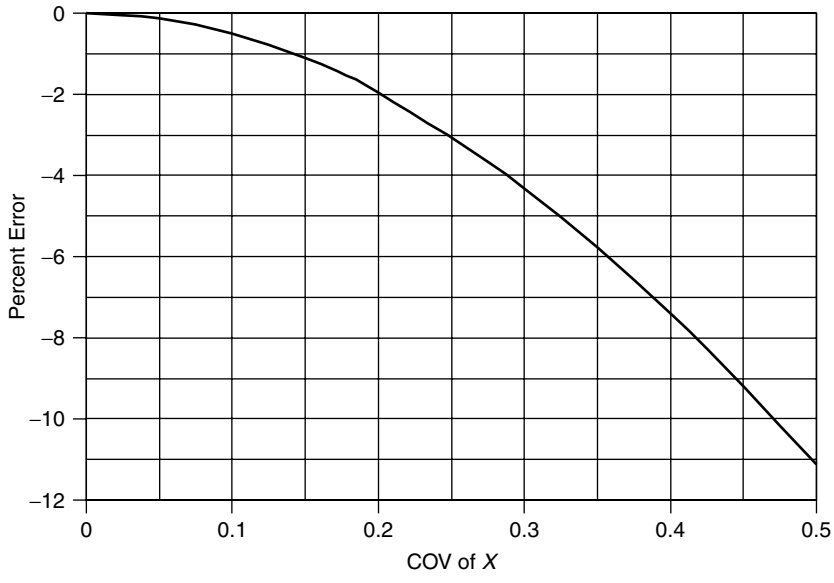


Figure 15.5 Error in the estimate of the variance for $Y = X^2$, X normally distributed. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

The values of the H_i's are determined by the function $f(z)$. Equations (15.32)–(15.36) deal with the condition that $f(z)$ is the pdf for the Normal distribution.

Rosenblueth's method establishes the locations of the integration points and the values of the H_i's for whatever pdf applies to the variable X. Gaussian quadrature with n points then guarantees that the integration will be exact for any polynomial function $g(x)$ of an order up to and including $(2n - 1)$. In the case of the Rosenblueth two-point method, this means the results will be exact for polynomials of cubic order or less. Thus, in the trivial case that Y is a linear function of X, its variance will be of order 2, and the point estimate method will calculate both the mean and the variance of Y exactly. If Y is a function of X^2, the mean will be calculated exactly, but the variance will be of order 4, so some error will occur. Figure 15.5 shows the error in the estimate of the variance as a function of the coefficient of variation (Ω) of X when X is Normally distributed. It can be seen that, the larger the Ω, the larger the error in the estimate.

15.7.2 Accuracy for Some Typical Functions

Figures 15.6–15.12 show the errors that arise when the point-estimate method is applied to several functions of X that typically arise in geotechnical reliability analyses. When $Y = \exp(X)$ with X Normally distributed, Y is logNormally distributed. We have already indicated for this case that, when X is the standard Normal variable, with mean = 0 and variance = 1, the point-estimate method is seriously in error and it is necessary to use five or more points in Gauss-Hermite quadrature to obtain reasonably accurate results. Figures 15.6 and 15.7 show the percent errors in the mean and variance as functions of

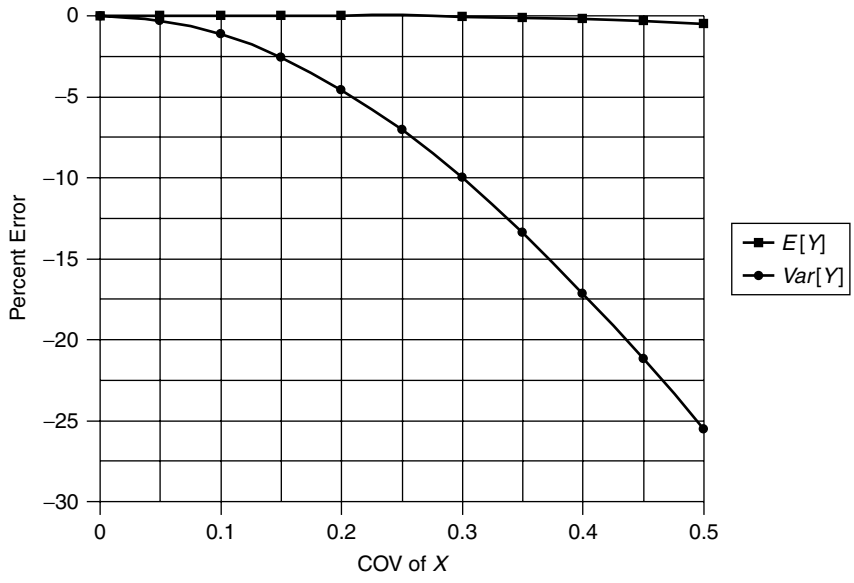


Figure 15.6 Error in the estimates of mean and variance for $Y = \exp X$, X normally distributed, mean value of $X = 1$. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

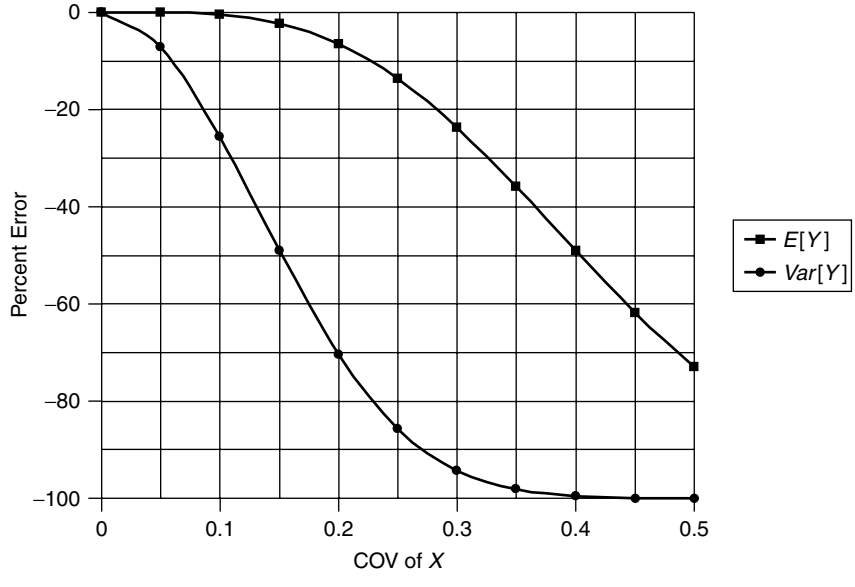


Figure 15.7 Error in the estimates of mean and variance for $Y = \exp X$, X normally distributed, mean Value of $X = 5$. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

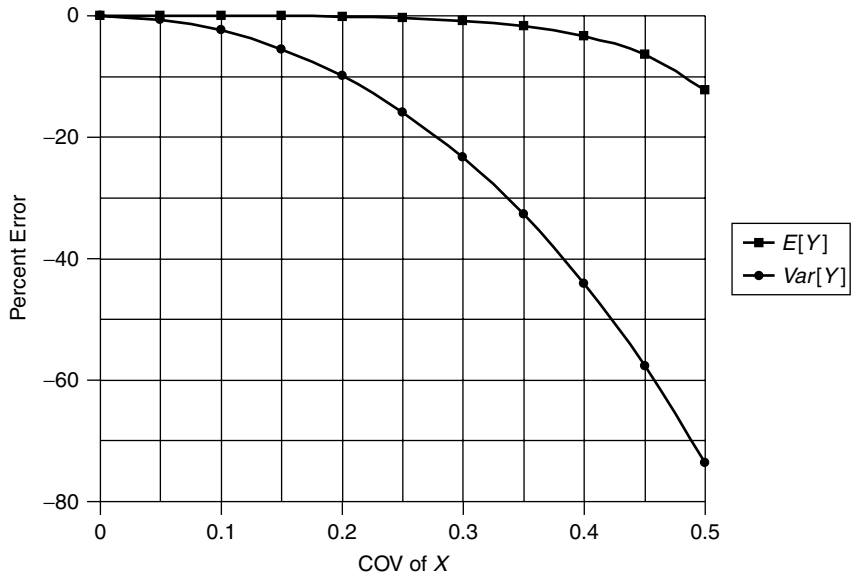


Figure 15.8 Error in the estimates of mean and variance for $Y = 1/X$, X uniformly distributed, mean value of $X = 20$. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

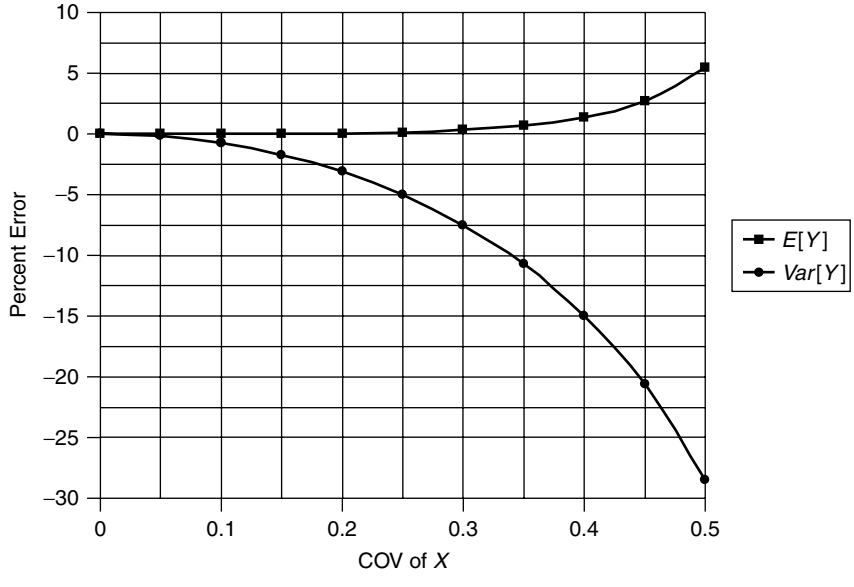


Figure 15.9 Error in the estimates of mean and variance for $Y = \ln X$, X uniformly distributed, mean value of $X = 2$. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

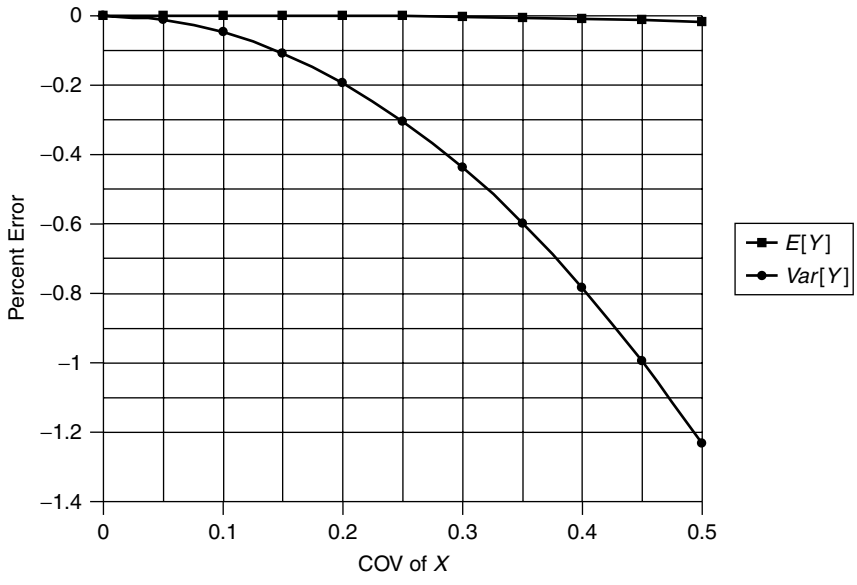


Figure 15.10 Error in the estimates of mean and variance for $y = \tan X$, X uniformly distributed, mean value of $X = 15$ degrees. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

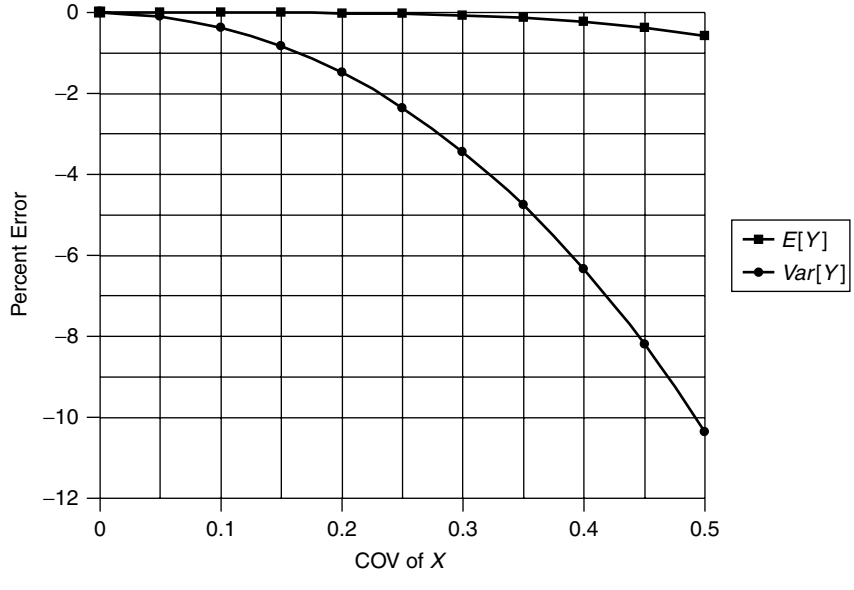


Figure 15.11 Error in the estimates of mean and variance for $y = \tan X$, X uniformly distributed, mean value of $x = 30$ degrees. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

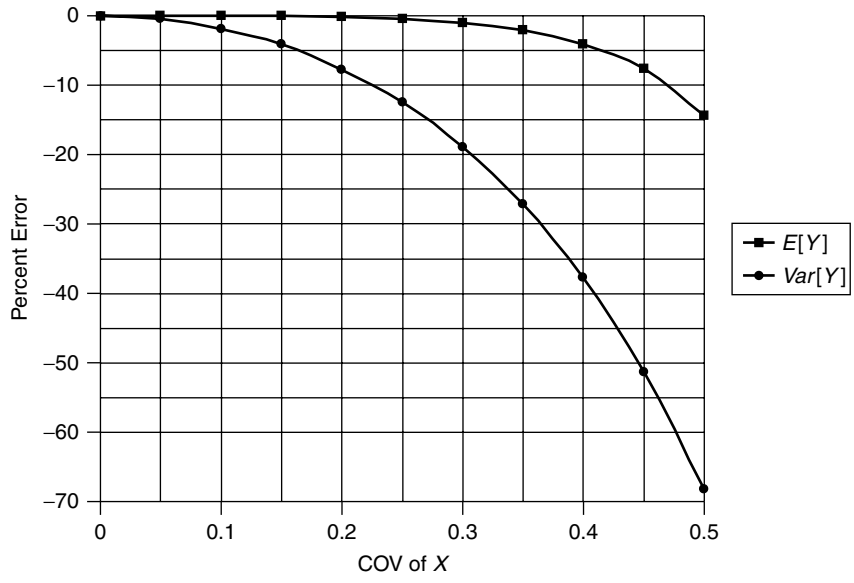


Figure 15.12 Error in the estimates of mean and variance for $y = \tan X$, X uniformly distributed, mean value of $x = 45$ degrees. (Christian, J. T. and Baecher, G. B., 1999, 'Point-Estimate Method as Numerical Quadrature,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 125, No. 9, pp. 779–786, reproduced by permission of the American Society of Civil Engineers.)

the Ω of X when the two-point estimate method is used. X is Normally distributed in both figures. When the mean of X is 1, the estimates of the mean of Y are accurate, but the estimates of the variance are seriously in error unless Ω is quite small. When the mean of X is 5, the results are unsatisfactory for all values of the Ω.

Unfortunately, it is seldom possible to compute the exact values of the mean and variance of a function $g(X)$ when X is Normally distributed, or even when it has almost any other practically interesting distribution. The integrations are simply too difficult. However, it is possible to make a coarse estimate of the accuracy of the method for other functions $g(X)$ by assuming that X is uniformly distributed between $\mu_X - \sqrt{3}\sigma_X$ and $\mu_X + \sqrt{3}\sigma_X$. Then Equation (15.39) reduces to

$$I = \frac{1}{2\sqrt{3}\sigma_X} \int_{\mu_X - \sigma_X}^{\mu + \sigma_X} g(x) \cdot dx \tag{15.41}$$

Equation (15.41) can be evaluated with $g(x)$ and with $[g(x)]^2$ under the integral sign. This will give exact values of the first two moments of $g(X)$ about the origin, and from these the exact values of the mean and variance can be calculated. In many practical cases such integrals are tabulated. The point estimate method is then used to estimate the first two moments of $g(X)$ about the origin and, from them the mean and variance.

Figure 15.8 shows the results of performing this calculation when $Y = 1/X$. The mean value of X is taken to be 20 kN/m^3, which is close to the value for the total unit weight of typical geotechnical materials. Unless Ω is small, the estimates of the variance are poor, but this is satisfactory because in most practical cases the uncertainty in the unit

weight is relatively small. Figure 15.9 shows the results when $Y = \ln X$ and the mean of X is 2. The pattern for the error in the mean of Y depends on the mean of X, but the errors in the variance of Y are not affected by the value of the mean of X. The errors are probably not acceptable for large values of Ω. Figures 15.10–15.12 show the results for $Y = \tan X$, with three values of the mean of X. When the mean of X is 15 or 30 degrees, the results range from excellent to satisfactory. When X is 45 degrees, the errors in the variance of Y become large when Ω exceeds approximately 0.2. However, it should be noted that a case with a mean of 45 degrees and Ω of 0.5 corresponds to an angle varying uniformly between 22.5 and 67.5 degrees, which seems an unlikely situation. Again, it must be emphasized that in all the analyses represented by Figures 15.8–15.12, X was assumed to vary uniformly.

The numerical results presented in this section demonstrate that the method works best when the Ω's are small and when the functions of X to be integrated can be reasonably approximated by a third-order polynomial. When these conditions do not apply, the results in Figures 15.5–15.12 can be used to estimate the errors, or the analyst can compare the results of the Rosenblueth approach to results from higher order Gaussian quadrature before embarking on an extensive set of calculations employing the point-estimate method with two points along each variable axis.

It should be noted that methods based on truncated Taylor series, such as FOSM, which can also be used to approximate the integrals that arise in reliability calculations, are generally less accurate than the point-estimate methods because they are based on lower order expansions. Harr (1989) observed that Rosenblueth's approach circumvented many of the shortcomings of Taylor series and Monte Carlo methods. Hong (1998) demonstrates improved accuracy for logNormal and Gamma distributions.

15.8 The Problem of the Number of Computation Points

A great limitation of the original point-estimate method for multiple variables is that it requires calculations at 2^n points. When n is greater than 5 or 6, the number of evaluations becomes too large for practical applications, even in the day of cheap computer time. As noted above, Rosenblueth (1975) himself proposed a technique for reducing the number of calculation points to $2n + 1$ when the variables are uncorrelated and when skewness can be ignored. Li's (1992) method applies when the function consists of polynomials of the individual variables and their cross-products. Lind (1983) proposed that, instead of using the points at the corners of the hypercube (Figure 15.2), one could select points near the centers of the faces of the hypercube, and he provided a procedure for finding those points and their weights. More recently, two relatively simple methods for reducing the number of points in the general case to $2n$ or $2n + 1$ have been proposed (Harr 1989; Hong 1996, 1998).

15.8.1 Multiple, Correlated, Unskewed Variables

Harr (1989) deals with the case in which the variables may be correlated but their skewness coefficients are zero. The procedure starts by computing the eigenvalues and eigenvectors of the correlation matrix. Then the hypercube is rotated into a coordinate system whose axes coincide with the eigenvectors. A hypersphere is imagined to pass through the corners of the hypercube. The calculation points are then the intersections between the rotated

coordinate axes and the surface of the hypersphere. This gives $2n$ points. Harr further recommends that the contribution of each calculation point be weighted in proportion to its eigenvalue.

Harr's method is best illustrated by an example. Consider the above example of the Culmann slope failure analysis, but now assume that the three uncertain variables are correlated with the following correlation coefficients:

$$\rho_{c,\tan\phi} = -0.5\rho_{c,\gamma} = 0.2\rho_{\gamma,\tan\phi} = 0.2 \tag{15.42}$$

This leads to the following correlation matrix, **K**:

$$\mathbf{K} = \begin{bmatrix} 1.0 & -0.5 & 0.2 \\ -0.5 & 1.0 & 0.2 \\ 0.2 & 0.2 & 1.0 \end{bmatrix} \tag{15.43}$$

Rosenblueth's case 3 inserts the correlation coefficients of Equations (15.42) into Equations (15.21) to obtain a new set of weights replacing those in the sixth column of Table 15.2. The eight-point calculations then give $\mu_M = 2.643$, $\mu_M = 0.447$, $\mu = 5.916$, and (if M is Normally distributed) $p_f = 1.7 \times 10^{-9}$.

Harr's method uses six points ($2n$ when $n = 3$). The calculations proceed in the following steps:

(1) Calculate the eigenvalues and normalized eigenvectors of K:

$$\lambda = \{\, 1.5 \quad 1.127 \quad 0.373 \,\} \tag{15.44}$$

$$\mathbf{E} = \begin{bmatrix} 0.707 & -0.291 & -0.645 \\ -0.707 & -0.291 & -0.641 \\ 0 & -0.912 & 0.411 \end{bmatrix} \tag{15.45}$$

The entries in each column of Equation (15.45) comprise the eigenvector for the corresponding eigenvalue.

(2) Compute the values of c, $\tan\phi$, and γ for each of the six points. Each pair of points corresponds to the intersections of one eigenvector with the hyperspherical surface. The easiest way to do this is to create three columns of six elements, one column for c, one for $\tan\phi$, and one for ϕ. The values in the $2i^{\text{th}}$ and $(2i+1)^{\text{st}}$ entries of the columns give the coordinates of the intersection of the i^{th} eigenvector. The value for a variable x at a point is found from $x = \mu_x \pm \sigma_x\sqrt{n} \cdot E_{i,j}$, where $E_{i,j}$ is the corresponding component of **E**. For example, the component in the fifth position of the c column (which corresponds to the third eigenvector) is $5.0 + 0.5 \cdot \sqrt{3} \cdot (-0.645) = 4.442$.
(3) Evaluate M and M^2 at each of the six points just defined.
(4) The weight associated with each point is equal to the corresponding eigenvalue divided by $2n$.
(5) The remainder of the calculations proceeds as in the earlier versions of the point-estimate method.

When this procedure is applied to the Culmann problem, the results are $\mu_M = 2.643$, $\mu_M = 0.445$, $\beta = 5.938$, and (if M is Normally distributed) $p_f = 1.4 \times 10^{-9}$. The

Table 15.5 Results from Harr's method for Culmann slope

Point	c (kPa)	tan (ϕ)	γ (kN/m^3)	M (kPa)	p
1	5.612	0.244977	22	2.680	0.250
2	4.388	0.290921	22	2.587	0.250
3	4.748	0.258509	18.526	2.559	0.188
4	5.252	0.277389	25.474	2.781	0.188
5	4.442	0.247007	23.566	1.354	0.062
6	5.558	0.288892	20.434	3.840	0.062
				Mest = Sum of Mp =	2.643
				Variance = Sum of $M^2p - (Mest)^2$ =	0.198
			$\sigma = 0.445$, $\beta = 5.938$		
			$H = 10\,\text{m}$, $\theta = 20\,\text{deg}$, $\psi = 26\,\text{deg}$		

differences between this solution and the eight-point method are small and are due in part to rounding errors. The detailed values are given in Table 15.5.

Lind's (1983) method starts with the covariance matrix, **C**:

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{bmatrix} \tag{15.46}$$

where the subscripts correspond to the variables. He then defines a series of vectors $\mathbf{z}_i$, one for each variable:

$$\begin{aligned} \mathbf{z}_1 &= [z_{11}, z_{12}, z_{13}, \ldots, z_{1n}] \\ \mathbf{z}_2 &= [0, z_{22}, z_{23}, \ldots, z_{2n}] \\ \mathbf{z}_3 &= [0, 0, z_{33}, \ldots, z_{3n}] \\ &\vdots \\ \mathbf{z}_n &= [0, \ldots, 0, z_{nn}] \end{aligned} \tag{15.47}$$

The terms are computed by from the terms in the **C** matrix by a set of recursive equations:

$$\begin{aligned} z_{ii} &= \left(n \cdot c_{ii} - \sum_{k=1}^{i-1} z_{ki}^2 \right)^{1/2} \quad i = 1, 2, \ldots, n \\ z_{ij} &= \left(n \cdot c_{ij} - \sum_{k=1}^{i-1} z_{ki} z_{kj} \right) \Big/ z_{ii} \begin{cases} i = 1, 2, \ldots, n-1 \\ j = i+1, i+2, \ldots, n \end{cases} \end{aligned} \tag{15.48}$$

Each of the **z** vectors can be made a row in a matrix, denoted **Z**. Each column of the matrix corresponds to one of the variables. The coordinates of each of the $2n$ evaluation

points are found by proceeding down each column, first adding and then subtracting the entry from the mean value of the corresponding variable. The weights are simply $1/2n$.

For the above example of the Culmann failure analysis with three correlated variables,

$$\mathbf{C} = \begin{bmatrix} 0.25 & -4.7 \times 10^{-3} & 0.22 \\ -4.7 \times 10^{-3} & 3.5 \times 10^{-4} & 8.2 \times 10^{-3} \\ 0.22 & 8.2 \times 10^{-3} & 4.84 \end{bmatrix} \tag{15.49}$$

$$\mathbf{Z} = \begin{bmatrix} 0.886 & -0.016 & 0.762 \\ 0 & 0.028 & 1.32 \\ 0 & 0 & 3.492 \end{bmatrix} \tag{15.50}$$

and the remainder of the analysis is summarized in Table 15.6 The results are almost identical to those from Harr's method though the specific computation points are different.

15.8.2 Multiple, Uncorrelated, Skewed Variables

Hong (1996, 1998) deals with the other problem of uncorrelated variables with significant skewness. He develops, for each variable, a two-point and a three-point estimation procedure. The two-point equations are similar to Equations (15.10)–(15.13) except that the term $\sqrt{1 + (\nu_x/2)^2}$ is replaced by $\sqrt{n + (\nu_x/2)^2}$ wherever it appears. The three point equations include both skewness and kurtosis and are chosen so that the central point is at the mean value of the variables. Since this point is common to all variables, the composite system has $2n + 1$ points. Skewness could have been included in the original Rosenblueth procedure by selecting the points and weights according to Equations (15.10)–(15.13), but this would still have led to 2^n points.

To illustrate the use of Rosenblueth's two-point method for multiple skewed variables and Hong's method, consider the Culmann problem extended so that the inclination of the failure surface is also uncertain. Also, the friction angle itself rather than the tangent of the angle will be taken as the uncertain variable. It will be assumed that all the uncertain variables are triangularly distributed. Table 15.8 lists the values at the two ends (a and b) and the peak (c) of the distributions.

Table 15.6 Results from Lind's method for Culmann slope

Point	c (kPa)	tan (ϕ)	γ (kN/m^3)	M (kPa)	p
1	5.866	0.251706	22.762	3.003	0.167
2	4.134	0.284193	21.238	2.236	0.167
3	5	0.296084	23.320	3.226	0.167
4	5	0.239815	20.680	2.123	0.167
5	5	0.267949	25.492	2.258	0.167
6	5	0.267949	18.508	3.009	0.167
				Mest = Sum of Mp =	2.643
				Variance = Sum of $M^2p - (Mest)^2$ =	0.198
			$\sigma = 0.445, \beta = 5.938$		
			$H = 10\,\text{m}, \theta = 20\,\text{deg}, \psi = 26\,\text{deg}$		

The extended form of Rosenblueth's method uses Equations (15.10)–(15.13) to compute the coordinates and the weights in each coordinate direction. The coordinates of the points at which the function is to be evaluated are found by forming each possible combination of the four variables identified at two points each. This gives sixteen points. The corresponding weights are found by multiplying the corresponding weights from the one-dimensional case. The calculations then proceed as in the usual multi-variable case. The results are that $\mu_M = 2.576$, $\beta_m = 1.096$, and $\sigma = 2.350$. If M is Normally distributed, then $p_f = 9.4 \times 10^{-3}$.

Hong's method employs eight points, two for each variable. At each pair of points three of the variables are fixed at their mean values, and the remaining one is allowed to take values above of below the mean value. For each of the n variables, the first calculation is to evaluate

$$\zeta = \sqrt{n + \left(\frac{\nu}{2}\right)^2} \tag{15.51}$$

Then the first coordinate value corresponding to this variable is

$$x_1 = \mu + \left(\frac{\nu}{2} - \zeta\right) \cdot \sigma \tag{15.52}$$

and the coordinate of the other point is

$$x_1 = \mu + \left(\frac{\nu}{2} + \zeta\right) \cdot \sigma \tag{15.53}$$

The corresponding weights are

$$p_1 = \frac{1}{2n} \cdot \left(1 + \frac{\nu}{2\zeta}\right) \tag{15.54}$$

and

$$p_2 = \frac{1}{2n} \cdot \left(1 - \frac{\nu}{2\zeta}\right) \tag{15.55}$$

Table 15.7 shows the coordinates and weights for the eight points. The results are that $\mu_M = 2.576$, $\mu_m = 1.096$, and $\beta = 2.350$. If M is Normally distributed, then $p_f = 9.4 \times 10^{-3}$. These are identical to those from the 16-point Rosenblueth method.

Once the calculations are set up, it is easy to perform parametric studies. For example, if the low value of c is made 0.0, he results are that $\mu_M = 1.910$, $\sigma_m = 1.484$, and $\beta = 1.287$. If M is Normally distributed, then $p_f = 0.1$. If the low value is raised to 4, the results are that $\sigma_M = 3.243$, $\mu_m = 0.804$, and $\beta = 4.032$. If M is Normally distributed, then $p_f = 2.8 \times 10^{-5}$.

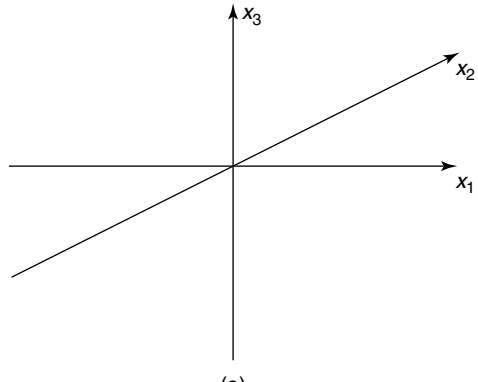
Figure 15.13 shows these methods graphically for the case of three uncorrelated and unskewed variables. Rosenblueth's original proposal for reducing the number or calculation points required $(2n + 1)$ calculations. Li's method requires $(n^3 + 3n + 2)/2$ calculations. Figure 15.14 describes how the number of calculations varies with the number of variables for each of these procedures.

Table 15.7 Results from Hong's method with four variables for Culmann slope

Point	c(kPa)	ϕ (deg)	γ(kN/m^3)	θ(deg)	M (kPa)	p
1	2.445	15.667	19.333	20.000	0.636	0.112
2	5.863	15.667	19.333	20.000	4.054	0.138
3	4.333	12.556	19.333	20.000	1.274	0.132
4	4.333	19.167	19.333	20.000	3.978	0.118
5	4.333	15.667	17.445	20.000	2.701	0.112
6	4.333	15.667	20.863	20.000	2.381	0.138
7	4.333	15.667	19.333	18.367	2.900	0.125
8	4.333	15.667	19.333	21.663	2.521	0.125
					Mest = Sum of M p =	2.576
					Variance = Sum of M^2 p − (Mest)^2 =	1.202
			$\sigma = 1.097$, $\beta = 2.350$			
			H = 10 m, $\psi = 26$ deg			

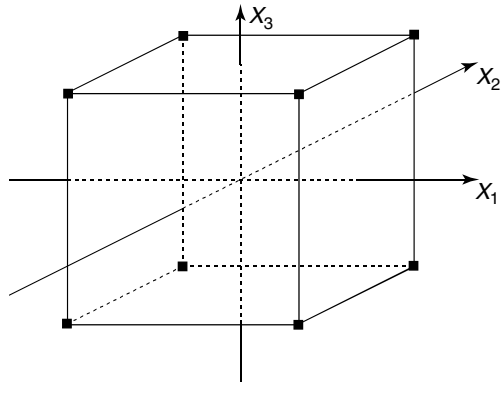
15.8.3 Points far from Mean for Large n

Lind's, Haar's, and Hong's methods give results superior to those from FOSM methods with little or no increase in computational effort. Because they use fewer points than the original point-estimate methods, they may be somewhat less accurate for some functions, but limited testing indicates that they are usually quite satisfactory. All the methods locate the points at or near the surface of a hypersphere or hyperellipsoid that encloses the points from the original Rosenblueth method. The radius of the hypersphere is proportional to

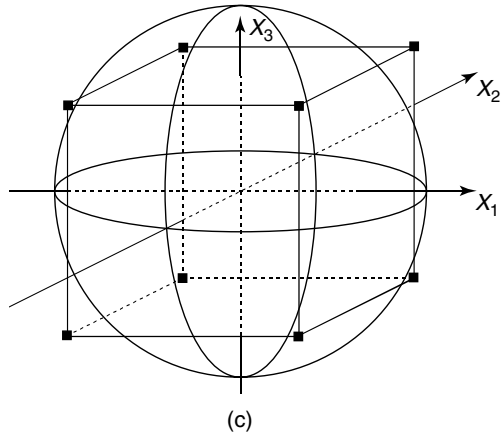


(a)

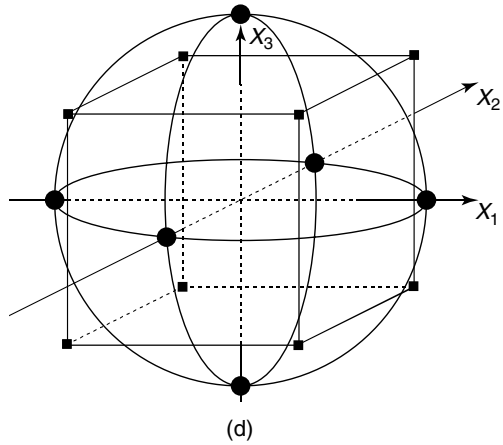
Figure 15.13 Distribution of evaluation points for three uncertain variables. The variables have been normalized by subtracting the mean and dividing by the standard deviation. (a) Coordinate system, (b) black squares are the eight points in the original Rosenblueth procedure, (c) ellipses are circular arcs defining a sphere circumscribed around the Rosenblueth points, (d) black dots are the intersections of the circumscribed sphere with the coordinate axes, which are the six points in the Harr and Hong procedures. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons.)



(b)



(c)



(d)

Figure 15.13 (*continued*)

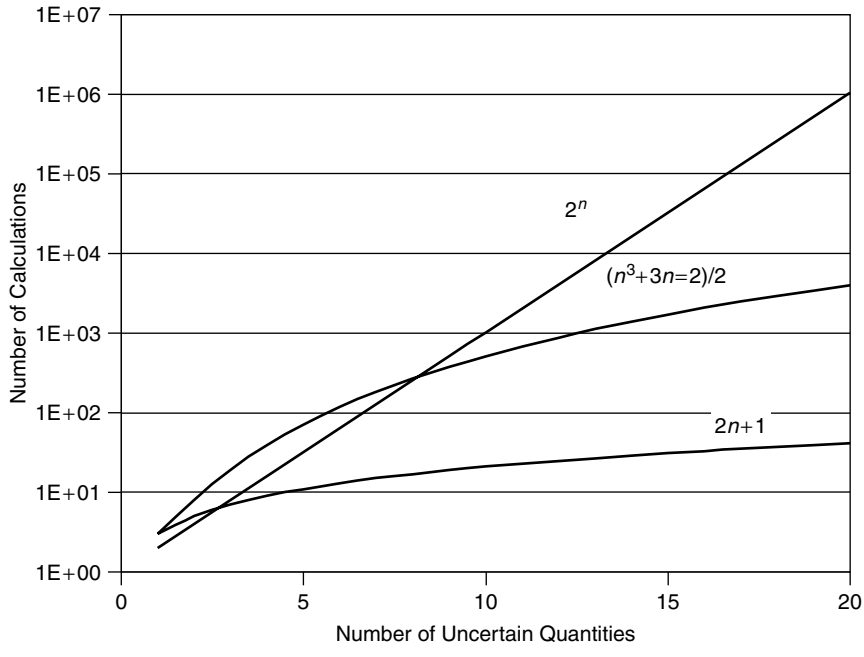


Figure 15.14 Numbers of calculations by various algorithms. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons.)

$\sqrt{n}$ in Harr's method and nearly so in Lind's and Hong's. It is possible that, when n is large, the values of the x_i's at which the evaluation is to take place may be so many standard deviations away from the means that they are outside the range of meaningful definition of the variables. Figure 15.15 shows what happens in a case of normalized uncorrelated and unskewed variables. The four black points are the points in the conventional Rosenblueth procedure for two variables. The four open points on the circle drawn through the black points are the evaluation points that would be used in the Harr, Hong, or Lind procedures. There is obviously no advantage in doing this for two variables as four points are needed in either case. However, if there were nine variables, a reduction from 512 to 18 evaluations would be well worth the effort. The outer circle in Figure 15.15 represents a two-dimensional section through the nine-dimensional hypersphere, and the evaluation points now lie three standard deviations form the mean. Thus, critical values of the variables will be located far from the region where the distributions are known best.

In Table 15.8, which gives the coordinates of the points to be used in the analysis with four uncertain variables triangularly distributed, it is clear that the values of some of the variables are quite close to the limits of their triangular distributions. If there were seven variables, Table 15.9 shows that the values of all the variables would fall outside the bounds of their distributions. Figure 15.16 is the same as Figure 15.4, except that the locations of the evaluation points for cases with 5 and 10 variables are shown. The evaluations fall far out along the tails of the distribution, where the values of the distribution are known poorly.

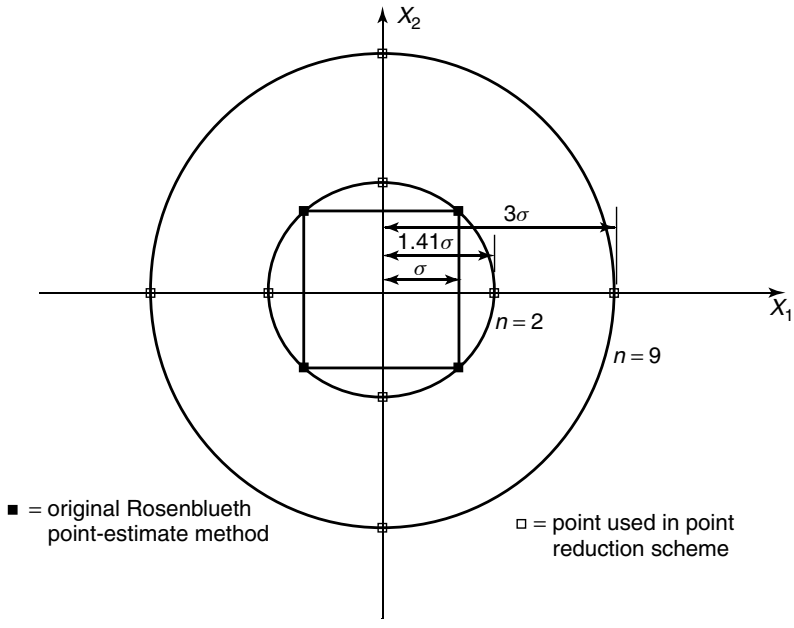


Figure 15.15 Evaluation points. Black points are the original Rosenblueth points for two uncorrelated, unskewed variables. Open points are those defined by a point reduction scheme on a circle circumscribed around the four points. Large circle is a two-dimensional cut through the nine-dimensional hypersphere for nine variables, and open points on this circle are the locations of points used in point reduction schemes. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons).

Table 15.8 Values of triangular distribution points for four-variable problem

	c(kPa)	ϕ (deg)	γ (kN/m^3)	θ (deg)
Point a	2	12	17	18
Point c	5	15	20	20
Point b	6	20	21	22
μ	4.333	15.64	19.333	20
σ	0.850	1.66	0.850	0.802
ν	−0.422	0.236	−0.422	0
		H = 10 m, ψ = 26 deg		

Table 15.9 Extreme values of coordinates for Culmann problem with $n = 7$

	c (kPa)	Φ (deg)	γ (kN/m^3)	θ (deg)
Minimum	1.898	11.492	16.898	17.840
Maximum	6.409	20.231	21.409	22.160

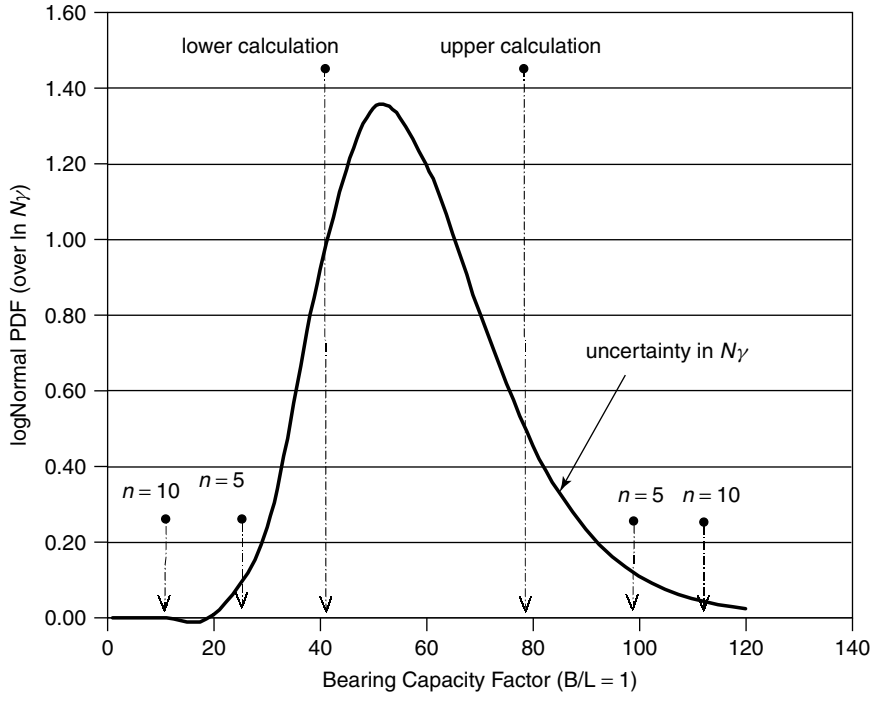


Figure 15.16 Points at which the bearing capacity is calculated using the point-estimate method with 5 and 10 variables for bearing capacity factor N_γ. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons.)

Because the distributions of many variables encountered in civil engineering are not well known, and because there may be physical reasons for limiting their ranges, it is common to use triangular or Beta distributions. Figure 15.17 shows what happens in the case of a triangular distribution between 0 and 1 with c defining the peak of the triangle. The solid lines describe, as a function of c, the number of uncertain variables necessary to move the coordinate of the lower evaluation point in Hong's method below zero, and the dashed lines depict the number of uncertain variables that move the upper point above one. Both evaluation points will fall outside the range of definition when $n \geq 7$. In the extreme case that the peak of the triangle occurs at one end of the distribution, one evaluation point will fall outside that end when the number of variables is 3.

The Beta distribution allows a large range of distribution shapes between the bounds. Table 15.10 shows that for combinations of parameters typical of those that are encountered in practice the evaluation points fall outside the boundaries of definition even for relatively small numbers of variables. Only in the strongly skewed case of $q = 2$ and $r = 8$ does the number of variables needed to push the evaluation point beyond the upper limit of the distribution become large (39), and in that case the number of variables required to locate the other evaluation point below the lower limit is only 5.

A value of a variable falling outside the limits of definition can be physically unreasonable. For example, negative values of cohesive strength or friction or construction time

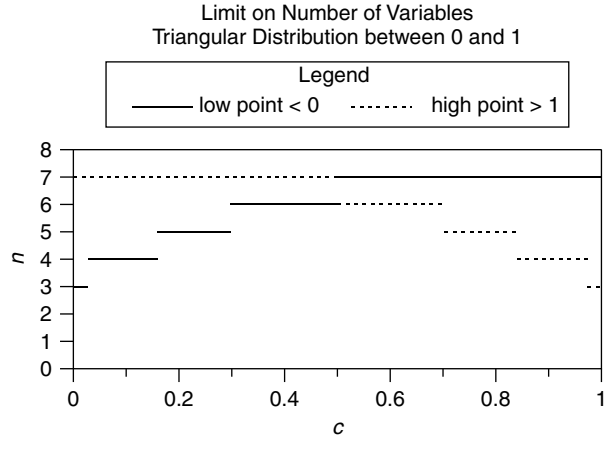


Figure 15.17 The number of uncertain variables necessary to move the evaluation points in Hong's method below 0 (solid lines) or above 1 (dashed lines) for a triangular distribution between 0 and 1. The parameter c is the coordinate of the third or peak point defining the triangular distribution. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons.)

Table 15.10 Number of variables causing evaluation points to fall outside the limits of a Beta distribution. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons)

Beta parameters		Number of variables for	
q	r	lower limit	upper limit
2	3	5	9
2	4	5	13
2	8	5	39
3	3	7	7
3	4	7	11
3	4.5	7	12
3.26	4.89	7	13
4	6	8	16

are unacceptable to most engineers. Values outside the limits or far from the means may also be unacceptable because they are used in functions that become unrealistic for the extreme values of the arguments. Some functions, such as $\ln X$ for $X < 0$, may not even have a real definition outside the range of definition of X.

Riela *et al.* (1999) in a study of the stability of mine slopes used a triangular distribution to describe the uncertainty in friction angle. In one case they had four significant uncertain variables, and in the other they had seven. Their analysis used the FOSM method

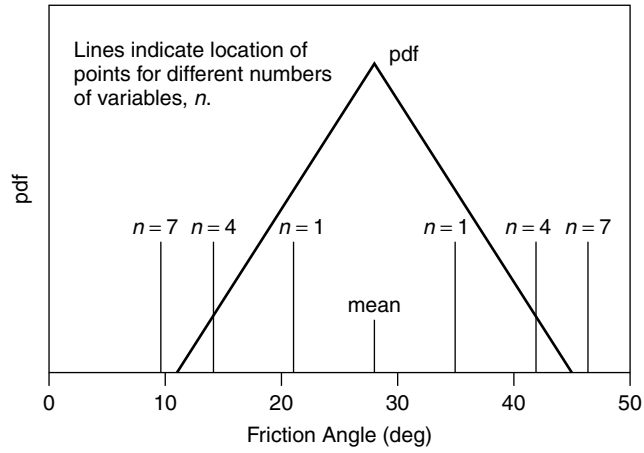


Figure 15.18 Probability density function (pdf) for triangularly distributed friction angle in Chuquicamata analysis and points for one, four, and seven variables. (Christian, J. T. and Baecher, G. B., 2002, 'The Point-Estimate Method with Large Numbers of Variables,' *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 126, No. 15, pp. 1515–1529, reproduced by permission of John Wiley & Sons.)

and the direct Rosenblueth method without trying to reduce the number of calculations. Figure 15.18 shows where the evaluation points would fall if one of the computation reduction schemes were employed. The lines labeled '$n = 1$' are the original points. Those labeled '$n = 4$' and '$n = 7$' identify the points that would be used in the computation reduction schemes for four and seven variables, respectively. For four variables the points fall near the ends of the distribution, and for seven variables they fall outside their range of definition.

It is clear that all three of the simple ways to modify the point-estimate method in order to reduce the number of evaluations cause the locations of the evaluation points to move far from the central region of the distributions. In the case of bounded distributions the evaluation points move beyond the range of definition of the distribution, and this can happen when $n = 6$ or 7. In some cases it occurs with as few as three variables. The phenomenon is primarily controlled by the number of variables and does not result from a poor choice of probability distributions. Since the main reason for adopting the modified point-estimate methods is to make it feasible to deal with large numbers of variables, this problem is a serious impediment.

Several alternatives suggest themselves. First, one could use the unmodified point-estimate method. The number of evaluations will be large (2^n), but in the days of cheap computer time, when people routinely do Monte Carlo simulations with hundreds of thousands of samples, the cost may be bearable. Secondly, when the variables are uncorrelated and unskewed, Rosenblueth's (1975) own modification could be used. Third, under the restrictive conditions that Y is the sum of polynomials and cross-products of each of the variables, Li's (1992) method is appropriate.

Finally, when faced with a large number of uncertain variables, the analyst would do well to ask whether all of them need to be carried along in the analysis. Relatively simple sensitivity studies will often reveal that uncertainties in many of the variables have little effect on the results and that other variables can be combined. The reduction in the number

of active variables not only makes the computations more tractable but also increases the chances that the results can be interpreted. However, Lind (1983) points out that, since removing some of the smaller contributors will reduce the overall variance, the variances of the remaining variables should be increased proportionately.

15.9 Final Comments and Conclusions

Rosenblueth's point estimate method, which is widely used in geotechnical engineering reliability analysis, but which has also been criticized for its simplicity, is a special case of the orthogonal polynomial approximation procedures leading to the Gaussian quadrature formulas that are well known in numerical analysis and finite element methods. Rosenblueth described the method to be used when there are two or three variables, and he showed how their correlation structure could be incorporated in the analysis. A simple extension of Rosenblueth's approach applies for more than three variables – correlated or uncorrelated.

The method is reasonably robust and is satisfactorily accurate for a range of practical problems, though computational requirements increase rapidly with the number of uncertain quantities of interest. The method is not Monte Carlo simulation and should not be confused with Taylor series methods. Caution should be used in applying the method to cases in which the transformation of uncertain quantities severely changes the distributional form, or to moments higher than order two. In particular, when the functions of the variable X to be integrated are not well represented by a third-order polynomial and when the Ω of X is large, the results can be seriously in error.

Despite these limitations, the method remains a simple, direct, and effective method of computing the low-order moments of functions of random variables. Its continued use in geotechnical reliability analysis is justified by experience and theory, but users familiar with the mathematical principles underlying it will be less likely to apply it outside the regions where it can be expected to yield good results.

16 The Hasofer–Lind Approach (FORM)

The previous two chapters describe the basic procedures for carrying out a first-order reliability analysis. The analysis starts by establishing the statistical properties of the variables – means, variances, and covariances – and then propagates these through the analytical model for some failure criterion such as the factor of safety (F) or the margin of safety (M). The results are expressed in the form of the mean and standard deviation of F or M, which are then used to compute the reliability index (β). With an assumption about the distribution of F or M, the value of β leads to a value of the probability of failure (p_f).

While the FOSM method and its extension by the Rosenblueth point-estimate method are powerful tools that usually give excellent results – or at least results that are as accurate as the underlying data – they do involve some approximations that may not be acceptable. One is the assumption that the moments of the failure criterion can be estimated accurately enough by starting with the mean values of the variables and extrapolating linearly. In other words, there is an implicit assumption that it makes little difference where the partial derivatives are evaluated. A second is that the form of the distribution of F or M is known and can be used to compute p_f from β. In practice these assumptions are seldom valid.

Hasofer and Lind (1974) addressed these concerns by proposing a different definition of the reliability index that leads to a geometric interpretation. The approach has been named after them, but it is also known as geometric reliability and as the First Order Reliability Method (FORM). To minimize confusion with the FOSM method, we have chosen to use the term FORM as little as possible here, but it is often seen, especially in the structural engineering literature. We begin by revisiting the simple problem of the stability of the vertical cut in cohesive soil from Chapter 13.

16.1 Justification for Improvement – Vertical Cut in Cohesive Soil

Example 13.1 described the reliability analysis of a vertical cut in cohesive soil expressed in terms of the margin of safety M

$$M = c - \gamma H/4 \tag{16.1}$$

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

where c is the cohesion, γ is the unit weight, and H is the height of the cut. For a 10 m deep cut and uncorrelated cohesion and unit weight, the following values were used:

$$\begin{aligned} H &= 10\,\text{m} \mu_c = 100\,\text{kPa} \qquad \sigma_c = 30\,\text{kPa} \\ \mu_\gamma &= 20\,\text{kN/m}^3 \qquad \sigma_\gamma = 2\,\text{kN/m}^3 \end{aligned} \tag{16.2}$$

Then

$$\begin{aligned} \mu_M &= 100 - (20)(10)/4 = 50\,\text{kPa} \\ \sigma_M^2 &= (30)^2 + (2)^2(10/4)^2 = 925(\text{kPa})^2 \\ \beta &= 50/\sqrt{925} = 1.64 \end{aligned} \tag{16.3}$$

If the cohesion and unit weight are both Normally distributed, any linear combination of them, such as M, must also be normally distributed, and

$$p_f = P[M \le 0] = \Phi(-\beta) = 5.01 \times 10^{-2} \tag{16.4}$$

Note that, because the variables are normally distributed and M is a linear combination of them, these results are exact.

Now, let us solve the problem using FOSM. The derivatives are

$$\frac{\partial M}{\partial c} = 1 \quad \text{and} \quad \frac{\partial M}{\partial \gamma} = -\frac{H}{4} \tag{16.5}$$

The value of μ_M is unchanged at 50 kPa. The FOSM estimate for σ_M is

$$\sigma_M^2 \approx \left(\frac{\partial M}{\partial c}\right)^2 \sigma_c^2 + \left(\frac{\partial M}{\partial \gamma}\right)^2 \sigma_\gamma^2 = (1)^2(30)^2 + \left(\frac{10}{4}\right)^2 (2)^2 = 925\,\text{kPa} \tag{16.6}$$

which is identical to the exact result. The rest of the calculation gives once again $\beta = 1.64$ and $p_f = 5.01 \times 10^{-2}$. This is what one would expect from a linear combination of Normal variables.

However, the problem could also have been stated in terms of the factor of safety

$$F = \frac{4c}{\gamma H} \tag{16.7}$$

$$\frac{\partial F}{\partial c} = \frac{4}{\gamma H} \quad \text{and} \quad \frac{\partial F}{\partial \gamma} = -\frac{4c}{H}\frac{1}{\gamma^2} \tag{16.8}$$

If the evaluations are done at the mean values of the variables, then

$$\frac{\partial F}{\partial c} = 0.020 \quad \frac{\partial F}{\partial \gamma} = -0.100 \tag{16.9}$$

$$\mu_F \approx \frac{4\mu_c}{H\mu_\gamma} = 2.00 \tag{16.10}$$

$$\sigma_F^2 \approx \left(\frac{\partial F}{\partial c}\right)^2 \sigma_c^2 + \left(\frac{\partial F}{\partial \gamma}\right)^2 \sigma_\gamma^2 = (0.020)^2(30)^2 + (0.100)^2(2)^2 = 0.4 \tag{16.11}$$

$$\beta = (2.0 - 1.0)/\sqrt{0.4} = 1.58 \tag{16.12}$$

$$p_f = 3.81 \times 10^{-2} \tag{16.13}$$

The calculation of the probability of failure assumes F is Normally distributed, which is certainly not the case. However, it is also not clear what the correct distribution of F is.

This example shows that, even for a very simple problem, the FOSM procedure gives different results for calculations based on M and on F even though the failure conditions ($M = 0$ and $F = 1$) are mathematically identical. Discrepancies and inaccuracies can be expected to become larger as problems become more complicated.

Figure 16.1 illustrates the differences between the two computational approaches. Values of c are plotted along the horizontal axis, and values of $\gamma H/4$ are plotted along the vertical axis. The large black dot represents the mean values of both variables. The heavy black line represents the failure condition ($M = 0$ or $F = 1$). Points above it and to the left are unsafe, and those below it and to the right are safe. The dashed lines represent constant values of M. The solid lines represent constant values of F. The lines for $M = 0$ and $F = 1$ coincide.

Reliability analysis is an attempt to quantify how close the system is to failure, that is how close the black dot is to the heavy failure line. The FOSM analysis based on M does this by treating M as a surface whose contours are the dashed lines. Then the most rapid line of descent to $M = 0$ is the negative gradient of the surface (the positive gradient is uphill). The negative gradient is identified on the figure, but, because the contours are parallel, the same gradient would be found anywhere on the M surface. When the same procedure is applied to the F surface, whose contours are the solid lines, the gradient does not represent the shortest distance to the failure condition. Furthermore, different gradients are found at different points. In particular, if we had been able to evaluate the gradient at a point on the line representing the failure condition ($F = 1$), it would have had the same orientation as the gradient from the M surface and the results from the two approaches would have been the same. The failure to achieve this result is a deficiency of the FOSM approach.

Furthermore, even if we did find the shortest distance from the black dot to the failure surface and if we knew the probability densities of the variables c and γ, we would not

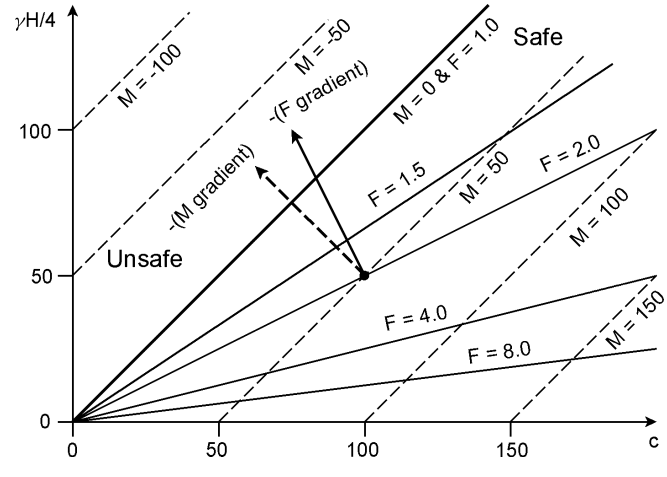


Figure 16.1 Factor of safety reliability compared to margin of safety reliability. Both methods start from the same point (shown as a large solid dot), but they go in different directions and meet the failure line at different points.

know the distribution of F. The computation of probability of failure requires that we assume that F has one of the common distributions, such as the Normal or LogNormal distributions. The Hasofer–Lind approach deals with both these problems.

16.2 The Hasofer–Lind Formulation

The first step in the Hasofer–Lind approach is to reformulate the problem with dimensionless variables. There are n uncertain variables, and each is identified by a subscript i. Each variable x_i is defined in terms of its mean μ_{x_i} and its standard deviation σ_{x_i}. We now define a primed variable

$$x_i' = \frac{x_i - \mu_{x_i}}{\sigma_{x_i}} \tag{16.14}$$

which is dimensionless and has the mean value of zero and unit standard deviation. It follows that

$$x_i = \sigma_{x_i} x_i' + \mu_{x_i} \tag{16.15}$$

We can rewrite any function of the unprimed variables as a function $g(x_1', x_2', \ldots, x_i', \ldots, x_n')$ of the primed variables. In the example of the vertical cut the failure criterion $F = 1$ or $M = 0$ gives

$$g(c', \gamma') = c'\sigma_c - \frac{H}{4}\gamma'\sigma_\gamma + \left(\mu_c - \frac{H}{4}\mu_\gamma\right) = 0 \tag{16.16}$$

We can express these relations in matrix form, using bold-face letters to indicate vectors and matrices. Then

$$\mathbf{x} \equiv \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \tag{16.17}$$

$$\mathbf{x}' \equiv \{x_1', x_2', \ldots, x_i', \ldots, x_n'\} \tag{16.18}$$

$$g(\mathbf{x}') = 0 \tag{16.19}$$

From the definition of the primed variables,

$$\frac{dx_i}{dx_i'} = \sigma_{x_i} \quad \frac{dx_i'}{dx_i} = \frac{1}{\sigma_{x_i}} \tag{16.20}$$

so, for any function f of the unprimed variables,

$$\frac{\partial f}{\partial x_i'} = \frac{\partial f}{\partial x_i}\frac{dx_i}{dx_i'} = \sigma_{x_i}\frac{\partial F}{\partial x_i} \tag{16.21}$$

Now let us return to the basic case of the reliability of a system with loading Q and resistance R, which was discussed in Chapter 13. From the above definitions we can

express reduced variables

$$R' = \frac{R - \mu_R}{\sigma_R}$$
$$Q' = \frac{Q - \mu_Q}{\sigma_Q} \tag{16.22}$$

If R and Q are uncorrelated, Equation (13.1) for the margin of safety becomes

$$M = R - Q = \sigma_R R' - \sigma_Q Q' + \mu_R - \mu_Q = 0 \tag{16.23}$$

Figure 16.2 is a plot of the failure criterion using the reduced variables as the axes. The origin is the point at which both R and Q equal their mean values, that is, the point at which each variable has its expected value. The distance d between the origin and the line $M = 0$ is

$$d = \frac{\mu_R - \mu_Q}{\sqrt{\sigma_R^2 + \sigma_Q^2}} \tag{16.24}$$

which is identical to the definition of the reliability index β. This result suggests that the reliability index can be interpreted *geometrically* as the distance between the point defined by the expected values of the variables and the closest point on the failure criterion.

Figure 16.3 presents another view of this geometric interpretation. The horizontal and vertical axes are the normalized variables, which are assumed to be Normally distributed and independent. Their joint distribution is the hill with its peak at the origin. The linear failure criterion ($M = 0$) is represented as a wall that cuts the failure region off from the rest of the distribution, so the probability of failure is the integral under the joint probability distribution surface in the failure region. If we now find the marginal distribution on a plane perpendicular to the wall, Figure 16.4 results. In effect, we are looking along the wall and determining how much of the distribution lies on each side. The probability

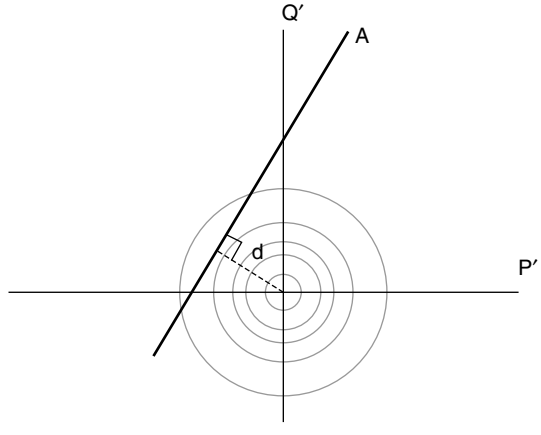


Figure 16.2 Plot of Resistance (R) and Load (Q) showing definition of reliability index.

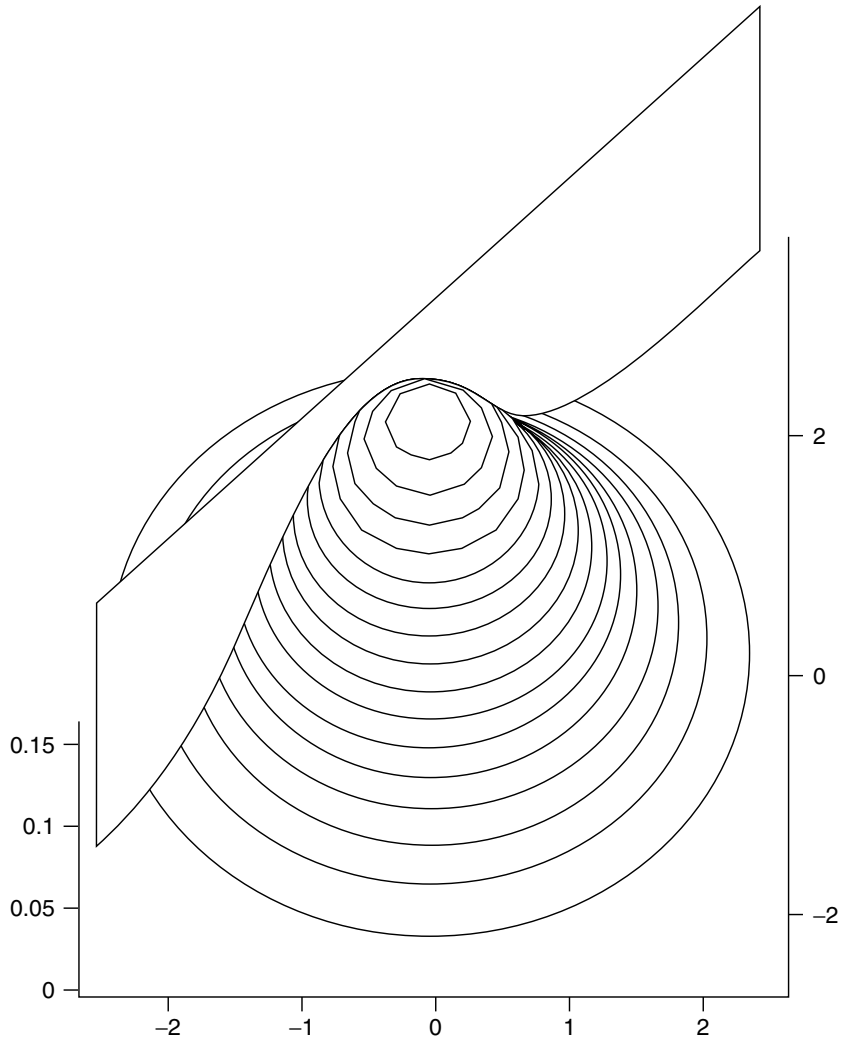


Figure 16.3 Joint probability distribution of two independent Normal variables and a linear failure criterion.

of failure is the area under the Normal curve and lying to the left of the failure line. This probability is a function of the dimensionless distance β.

Note that the joint Normal distribution of the two variables and the Normal marginal distribution in Figure 16.4 represent the distributions of the variables – not the distribution of the failure function. In the development so far, the failure function can have any shape as long as the criterion ($F = 1$ or $M = 0$) is a straight line. This contrasts with the FOSM and Point-Estimate methods, in which the computations are directed at establishing statistical parameters for the failure functions. One of the arguments in favor of the Hasofer–Lind formulation is that it does not require that the distribution of the failure function be found.

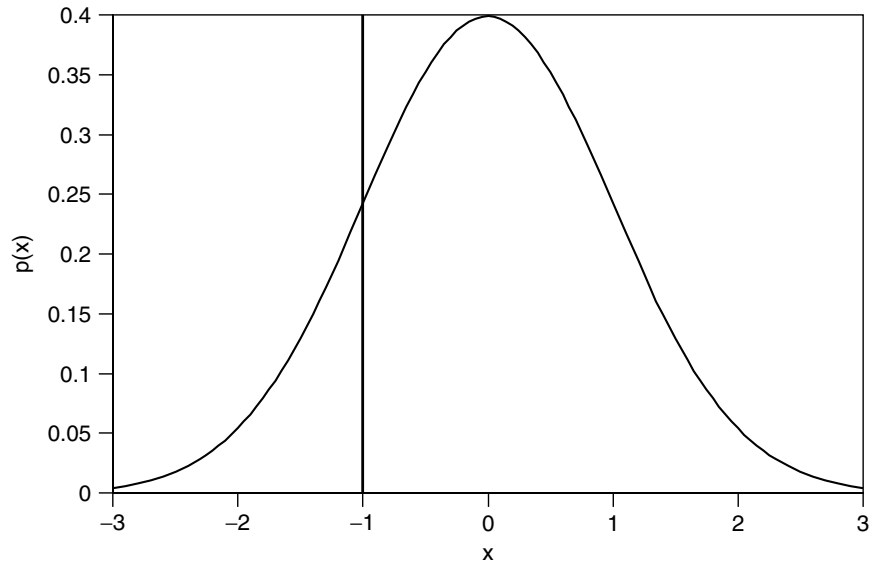


Figure 16.4 Marginal distribution of margin of safety from Figure 3.

This result can be generalized when there are more than two variables. The failure criterion is Equation (16.19). In the multidimensional space, the distance from the origin to a point on the failure criterion is

$$d = \sqrt{x_1'^2 + x_2'^2 + \cdots + x_n'^2} = (\mathbf{x}'^T \mathbf{x})'^{1/2} \tag{16.25}$$

where $\mathbf{x}'$ is the vector of the x_i' s and the superscript T indicates the transpose of a matrix or vector. The problem is then a constrained minimization that can be summarized as the requirement to minimize d with the constraint that Equation (16.19) is satisfied. Two practical techniques recommend themselves:

1. Express the constraint (Equation (16.19)) and the definition of d in a spreadsheet such as EXCEL or QUATTRO-PRO or other mathematical software such as MATLAB or MATHCAD. These systems have commands, usually named 'solve' or 'minimize,' that direct the computer to perform the minimization. Some of the examples in this chapter were solved using this feature in MATHCAD, and Low and his colleagues (Low 1996; Low 1997; Low and Tang 1997a, 1997b; Low *et al*. 1998) have used the optimization features of the spreadsheet EXCEL.
2. Develop a computational algorithm suited to the particular problem at hand and implement it either in a special purpose computer program or by hand. Since it is always a good idea for the analyst to know what is going on rather than simply to use a black box, the following section describes how the calculations can be carried out without relying on a canned computer program.

16.3 Linear or Non-linear Failure Criteria and Uncorrelated Variables

To describe the best-known procedures for finding the minimum of Equation (16.25), we will start with a generalized approach using the technique of the Lagrangian multiplier. We will then show that this gives results identical to a revised FOSM formulation. The section ends with a discussion of the Rackwitz (Rackwitz and Fiessler 1978) algorithm, which is the most widely used approach.

16.3.1 The Lagrangian Multiplier Approach

The problem is to minimize Equation (16.25) subject to the constraint that Equation (16.19) is satisfied. First, we define the *gradient* of the failure function with respect to the primed variables

$$\mathbf{G} = \left(\frac{\partial g}{\partial x_1'}, \frac{\partial g}{\partial x_2'}, \ldots, \frac{\partial g}{\partial x_n'} \right) \tag{16.26}$$

Since $g(\mathbf{x}') = 0$ at the point to be found, minimizing d is equivalent to minimizing L in the following equation:

$$L = d + \lambda g(\mathbf{x}') = (\mathbf{x}'^T \mathbf{x}')^{1/2} + \lambda g(\mathbf{x}') \tag{16.27}$$

The term λ is called a *Lagrangian Multiplier*. It is zero at the solution point, so the minimum of L and d must be the same, but including λ makes the rest of the development easier.

For L to be a minimum all its partial derivatives must be zero. Therefore,

$$\frac{\partial L}{\partial x_i'} = \frac{x_i'}{(\mathbf{x}'^T \mathbf{x}')^{1/2}} + \lambda \frac{\partial g}{\partial x_i'} = 0 \tag{16.28}$$

and

$$\frac{\partial L}{\partial \lambda} = g(\mathbf{x}') = 0 \tag{16.29}$$

which is nothing more than a restatement of Equation (16.19).

Rewriting Equations (16.28) in matrix form gives

$$\begin{bmatrix} \frac{1}{(\mathbf{x}'^T \mathbf{x}')^{1/2}} & 0 & \cdots & 0 \\ 0 & \frac{1}{(\mathbf{x}'^T \mathbf{x}')^{1/2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{(\mathbf{x}'^T \mathbf{x}')^{1/2}} \end{bmatrix} \begin{Bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{Bmatrix} = - \begin{Bmatrix} \lambda \frac{\partial g}{\partial x_1'} \\ \lambda \frac{\partial g}{\partial x_2'} \\ \vdots \\ \lambda \frac{\partial g}{\partial x_n'} \end{Bmatrix} \tag{16.30}$$

Hence,

$$x_i' = -\lambda d \frac{\partial g}{\partial x_i'} \tag{16.31}$$

$$\mathbf{x}' = -\lambda d \mathbf{G} \tag{16.32}$$

$$d = (\mathbf{x}'^T \mathbf{x}')^{1/2} = [(\lambda d \mathbf{G})^T (\lambda d \mathbf{G})]^{1/2} = \lambda d (\mathbf{G}^T \mathbf{G})^{1/2} \tag{16.33}$$

$$\lambda = (\mathbf{G}^T \mathbf{G})^{-1/2} \tag{16.34}$$

$$\mathbf{x}' = -\frac{d\mathbf{G}}{(\mathbf{G}^T \mathbf{G})^{1/2}} \tag{16.35}$$

$$\mathbf{G}^T \mathbf{x}' = -\frac{d\mathbf{G}^T \mathbf{G}}{(\mathbf{G}^T \mathbf{G})^{1/2}} = -d(\mathbf{G}^T \mathbf{G})^{1/2} \tag{16.36}$$

$$d = -\frac{\mathbf{G}^T \mathbf{x}'}{(\mathbf{G}^T \mathbf{G})^{-1/2}} \tag{16.37}$$

This is the minimum value of d, and when we impose the condition that the minimum must satisfy the failure criterion ($g = 0$), we obtain the reliability index β:

$$\beta = d_{\min} = -\frac{\mathbf{G}^{*T} \mathbf{x}'^*}{(\mathbf{G}^{*T} \mathbf{G}^*)^{1/2}} = -\frac{\sum x_i'^* \left(\dfrac{\partial g}{\partial x_i'}\right)_*}{\sqrt{\sum \left(\dfrac{\partial g}{\partial x_i'}\right)_*^2}} \tag{16.38}$$

In this equation, the superscript or subscript star indicates that the term or derivative is evaluated at the nearest point on the failure criterion.

16.3.2 The Taylor Series Approach

We start from the starred point on the failure criterion even though we have not developed a method for calculating the location of this point. The Taylor series for finding the value of g at some other point is

$$g(x_1, x_2, \ldots, x_n) = g(x_1^*, x_2^*, \ldots, x_n^*) + \sum (x_i - x_i^*) \left(\frac{\partial g}{\partial x_i}\right)_* + \sum \sum (x_i - x_i^*)(x_j - x_j^*) \left(\frac{\partial^2 g}{\partial x_i \partial x_j}\right)_* + \cdots \tag{16.39}$$

This expression is in terms of the unprimed variables. After removing the higher order terms and recognizing that $g = 0$ at the failure criterion, we get

$$\begin{aligned} g(x_1, x_2, \ldots, x_n) &\approx \sum (x_i - x_i^*) \left(\frac{\partial g}{\partial x_i}\right)_* \\ &= \sum (\sigma_{x_i} x_i' + \mu_{x_i} - \sigma_{x_i} x_i'^* - \mu_{x_i}) \left(\frac{\partial g}{\partial x_i'}\right)_* \frac{1}{\sigma_{x_i}} \\ &= \sum (x_i' - x_i'^*) \left(\frac{\partial g}{\partial x_i'}\right)_* \end{aligned} \tag{16.40}$$

The primed variables are, by definition, zero when the unprimed variables have their mean values. Hence,

$$\mu_g \approx -\sum x_i^{'*} \left(\frac{\partial g}{\partial x_i'}\right)_* \tag{16.41}$$

and

$$\sigma_g^2 \approx \sum \sigma_{x_i}^2 \left(\frac{\partial g}{\partial x_i}\right)_*^2 = \sum \sigma_{x_i}^2 \left(\frac{\partial g}{\partial x_i'}\right)_*^2 \left(\frac{1}{\sigma_{x_i}}\right)^2 = \sum \left(\frac{\partial g}{\partial x_i'}\right)_*^2 \tag{16.42}$$

Therefore,

$$\beta = \frac{\mu_g}{\sigma_g} = -\frac{\sum x_i^{'*} \left(\frac{\partial g}{\partial x_i'}\right)_*}{\sqrt{\sum \left(\frac{\partial g}{\partial x_i'}\right)_*^2}} \tag{16.43}$$

which is identical to the result found using the Lagrangian Multiplier.

16.3.3 Solving the Equations

Both the above approaches give the same final equation for the reliability index, but they do not describe how to find the starred point at which Equations (16.38) or (16.43) are to be evaluated. As in any other non-linear minimization problem, the choice of algorithm depends on the specific function to be minimized and the constraints, and these will vary from problem to problem. There is a large literature on minimization problems; Press *et al*. (1992) give a good general introduction to the subject. Lin and Der Kiureghian (1991) examine a number of algorithms that have been proposed for the reliability problem. Rackwitz and his colleagues (Rackwitz 1976; Rackwitz and Fiessler 1978) proposed a technique that remains widely used. Ang and Tang (1990) included it in the second volume of their book, and it is described below.

Equation (16.26) in Section 16.3.1 defines $\mathbf{G}$ as the gradient of the failure function with respect to the primed variables. This can be normalized into a unit vector $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = \frac{\mathbf{G}}{(\mathbf{G}^T\mathbf{G})^{1/2}} \quad \alpha_i = \frac{\left(\frac{\partial g}{\partial x_i'}\right)}{\sqrt{\sum \left(\frac{\partial g}{\partial x_i'}\right)^{1/2}}} \tag{16.44}$$

We can use a superscript star to indicate that the unit vector is evaluated at the failure point. It then follows that the coordinates of the failure point must be

$$x_i^{'*} = -\alpha_i^* \beta \tag{16.45}$$

The Rackwitz algorithm then proceeds in six iterative steps (Ang and Tang 1990):

1. Assume initial values of x_i^* and compute the corresponding values of $x_i^{'*}$.

2. Compute **G** and $\boldsymbol{\alpha}$ at $x_i'^*$.
3. Form the expressions for the new $x_i^* = \mu_{x_i} - \alpha_i \sigma_{x_i} \beta$.
4. Substitute these expressions into $g(x_1^*, x_2^*, \ldots, x_n^*) = 0$, and solve for β.
5. With this value of β, calculate new values of $x_i'^* = -\alpha_i \beta$.
6. Repeat steps 2 through 5 until the process converges.

The process is best understood from examples, two of which follow. The first is the failure of the vertical cut with which the chapter started, and the second is the Culmann single plane slope failure analysis discussed in Chapter 15.

Example 16.1 – Reliability of Vertical Cut Using Factor of Safety

The failure criterion based on the factor of safety is $F - 1 = 0$, so

$$g(c, \gamma) = \frac{4c}{\gamma H} - 1 = 0$$

$$\frac{\partial g}{\partial c} = \frac{4}{\gamma H} \qquad \frac{\partial g}{\partial c'} = \frac{4}{\gamma H}\sigma_c$$

$$\frac{\partial g}{\partial \gamma} = -\frac{4c}{\gamma^2 H} \qquad \frac{\partial g}{\partial \gamma'} = -\frac{4c}{\gamma^2 H}\sigma_\gamma$$

First iteration:

Step 1 – assume $c^* = \mu_c = 100$ and $\gamma^* = \mu_\gamma = 20$, so $c' = \gamma' = 0$.

Step 2 – $\dfrac{\partial g}{\partial c'} = \dfrac{(4)(30)}{(20)(10)} = 0.6 \qquad \dfrac{\partial g}{\partial \gamma'} = -\dfrac{(4)(100)(2)}{(20)^2(10)} = -0.2$

$$\mathbf{G} = \begin{Bmatrix} 0.6 \\ -0.2 \end{Bmatrix} \boldsymbol{\alpha} = \begin{Bmatrix} 0.948683 \\ -0.316228 \end{Bmatrix}$$

Step 3 – $c^* = 100 - (0.948683)(30)\beta = 100 - 28.460490\beta$

$$\gamma^* = 20 + (0.316228)(2)\beta = 20 + 0.632456\beta$$

Step 4 – $[4(100 - 28.460490\beta)]/[10(20 + 0.632546\beta)] = 1$

$$\beta = 1.664357$$

Step 5 – $c'^* = -(1.664357)(0.948683) = -1.578947$

$$\gamma'^* = +(1.664357)(0.316228) = +0.526316$$

Second iteration:

Step 1 – from first iteration $c^* = 52.631584$ and $\gamma^* = 22.052633$

Step 2 –

$$\frac{\partial g}{\partial c'} = \frac{(4)(30)}{(21.052633)(10)} = 0.570000 \qquad \frac{\partial g}{\partial \gamma'} = -\frac{(4)(52.631584)(2)}{(21.052633)^2(10)} = -0.095000$$

$$\mathbf{G} = \begin{Bmatrix} 0.570 \\ -0.095 \end{Bmatrix} \qquad \boldsymbol{\alpha} = \begin{Bmatrix} 0.986394 \\ -0.164399 \end{Bmatrix}$$

Step 3 – $c^* = 100 - (0.986394)(30)\beta = 100 - 29.591820\beta$

$\gamma^* = 20 + (0.164399)(2)\beta = 20 + 0.328798\beta$

Step 4 – $[4(100 - 29.591820\beta)]/[10(20 + 0.328798\beta)] = 1$

$\beta = 1.64399$

Step 5 – $c'^* = -(1.64399)(0.986394) = -1.62162$

$\gamma'^* = +(1.64399)(0.164399) = +0.27027$

Further iterations will not change these results, within the limits of numerical accuracy. The value of β is identical to that found directly from the margin of safety. This demonstrates that the Hasofer–Lind method gives consistent results whether the failure criterion is expressed as a margin of safety or as a factor of safety, so long as the actual failure conditions are the same.

Example 16.2 – Reliability of Single Plane Failure (Culmann Analysis)

One of the examples in Chapter 15 was the calculation of the reliability of a slope when failure occurs along a single plane; this is known as the Culmann analysis. The margin of safety M is

$$M = c + \left[\frac{1}{2}\frac{H}{\sin\psi}\sin(\psi - \theta)\cos\theta\right][\tan\phi - \tan\theta]\gamma$$

so the failure criterion is $M = 0$. It is assumed that only γ, c, and $\tan\phi$ are uncertain; the other parameters have fixed values. The parameters are listed at the bottom of Table 16.1.

In most cases, it is necessary to resort to central difference numerical approximations to obtain the partial derivatives needed to carry out the Hasofer–Lind analysis, but the equation for M is simple enough that they can be evaluated analytically. First, to simplify the expressions, the first term in brackets is replaced by the symbol A. For the values in Table 16.1, $A = 1.120336$. Then

$$\frac{\partial M}{\partial \gamma} = A(\tan\phi - \tan\theta) \qquad \frac{\partial M}{\partial \gamma'} = A(\tan\phi - \tan\theta)\sigma_\gamma$$

$$\frac{\partial M}{\partial c} = 1 \qquad \frac{\partial M}{\partial c'} = \sigma_c$$

$$\frac{\partial M}{\partial \tan\phi} = A\gamma \qquad \frac{\partial M}{\partial \tan\phi'} = A\gamma\sigma_{\tan\phi}$$

Table 16.1 shows the numerical results for each step in the first four iterations, starting with the mean values of the uncertain parameters. The last column of the table shows the results of using the built-in minimization function of either MATHCAD or a spreadsheet. The Rackwitz iteration converges satisfactorily in two iterations, and the iterated value of β agrees with the value computed from the built-in minimization to seven significant figures after four iterations. Although not shown here, the values of the primed variables agree to seven significant figures after eight iterations.

Table 16.1 Parameters and Hasofer–Lind analysis for Culmann failure, Case 1

		Iteration Number				
Step	Parameter	1	2	3	4	Minimize
1	γ	22 kN/m^3	24.519433	25.261950	25.223633	
	c	5 kPa	3.790286	3.910041	3.924214	
	$\tan\phi$	0.267949	0.225991	0.225815	0.225104	
2	$\partial g/\partial\gamma'$	−0.236667	−0.340082	−0.340516	−0.342269	
	$\partial g/\partial c'$	0.500000	0.500000	0.500000	0.500000	
	$\partial g/\partial \tan\phi'$	0.462298	0.515240	0.530843	0.530037	
	α_γ	−0.328283	−0.428081	−0.423092	−0.425158	
	α_c	0.693555	0.629378	0.621252	0.621088	
	$\alpha_{\tan\phi}$	0.641258	0.648561	0.659574	0.658400	
3	$\alpha_\gamma\sigma_\gamma$	−0.722223	−0.941777	−0.930803	−0.935349	
	$\alpha_c\sigma_c$	0.346778	0.314689	0.310626	0.310544	
	$\alpha_{\tan\phi}\sigma_{\tan\phi}$	0.012028	0.012165	0.012371	0.012349	
4	β	3.4884	3.4636	3.4633	3.4633	3.4633
5	γ'	1.145197	1.482704	1.465288	1.472440	1.471572
	c′	−2.419428	−2.179919	−2.151572	−2.150997	−2.150653
	$\tan\phi'$	−2.236991	−2.246362	−2.284292	−2.280218	−2.281102

Values of Statistical Parameters

$\mu_\gamma = 22$ kN/m^3	$\sigma_\gamma = 2.2$ kN/m^3 ($\Omega = 0.1$)	H = 10 m
$\mu_c = 5$ kPa	$\sigma_c = 0.5$ kPa ($\Omega = 0.1$)	$\theta = 20^\circ$
$\mu_{\tan\phi} = 0.267949$ ($\phi = 15^\circ$)	$\sigma_{\tan\phi} = 0.018756$ ($\Omega = 0.07$)	$\psi = 26^\circ$

If the variances in the uncertain variables are increased and the mean of $\tan\phi$ is set at 0.577350 (corresponding to $\phi = 30^\circ$), the results are as shown in Table 16.2. The convergence is slower, but the results of the Rackwitz iteration and the built-in minimization again agree to seven significant figures after eight iterations.

When the mean frictional contribution corresponds to $\phi = 15^\circ$, the factor of safety computed from the mean values of the variables is 1.294, and, when it corresponds to $\phi = 30^\circ$, the factor of safety becomes 2.144. However, Tables 16.1 and 16.2 show that the larger uncertainties in the latter case lower the reliability. If one calculates the exact values of the mean and standard deviation of the margin of safety and the reliability index as their ratio, the results are

Case of Table 16.1

$$\mu_M = 2.6333 \quad \sigma_M = 0.7224 \quad \beta = 3.6452$$

Case of Table 16.2

$$\mu_M = 10.2593 \quad \sigma_M = 4.1517 \quad \beta = 2.4711$$

Table 16.2 Parameters and Hasofer–Lind analysis for Culmann failure, Case 2

Step	Parameter	Iteration Number 1	2	3	4	Minimize
1	γ 22 kN/m^3	18.856275	23.692739	23.056535		
	c 5 kPa	2.282954	2.111294	2.550344		
	$\tan\phi$ 0.577350	0.255903	0.284430	0.265239		
2	$\partial g/\partial\gamma'$	1.051853	−0.532714	−0.392090	−0.486696	
	$\partial g/\partial c'$	2.000000	2.000000	2.000000	2.000000	
	$\partial g/\partial \tan\phi'$	3.415244	2.97217	3.678022	3.579259	
	α_γ	0.256853	−0.148594	−0.093245	−0.117875	
	α_c	0.488382	0.557877	0.475630	0.484388	
	$\alpha_{\tan\phi}$	0.833972	0.816513	0.874689	0.866876	
3	$\alpha_\gamma\sigma_\gamma$	1.130153	−0.653815	−0.410278	−0.518649	
	$\alpha_c\sigma_c$	0.976764	1.115753	0.951260	0.968777	
	$\alpha_{\tan\phi}\sigma_{\tan\phi}$	0.115559	0.113139	0.121200	0.120118	
4	β	2.7817	2.5890	2.5752	2.5744	2.5744
5	γ'	−0.714483	0.384713	0.240122	0.303460	0.295097
	c'	−1.358523	−1.444353	−1.224828	−1.247023	−1.237608
	$\tan\phi'$	−2.319844	−2.113967	−2.252472	−2.231708	−2.238007

Values of Statistical Parameters

$\mu_\gamma = 22$ kN/m^3	$\sigma_\gamma = 4.4$ kN/m^3 ($\omega = 0.2$)	H = 10 m
$\mu_c = 5$ kPa	$\sigma_c = 2.0$ kPa ($\omega = 0.4$)	$\theta = 20^\circ$
$\mu_{\tan\phi} = 0.577350$ ($\phi = 30^\circ$)	$\omega_{\tan\phi} = 0.138564$ ($\omega = 0.24$)	$\psi = 26^\circ$

The values of β are close but not identical to those in Tables 16.1 and 16.2. This is not an unexpected result. The definitions of reliability index are different. When β is defined as the ratio of the mean and standard deviation of the failure function, it is a measure of the location and shape of the failure function. If we assume or know the shape of the distribution of the failure function (e.g. Normal or LogNormal) the reliability index can be used to compute the probability of failure. In the Hasofer-Lind analysis the reliability index defines the distance from the means of the variables to the failure point. The probability of failure is computed by evaluating how much of the distribution of the variables falls on the unsafe side. The probability distribution of the failure function does not enter the calculations, and the shape of the failure function enters only insofar as it defines the boundary between the regions. The two definitions agree only when the failure function is linear.

Many non-linear problems have multiple solutions, and a particular iterative scheme may not converge or may converge on a solution other than the optimal one. In the present cases three iterative procedures – the Rackwitz algorithm, the MATHCAD minimization, and a spreadsheet minimization – give the same results, so we can have some confidence in the results. In other cases the analyst must examine the results to be sure that they are reasonable.

16.4 Higher Order Reliability

Most practical cases involve non-linear failure functions. When there are only two variables X and Y, the relations can be examined on a two-dimensional plot as in Figure 16.5. The joint probability density functions plots upwards out of the paper. Line A is the case of the linear failure criterion, already illustrated in Figures 16.2 and 16.3. The probability of failure is the volume under the joint probability density function on the unsafe side of the failure criterion.

Curve B is a convex failure criterion. Since the linear failure line is tangent to curve B, the volume on the unsafe side of curve B must be less than that on the unsafe side of curve A. The probability of failure computed on the assumption that the failure criterion is linear is, therefore, a conservative estimate. Figure 16.6 presents an oblique view of this case. Curve C is a concave failure criterion. Some of the volume under curve C lies on the safe side of curve A, so the linear estimate is unconservative. Of course, the failure criterion could be convex on one side of the point of tangency and concave on the other. It could be a complicated wavy curve. The relations become more complicated when more than two variables are involved.

Two issues have to be addressed. First is the question of how adequate is the linear approximation. Figure 16.7 shows the failure criterion for the Culmann problem. It as a hyperbolic paraboloid, so it is curved everywhere. However, comparisons between the results of Hasofer-Lind analyses and Monte Carlo simulation for the example problems show that the differences between the probabilities of failure computed from the linearized Hasofer–Lind approach (FORM) and from simulation are not large. For most geotechnical problems the Hasofer-Lind approach gives satisfactory results.

Secondly, if the Hasofer–Lind approach does not give adequate results, what is to be done next? As suggested above and described in more detail in Chapter 17, Monte Carlo simulation with variance reduction can be used to improve the estimate of the probability of failure. A second approach is to compute directly the probability volume

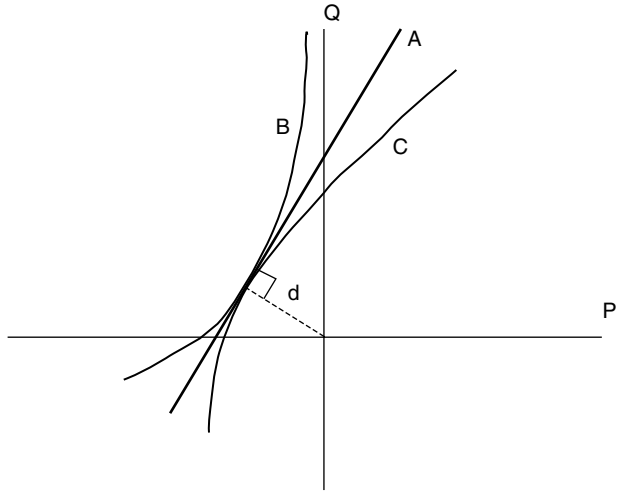


Figure 16.5 Linear, convex, and concave failure criteria for two independent variables.

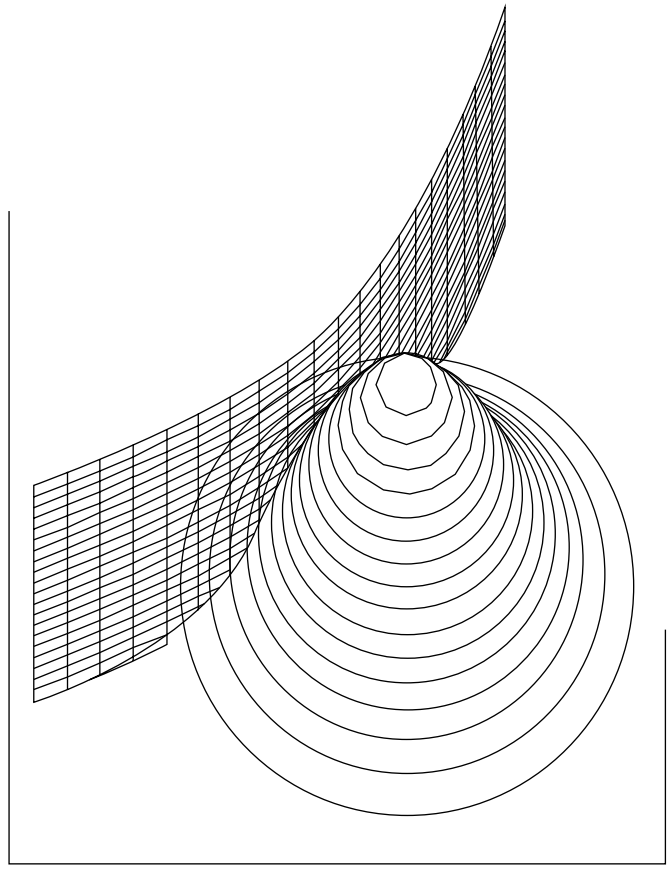


Figure 16.6 Joint probability distribution of two independent Normal variables and a non-linear failure criterion.

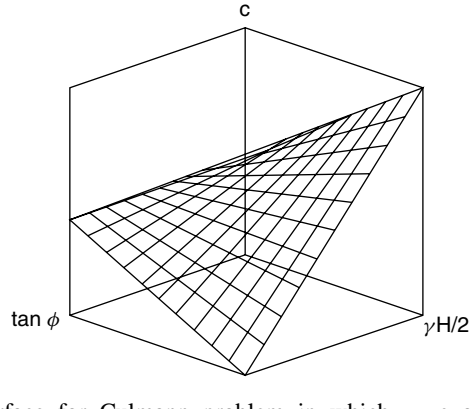


Figure 16.7 Failure surface for Culmann problem in which γ, c and tan Φ are the uncertain variables.

in the unsafe region. Der Kiureghian and his colleagues (1987; 1991), and many others, have developed methods for solving the problem of the Second Order Reliability Method (SORM). These techniques may involve numerical integration of some formidable integrals (Breitung 1984). In any case, the technique to be used depends on the form of the non-linearities, so they are most applicable in situations like the development of structural codes where the forms of the non-linearities are known. Since this is not the case for broad classes of geotechnical problems, SORM has not been widely used in geotechnical reliability applications.

16.5 Correlated Variables

The previous discussion has concentrated on cases in which the uncertain variables were not correlated. The Hasofer–Lind approach can also be used effectively when the variables are correlated, but some adjustments are necessary. Two techniques are widely used – one based on the Cholesky decomposition of the correlation matrix and the other on its eigenvectors and eigenvalues.

16.5.1 Cholesky Approach

Let $\mathbf{K}$ be the correlation matrix:

$$\mathbf{K} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{bmatrix} \tag{16.46}$$

Since $\mathbf{K}$ must be symmetric and positive definite, it can be factored into two matrices that are the transposes of each other:

$$\mathbf{K} = \mathbf{S}\mathbf{S}^T \tag{16.47}$$

$\mathbf{S}$ is a lower triangular matrix, and its transpose $\mathbf{S}^T$ is an upper triangular matrix. The details of how to compute $\mathbf{S}$ are found in most books on numerical methods and are summarized in Chapter 17. Now, let $\mathbf{z}$ be a vector of n independent variables $\{z_1, z_2, \ldots, z_n\}$, each of which has a standard Normal distribution. That is, each has zero mean and unit standard deviation. We want to find a vector $\mathbf{w}$ of n variables, each of which has a standard Normal distribution, with correlation matrix $\mathbf{K}$. It can be shown that

$$\mathbf{w} = \mathbf{S}\mathbf{z} \quad \text{or} \quad w_i = \sum_{j=1}^{n} s_{ij} z_j \tag{16.48}$$

The $\mathbf{z}$ vector consists of uncorrelated primed variables that are the basis of the Hasofer–Lind approach. The $\mathbf{w}$ vector is another set of primed variables, but this set is correlated.

If we could work with the uncorrelated variables, all of the previous developments would be applicable, and in particular the distance d would be $(\mathbf{z}^T\mathbf{z})^{1/2}$. Then

$$\begin{aligned} d = (\mathbf{z}^T\mathbf{z})^{1/2} &= [(\mathbf{S}^{-1}\mathbf{w})^T(\mathbf{S}^{-1}\mathbf{w})]^{1/2} = [\mathbf{w}^T\mathbf{S}^{-1,T}\mathbf{S}^{-1}\mathbf{w}]^{1/2} \\ &= [\mathbf{w}^T\mathbf{S}^{T,-1}\mathbf{S}^{-1}\mathbf{w}]^{1/2} = [\mathbf{w}^T(\mathbf{S}\mathbf{S}^T)^{-1}\mathbf{w}]^{1/2} \\ &= (\mathbf{w}^T\mathbf{K}^{-1}\mathbf{w})^{1/2} \end{aligned} \tag{16.49}$$

In other words, correlation can be incorporated into the analysis by using the same failure criterion but replacing the distance of Equation (16.25) with

$$d = (\mathbf{x}'^T\mathbf{K}^{-1}\mathbf{x}')^{1/2} \tag{16.50}$$

Low *et al*. (1998) observe that, if $\mathbf{m}$ is the vector of means of the variables and $\mathbf{C}$ is the covariance matrix, Equation (16.50) can also be written

$$d = [(\mathbf{x} - \mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{m})]^{1/2} \tag{16.51}$$

In summary, defining the constraint as before and minimizing the distance expressed by either Equation (16.50) or Equation (16.51) results in the correlated version of the Hasofer–Lind analysis.

16.5.2 Eigenvalue and Eigenvector Approach

Shinozuka (1983) proposed another approach, which Ang and Tang (1990) also describe. The correlation matrix $\mathbf{K}$ has n positive eigenvalues, designated λ_i. For each eigenvalue there is a normalized eigenvector, designated $\boldsymbol{\phi}_i$. The matrix of the eigenvectors is

$$\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1\boldsymbol{\phi}_2\cdots\boldsymbol{\phi}_n\right] \tag{16.52}$$

From linear algebra we know that $\boldsymbol{\Phi}$ is orthonormal; that is

$$\boldsymbol{\Phi}^T = \boldsymbol{\Phi}^{-1} \tag{16.53}$$

It then follows that a vector $\mathbf{y}$ of uncorrelated variables can be obtained from $\mathbf{x}'$

$$\mathbf{y} = \boldsymbol{\Phi}^{-1}\mathbf{x}' = \boldsymbol{\Phi}^T\mathbf{x}' \qquad \mathbf{x}' = \boldsymbol{\Phi}\mathbf{y} \tag{16.54}$$

The variances of the $\mathbf{y}$ variables are the eigenvalues of $\mathbf{K}$. Furthermore,

$$\mathbf{x} = [\boldsymbol{\sigma}]\mathbf{x}' + [\boldsymbol{\mu}] = [\boldsymbol{\sigma}]\boldsymbol{\Phi}\mathbf{y} + [\mu] \tag{16.55}$$

where $[\boldsymbol{\sigma}]$ and $[\boldsymbol{\mu}]$ are diagonal matrices with the standard deviations and means of the $\mathbf{x}$ variables on the diagonals.

The procedure is first to calculate the eigenvalues and eigenvectors and express the $\mathbf{x}$ variables in terms of the uncorrelated $\mathbf{y}$ variables from Equation (16.55). Then these are substituted into the failure function, which is then expressed as a function of the

uncorrelated variables. The remainder of the calculation proceeds as in the uncorrelated case, using the new variables. When the iterations have converged, the **x** variables are back-calculated.

16.5.3 Low and Tang's Approach

Low and Tang (1997a) proposed a very efficient procedure that takes advantage of the optimization and solution techniques that are available in modern spreadsheets and mathematical software packages. They work directly with the correlated variables themselves without any rotations or transformations. If **x** is the vector of uncertain variables and **C** is their covariance matrix (not the correlation matrix) and $\boldsymbol{\mu}$ is the vector of their means, they show that the Hasofer–Lind reliability index is

$$\beta = \min \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \tag{16.56}$$

subject to the constraint that **x** must satisfy the failure criterion that $M = 0$. The procedure is as follows:

1. Values of the means and covariance matrix are defined and $\mathbf{C}^{-1}$ is calculated.
2. The functional form of the following equation is entered

$$\beta^2 = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \tag{16.57}$$

3. The failure criterion ($M = 0$) is expressed as a constraint in terms of the variables in **x**.
4. The 'minimize' or 'solve' command (depending the spreadsheet or mathematical software used) is invoked to minimize β^2 by changing the values of **x** subject to the constraint that the failure criterion is satisfied.
5. The result consists of values of the terms in **x** at the failure point and a corresponding value of β^2. Computing β and the probability of failure follows easily.

This technique is extremely easy to use. It can deal with correlated as well as uncorrelated variables. Low and Tang also show how distributions other than Normal can be employed.

16.6 Non-normal Variables

The Hasofer–Lind approach is based on independent, Normal variables. We have shown how the procedure can be modified when the variables are correlated. When some of the variables are logNormally distributed, the problem can be reformulated in terms of the logarithms of those variables, which are Normally distributed. The mathematical manipulations will become more complicated, but the basic procedures will be those described in this chapter.

For other distributions there is a procedure known as the Rosenblatt transformation (Rosenblatt 1952), which is also described by Ang and Tang (1990). This presentation follows Ang and Tang's. The transformation considers a set of n random variables

$(X_1, X_2, \ldots, X_n)$. If their joint CDF is $F(x_1, x_2, \ldots, x_n)$ and $\Phi(u)$ is the standard normal CDF, then we can develop a set of relations:

$$\begin{aligned}
\Phi(u_1) &= F_1(x_1) \\
\Phi(u_2) &= F_2(x_2|x_1) \\
&\vdots \\
\Phi(u_n) &= F_n(x_n|x_1, \ldots, x_{n-1})
\end{aligned} \tag{16.58}$$

In words, this means that the u's are scaled so that their standard Normal CDFs correspond to the conditional CDFs of the x's. Equation (16.58) implies that

$$\begin{aligned}
u_1 &= \Phi^{-1}[F_1(x_1)] \\
u_2 &= \Phi^{-1}[F_2(x_2|x_1)] \\
&\vdots \\
u_n &= \Phi^{-1}[F_n(x_n|x_1, \ldots, x_n)]
\end{aligned} \tag{16.59}$$

Figure 16.8 shows how this works in the case of one variable. On the right is the CDF of a function x that has an arbitrary distribution. For each value of x, there corresponds a value of the CDF. On the left is the CDF of the standard Normal variable, and for the particular value of the CDF there corresponds a value u. In most cases the failure functions and its derivatives are expressed in terms of the x's, but the analysis is driven by the u's. During the analysis one moves from a set of the u's to corresponding values of the x's by the following procedure:

1. For all the u_i and find $\Phi(u_i)$.
2. Find $x_1 = F_1^{-1}[\Phi(u_1)]$.
3. Find $F_2^{-1}[\Phi(u_2)]$ and from the result and the value of x_1 calculate x_2.
4. Find $F_3^{-1}[\Phi(u_3)]$ and from the result and the values of x_1 and x_2 calculate x_3.
5. Repeat for the rest of the variables.

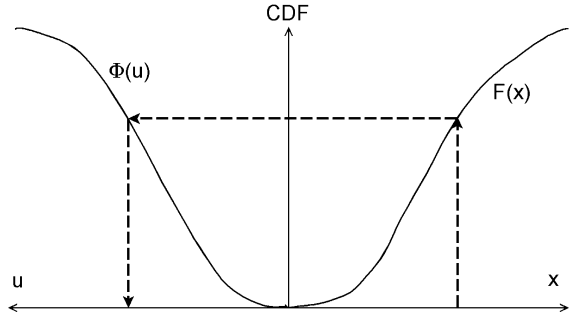


Figure 16.8 Conceptual basis for the Rosenblatt transformation.

Except for variables whose distribution functions are mathematically quite tractable, this procedure can only be implemented effectively on a computer. However, it does provide a generally applicable way to convert almost all distributions into Normal distributions and thus makes the Hasofer–Lind approach widely applicable.

17 Monte Carlo Simulation Methods

A wide range of engineering and scientific disciplines use simulation methods based on randomized input, often called Monte Carlo methods. They have been employed to study both stochastic and deterministic systems. Because developments are often bound to the application disciplines, notation and nomenclature sometimes reflect the preferences of a particular field.

17.1 Basic Considerations

Monte Carlo methods can be divided into two broad and sometimes overlapping areas. First, there is the simulation of a process that is fundamentally stochastic. For example, we might want to study the patterns that develop when vehicles arrive at a set of toll barriers, pay their tolls, and proceed through the barrier. The intervals between the arrivals, the drivers' choices of lanes, and the times to service each vehicle are all random variables, whose statistical parameters can be defined but whose actual values are uncertain. Traffic engineers have used Monte Carlo simulation methods to study such problems for many years.

17.1.1 A Simple Integration Problem

The second area involves problems that are not inherently stochastic but can be solved by simulation with random variables. An example is the definite integration of a function, say $\cos^2\theta$ between 0 and π. In this case, we know that

$$I = \int_0^{\pi} \cos^2\theta d\theta = \pi/2 \tag{17.1}$$

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

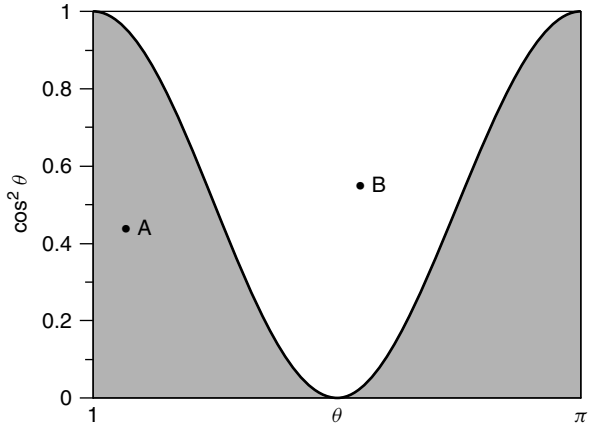


Figure 17.1 Function $\cos^2\theta$ to be integrated between 0 and π by Monte Carlo simulation. Integral is shaded area.

If we did not, we could approximate the integral by Monte Carlo simulation. We start by imagining that the function and the limits are plotted and enclosed in a box, as shown in Figure 17.1. The area of the box is π. The area we are looking for is the shaded area under the curve. We proceed by generating a series of points such as A and B. Both the horizontal and the vertical coordinates of each point are random. If the point falls in the shaded area (A), it is a 'Hit'; if not (B), it is a 'Miss.' We count the number of hits and misses and compute the ratio between the number of hits and the total number of points generated. This ratio should approximate the ratio between the shaded area and the area of the enclosing rectangle. Of course, the exact answer is 0.5.

Table 17.1 presents the results for a series of ten runs. Each run consists of 100 points. In the Hit-or-Miss procedure (second column) each point consists of a random value of θ generated uniformly between 0 and π and a random value of y generated uniformly between 0 and 1. If the value of $\cos^2\theta$ is greater than y, the point is counted as a 'Hit,' otherwise it is a 'Miss.' After 100 points are tested, the number of "Hits" divided by 100 gives the entry in Table 17.1. The entire procedure is repeated ten times, with the results shown in the entries for trials 2 through 9.

In the alternative 'Sample-Mean' procedure a random value of θ is generated uniformly between 0 and π; actually the same values as in the Hit-or-Miss procedure were used. Then $\cos^2\theta$ is computed. The average of 100 values of $\cos^2\theta$ is entered in the third column of Table 17.1. Again the entire procedure is repeated ten times.

Finally, the last line of Table 17.1 contains the averages of the ten trials for both methods, which can also be regarded as the averages of 1000 random points. Deferring a discussion of error analysis for the present, we can draw several conclusions:

- The results are close to the exact value, but not always. Trial 9 in the Hit-or-Miss approach gave a value of 0.38, which is a poor estimate of the correct value.
- The results tend to straddle the exact value so that using more points improves the accuracy of the estimate.
- While the procedure tends to converge on the exact value, convergence is slow.

Table 17.1 Results of Monte Carlo integration of $\cos^2\theta$

	Mean of 100 Points	
Trial No.	Hit or Miss	Sample Mean
1	0.50	0.459510
2	0.51	0.576580
3	0.54	0.546674
4	0.44	0.476346
5	0.42	0.482850
6	0.50	0.481759
7	0.47	0.492535
8	0.52	0.501243
9	0.38	0.488337
10	0.51	0.527823
Average	0.479	0.503366

Note: Numbers in the table must be multiplied by π to obtain the value of the integral.

- Results are more accurate from the Sample-Mean approach than from the Hit-or-Miss method. Rubinstein (1981) proves this result in the general case.
- The technique is easy to use and does not require skill in evaluating definite integrals.

This example is intended only to illustrate the concepts of Monte Carlo techniques. Evaluating a definite integral of one variable or, equivalently, finding the area of a plane figure is not a good application for the method. To integrate $\cos^2\theta$ numerically, one could select a series of equally spaced values of θ, calculate $\cos^2\theta$ at each point, average the results, and multiply by π. This is simply repeated use of the trapezoidal rule. If each value of θ is in the center of one of n equal intervals, the exact result of 0.5 is recovered using *any* value of n greater than 1. If the values of θ are at the ends of n equal intervals (so there are $n + 1$ points), the results converge rapidly to the correct value so that the results using 100 points are as good as or better than those from 1000 points of Monte Carlo simulation. Monte Carlo methods are best suited for cases in which (a) there are many independent variables or (b) the function to be integrated is strongly non-linear so that it is difficult to determine how to divide the region of integration efficiently.

17.1.2 Accuracy and Convergence

The analyst must determine how many trials are necessary to ensure a desired level of accuracy in the results. The following summary is based on the books by Rubinstein (1981), Morgan and Henrion (1990), and Fishman (1995), but many other authors present similar material. For a more thorough coverage the reader is directed to a text, such as Fishman's (1995), that is devoted entirely to Monte Carlo methods.

Each point in the Hit-or-Miss strategy is an independent sampling of the total area in Figure 17.1. It is completely analogous to withdrawing balls from an urn that contains a large number of black and white balls and replacing the ball after each test. This is a *Bernoulli trial*, and the probability of success on a single trial, which is in this case the probability of hitting the shaded area, is the ratio between the shaded area and the total

area. Let us call this p. If A is the area of the region over which the sampling is done, N is the number of sampling points, and N_H is the number of hits, then

$$\begin{aligned} p &= I/A \\ \hat{p} &= N_H/N \end{aligned} \tag{17.2}$$

This is the classic case of $\hat{p}$ as an unbiased estimator of p. N_H must be Binomially distributed. If a large number of samples are taken and Q is defined as

$$Q = A \cdot \frac{N_H}{N} = A \cdot \hat{p} \tag{17.3}$$

then

$$I = p \cdot A \approx \hat{p} \cdot A = Q \tag{17.4}$$

Furthermore, the expected value of Q is

$$E[Q] = A \cdot E[\hat{p}] = A \cdot p = I \tag{17.5}$$

so that Q is an unbiased estimator of I.

Because N_H is binomially distributed,

$$Var[N_H] = N \cdot p \cdot (1 - p) \tag{17.6}$$

It follows that

$$\begin{aligned} Var[\hat{p}] = Var\left[\frac{N_H}{N}\right] &= \frac{1}{N^2} \cdot Var[N_H] = \frac{1}{N} \cdot p \cdot (1 - p) \\ &= \frac{1}{N} \cdot \frac{I}{A} \cdot \left(1 - \frac{I}{A}\right) = \frac{1}{N} \cdot \frac{I}{A^2} \cdot (A - I) \end{aligned} \tag{17.7}$$

and

$$Var[Q] = A^2 \cdot Var[\hat{p}] = \frac{I}{N} \cdot (A - I) \tag{17.8}$$

$$\sigma_Q = \frac{1}{\sqrt{N}} \cdot I \cdot \left(\frac{A}{I} - 1\right)^{1/2} \tag{17.9}$$

The last result indicates that the error in the estimate of the integral decreases as the *square root* of the number of trials. This is a slow rate of convergence, and much of the modern work on Monte Carlo methods has been directed at either reducing the error in the sampling process or decreasing the number of trials necessary to achieve a desired accuracy.

In most cases, the analyst is less interested in estimating the error than in determining how many trials must be made to achieve a desired accuracy. Fishman (1995) shows that, starting from Chebychev's inequality

$$pr\left[\frac{|Z|}{\sigma_Z} \geq \beta\right] \leq \frac{1}{\beta^2} \tag{17.10}$$

defining $Z = \hat{p} - p$ and $\beta = \varepsilon/\sigma$, and recalling that $Var[Z] = Var[\hat{p}] = p \cdot (1 - p)/N$, we obtain

$$pr[|\hat{p} - p| < \varepsilon] \geq 1 - p \cdot (1 - p)/(N\varepsilon^2) \tag{17.11}$$

We must specify a *confidence level*, which is to say that the probability in Equation (17.11) must be greater than some desired level, say 90% or 95%. This is usually stated in the form $(1 - \delta)$ with $\delta = 0.1$ or 0.05, respectively. It follows that the number of trials needed to achieve a desired confidence level is

$$n_C(\varepsilon, \delta, p) = p \cdot (1 - p)/(\delta\varepsilon^2) \tag{17.12}$$

where any real result is rounded up to the next integer.

Equation (17.12) cannot be used as written because the value of p is not known *a priori*. One alternative is to use $\hat{p}$ as an estimate of p. An upper limit on n_C follows from the recognition that $p(1 - p)$ has its maximum when $p = 0.5$. This gives

$$n_C(\varepsilon, \delta) = 1/(4\delta\varepsilon^2) \tag{17.13}$$

The subscript C indicates that it is based on the Chebychev approximation.

An improved estimate follows from the Central Limit Theorem, which states that as $N \to \infty$ the quantity $(N_H - Np)/[Np(1 - p)]^{1/2}$ converges to the standard Normal distribution with mean of zero and standard deviation of 1. Fishman (1995) shows that the corresponding requirements for achieving a desired accuracy with a given confidence level are

$$n_N(\varepsilon, \delta, p) = p \cdot (1 - p) \cdot [\Phi^{-1}(1 - \delta/2)/\varepsilon]^2 \tag{17.14}$$

$$n_N(\varepsilon, \delta) = [\Phi^{-1}(1 - \delta/2)/(2\varepsilon)]^2 \tag{17.15}$$

The expression $\Phi^{-1}(x)$ is the inverse of the CDF for the standard Normal distribution; that is, it is the value of the variate that gives the cumulative distribution x. Any real results in Equations (17.14) and (17.15) are rounded up to the next integer. The subscript N indicates that the result is based on a Normal approximation. Fishman (1995) observes that this estimate does not converge uniformly so that many workers prefer to multiply its estimates by 2.

Table 17.2 lists the maximum values of n_C and n_N for various error bounds and for 90% and 95% confidence levels. The values for n_N are also plotted in Figure 17.2. The results are consistent with the results for the Hit-or-Miss strategy in Table 17.1. For 100 trials Table 17.2 and Figure 17.2 indicate that we can have 90% confidence that the error is less than 0.08 and 95% confidence that it is less than 0.1. Table 17.1 shows that in one out of the ten sets of 100 trials (Number 9) the mean differed from the true value by

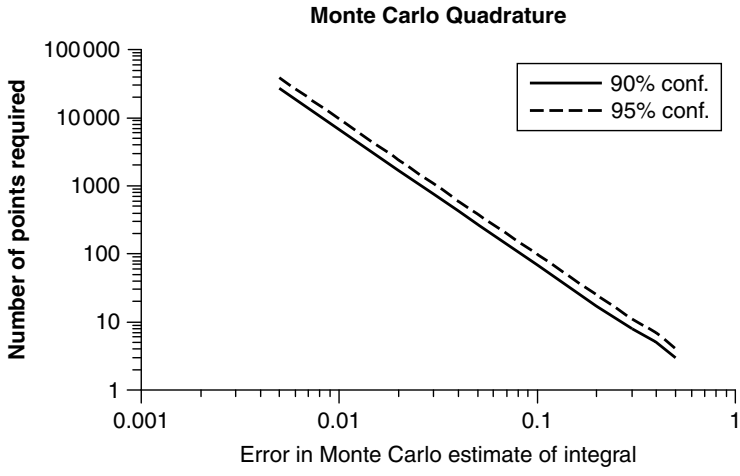


Figure 17.2 Error estimates for Hit-or-Miss Monte Carlo integration.

Table 17.2 Number of trials required to achieve desired accuracy

	90% confidence		95% confidence	
ε	n_C	n_N	n_C	n_N
0.005	100,000	27,056	200,000	38,415
0.01	25,000	6,764	50,000	9,604
0.02	6,250	1,691	12,500	2,401
0.03	2,778	752	5,556	1,068
0.04	1,563	423	3,125	600
0.05	1000	271	2000	385
0.1	250	68	500	97
0.2	63	17	125	25
0.3	28	8	56	11
0.4	16	5	32	4
0.5	10	3	20	4

more than 0.1. For 1000 trials one would expect 90% confidence that the error is less than 0.025 and 95% confidence that it is less than 0.03. In fact, the error in the Hit-or-Miss strategy is 0.021 and in the 'Sample Mean' strategy is about 0.004.

It should be noted that Trial Number 9 in Table 17.1 presents an unusual result. It has the largest error observed in several repeated runs of the numerical experiment. However, it does represent the sort of result that can be expected in Monte Carlo simulation and is consistent with the calculated error bounds.

17.2 Computer Programming Considerations

Most practical applications of the Monte Carlo method require that a function be evaluated at a large number of points, so a computer is essential. Indeed, the popularity of Monte

Carlo methods coincides historically with the availability of computers. Today the analyst can use many computer systems specifically written for stochastic simulation, Monte Carlo capabilities are in most general-purpose statistical software, and even spreadsheets include the functions necessary to perform Monte Carlo analyses. Thus, software developers have now resolved many issues with which the analyst had to wrestle in the past. Nevertheless, the analyst should be aware of some of the issues that lie behind the Monte Carlo method because it may be necessary to perform analyses that are beyond the capabilities of canned software packages, and it may be necessary to check that the software is functioning properly.

17.2.1 Random Number Generation

Any simulation that relies on random numbers requires that there be some way to generate the random numbers. There is a large literature on the subject (Kahaner *et al*. 1989; Press *et al*. 1992; Fishman 1995; Knuth 1997), and readers interested in the details of random number generation should consult those references. The following is a brief summary of the highlights.

The very concept of a series of uniformly distributed random numbers is a difficult one to formulate, even though it seems intuitively simple. Statisticians have developed a set of criteria that must be satisfied by a sequence of truly random numbers. First, they must be uniformly distributed over the interval of definition. Statisticians use tests such as the χ^2 and Kolmogorov–Smirnov tests to determine whether the numbers are uniformly distributed. Rubinstein (1981) and Press *et al*. (1992) give the details. Second, the value of any number in the sequence must be statistically independent of the other numbers. In practice it is difficult to satisfy these criteria.

Some commercially available packages include random number generators that rely on the architecture of a particular computer and are, therefore, not portable to different machines. Kahaner *et al*. (1989), Press *et al*. (1992) and Fishman (1995) provide detailed programming information or actual Fortran or C code that can be implemented on different machines. Some analytical systems, such as MATHCAD, give references for the algorithms used to generate random numbers, but most spreadsheets do not. In either case, the user should be aware of some of the considerations in the following paragraphs.

Most random number generators employ a *linear congruential* algorithm, in which a sequence of uniformly distributed random integers $(I_1, I_2, \ldots, I_i, I_{i+1}, \ldots, I_n)$ is generated from

$$I_{i+1} = a\, I_i + c \,(\text{mod } m) \tag{17.16}$$

In this equation, a and c are constants and (mod m) indicates that the calculations are to be modulo m; that is, the result is divided by m and only the remainder retained. Since integer operations are faster than real calculations, the usual procedure is to perform the calculations in integer mode and then divide the results by m to obtain real random numbers Z_i between 0 and 1. Equation (17.16) implies that

- each random variable depends only on the previous value in the sequence,
- if any value recurs, the sequence starts over,
- the length of the non-repeating sequence cannot exceed m,
- but it may be a great deal shorter.

Therefore, m is usually taken to be as large as practicable. In most applications m is 2^{32} or 2^{31}, or close to it. In many applications $c = 0$, and the algorithm is simplified. It can be shown that, with the appropriate choice of a and m, the results are as good as they would be with $c \neq 0$. However, it is critically important that the first value of the sequence of random number, called the *seed*, not be set to zero, even though many users of random numbers are tempted to do so. Park and Miller (1969) and Press *et al*. (1992) endorse using $a = 7^5 = 16{,}807$ and $m = 2^{31} - 1 = 2{,}147{,}483{,}647$.

The most commonly used sequences of random numbers are not random at all but are determined exactly by the computational algorithm. They are sometimes called 'pseudo-random' numbers. However, the sequence must be sufficiently long that it does not repeat during the simulation. If the parameters of the sequence are chosen with care, the sequence will appear random and can be used as though it were. A linear congruential algorithm is very fast if implemented properly.

One difficulty with pseudo-random number generators of this type is well known in the simulation community but is not so familiar to casual users of simulation software. As Press *et al*. (1992) state,

> [The generator] is not free of sequential correlation on successive calls. If k random numbers at a time are used to plot points in k dimensional space (with each coordinate between 0 and 1), then the points will not 'fill up' the k-dimensional space, but rather will lie on $(k-1)$-dimensional 'planes.' There will be *at most* about $m^{1/k}$ such planes. If the constants m, a, and c are not very carefully chosen, there will be *many fewer than that*. The number m is usually chosen close to the machine's largest representable integer, e.g. $\sim 2^{32}$. So, for example, the number of planes on which triples of points lie in three-dimensional space is usually no greater than about the cube root of 2^{32}, about 1600. You might well be focusing attention on a physical process that occurs in a small fraction of the total volume, so that the discreteness of the planes can be very pronounced. (Italics in the original.)

The parameters in a linear congruential generator must be selected carefully to maximize the number of planes – or hyperplanes – along which the sets of points cluster. It is surprising that some of the commercially developed generators behave very badly. The best known horror story concerns the RANDU routine that was part of the IBM Scientific Subroutine Package for the 360 and 370 computers in the 1960s and 70s (Press *et al*. 1992; Fishman 1995). This used $a = 2^{16} + 3 = 65{,}539$, $c = 0$, and $m = 2^{31}$. Although $m^{1/k}$ gives about 1290 as the number of planes that should exist in three dimensions, this generator actually had all the points falling on 15 planes in three dimensions. Figure 17.3 shows in part (a) that the points seem to be fully distributed when the space is viewed from an arbitrary angle but in part (b) that the 15 planes are clearly visible from the right perspective. This random number generator was widely used, and its performance became notorious.

Knuth (1997) gives the following recommendation for dealing with these issues in important applications:

> The most prudent policy for a person to follow is to run each Monte Carlo program at least twice using quite different sources of random numbers, before taking the answers of the program seriously; this will not only give an indication of the stability of the

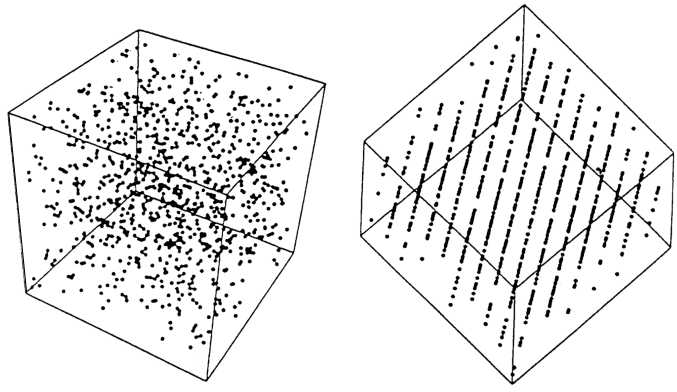


Figure 17.3 Pattern of sequential values from RANDU. (a) Arbitrary view; (b) View along planes. (Fishman, G. S.; *Monte Carlo: Concepts, Algorithms, and Applications*, p. 620, Fig. 7.7, 1995, New York, © Springer-Verlag, reproduced with permission.)

results, it will also guard against the danger of trusting in a generator with hidden difficulties. (Every random number generator will fail it at least one application.)

17.2.2 Improvements

Several improvements have been proposed, but most of these have not found extensive use because they are more time-consuming than the simple linear congruential generators. Perhaps the simplest approach is adopt a two-step approach. First a series of pseudo-random numbers is generated in the usual way. Then this array is sampled randomly by choosing the location with *another* random number generator. After each sample, the number is replaced with the next number in the original random sequence. This process has the effect of shuffling the arrangement of random numbers. Press *et al*. (1992) give the Fortran code for such a generator, which is both fast and easy to use. The C version of the code is available in the other editions of their book.

17.2.3 Non-uniform Distributions

The standard techniques described in the preceding sections provide a series of random numbers uniformly distributed between 0 and 1. However, what we usually need is a sequence of random numbers satisfying some other distribution, such as a Normal distribution. It is easy to formulate a way to do this. If Y is uniformly distributed between 0 and 1, X is distributed in some known way, and $\Phi(x)$is the CDF of X, it is easy to write

$$\begin{aligned} y &= \Phi(x) \\ x &= \Phi^{-1}(y) \end{aligned} \tag{17.17}$$

The procedure would then be to generate a sequence of uniformly distributed random numbers and then to find the inverse of the CDF for each number in the sequence. The result would be a sequence of random numbers with the desired distribution. For example,

if we generate a series of random numbers x_i that are uniformly distributed between 0 and 1, then $-\ln(x_i)/\lambda$ will be a series of exponentially distributed numbers with mean of $1/\lambda$. The trouble with this approach is that we almost never have an analytical form for the inverse CDF, even for well-known distributions. The difficulty is sometimes resolved by approximating the CDF with a simpler function.

For some distributions mathematical procedures exist for avoiding approximating the inverse CDF. It can be shown (Press *et al.* 1992; Fishman 1995) that, if X and Y are uniform variates between 0 and 1 and if we define two new variables

$$\begin{aligned} U &= \sqrt{-2\ln X}\cdot\cos(2\pi Y) \\ V &= \sqrt{-2\ln X}\cdot\sin(2\pi Y) \end{aligned} \tag{17.18}$$

then U and V are independent standard normal variates. Another technique is the acceptance-rejection algorithm, in which one generates a random variate uniformly distributed over the area under some analytically simple curve that envelopes the desired distribution and accepts or rejects the value depending on whether it falls under the target distribution.

Abramowitz and Stegun (1964), Press *et al.* (1992), Evans *et al.* (1993) and Fishman (1995) give approximations or algorithms for many distributions. Many canned software packages and spreadsheets are able to generate random number sequences satisfying several prescribed distributions, so the user need only specify that the random numbers must satisfy a prescribed distribution. However, once again the analyst should be aware of what is going on, especially for important applications involving large numbers of variables or many dimensions.

17.2.4 Correlated Random Numbers

In many practical applications the uncertain variables are correlated, and the random number systems must be capable of generating correlated sequences. This is relatively easy to do if the uncertain variables are Normally distributed. Consider a vector $\mathbf{X}$ consisting of n random variables, each of which is statistically independent and has a standard Normal distribution with zero mean and unit standard deviation. (There are n variables, but there may be thousands of randomly generated instances of each variable.) Consider also another vector $\mathbf{Y}$ consisting of n random variables, also with standard Normal distribution and related to $\mathbf{X}$ by a correlation matrix $\mathbf{K}$:

$$\mathbf{K} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{bmatrix} \tag{17.19}$$

in which ρ_{ij} is the correlation coefficient between variables i and j. Since in any physically realizable case $\mathbf{K}$ must be positive definite, the Cholesky algorithm can be used to factor it into an upper triangular matrix $\mathbf{S}$ and its lower triangular transpose $\mathbf{S}^T$:

$$\mathbf{S}^T\mathbf{S} = \mathbf{K} \tag{17.20}$$

We summarize the Cholesky decomposition shortly, but let us first follow its implications.

The definitions of variance and covariance can be summarized by

$$\int \mathbf{X} \cdot \mathbf{X}^T f(x_1, x_2, \ldots, x_n)\, dx = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I \tag{17.21}$$

and

$$\int \mathbf{Y} \cdot \mathbf{Y}^T f(y_1, y_2, \ldots, y_n)\, dy = \mathbf{K} \tag{17.22}$$

The notation in these equations is somewhat imprecise, but it is adopted to reduce clutter. The intention is that the f's are the joint probability density functions and the integrations, whether single or double, are carried out between $-\infty$ and $+\infty$. Now, let the vectors be related by

$$\mathbf{Y} = \mathbf{S}^T \mathbf{X} \tag{17.23}$$

Then

$$\begin{aligned} \int \mathbf{Y} \cdot \mathbf{Y}^T f(y_1, y_2, \ldots, y_n) dy &= \int \mathbf{S}^T \mathbf{X} \cdot \mathbf{X}^T \mathbf{S} f(x_1, x_2, \ldots, x_n) dx \\ &= \mathbf{S}^T \int \mathbf{X} \cdot \mathbf{X}^T f(x_1, x_2, \ldots, x_n) dx\ \mathbf{S} \\ &= \mathbf{S}^T \cdot \mathbf{I} \cdot \mathbf{S} = \mathbf{S}^T \cdot \mathbf{S} = \mathbf{K} \end{aligned} \tag{17.24}$$

In other words, if $\mathbf{Y}$ is computed from the independent $\mathbf{X}$ according to Equation (17.23), it follows that $\mathbf{Y}$ has the correlation described by $\mathbf{K}$.

The Cholesky decomposition is a well-known technique in numerical analysis. The algorithm is

$$\begin{aligned} S_{ii} &= \left(K_{ii} - \sum_{k=1}^{i-1} S_{ki}^2 \right)^{1/2} \qquad i = 1, \ldots, n \\ S_{ij} &= \left(K_{ij} - \sum_{k=1}^{i-1} S_{ki} S_{kj} \right) / S_{ii} \quad j = i+1, \ldots, n \\ S_{ji} &= 0 \qquad\qquad\qquad j = i+1, \ldots, n \end{aligned} \tag{17.25}$$

The calculations start at the upper left corner and proceed by first calculating the diagonal term, calculating the remaining terms in that row and column, and proceeding to the next diagonal terms. Fishman (1995) lists the algorithm, without identifying it as Cholesky decomposition. Morgan and Henrion (1990) provide a simple form for the special case that $n = 2$. The correlation coefficient between the two Normally distributed random variables is ρ. If x_1 and x_2 are a pair of uncorrelated random numbers drawn from the

standard Normal distribution. Then a pair of correlated random numbers, y_1 and y_2, from the standard Normal distribution are

$$y_1 = x_1$$
$$y_2 = \rho x_1 + \sqrt{1 - \rho^2} x_2 \tag{17.26}$$

In summary, to generate correlated sequences of random numbers when the random variables have standard Normal distributions, one proceeds as follows:

1. Calculate the correlation matrix, **K**.
2. Perform the Cholesky decomposition to find **S**.
3. Generate n uniform random numbers between 0 and 1.
4. Modify each number so that it corresponds to the standard Normal distribution with mean 0 and standard deviation 1. This is one instance of the vector **X**.
5. Compute the vector **Y** from Equation (17.23). This is an instance of the correlated variables.
6. To obtain the random variables with the correct means and standard deviations, for each variable compute:
$$z_i = \mu_{z_i} + \sigma_{z_i} \cdot y_i \tag{17.27}$$
The z_i comprise one instance of the vector **Z** of correlated random variables with the desired means and standard deviations.
7. Repeat steps 2 through 6 for each additional randomly generated instance of **Z**.

17.3 Simulation of Random Processes

So far, this chapter has described the use of Monte Carlo methods to evaluate integrals. The other application is to study the behavior of random processes. These are problems in which the input is stochastic and the results uncertain. The Monte Carlo method is particularly effective when the process is strongly nonlinear or involves many uncertain inputs, which may be distributed differently.

To perform such a study, the analyst generates a random value for each uncertain variable and performs the calculations necessary to yield a solution for that set of values. This gives one sample of the process. The trials are repeated many times, giving many samples of the process. Once a large number of runs have been completed, it is possible to study the output statistically and to obtain values of means, variances, probabilities of various percentiles, and other statistical parameters. Two important points should be noted:

1. Regardless of the number of stochastic variables, each run gives one sample of the process. Hence, increasing the number of stochastic input variables does not increase the number of runs for the same level of accuracy.
2. The technique is essentially a repeated sampling of the stochastic process, and the methods of Part 2 and particularly Chapter 11 can be used to examine the accuracy and precision of the results.

17.3.1 Some Useful Results

The unbiased estimate of the mean of a random process is simply the mean of the values returned by the process

$$E[\mu_X] = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{17.28}$$

and the standard deviation of this estimate is related to the standard deviation of all the values by

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} \tag{17.29}$$

If the process is Normally distributed, the variance is χ^2 distributed. It can be shown that the upper confidence limit at the $(1-\alpha)$ level is

$$(\sigma^2)_{1-\alpha} = \frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \tag{17.30}$$

in which s^2 is the variance computed from the sample, and the denominator is the cumulative χ^2 distribution with the parameters α and $(n-1)$. These equations can be used to obtain estimates of the number of samples necessary to achieve a desired level of accuracy.

17.3.2 An Example of the Monte Carlo Method

Example 13.1 described the analysis of the reliability of a vertical cut in a cohesive soil. Direct analytical techniques are adequate to treat this problem, whether or not the variables are correlated. If the input variables are Normally distributed, the margin of safety is Normal as well, but the factor of safety F is not. Since we have the exact analysis of this problem, we use it to examine the behavior of Monte Carlo simulation applied to the same problem.

Example 17.1 – Vertical Cut in Cohesive Soil

The equation for the margin of safety of a vertical cut is

$$M = c - \gamma H/4 \tag{17.31}$$

In the present problem $H = 10\,\text{m}$, and the other quantities are Normally distributed with the following parameters

$$\begin{aligned} &\mu_c = 100\,\text{kPa}\ \ \sigma_c = 30\,\text{kPa} \\ &\mu_\gamma = 20\,kN/m^3\ \ \sigma_\gamma = 2\,kN/m^3\ \ \rho_{c\gamma} = 0.5 \end{aligned} \tag{17.32}$$

For each sample, the analysis generated two random numbers, x and y, with standard Normal distribution. The Cholesky decomposition of Equation (17.26) can be used to give a third random number z that is correlated to x

$$z = \rho_{c\gamma} \cdot x + \sqrt{1-\rho^2_{c\gamma}} \cdot y \tag{17.33}$$

The values of c and γ for this sample are then

$$c = \mu_c + \sigma_c \cdot x$$
$$\gamma = \mu_\gamma + \sigma_\gamma \cdot z \tag{17.34}$$

These are used to generate one sample of the margin M.

It will be recalled from Example 13.1 that M is Normally distributed and the exact values of the results are

$$\mu_M = 50\,\text{kPa}\ \sigma_M = 27.84\,\text{kPa}$$
$$\beta = 1.80 p_f = 3.62 \times 10^{-2} \tag{17.35}$$

The Monte Carlo simulation was carried out with 100, 1000, and 10,000 samples. Table 17.3 gives the results for one set of runs. Figures 17.4, 17.5 and 17.6 show the corresponding plots of the ordered results. In each plot the dashed line is the result of the Monte Carlo simulation and the solid line is the exact cumulative distribution for the margin of safety. The table and plots show that, while the results from a simulation using 100 points give a general picture of the true answer, 1000 points or more are necessary to obtain accurate results. An additional benefit of the plots is that, generally, the smoother the curves the more accurate the results.

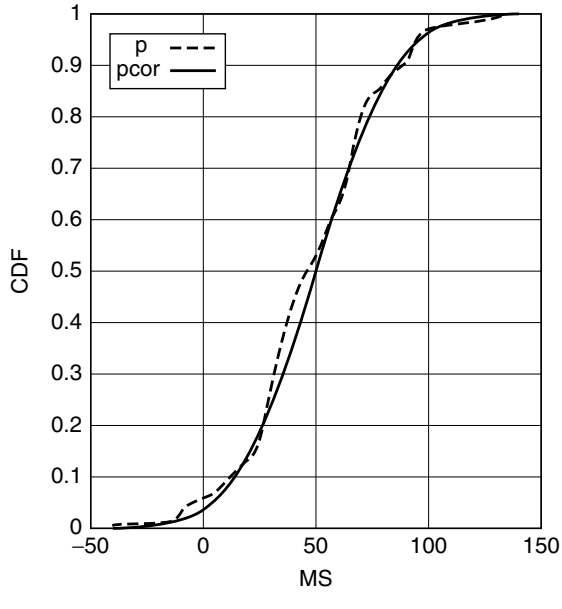


Figure 17.4 Ordered sequence of results for 100 simulations of vertical cut problem (dashed line). The solid line is the exact answer.

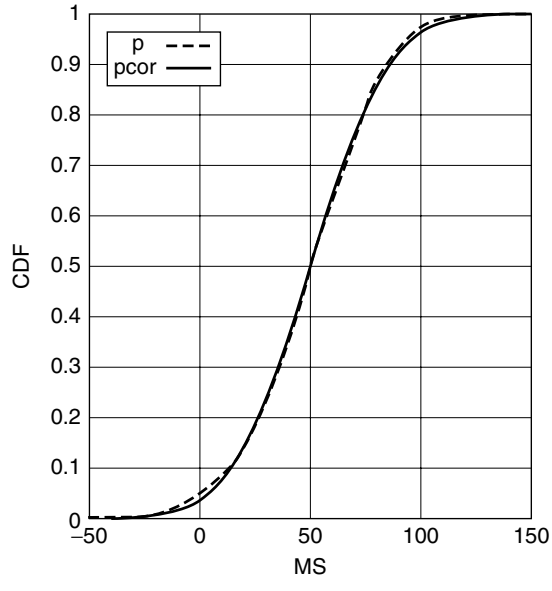


Figure 17.5 Ordered sequence of results for 1000 simulations of vertical cut problem (dashed line). The solid line is the exact answer.

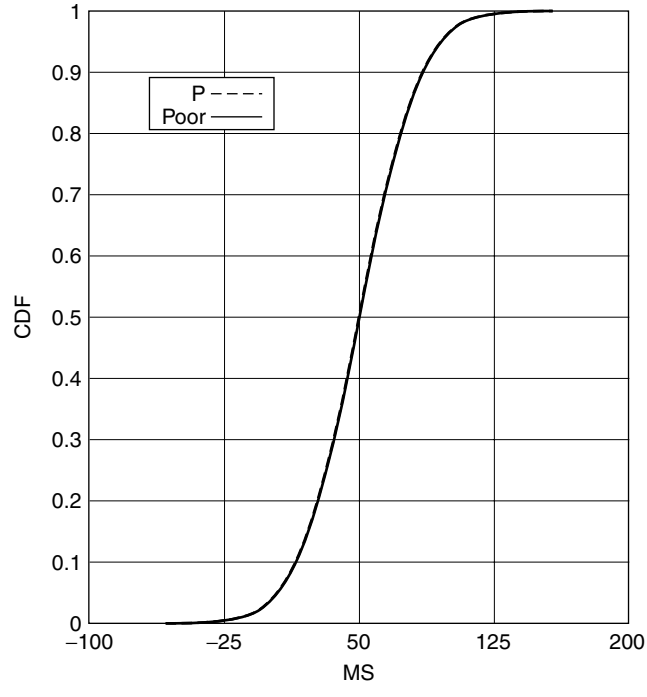


Figure 17.6 Ordered sequence of results for 10,000 simulations of vertical cut problem (dashed line). The solid line is the exact answer.

Table 17.3 Results from Monte Carlo simulation of vertical cut

	Results			
n	μ_M	σ_M	β	p_f
100	47.90 kPa	30.05 kPa	1.594	5.55×10^{-2}
1,000	49.71 kPa	28.03 kPa	1.773	3.81×10^{-2}
10,000	49.95 kPa	27.66 kPa	1.806	3.55×10^{-2}
exact	50.00 kPa	27.84 kPa	1.796	3.62×10^{-2}

17.4 Variance Reduction Methods

The previous section shows what might be called the brute-force Monte Carlo method. A set of randomly distributed points is generated; they are used to compute a sample of values of the desired function or integral; and the statistical properties of the result are calculated from the sample. As the examples show, obtaining satisfactory accuracy requires that a large number of random points be used. Is there any way to reduce the computational burden or, equivalently, to increase the accuracy for the same number of points? It turns out that this is indeed possible, and the techniques as usually described as *variance reduction methods*.

Variance reduction is something of a misnomer. The techniques are actually methods of obtaining the same level of accuracy with fewer simulations. Rubinstein (1981) summarizes the goals and conditions as well as anybody:

> Variance reduction can be viewed as a means to use known information about the problem. In fact, if nothing is known about the problem, variance reduction cannot be achieved. At the other extreme, that is, complete knowledge, the variance is equal to zero and there is no need for simulation. Variance reduction cannot be obtained from nothing; it is merely a way of not wasting information. One way to gain this information is through a direct crude simulation process. Results from this simulation can then be used to define variance reduction techniques that will refine and improve the efficiency of a second simulation. Therefore the more that is known about the problem, the more effective the variance reduction techniques that can be employed. Hence it is always important to clearly define what is known about the problem. Knowledge of a process to be simulated can be qualitative, quantitative, or both.

Variance reduction always involves some effort at the beginning of the problem to identify ways in which the number of simulations can be reduced. It may also involve some additional computational effort for each simulation. Therefore, we must first ask whether the improved efficiency is worth the effort, and a brief review of the time consumed in the parts of a simulation analysis is helpful. A simulation analysis consists of three parts:

1. Analyzing the problem and determining the appropriate computational methods. The more sophisticated the computational approach, the more time-consuming this part of the process is likely to be.
2. Computation: This includes three components:

(a) Setting up the computation, initializing variables, and so on.
(b) Performing the simulation n times. Each simulation involves generating the random variables and computing the thing to be simulated.
(c) Summarizing the results, making plots, computing variances, and so on.

3. Evaluating the results.

Variance reduction reduces the cost of the simulation by reducing the number of simulations that have to be done in part 2b. If the computations in each simulation are simple and cheap but the initial costs in parts 1 and 2a are large, there is little benefit is devoting much effort to variance reduction. On the other hand, if the costs for each simulation are high, there is a lot to be gained by reducing their number. In this area, as in much other numerical work, the analyst must weigh benefits against costs and exercise judgment. It is worth bearing in mind that, in practice, most analyses are repeated far more often than the analysts originally planned, so improved efficiency may be worth some attention early in an analytical program.

Some improvements are relatively simple. For example, the simple integration problem introduced in Section 17.1.1 involves a function that is symmetrical about $\pi/4$. The simulation effort could have been cut in half by distributing the points between 0 and $\pi/2$ instead of between 0 and π.

17.4.1 Importance Sampling

Importance sampling starts with the observation that, if we are going to sample randomly, we should distribute the points to put most of them in the region that contains information and to waste as few as possible. For example, in the hit-or-miss attack on the integration problem discussed in the first part of this chapter, half the random points are expected to fall outside of the area to be integrated; changing the search area so that more of the points are useful should improve the efficiency of the computation. In the analysis of the vertical cut, we used Monte Carlo simulation to evaluate the mean and standard deviation of the margin of safety and then combined these estimates with our knowledge that M is normally distributed to determine the probability of failure. If the distribution of the failure criterion were not known (as, say, when it is defined by the factor of safety), the probability of failure could be estimated by comparing the number of simulated points that fall in the unsafe region to the total number of points. Since the probability of failure in this example is on the order of 10^{-3}, it is obvious that several thousand points must be used to obtain reasonable accuracy unless the sampling can be concentrated in the failure region.

The central problem is to find the integral of a function $g(x)$

$$I = \int g(x)dx \tag{17.36}$$

where x can be a vector of several variables, the integration can be multidimensional, and the limits of integration are appropriate to the specific problem. Now, let us assume that we can find some probability distribution function $f(x)$ concentrated in the region of interest. Then

$$I = \int \frac{g(x)}{f(x)} f(x)dx = E\left[\frac{g(x)}{f(x)}\right] \tag{17.37}$$

If we generate points from the distribution $f(x)$, then the Monte Carlo estimate of the integral is

$$I \approx \frac{1}{n} \sum_{i=1}^{n} \frac{g(x_i)}{f(x_i)} \tag{17.38}$$

It turns out (see Rubinstein (1981), Press *et al.* (1992) or Fishman (1995)) that the optimal choice of $f(x)$ is

$$f(x) = \frac{g(x)}{I} \tag{17.39}$$

This states essentially that, if we know the value of the integral, we can then find the best importance function, but the whole purpose of the exercise is to find the value of that integral. Nevertheless, Equation (17.39) does indicate that a good choice of $f(x)$ is one that closely approximates the function to be integrated in the significant region. The trick is to find $f(x)$ without expending more effort than is saved by the increased efficiency of the simulation. Two examples demonstrate how this is done.

Example 17.2 – Importance Sampling for Integration of a Function

Section 17.1.1 showed how Monte Carlo simulation could be used to estimate the integral of $\cos^2\theta$ between 0 and π. We have already observed that the effort could be cut in half by integrating between 0 and $\pi/2$ and doubling the result. Figure 17.7 shows two possible choices for $f(x)$ compared with $\cos^2\theta$. The solid line is $\cos^2\theta$, the dotted line is a linear

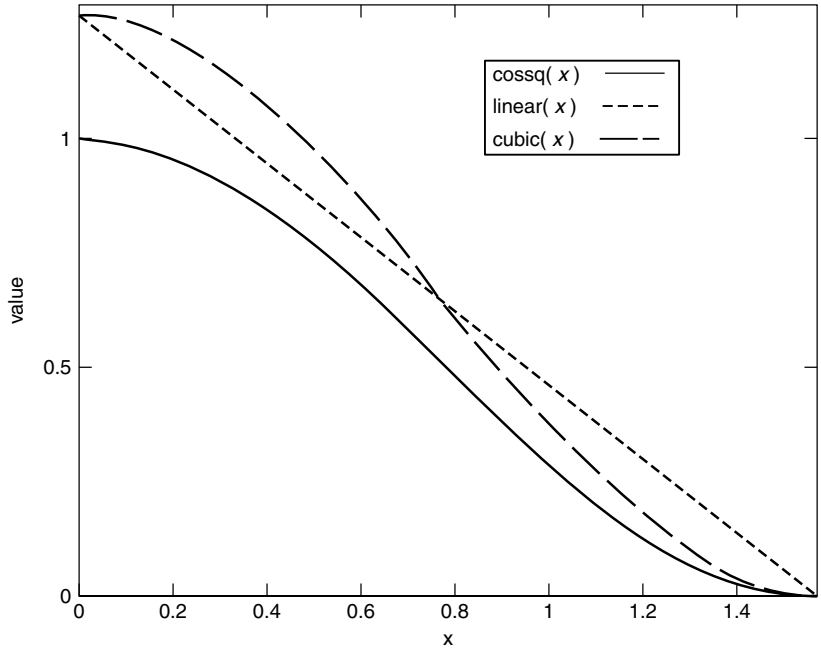


Figure 17.7 Function ($\cos^2\theta$) to be integrated (solid line), linear importance function (dotted line), and cubic importance function (dashed line).

function, and the dashed line is a cubic function. The linear and cubic functions are scaled to enclose unit areas. The functional forms are

$$linear\, f(\theta) = \frac{4}{\pi}\left(1 - \frac{2}{\pi}\theta\right)$$
$$cubic\, f(\theta) = \frac{4}{\pi}\left(1 - 3\left(\frac{2}{\pi}\right)^2\theta^2 + 2\left(\frac{2}{\pi}\right)^3\theta^3\right) \tag{17.40}$$

The computation proceeds as follows:

1. Establish n = number of simulation points to be used.
2. For each point, generate a random variable uniformly distributed between 0 and 1.
3. From the cumulative distributions corresponding to Equations (17.40) find θ corresponding to the value found in step 2. This requires solving a quadratic or quartic equation, but modern software packages include efficient solving routines.
4. Compute $\cos^2\theta$, $f(\theta)$, and their ratio. Keep a running total of the ratios.
5. Compute the average value of the ratio. Multiply this by 2 to get the integral from 0 to π.

To compare the results with those from the brute-force approach, the computed values of the integrals were divided by π. The correct answer is then 0.5. The number of simulation points, n, varied from 10 to 10,000. For each value of n, ten simulations were run. An estimate of the standard deviation of the error for each n is then

$$\sigma_{error} \approx \left(\sum_{i=1}^{k} error_i/(n-1)\right)^{1/2} \tag{17.41}$$

Figure 17.8 shows the error plotted against the number of points in the simulations. Squares correspond to the basic case, solid upward triangles to linear importance, and solid downward triangles to cubic importance. The open triangles are the results from using control variates, which are discussed in Section 17.4.4.

Figure 17.8 demonstrates that importance sampling dramatically reduces the number of simulations necessary to achieve a given level of accuracy, or, alternatively, reduces the error for a given number of simulations. The figure shows that, even when 10,000 simulations are used, the estimated error in the brute-force case is about 0.003, and this is nearly the same as the error in the cubic case with 10 simulations. There is about a one-thousandfold reduction in the number of points necessary to achieve a particular accuracy when going from the basic case to the cubic case. The ratio between $\cos^2\theta$ and the cubic importance function varies between 0.79 and 0.64 and is nearly constant over the region where $\cos^2\theta$ has the largest value. Thus, cubic importance sampling generates sampling points that are almost entirely within the area to be integrated.

Example 17.3 – Importance Sampling for Failure of Vertical Cut

Example 17.1 described the use of Monte Carlo simulation to estimate the mean and standard deviation of the margin of safety of the vertical cut. From these we can compute

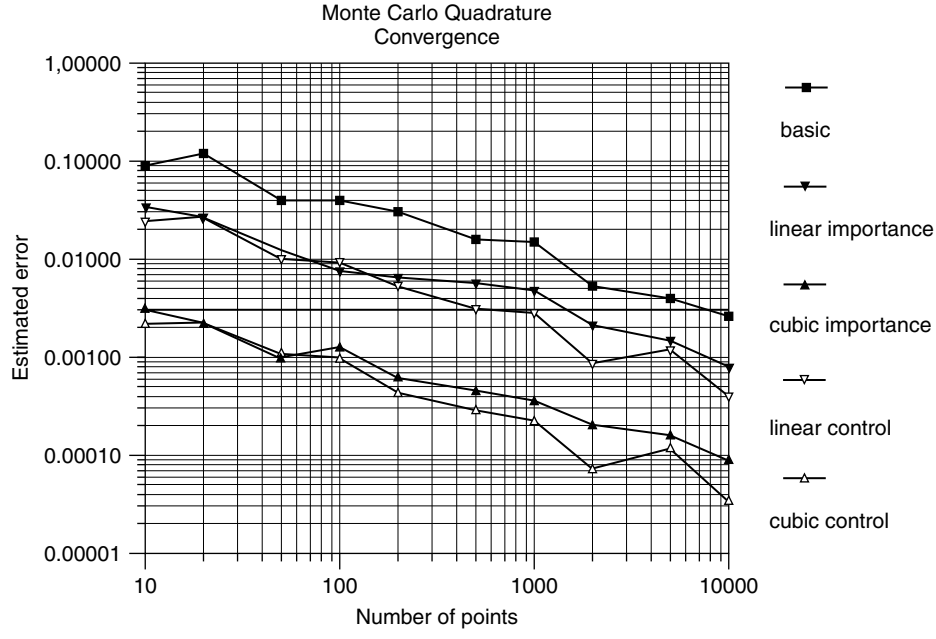


Figure 17.8 Estimated error in Monte Carlo simulation of integration of $\cos^2\theta$ with basic method, linear importance function, cubic importance function, linear control variate, and cubic control variate.

the probability of failure for a normally distributed M, but we might need to compute the probability of failure directly when the distribution is not known. Many simulations will be needed if we estimate the probability by comparing the number of failure states with the total number of simulations. Importance sampling is one way to reduce this number.

Hasofer and Wang (1993) and Melchers (1993) describe the theoretical basis for importance sampling; the specific technique follows from Melchers' approach. There are two uncertain variables, c and γ, and they have a joint probability distribution function $f(c, \gamma)$, which in this case is a bivariate Normal distribution. We want an importance function of two variables that is concentrated in the region of failure. First we find values of c and γ that satisfy the failure condition ($M = 0$ or $FS = 1$) and have the maximum value of the joint probability distribution function. This is actually easy to do using the 'solve' feature in a spreadsheet or a mathematical simulation system; it is also the failure point found in the Hasofer-Lind first-order reliability method described in Chapter 16. We then assume that the importance function is a joint Normal distribution with means at the failure point and variances equal to the variances of the original variables c and γ. Let us call this $h(c, \gamma)$.

The probability of failure is

$$p_f = \iint_D f(c, \gamma) dc d\gamma = \iint_{c,\gamma} J f(c, \gamma) dc d\gamma \tag{17.42}$$

where D is the region of failure ($M \leq 0$ or $FS \leq 1$) and J is a function that is 0 outside of D and 1 inside it. The basic Monte Carlo simulation for calculating p_f is

1. Establish n = the number of simulation points to be used.
2. For each point i generate a pair of values c_i and γ_i from their joint distribution $f(c, \gamma)$, including the correlation.
3. If $M(c_i, \gamma_i) \leq 0$, add 1 to the number of hits and go to the next i. If not, skip to the next i.
4. After all n points have been evaluated, p_f is approximately the sum divided by n.

This can be stated also in the form

$$p_f \approx \frac{1}{n} \sum_{i=1}^{n} J_i \quad \text{where } J_i = 1 \text{ if } M(c_i, \gamma_i) \leq 0$$
$$J_i = 0 \text{ if } M(c_i, \gamma_i) > 0 \tag{17.43}$$

Importance sampling changes the second part of Equation (17.42) into

$$p_f = \iint_{c,\gamma} J \frac{f(c, \gamma)}{h(c, \gamma)} h(c, \gamma) dc d\gamma \tag{17.44}$$

which leads to the Monte Carlo approximation

$$p_f \approx \frac{1}{n} \sum_{i=1}^{n} J_i \frac{f(c_i, \gamma_i)}{h(c_i, \gamma_i)} \tag{17.45}$$

where J_i is as defined for the basic Monte Carlo method. This leads to the following computational procedure:

1. Establish n = the number of simulation points to be used.
2. For each point i generate a pair of values c_i and γ_i from the distribution $h(c_i, \gamma_i)$. Note that this is not $f(c_i, \gamma_i)$ and does not include the correlation.
3. If $M(c_i, \gamma_i) > 0$, skip to the next i. If $M(c_i, \gamma_i) \leq 0$, calculate $f(c_i, \gamma_i)/h(c_i, \gamma_i)$, add the result to a running sum, and go to the next i.
4. After all n points have been evaluated, the estimate of p_f is the running sum divided by n.

The correct probability of failure for the parameters in Example 17.1 is 0.0362. Both the basic Monte Carlo simulation and the importance-sampled simulation were run for $n = 10$, 100, and 1000, and each simulation was repeated ten times. Figure 17.9 shows the results. There are two vertical lines above each value of n. The right one represents the results for the basic simulation, and the left one corresponds to the importance sampled simulation. The top of the line is the largest estimate of p_f, the bottom is the smallest value, and the horizontal line near the middle of the vertical line represents the mean result. It is not surprising that results for basic simulation with $n = 10$ are poor, but the mean result for the importance sampled simulation with $n = 10$ is remarkably good. The results

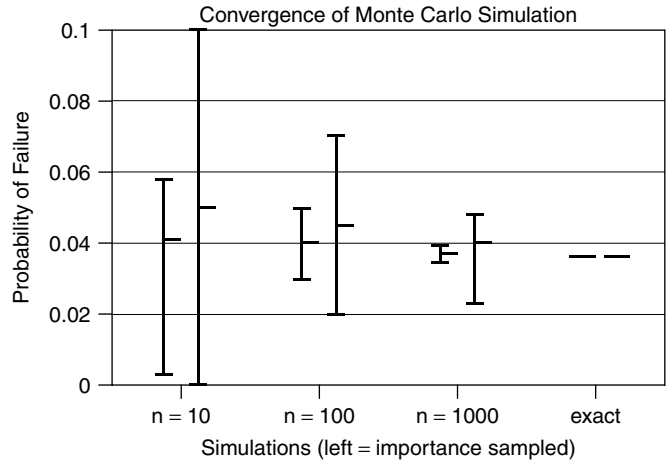


Figure 17.9 Probability of failure of vertical cut 10 m high evaluated by Monte Carlo simulation with and without importance sampling with different numbers of simulations. Results on the left are importance sampled. Lines span between the smallest and largest values from 10 sets of simulations; line near middle is average of 10 results.

get better as n increases, and in each case the importance-sampled results are significantly superior. When $n = 1000$, the results with importance sampling are nearly exact.

These results demonstrate the particular usefulness of importance sampling when the probability to be calculated is small. The convergence could be improved further by choosing a better importance function. Hasofer and Wang (1993) propose a more complicated function that is exponentially distributed in a direction normal to the failure criterion and normally distributed in the remaining directions. While this would probably give better results, it is more complicated, and the present results demonstrate the effectiveness of importance sampling even with a relatively crude importance function.

Another point to be emphasized is that the same calculation could be done using the factor of safety instead of the margin of safety. The criterion that $FS \leq 1$ is identical to the requirement that $M \leq 0$. Thus, the same center and form for the importance function are defined, and the computational results are identical.

17.4.2 Antithetic Sampling

Antithetic sampling can be useful when there are reasons to believe that different senses of deviation from the mean values of the variates will have different effects on the value of variable to be integrated. Suppose Z_A and Z_B are two unbiased estimators of I. Then

$$Z = \tfrac{1}{2}(Z_A + z_B) \tag{17.46}$$

is an unbiased estimator of I. The variance of Z is

$$Var(Z) = \tfrac{1}{4}[Var(Z_A) + Var(Z_B) + 2\,Cov(Z_A, Z_B)] \tag{17.47}$$

If Z_A and Z_B are uncorrelated, then

$$Var(Z) = \tfrac{1}{4}[Var(Z_A) + Var(Z_B)] \tag{17.48}$$

Therefore, if $Cov(Z_A, Z_B) < 0$, the variance in Equation (17.47) is less than the variance when the two samples are uncorrelated, and the simulation will be more efficient.

One widely used technique for making Z_A and Z_B negatively correlated is to create Z_A from n simulations and Z_B from n complementary simulations, making a total of $2n$ simulations. By the complement we mean that, if X is uniformly distributed between 0 and 1, each x_B is equal to the corresponding x_A subtracted from 1, and, if X has a standard Normal distribution, each x_B is equal to the negative of the corresponding x_A. It is easy to derive the corresponding relations for other distributions. The algorithm works as follows:

1. Perform one simulation.
2. Compute the complementary random numbers and perform another simulation.
3. Repeat steps 1 and 2 a total of n times.
4. Combine the results as though this had been one simulation with $2n$ samples.

Obviously, this will work only if the complementary random numbers do actually generate negatively correlated results. The following example shows what happens when this condition is satisfied and when it is not.

Example 17.4 – Antithetic Sampling for Culmann Failure along Single Plane

The Culmann analysis applies to the failure of a slope along a single plane. The following parameters appear in the solution:

H = height of slope = 10 m in this case
ψ = inclination of failure plane = 20° in this case
θ = inclination of face of slope = 26° in this case
c = cohesion, $\mu_c = 5$ kPa, $\sigma_c = 2$ kPa in this case
$\tan\phi$ = tangent of friction angle, $\mu_{\tan\phi} = 0.5774$, $\sigma_{\tan\phi} = 0.1386$ in this case
γ = unit weight, $\mu_\gamma = 5\,\text{kN/m}^3$, $\sigma_\gamma = 2\,\text{kN/m}^3$ in this case

All variables are uncorrelated. The mean value of $\tan\phi$ corresponds to $\phi = 30°$.

The factor of safety, FS, is

$$FS = \frac{c + \dfrac{1}{2}\dfrac{\gamma H}{\sin\psi}\sin(\psi - \theta)\cos\theta\tan\phi}{\dfrac{1}{2}\dfrac{\gamma H}{\sin\psi}\sin(\psi - \theta)\sin\theta} \tag{17.49}$$

and the margin of safety is

$$M = c + \frac{1}{2}\frac{\gamma H}{\sin\psi}\sin(\psi - \theta)\cos\theta(\tan\phi - \tan\theta) \tag{17.50}$$

Since only c, γ, and $\tan\phi$ are uncertain, exact values of μ_M and σ_M can be calculated. The results are $\mu_M = 10.26$ kPa and $\sigma_M = 4.15$ kPa.

The analysis consisted of Monte Carlo simulation of both M and FS. The basic simulations used 10, 100 and 1000 points. For each value of n (number of points), ten independent simulations were run. All the uncertain variables were assumed to be Normally distributed. Then, the simulations were repeated with the same first $n/2$ points and a second group of $n/2$ points using the complements of these values. Mean values of M and FS were calculated for each sampling. The results are shown in Figure 17.10 for the factor of safety. The plot shows the maximum and minimum and the mean of the results for each group of simulations. Antithetic sampling clearly improves the convergence. Indeed, the results using ten points with antithetic simulation are better than those from the simulations using 1000 points. Similar results occur for the margin of safety.

When the results of these simulations are used to compute the standard deviations of M or FS, the results are completely different, as Figures 17.11 shows. The results from antithetic simulation are worse than those from direct simulation. Study of Equations (17.49)

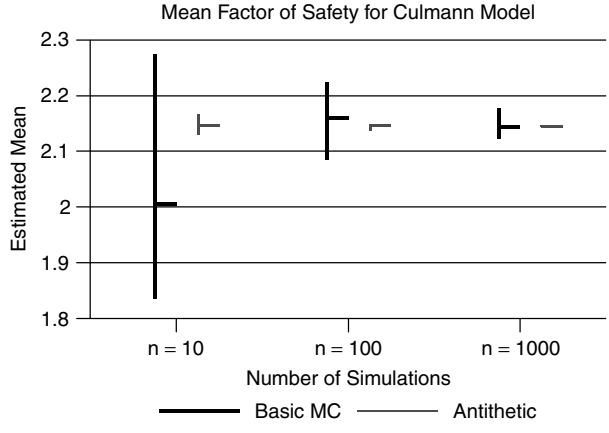


Figure 17.10 Results for basic Monte Carlo simulation and simulation with antithetic sampling of mean factor of safety for Culmann single plane failure model.

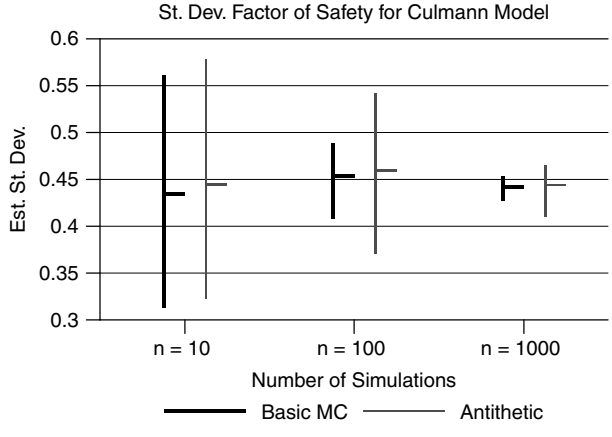


Figure 17.11 Results for basic Monte Carlo simulation and simulation with antithetic sampling of standard deviation of factor of safety for Culmann single plane failure model.

and (17.50) reveals why. Changes in c and tan ϕ affect M or FS directly, although the effects of changing γ are somewhat more complicated. The value of M or FS calculated with one set of values should lie on the other side of the mean result from the results using the complementary values. Therefore, antithetic sampling improves convergence because differences tend to cancel out in calculating the mean. However, the variance is computed from the square of the difference between the values and the mean, so that values on opposite sides of the mean both contribute positively. In effect, the antithetic portion of the simulation emphasizes rather than canceling the earlier portion of the calculation of the standard deviation, and the calculation is based on half the number of independent points as the basic simulation. Therefore, antithetic simulation givers worse results than basic simulation.

17.4.3 Correlated Sampling

Correlated sampling is useful when we are interested in the relative effects of different treatments. For example, suppose there is a deposit of clay that is to be consolidated under a preload to improve shear strength. Alternative designs of sand drains and wicks have been proposed. The properties and geometry of the soil are uncertain, and so are the properties of the different design alternatives. One way to study this problem would be to carry out independent Monte Carlo simulations of the soil with each alternative design. Then the results of one simulation could be subtracted from the other to estimate the relative effects of the different treatments.

A significant improvement follows after recognizing that the properties and geometry of the soil deposit are the same for each alternative design. Therefore, the simulation should be run using the same random numbers to describe the soil in each alternative design. Of course, the uncertain parameters that describe the different designs should be varied independently. The results are expressed in terms of the differences between the results for the several designs.

Such a procedure is called *correlated sampling* because using the same models for the soil in each design correlates the results. The amount of the improvement varies, depending on how much of the uncertainty is in the design parameters and how much in the soil.

17.4.4 Control Variates

The technique of the control variate can be used when there is some function that closely approximates the function we have to deal with and that we can integrate exactly. To be consistent with the notation used earlier in the chapter, let us assume that the function is $g(x)$ and the integral is I. The integral I could be an area, a volume, a mean value, a variance, or some other quantity. Now, let the function that closely approximates $g(x)$ – also known as the *control variate* – be $C(x)$, and let the integral of $C(x)$ be denoted I_C. We can always change the variables so that the integral over the range of the problem is unity. Then we can define a new function $Z(x)$ as

$$Z(x) = g(x) - \beta[C(x) - I_C] \tag{17.51}$$

where β is a factor to be determined. Because

$$\int C(x)dx = I_C \tag{17.52}$$

it follows that, regardless of the value of β,

$$\int Z(x)dx = \int g(x)dz \tag{17.53}$$

If the integrations are approximated by Monte Carlo simulation, it similarly follows that the expected values of the simulation of the integral of Z will be the expected value of the integral of g.

The basic idea of the control variate approach is to simulate the integration of Z rather than g. Why this should work is illustrated in Figure 17.12, which shows two control variates that approximate the function $\cos^2\theta$, whose integration was discussed in earlier sections. One control variate is a linear approximation; the other is cubic. If we let $\theta = \pi x/2$, the equations for the control variates become

$$\begin{aligned} C(x) &= 1 - x && \text{linear} \\ C(x) &= 1 - 3x^2 + 2x^3 && \text{cubic} \end{aligned} \tag{17.54}$$

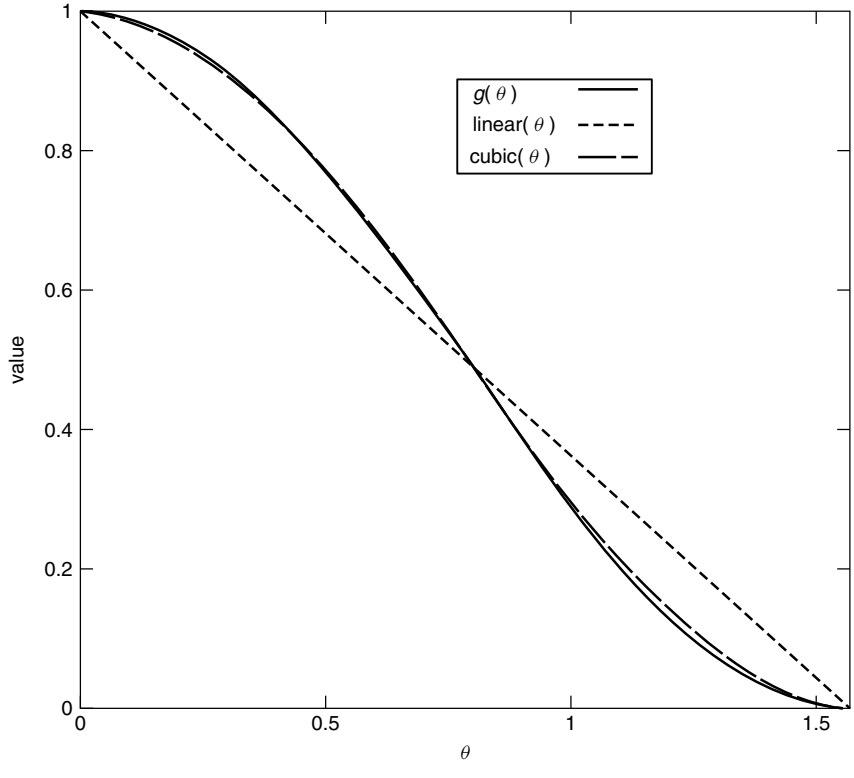


Figure 17.12 Linear and cubic control variates compared to $\cos^2\theta$.

The area under each of the control variates is $\pi/4$, which equals the area under $\cos^2\theta$. If we set $\beta = 1$ and rearrange Equation (17.51) into

$$Z(x) = I_C + [g(x) - C(x)] \tag{17.55}$$

we can see that the procedure is equivalent to starting with a good estimate of the integral I_C and doing the simulation on the difference between the functions. If $C(x)$ is a good approximation, the correction term should be small.

The open triangles in Figure 17.8 show the results of using control variates to improve convergence of the integral with the factor β set to unity. The calculations were done the same way as for importance sampling. The figure shows that the variance reduction in this case is even more effective than it was with importance sampling. A further advantage of the control variate approach in this case is that the importance function approach requires that the random variables conform to a cubic probability distribution function, which involves finding the inverse of a quartic function, while the control variable is a simple cubic function that is not inverted. Thus, the programming and computational efforts are both reduced.

The control variate does not have to match the function to be integrated exactly. A more general statement of control variate theory is that, if $\overline{Y}$ is an estimator to be evaluated by simulation, then the modified simulation is governed by

$$\overline{Z} = \overline{Y} - \beta(C - \mu_C) \tag{17.56}$$

The optimum value of β is

$$\beta^* = \frac{Cov(\overline{Y}, C)}{Var(C)} \tag{17.57}$$

and the variance in the resulting estimate is

$$Var(\overline{Z}) = Var(\overline{Y})[1 - \rho^2_{\overline{Y},C}] \tag{17.58}$$

In other words, the requirement is that the control variate be well correlated with the uncontrolled estimator.

17.4.5 Stratified Sampling and the Latin Hypercube

The *stratified sampling* technique is widely used in statistical sampling, and it can be applied to Monte Carlo simulation as well. Whereas importance sampling is an attempt to place most of the sampling points in regions that contribute most to the integral, stratified sampling tries to place the points so that more of them will be found in regions where the variance of the function $g(x)$ is largest. Following Rubinstein's (1981) notation, we consider that the sampling is to be done with N points in m intervals or regions. The probability that the random variable lies in the region i is P_i, and of course the P_i's must add up to unity. If then the standard deviation of the variable in region i is σ_i, the number of sampling points to be used in region i should be

$$N_i = N\frac{P_i\sigma_i}{\sum_{j=1}^{m} P_j\sigma_j} \tag{17.59}$$

A common procedure is to select intervals so that the P_i's are all the same. Then the distributions are weighted according to the variances in each interval. A special case occurs when $P_i = 1/m$ and $N_i = N/m$. Rubinstein (1981) calls this *systematic sampling*. It can also be shown that, when one point is used in each interval, the most efficient place for that point is in the middle of the interval.

Although stratified sampling can reduce the number of sampling points, it is subject to one major problem. If there are D variables to be sampled and each is to be sampled in k regions, the total number of simulation points becomes k^D, which can grow into a very large number. To avoid this so-called 'k^D problem,' the *Latin hypercube* method is often employed. For each variable the k sampling points are placed in independent random order. Then k samplings are made by choosing the first randomized value for each variable, the second randomized value for each variable, and so on until k sets of randomized variables have been chosen. This ensures that each value is used once and that their combination is randomized. The effects of these approaches are illustrated in the following example.

Example 17.5 – Estimated Settlement with Maximum Past Pressure

Duncan (2000) describes a problem of estimating the settlement of a soft soil under a surface load. Figure 17.13 shows the geometry of the problem and the means and standard deviations of the consolidation parameters. Figure 17.14 shows the estimated initial stress conditions and the maximum past pressures. It is assumed that the uncertain variables are Normally distributed. Dividing the consolidating layer into six layers, each 5 ft. thick, and using the mean values of the parameters gives an estimated final settlement of 1.024 ft. The FOSM method, as used by the U. S. Army Corps of Engineers with the increment in each uncertain variable taken as its standard deviation, yields 0.254 ft. for the standard deviation of the final settlement. Corresponding estimates calculated by the Rosenblueth point-estimate method are 1.013 ft. and 0.256 ft. In these and subsequent calculations the

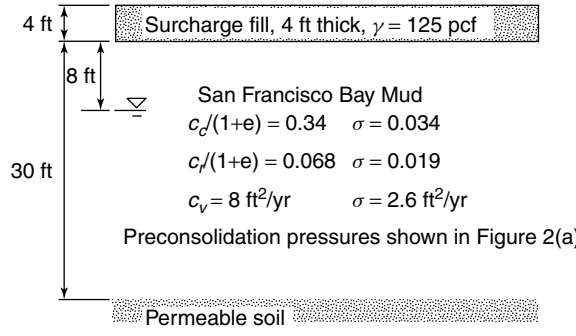


Figure 17.13 Geometry of settlement problem. (Duncan, J. M., 2000, 'Factors of Safety and Reliability in Geotechnical Engineering,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE,* Vol. 126, No. 6, pp. 307–316, reproduced by permission of the American Society of Civil Engineers.)

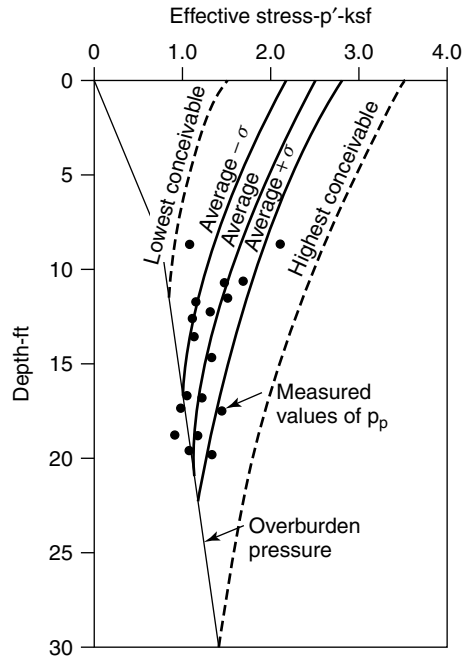


Figure 17.14 Stress distribution parameters for settlement problem. (Duncan, J. M., 2000, 'Factors of Safety and Reliability in Geotechnical Engineering,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE,* Vol. 126, No. 6, pp. 307–316, reproduced by permission of the American Society of Civil Engineers.)

final settlement of a layer is calculated as

$$\rho_i = \begin{cases} C_R h \log\left(\dfrac{p_f}{p_0}\right) & \text{if } p_f \le p_m \\ C_C h \log\left(\dfrac{p_f}{p_0}\right) & \text{if } p_m = p_0 \\ C_R h \log\left(\dfrac{p_m}{p_0}\right) + C_c h \log\left(\dfrac{p_f}{p_m}\right) & \text{otherwise} \end{cases} \tag{17.60}$$

where ρ_i is the settlement of layer i, C_C and C_R are understood to be divided by $(1 + e)$, h is the thickness of the layer, p_f is the final vertical effective stress, p_0 is the initial vertical effective stress, and p_m is the maximum past vertical effective stress. These results differ slightly from those reported by Duncan because of minor differences in computational procedures and in scaling values from figures.

Figures 17.15, 17.17 and 17.18 show the results of direct Monte Carlo simulation using 100, 1,000, and 10,000 points, respectively. In each case each variable was sampled at random with no attempt at variance reduction. For comparison the CDFs for Normal distributions with the parameters from the FOSM and point-estimate methods are included on the plots. While the results with 100 points are ambiguous, those for 1000 and 10,000

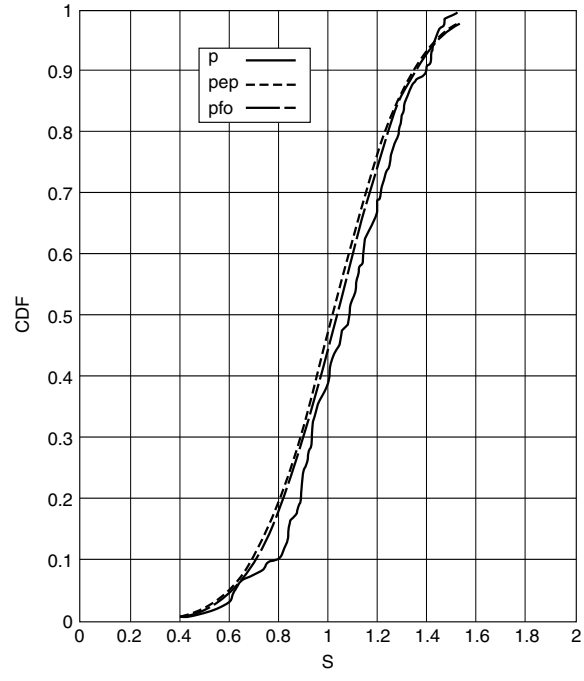


Figure 17.15 Results from Monte Carlo simulation of settlement problem with 100 points. p = simulation, pep = point-estimate, pfo = FOSM.

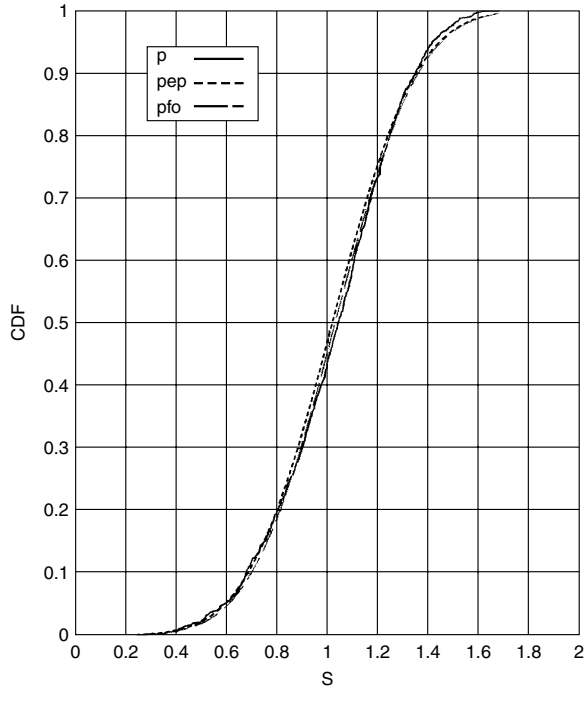


Figure 17.16 Results from Monte Carlo simulation of settlement problem with 1000 points. p = simulation, pep = point-estimate, pfo = FOSM.

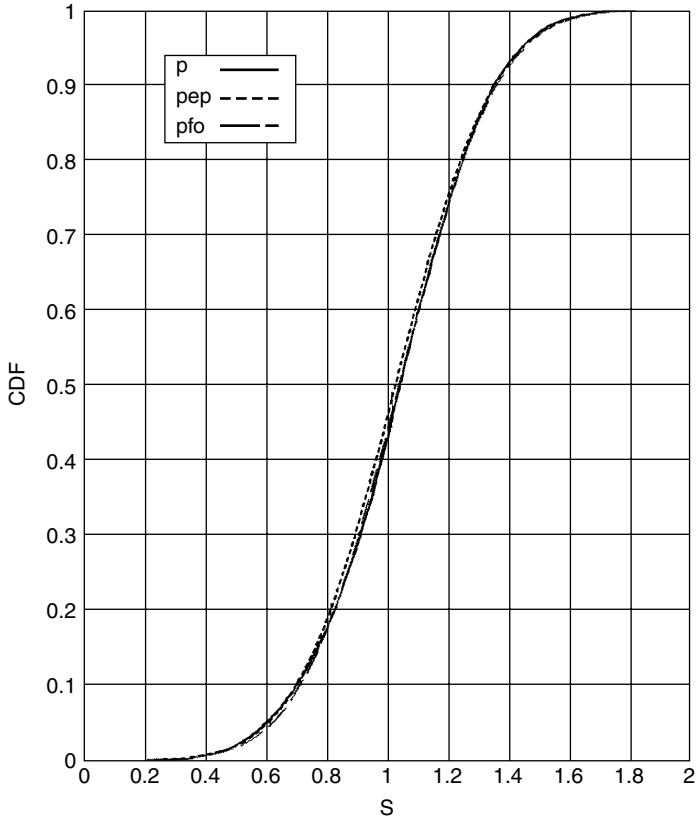


Figure 17.17 Results from Monte Carlo simulation of settlement problem with 10,000 points. p = simulation, pep = point-estimate, pfo = FOSM.

points show that a Normal distribution fits the results quite well. Though not plotted here, the comparison with a LogNormal distribution is much less satisfactory.

Figure 17.18 shows the results from Monte Carlo simulation with stratified sampling. In the simulation, 10 points were used for each variable, for a total of 1000 points. The results are much better than those from direct Monte Carlo simulation with the same number of points but not quite as good as those with ten times as many. The procedure for each simulation was as follows:

1. Divide the probability range from 0 to 1 into 10 intervals.
2. Identify the middle of each range.
3. From the inverse of the Normal CDF find the value of the variable corresponding to the middle of the range.
4. Use all combinations of these values in the simulation.

It should be noted that an error exists in this procedure because the middle of the probability range is not necessarily the same as the middle of the range of the variable in that portion of the distribution. The effect is most severe at the ends of the distribution.

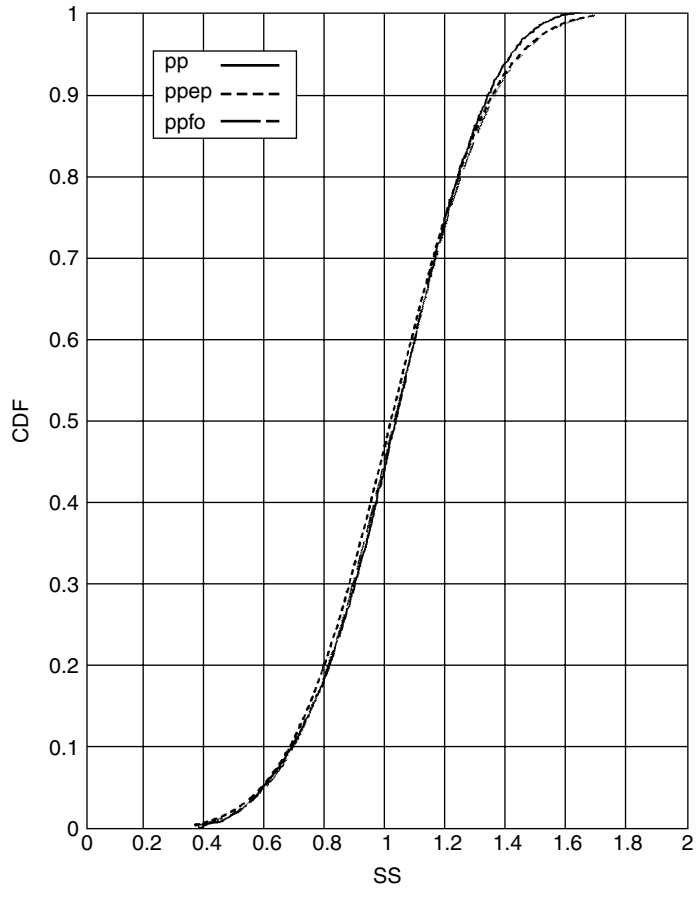


Figure 17.18 Results from Monte Carlo simulation of settlement problem with 1000 total points and stratified sampling. p = simulation, pep = point-estimate, pfo = FOSM.

Finally, the Latin hypercube method gave the results in Figure 17.19, for 1000 points. The procedure described above for stratified sampling was applied to each variable, only using 1000 points instead of 10. The orders of the values were then randomized to give 1000 points in a Latin hypercube. The results are somewhat better than those for a comparable number of points in direct simulation. The number of simulations required to achieve a desired level of accuracy is less than that required by direct Monte Carlo simulation, but not by an order of magnitude.

17.5 Summary

Monte Carlo simulation uses randomly generated points to cover the range of values that enter into a calculation. In most cases the problem involves integration, but it can arise in deterministic or probabilistic circumstances. Some of the computations in probabilistic applications are in fact evaluations of integrals, rather than simulations of stochastic processes.

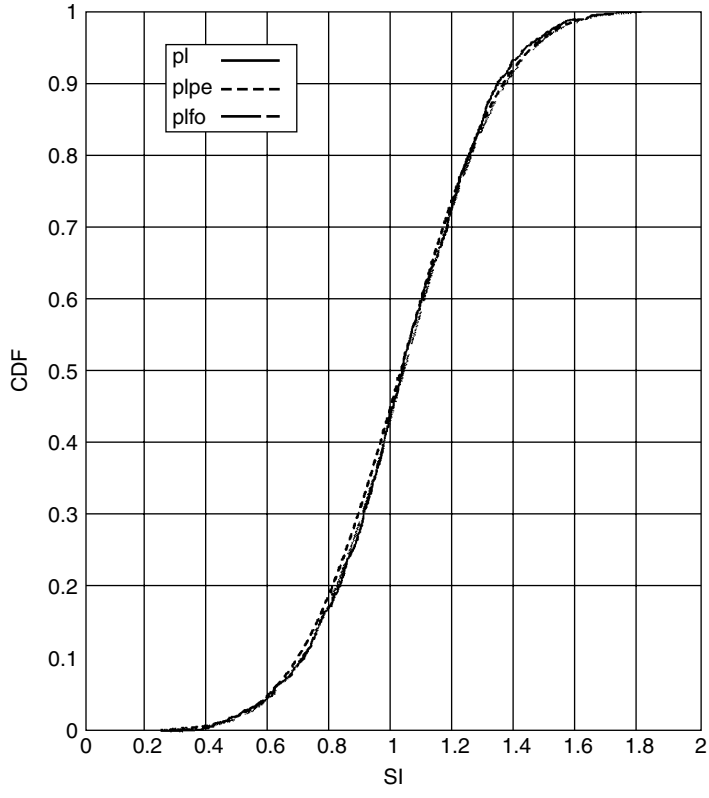


Figure 17.19 Results from Monte Carlo simulation of settlement problem with 1000 points and stratified sampling combined with the Latin hypercube. pl = simulation, plpe = point-estimate, plfo = FOSM.

The technique has the advantage that it is relatively easy to implement on a computer and can deal with a wide range of functions, including those that cannot be expressed conveniently in explicit form. The major disadvantage is that it can converge slowly.

Several techniques are available to accelerate convergence or, equivalently, to reduce variance. Each can be effective for some problems and counter-productive for others. The choice of which method to employ depends on knowledge of the problem and its components. Sometimes a preliminary calculation with a limited number of simulation points will indicate which variance reduction method can be expected to work best. In many cases two or more techniques can be combined effectively.

Finally, accurate Monte Carlo simulation depends on reliable random numbers. Generating random numbers is an art, and one method will not work for all circumstances. There is a substantial literature on the subject, and a person contemplating an important study involving random numbers would be well advised to be satisfied that the random number generators are appropriate for the job.

18 Load and Resistance Factor Design

Most of this book discusses conceptual issues of uncertainty analysis, statistics, and reliability theory and deals with design examples on their own merits, independent of building codes and other aspects of professional practice. In recent years, however, these concepts of uncertainty, statistics, and reliability have begun to enter the public debate over how new structural and geotechnical building codes ought to be written. An important result of this debate is the evolution of building codes based on reliability considerations. The most important of these developments at present is the appearance of *Load and Resistance Factor Design* (LRFD) as the basis for new codes. This chapter discusses the evolution of LRFD-based codes and how reliability theory is used to calibrate and optimize these new codes.

18.1 Limit State Design and Code Development

Allowable Stress Design (ASD) – sometimes called *Working Stress Design* (WSD) – has been the traditional design basis in civil and mechanical engineering for over a century. Working stress design attempts to ensure that loads acting upon a structure or foundation do not exceed some allowable limit, often taken to be the limit of elastic resistance. Typically, the level of assurance that loads do not exceed this allowable limit is reflected in a factor of safety on the ratio of allowable to predicted loads, stresses, or other demands. The safety factor commonly used in ASD is not the *central safety factor* discussed in earlier chapters that uses the ratio of expected resistance to expected load, but the *nominal safety factor* that uses some higher than expected value of load and some lower than expected value of resistance. That is, the load might be taken as that value with an exceedance probability of, say, 0.05; while the resistance might be taken as that value with an exceedance probability of, say, 0.95.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

The geological uniqueness of individual construction sites, along with the demand for economical design, caused the geotechnical community at a very early stage in the development of the field to adopt *Limit State Design* (LSD). Limit state design rather than being based on allowable stresses, is based on predictions how designs actually perform near failure. 'Failure' in this meaning is typically taken either as ultimate failure (collapse) or as conditions affecting serviceability (e.g. excessive deformation). Although limit state design has become widespread in modern structural codes, in fact, geotechnical design led the civil engineering profession in the move toward performance-based design. Limiting equilibrium approaches to the design of slopes, deformation-based approaches to the design of underground openings, and plasticity criteria for the foundations of offshore structures are all examples of limit state design in traditional geotechnical practice. With limit state design, the relation between uncertainties in loads and resistances can be more directly related to the probabilities of 'failure' than they can with ASD. This has led to new directions in North American and European building codes.

In the 1970s and 1980s, the field of structural engineering was dramatically changed by the rapid development of structural reliability theory (Melchers 1987), and the resulting incorporation of probabilistic concepts in structural codes (AASHTO 1994; Ellingwood *et al*. 1980).[1] This was most notable in the development and implementation of *load and resistance factor design* (LRFD) as the basis for new code development in the 1980s and 1990s (Ravindra and Galambos 1978). Among the new structural codes using the LRFD approach are those of the American Institute of Steel Construction (AISC 1994), the American Concrete Institute (ACI 2002), the American Association of State Highway and Transportation Officials (AASHTO 1994), and the American Forest and Paper Association.

LRFD replaces single factors of safety on the ratio of total resistance to total load with a set of partial safety factors on individual components of resistance and load and uses limiting states as the checking points for design. The partial safety factors can be chosen to reflect differing uncertainties associated with the individual components of resistance and load and thus to improve upon the undifferentiated approach of simple factors of safety. This approach is not new. Freudenthal (1951, 1956) proposed partial safety factors in his seminal early papers on structural safety, and Taylor (1948) discusses partial safety factors (specifically with respect to the soil strength parameters c and ϕ for analysis of the stability of embankments) in his classic soil mechanics text.

18.1.1 Allowable Stress Design

Allowable Stress Design (ASD) attempts to ensure that for some applied service load on a structure or foundation the stresses induced in the soil mass are less than some set of specified allowable stresses. ASD combines uncertainties in loads and soil strengths (or deformations) into a single factor of safety, *FS*. That is, a set of design loads, Q_i, comprising the actual forces estimated to be applied directly to a structure is balanced against a set of resistances, R, from the soil mass, such that

$$R/FS = \sum Q_i \tag{18.1}$$

[1] This period also saw a good deal of work on the philosophy of code writing, some of which is captured in the publications of Lind (1972), Veneziano (1976), Turkstra (1970), Ditlevsen (1997), and others.

in which, Q_i = individual design loads, $i = 1, \ldots, n$, R = ultimate resistance, and FS = factor of safety. Equation (18.1) can also be written as

$$R/Q = FS \tag{18.2}$$

in which $Q = \Sigma Q_i$ is the total load. The factor of safety is commonly defined as the ratio of the resistance of the structure to the sum of the loads acting on the structure.

The factor of safety of Equation (18.2) can be thought of as either a variable or a constant. As a variable, the factor of safety reflects uncertainty in both resistance and load. As a constant, the factor of safety reflects a target value for the ratio of predicted resistances and predicted loads. In practice, the factor of safety depends mostly on the level of control that is used in design and construction phases, and thus on resistances. As a result, traditionally, the factor of safety is considered to apply to the resistance side of Equation (18.1), in that uncertainties in material strength are the largest uncertainties affecting geotechnical design.

As an example of ASD, AASHTO (1998) recommends factors of safety for pile foundations reflecting different levels of control in the prediction of ultimate resistance. When a more reliable and consistent level of construction control is used, a smaller factor of safety can be used, leading to a less expensive design.

Because uncertainty is involved in both the resistance and the loads, it is important to distinguish the two ways referred to above in which the factor of safety is sometimes defined. The first is the *central factor of safety*

$$FS = \frac{E[R]}{E[Q]} \tag{18.3}$$

in which $E[R]$ is the expected value of total resistance and $E[Q]$ is the expected value of total loads; and the second is the *nominal factor of safety*, which is the definition more common in traditional practice

$$FS = \frac{R_k}{Q_k} = \frac{E[R] - k_R\sigma_R}{E[Q] + k_Q\sigma_Q} \tag{18.4}$$

in which R_k and Q_k are the nominal values of resistance and load, respectively; σ_R, σ_Q are the standard deviations; and k_R, k_Q are constants. In the present discussion, unless otherwise specified, factor of safety refers to the *central* meaning of Equation (18.3).

Serviceability limit states for working stress design typically apply to deformation criteria. Following the example of pile design, in addition to the predicted axial geotechnical and structural capacity of a pile, the design of driven piles evaluates pile deflections and compares them with deformation criteria using the following:

$$\delta_i \leq \delta_n \tag{18.5}$$

where δ_i is the estimated displacement; and δ_n is the tolerable displacement established by designer, and tolerable displacement depends on the type of structure (e.g. serviceable of bridge foundations depend on the type of supports used for the superstructure as well as the superstructure design itself).

18.1.2 Limit states

Limit states, as discussed in previous chapters, are defined as those conditions under which a structure or its components no longer perform an intended function (Allen 1994; Simpson *et al.* 1981). Whenever a structure or a part of a structure fails to satisfy one of its designated operational criteria, it is said to have reached a limit state. The two limit states of interest for most foundations are (1) ultimate limit state (strength limit state), and (2) serviceability limit state. Ultimate limit states pertain to structural safety and collapse. For foundations, the ultimate limit state is typically taken to be ultimate bearing capacity of the soil. Serviceability limit states pertain to conditions under which performance requirements under normal service loads are exceeded. Such conditions might include excessive deformations or deterioration of a structural foundation system such as piles. Serviceability limit states are typically checked using a partial factor of unity on all specified or characteristic service loads and load effects (Meyerhof 1994).

In ASD, the concept of allowable soil bearing pressure may be controlled either by bearing capacity (ultimate state) or by settlement (serviceability limit) considerations. ASD implicitly accounts for both of these, but generally does so only implicitly (Becker 1996). In other words, the ASD method usually does not require that calculations be made to check both limit states but instead resorts to charts that show whether a design needs to be checked for ultimate limit state or for serviceability limit state.

Common wisdom holds that serviceability limit states have a higher probability of occurrence than do ultimate limit states, principally because they occur at lower absolute loads (Duncan *et al.* 1989). As a result, the allowable settlement of a structure rather than the ultimate bearing capacity of the soil generally controls the design of shallow foundations. Accordingly, most routine design is based on specific serviceability limits, and ultimate limits are checked subsequently (Becker 1996). However, this in not always the case; in the design of piles ultimate limit state often controls the design.

18.2 Load and Resistance Factor Design

In contrast to working stress design with its single safety criterion – the factor of safety, FS – Load and Resistance Factor Design (LRFD) takes into consideration the variability in loads and resistances separately by defining separate factors on each. A load factor, γ, is assigned to variability or uncertainty in loads, while a resistance factor, ϕ, is assigned to variability or uncertainty in resistances. Thus, in LRFD the comparison of loads and resistances is formulated for strength limit states in an equation of the form (Withiam *et al.* 1997)

$$R_r = \phi R_n \geq \eta \sum \gamma_i Q_i \tag{18.6}$$

in which R_r is factored resistance, R_n is ultimate resistance, Q_i is force effect, stress, or stress resultant, ϕ is resistance factor, γ_i is load factor, and $\eta = \eta_D \eta_R \eta_I > 0.95$ is a series of dimensionless factors (taken from the structural code) accounting for the effects of ductility (η_D), redundancy (η_R), and operational importance (η_I). For the serviceability limit states,

$$\eta \sum \gamma_i \delta_i \leq \phi \delta_n \tag{18.7}$$

in which δ_i is the estimated displacement, and δ_n is the tolerable displacement. The method for determining the ultimate geotechnical resistance (R_n) is basically the same for both the LRFD and ASD methods. The difference between the methods is the separation of factors for load and resistance in the LRFD method.

The use of LRFD in geotechnical engineering is a relatively recent development, which has been in large part driven by the development of LRFD codes for structural engineering, especially in application to highway structures. To date, the principal use of LRFD in geotechnical engineering has been in foundation design. The approach has not enjoyed extensive use in the design of earth structures, such as dams, or of underground works.

Early development of the LRFD codes for structural design grew out of efforts at the Nationals Standards Institute to develop probability-based load criteria for buildings (Ellingwood *et al.* 1980); some of this was summarized and later republished by ASCE (1993). The American Petroleum Institute sponsored research on LRFD methods for the design of offshore structures during the 1980s. Much of the publicly available parts of this work is reported in API sources (API 1989; Moses 1985; Moses and Lind 1970). Other contemporary efforts to incorporate LRFD in codes for the structural design of buildings are reported by Sui *et al.* (1975), the National Research Council of Canada (1977) and CIRIA (1977). In more recent years, LRFD has begun to be introduced to structural designs in traditionally non-civil engineering applications, such as ship structures (Ayyub *et al.* 1997, 1995), although its introduction in machine design and other traditionally mechanical engineering applications continues to lag behind.

Early applications of LRFD to transportation structures (e.g. highway bridges) – which in recent years has led to more attention paid to the use of LRFD for the foundations of those structures – were encouraged by the American Association of State Highway and Transportation Officials (AASHTO), and by 1994 the AISC *Manual of Steel Construction* itself had adopted LRFD (AASHTO 1994; AISC 1994; Galambos and Ravindra 1978). This move to LRFD appears to have allowed significant savings in structural system costs on highway bridges, with some studies suggesting savings of from 5–30% in steel weight, with average savings of about 10% (Stenersen 2001). In an effort to harmonize design between structural systems and the foundations on which those systems rest, AASHTO and the Federal Highway Administration again led the effort to develop geotechnical codes similarly based on LRFD (AASHTO 1997; Withiam *et al.* 1998). These codes have primarily focused on the design of shallow foundations, piles and drilled shafts, and earth retaining works.

The benefits of moving to LRFD for foundations has been a more efficiently balanced design with corresponding improvements to reliability, a more rational approach to accounting for load and resistance uncertainties in design codes, and a broadening of the benefits of reliability-based design into routine design problems where specific reliability calculations are not cost-effective (i.e. LRFD-based codes deliver optimized design formulas without the need for repeating reliability calculations for each particular design situation). Other applications of LRFD to geotechnical engineering – principally for foundation design – have been published by Barker *et al.* (1991), Fellenius (1994), Meyerhof (1994), O'Neill (1995), Kuhlhawy and Mayne (1990) and Goble (1998), among others.

18.2.1 Calibration to allowable stress design

Determinating load and resistance factors for a particular case is done either by calibrating Equation (18.6) to past practice, that is, to typical or codified factors of safety used in working stress design, or by optimizing Equation (18.6) for a chosen reliability index

β using the methods of reliability theory of the previous chapters. Typically, this optimization has used either first-order second-moment reliability (FOSM) or Hasofer–Lind reliability (FORM).

Calibration through fitting to the ASD method is used when insufficient statistical data exist to perform a calibration by optimization (Withiam *et al.* 1997). The method of calibrating by fitting to ASD is a simple process in which the LRFD Equation (18.6) is divided by the fundamental ASD Equation (18.1) to determine an equation for the resistance factor (ϕ). If an average value of 1.0 is used for the load modifier η, then

$$\phi \geq \frac{\sum \gamma_i Q_i}{FS \sum Q_i} \tag{18.8}$$

If only the dead loads and live loads are considered, this becomes

$$\phi = \frac{\gamma_D Q_D + \gamma_L Q_L}{FS(Q_D + Q_L)} \tag{18.9}$$

Dividing both the numerator and denominator by Q_L (O'Neill 1995) yields

$$\phi = \frac{\dfrac{\gamma_D Q_D}{Q_L} + \gamma_L}{FS\left(\dfrac{Q_D}{Q_L} + 1\right)} \tag{18.10}$$

Equation (18.10) can be used to determine the resistance factors that need to be used in the LRFD equations to obtain a factor of safety equal to that of the ASD method.

While the load and resistance factor design method represented by Equation (18.6) as calibrated using Equation (18.10) provides a clearer separation of load and resistance uncertainties than does a single factor of safety, the use of the load and resistance factors based on earlier, non-reliability-based procedures does little to improved design. Table 18.1 shows values of the resistance factors using Equation (18.10) for factors of safety varying between 1.5 to 4 and average dead and live load factors of 1.25 and 1.75, respectively. The table gives resistance factors that would be used to obtain factors of safety of 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 for dead to live load ratios of 1, 2, 3 and 4. Clearly,

Table 18.1 Values of resistance factors from Equation (18.10) corresponding to different values of safety factor and dead to live load ratios for $\gamma_D = 1.25$ and $\gamma_L = 1.75$ (after Withiam *et al.* 1998)

	Resistance factor, ϕ			
Safety factor	$Q_D/Q_L = 1$	$Q_D/Q_L = 2$	$Q_L/Q_D = 3$	$Q_L/Q_D = 4$
1.5	1.00	0.94	0.92	0.90
2.0	0.75	0.71	0.69	0.68
2.5	0.60	0.57	0.55	0.54
3.0	0.50	0.47	0.46	0.45
3.5	0.53	0.40	0.39	0.39
4.0	0.38	0.35	0.34	0.34

calibrating the LRFD method to existing practice leads to a design much the same as that obtained using ASD.

18.2.2 Calibration using FOSM reliability

The resistance factor chosen for a particular limit state must take into account (i) the variability of the soil and rock properties, (ii) the reliability of the equations used for predicting resistance, (iii) the quality of the construction workmanship, (iv) the extent of soil exploration (little versus extensive), and (v) the consequences of failure.

Replacing the ultimate or nominal resistance R_n in Equation (18.6) by $E[R]/\lambda_R$ (the product of the expected value of resistance $E[R]$ and the statistical bias in the resistance λ_R) and ignoring the correction terms η leads to a resistance factor that satisfies the inequality

$$\phi_R \geq (\lambda_R / E[R]) \sum \gamma_i Q_i \tag{18.11}$$

For the case in which R and Q are each logNormally distributed and probabilistically independent (an assumption often made in LRFD codes for geotechnical design), setting Equation (18.11) to an equality produces a simple relationship between resistance factor ϕ and the uncertainties in resistance and loads as expressed in coefficients of variation. For logNormal variables, it is convenient to define the limiting condition on the safety margin involving the logarithms, $\ln R - \ln Q$, such that

$$g(R, Q) = \ln R - \ln Q = 0 \tag{18.12}$$

Then the reliability index β is the number of standard deviations of the derived random variable $(\ln R - \ln Q)$ separating its mean value from the limiting state or failure condition of 0

$$\beta = \frac{E[\ln R] - E[\ln Q]}{\sqrt{Var[\ln R - \ln Q]}} \tag{18.13}$$

That standard deviation is related to the marginal standard deviations by the sum of variances

$$Var[\ln R - \ln Q] = Var[\ln R] + Var[\ln Q] \tag{18.14}$$

Recall from Appendix A that the first two moments of a logNormal variable can be simply related to the first two moments of the logarithms of the variable by

$$E[\ln R] = \ln \left(\frac{E[R]}{\sqrt{1 + \Omega_R^2}} \right) \tag{18.15}$$

$$Var[\ln R] = \ln(1 + \Omega_R^2) \tag{18.16}$$

in which Ω is the coefficient of variation; similarly for Q. Substituting into Equation (18.13) yields the expression (Barker *et al.* 1991)

$$\phi_R = \frac{\lambda_R \left(\sum \gamma_i Q_i \right) \sqrt{\dfrac{1 + COV(Q)^2}{1 + COV(R)^2}}}{E(Q) \exp\{\beta_T \sqrt{\ln[(1 + COV(R)^2)(1 + COV(Q)^2)]}\}} \tag{18.17}$$

in which λ_R is the overall bias for resistance, $COV(Q)$ is coefficient of variation of the load, $COV(R)$ is coefficient of variation of the resistance, β_T is target reliability index, and $E(Q)$ is the expected value of total load.

When dead and live loads are considered separately, Equation (18.17) becomes (Yoon and O'Neill, 1997)

$$\phi_R = \frac{\lambda_R \left(\dfrac{\gamma_D E(Q_D)}{E(Q_L)} + \gamma_L \right) \sqrt{\left[\dfrac{(1 + COV(Q_D)^2 + COV(Q_L)^2)}{(1 + COV(R)^2)} \right]}}{\left(\dfrac{\lambda_{Q_D} E(Q_D)}{E(Q_L)} + \lambda_{Q_L} \right)}$$

$$\times \exp\{\beta_T \sqrt{\ln[(1 + COV(R)^2)(1 + COV(Q_D)^2 + COV(Q_L)^2)]}\} \quad (18.18)$$

in which λ_R, λ_{QD} and λ_{QL} are bias terms for resistance, dead load (Q_D) and live load (Q_L), and γ_D and γ_L are dead load and live load factors, respectively.

Because the levels of uncertainty in dead loads and live loads differ from one another, so, too, do their corresponding load factors. Table 18.2 shows typical bias terms and coefficients of variation for dead load components and for live loads on highway bridges. The AASHTO (1994) load factors combine dead loads into a single factor.

A typical example of calibrated resistance factors is shown in Table 18.3, taken from Barker *et al.* (1991) for axially loaded piles, calibrated both against ASD and by using FOSM analysis with the specified target reliability indices. For use in codes, recommended values of load and resistance factors are typically rounded to the nearest increment of 0.05, and judgment on the historical development of earlier code factors of safety is taken into account in balancing differences between the two calibrations. Also, for most common situations, the dependence of reliability index, β, on the ratio of dead load to live load, Q_D/Q_L, is relatively small. Thus, a representative value of Q_D/Q_L is typically used, and a single resistance factor chosen, irrespective of load ratio.

18.2.3 Calibration using FORM reliability

The currently preferred way to perform the calibration to obtain load and resistance factors is through first-order reliability (FORM), that is, using the Hasofer–Lind procedure discussed in Chapter 16. This procedure is based on choosing a checking point, called,

Table 18.2 Statistics for structural load components for highway bridges. (Nowak, A. S., 1995, 'Calibration of LRFD Bridge Code,' *Journal of Structural Engineering, ASCE* Vol. 121, No. 8, pp. 1245–1251, reproduced by permission of the American Society of Civil Engineers)

Load component	Bias, λ	*COV*, Ω	AASHTO Load factor, γ
Dead load			1.25
Factory-made	1.03	0.08	
Cast-in-place	1.05	0.10	
Asphaltic wearing surface	1.00	0.25	
Live load	1.10–1.20	0.18	1.75

Table 18.3 Resistance factors for axial loaded driven piles in sand calibrated to ASD and FOSM ($D/L = 3.7$, $\gamma_D = 1.3$ and $\gamma = 2.17$) (after Barker *et al.* 1991)

Soil test	Pile length (m)	Factor of safety, *FS*	Target reliability, β_T	Resistance factor*, ϕ		
				Reliability analysis,	Fitting with ASD	Recommended
SPT	10	4.0	2.0	0.48	0.33	0.45
SPT	30	4.0	2.0	0.51	0.33	0.45
CPT	10	2.5	2.0	0.59	0.53	0.55
CPT	30	2.5	2.0	0.62	0.53	0.55

*Notes: Dead load to live load ratio assumed to be, $D/L = 3.7$; load factors taken as $\gamma_D = 1.3$ and $\gamma = 2.17$. Were these calculations performed with the current AASHTO (1997) load factors of $\gamma_D = 1.25$ and $\gamma = 1.75$, the corresponding resistance factors would be 5–20% lower (Withiam *et al.* 1998).

the design point, at a particular point on the limiting state surface, and calculating the reliability index, β, separating that point from the joint mean of the uncertain load and resistance variables. The design point is taken as that point on the limiting state surface at which the probability density of the joint random variables is greatest. To simplify the procedure, the calculation of β is made in a normalized space of derived random variables, in which each derived variable has unit variance and is independent of all other derived variables. In this derived space, the shortest distance between the joint mean of the derived variables and the state limit function identifies both the design point and the reliability index. The method yields partial safety factors for loads and resistances at a specified target reliability index, β_T. As discussed in Chapter 16, this procedure has the advantage over FOSM of greater invariance with respect to the mathematical definition of limiting state.

Define the *limit state function* as,

$$g(\mathbf{X}) = g(X_1, X_2, \ldots, X_n) = 0 \tag{18.19}$$

in which $\mathbf{X}$ is a vector of random variables $(X_1, X_2, \ldots X_n)$ of loads and resistances, for which the limit state is $g(\mathbf{X}) = 0$. Failure occurs for $g(\mathbf{X}) < 0$, non-failure otherwise (Figure 18.1). The n-dimensional space of the variables $\mathbf{X}$ is first transformed to $\mathbf{X}'$ by the normalization

$$x_i' = \frac{x_i - m_{x_i}}{\sigma_{x_i}} \tag{18.20}$$

to create zero-mean, unit-variance variables. The space $\mathbf{X}'$ is then rotated to remove correlations among the derived variables, yielding zero-mean, unit-variance, independent variables. Finally, the space is transformed for non-Normal distributions (e.g. for logNormal variables the logarithm of each is taken), yielding a normalized space in which the shortest distance between the origin (i.e. the joint means of the derived random variables x_i') and the design point – the *most probably failure point* – on the transformed limit state gives the reliability index for the design (Figure 18.2).

For a given target reliability index β_T, the mean values of basic variables can be used to compute partial safety factors required to provide the target reliability

$$\gamma_i = \frac{x_i^*}{E[x_i]} \tag{18.21}$$

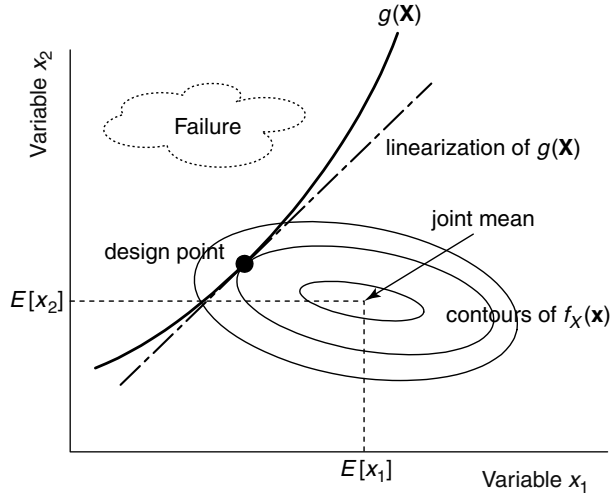


Figure 18.1 Limit state function and pdf of basic random variables.

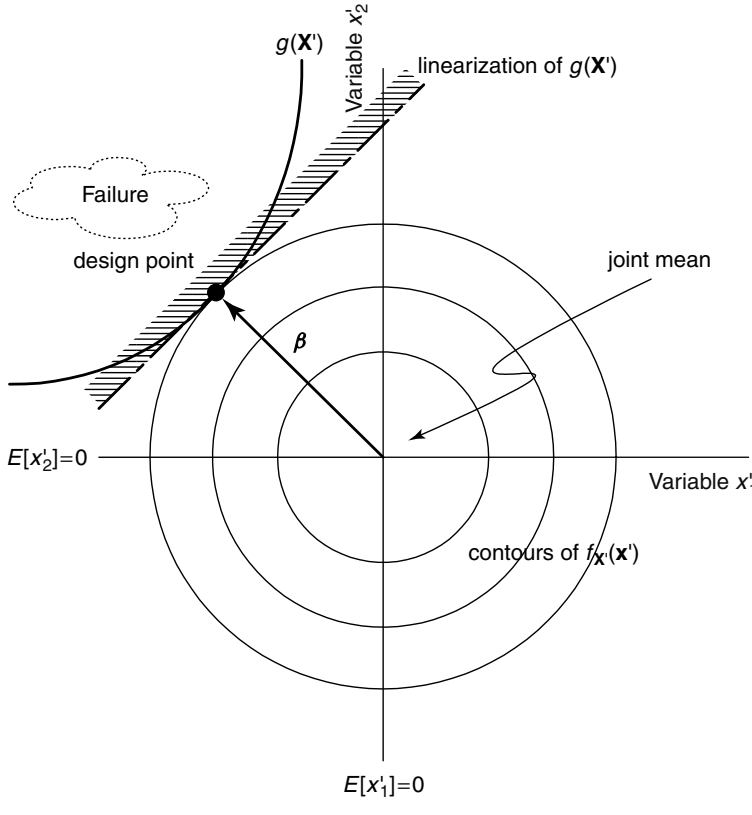


Figure 18.2 Transformed basic variable space.

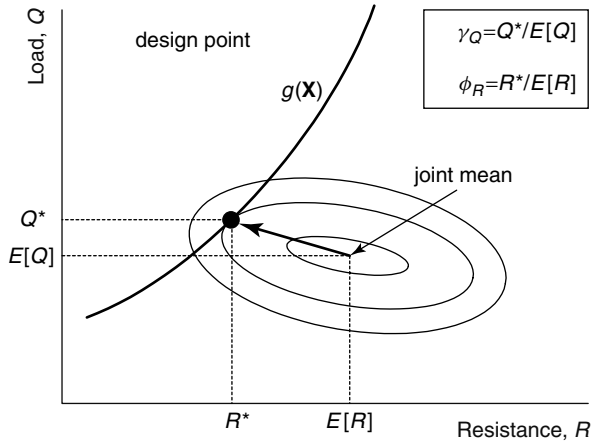


Figure 18.3 Relationship of partial factors to design point.

in which γ_i is the partial factor for variable x_i, and x_i* is the value of the transformed variable x_i at the design point. Considering the usual case (Figure 18.3) in which there are more than one load variables, Q_i, but only a single aggregated resistance variable, R, the corresponding load and resistance factors become

$$\gamma_i = \frac{Q_i^*}{E[L_i]} \tag{18.22}$$

$$\phi_R = \frac{R^*}{E[R]} \tag{18.23}$$

in which the γ_i are the load factors and ϕ is the resistance factor. Note, this case of multiple load variables and a single resistance variable is strongly influenced by structural (as opposed to, geotechnical) reliability considerations, in which the uncertainties of load conditions dominate those of resistance. As the practice of geotechnical reliability continues to evolve, it is likely that geotechnical applications of LRFD will begin to differentiate resistance factors, while consolidating the multiplicity of load factors. This would be in line with Taylor's early work on separating partial safety factors for cohesion and friction components of soil strength (Taylor 1948). For the present, however, load factors for superstructure design, for practical purposes, need to harmonized with load factors for foundation design, and thus the formulation of Equations (18.22) and (18.23) remains.

18.2.4 Choice of reliability index β

Studies by a number of workers suggest ASD methods for foundation design result in nominal probabilities of failure that range from 0.0001 to 0.01 (Meyerhof 1994). These correspond to reliability indices that range from 2.5 to 3.5 (Table 18.4). In modern foundation codes, target reliability indices ranging from 2.0 to 3.0 are common, the former for cases of non-essential designs with high redundancy, the latter for critical designs with little redundancy. Unlike the case of a unique site and facility with a known loss function

Table 18.4 Relationship between probability of failure and reliability index for lognormal distribution (after Withiam *et al*. 1998)

Reliability index, β	Probability of failure, p_f	
	Normal distribution	logNormal distribution
2.5	6.21×10^{-3}	0.99×10^{-2}
3.0	1.35×10^{-3}	1.15×10^{-3}
3.5	2.33×10^{-4}	1.34×10^{-4}
4.0	3.17×10^{-5}	1.56×10^{-5}
4.5	3.40×10^{-6}	1.82×10^{-6}
5.0	2.87×10^{-7}	2.12×10^{-7}
5.5	2.90×10^{-8}	2.46×10^{-8}

for failures, a code must be developed for a broad spectrum of potential conditions, and thus the optimization of reliability level against cost is more qualitative.

As an approximation, the relationship between probability of failure and reliability index for Normal variables is often taken from Rosenblueth and Esteva (1972) as

$$p_f = 460e^{(-4.3\beta)} \text{ for } 2 \leq \beta \leq 6 \tag{18.24}$$

$$\beta = \frac{\ln\left(\frac{460}{p_f}\right)}{4.3} \text{ for } 10^{-1} \leq p_f \leq 10^{-9} \tag{18.25}$$

18.3 Foundation Design Based on LRFD

Two areas of geotechnical engineering where LRFD approaches have enjoyed popularity are the design of foundations and the development of codes for foundations. This has been driven in part by the desire to harmonize the design of structural elements of bridges and buildings with the design of the foundations that support those superstructures. In the United States the move to LRFD codes for both shallow and deep foundations has been fostered by AASHTO.

The design of foundation systems has benefited more from LRFD than have other areas of geotechnical engineering because it is one of the few areas to be codified, especially in European practice. Foundation performance is also an area for which the extensive statistical data exist for comparing observed with predicted performance that are required with which to calibrate partial factors.

18.3.1 Spread footings

Spread footings are typically designed against six (6) limiting states, one pertaining to service limits, i.e. settlement deformations, and the remainder to strength limits (Table 18.5). Suggested or codified factors of safety in ASD typically are expressed as a range, depending on the type and extent of site characterization and the analytical method used in making performance predictions. Higher factors of safety have been recommended for use with

Table 18.5 Limiting states for spread footing design

Limiting state	Limit state equation	ASD Factor of safety
Settlement	$\delta_i \leq \delta_n$	
Bearing capacity	$Q \leq R_n FS$	2.0–3.0
Sliding	$Q \leq R_n FS$	2.0–3.0
Overturning	Resultant load w/in B/6 of centerline	
Overall stability or structural capacity	$\sum T \leq \sum RFS$	1.25–2.0

In which, δ_i = differential settlement, δ_n = tolerable differential settlement, Q = allowable load, R_n = ultimate capacity of the soil subgrade, FS = factor of safety, ΣT = net driving forces, ΣR = net resisting forces.

empirical measures of soil properties, such as the Standard Penetration Test (SPT), rather than with measures based on the principles of mechanics, such as the Cone Penetration Test (CPT) or laboratory measurements. Whereas, the summation of all uncertainties in loads and resistances for ASD is accounted for in the factor of safety, in LRFD the uncertainties in loads and resistances are treated separately to account for the differing magnitude of those uncertainties.

In U.S. codes for LRFD, the load factors γ_i are typically set by the structural sections of the code (e.g. Table 18.2), which historically have preceded the geotechnical sections. In the AASHTO codes, the current dead load and live load factors are

$$\gamma_{DQ} = 1.25 \tag{18.26}$$

$$\gamma_{LQ} = 1.75 \tag{18.27}$$

In the case of highway bridge structures, for example, load factors are optimized within the calibration of the structural design procedure and then imposed on the geotechnical design procedure. Obviously, this produces a sub-optimal calibration of resistance factors in the geotechnical side of the LRFD code, because load and resistance factors should be jointly determined or optimized in the calibration procedure. On the other hand, adopting the structural load factors in the geotechnical design provides consistent load factors across the structural and geotechnical parts of the codes, since, after all, the load factors in both the structural and geotechnical parts of the code apply to the very same live and dead loads.

Load factors associated with earth pressures on structural elements, taken from the AASHTO code are shown in Table 18.6. It is important to note that most work to date on LRFD for geotechnical design has focused on foundation design. This focus has been driven by the desire to develop geotechnical sections within codes primarily concerned with structural engineering, for example the AASHTO codes for highway bridges or the National Institute of Standards and Technology model codes for building structures (Ellingwood *et al*. 1980). Thus, most of the work on LRFD for geotechnical applications has enjoyed a more or less clear separation – and presumably a probabilistic independence – between (structural) loads and (soil or rock) resistances. In the arena of earth structures, such as retaining structures and embankments, this separation is no longer so obvious. Loads depend on soil properties such as density and groundwater levels, but so, too, do resistances. The development of LRFD procedures for these other design considerations in geotechnical engineering will be more complex.

Table 18.6 Load factors associated with earth pressures on structural elements (AASHTO 1994)

Load		Load factor, γ	
		Maximum	Minimum
Components and attachments		1.25	0.90
Downdrag		1.80	0.45
Wearing surfaces and utilities		1.50	0.65
Horizontal earth pressures	Active	1.50	0.90
	At-rest	1.35	0.90
Vertical earth pressures	Overall stability	1.35	N/A
	Retaining structure	1.35	1.00
	Rigid buried structure	1.30	0.90
	Rigid frame	1.30	0.90
	Flexible buried structures (other than metal box culverts)	1.95	0.90
	Flexible metal box culverts	1.50	0.90
Earth surcharge		1.50	0.75

Example

The bearing capacity of shallow footings is usually predicted using Terzaghi's superposition method (Terzaghi 1943) combined with a good deal of empirical calibration. The literature contains many theoretical derivations, as well as model tests and field observations, which may be combined to assess the mean and variance of bearing capacity predictions as a function of soil properties. Using Terzaghi's superposition, the ultimate bearing capacity of a shallow, concentrically loaded strip footing on a homogeneous soil is commonly determined from

$$q_v = cN_c + qN_q + \tfrac{1}{2}BN_\gamma \tag{18.28}$$

in which q_v is the ultimate bearing capacity for a vertical concentric load, N_c, N_q, and N_γ are bearing capacity factors, B is the foundation width, q is a uniform surcharge around the foundation, and c and γ are effective soil cohesion and unity weight, respectively. Note that the N's are functions of the friction angle ϕ.

For the special case of a footing initially on the surface of a cohesionless soil, $c = q = 0$, and Equation (18.28) becomes $q_v = \frac{1}{2}BN_\gamma$. Empirical data summarizing the relationship of this equation (and other theoretical solutions) to empirically observed ultimate bearing capacities are shown in Figure 18.4 (from Ingra and Baecher 1983). Regression analysis on the data yield the predictive relationships,

$$\begin{aligned} E[N_\gamma]_{L/B=6} &= \exp\{-1.646 + 0.173\phi\} \\ Var[N_\gamma]_{L/B=6} &= (0.0429)\exp\{-3.292 + 0.345\phi\} \end{aligned} \tag{18.29}$$

so, for the case of, say, $\phi = 35^\circ$, the mean and standard deviation of predicted bearing capacity factor would be, $E[N\gamma] = 82$ and $\sqrt{Var[N\gamma]} = 17$. Thus, using a first-order second-moment approximation (Chapter 14) to find the mean and standard deviation of

the bearing capacity from Equation (18.28)

$$\begin{aligned} E[q_\gamma] &= \tfrac{1}{2}\gamma B E[N_\gamma] = \tfrac{1}{2}(18.5\,kN/m^3)(1m)(82) = 760\,\text{kPa} \\ Var[q_\gamma] &= \{\tfrac{1}{2}\gamma B\}^2\, Var[N_\gamma] = \{\tfrac{1}{2}(18.5\,kN/m^3)(1m)\}^2(17)^2 = (157)^2\,\text{kPa}^2 \end{aligned} \tag{18.30}$$

which yields a coefficient of variation of, $\Omega_q = 157/760 = 0.21$.

In design, the prediction of bearing capacity using this empirically calibrated formula would have no bias, that is, the mean value of the ratio of observed to predicted bearing capacity is by definition, 1.0, since Figure 18.4 calibrates the method to experience. The coefficient of variation would be, as above, 21%. For these values, and using dead and live load factors of 1.25 and 1.75, respectively, the corresponding resistance factor using the FORM methodology is about $\phi_R = 0.62$ for a reliability index $\beta = 3$. The design equation then becomes

$$\begin{aligned} R_r &= \phi R_n \geq \eta \sum \gamma_i Q_i \\ &= (0.62) R_n \geq (1.0)[1.25 Q_{DL} + 1.75 Q_{LL}] \end{aligned} \tag{18.31}$$

completing the example.

Barker *et al.* (1991) developed resistance factors for bearing capacity and sliding resistance of shallow footings supporting highway bridges, based on calibration using the FOSM approach. This procedure involved four steps: (1) estimating the inherent level of reliability in various methods of calculating footing capacity; (2) observing variations in reliability with bridge span length; (3) choosing target reliabilities based on the safety margin used in ASD; and (4) back calculating resistance factor using Equation (18.17). Partial results are shown in Table 18.7, suggesting the differences between the ASD and FOSM calibrations, and the effect of bridge span length. They suggest the values of resistance factors in the last column, in part to be more consistent with ASD experience.

Resistance factors for overall slope stability under a shallow footing were also calculated by Barker *et al.* (1991) by direct calibration to the AASHTO code ASD requirements for factor of safety, using Equation (18.17). ASD calibration was used because it was thought that computer methods of slope stability analysis did not easily lend themselves to FOSM calibration procedures. More recent developments by Low (1996) using the Hasofer–Lind method for slope stability analysis, however, suggest that a reliability-based calibration of resistance factor for slope stability design is straightforward, although yet to be reported in the literature.

Table 18.7 Resistance factors for semi-empirical evaluation of bearing capacity for spread footings on sand using FOSM. (After Barker *et al.* (1991), modified by Withiam *et al.* (1998))

					Resistance factor, ϕ		
Data	FS	Average reliability index, β	Target reliability index, β_T	Bridge span (m)	Fitted ASD	Calibrated FOSM	Suggested
SPT	4.0	4.2	3.5	10	0.37	0.49	0.45
				50	0.37	0.53	0.45
CPT	2.5	3.2	3.5	10	0.60	0.52	0.55
				50	0.60	0.57	0.55

Resistance factors from the AASHTO bridge design code for all limiting states of shallow footings are given in Table 18.8. For this same set of resistance factors in the AASHTO code, Withiam *et al*. (1997) evaluate 'equivalent factors of safety' using the relationship,

$$FS_{\text{LRFD}} = \overline{\gamma}/\phi \tag{18.32}$$

in which $\overline{\gamma}$ is the average load factor, assuming $\gamma = 1.45$ for bearing capacity and sliding and $\gamma = 1.35$ for overall stability, and compares these implicit values to the older ASD factors of safety common in bridge design practice. Withiam goes on to recommend that, for different ASD factors of safety or for load factors different than those assumed in the development of Table 18.8, the following relationship be used:

$$\phi_{\text{modified}} = \phi_{\text{tabulated}} \left(\frac{FS_{\text{tabulated}}}{FS_{\text{desired}}} \right) \left(\frac{\overline{\gamma}_{\text{tabulated}}}{\overline{\gamma}_{\text{actual}}} \right) \tag{18.33}$$

Note that, this relationship is inherently an ASD calibration approach, and, as Withiam notes, may be inconsistent with the goal of achieving equal reliabilities for different limiting states. A modified resistance factor is more appropriately calculated using a

Table 18.8 Resistance factors for geotechnical strength limiting state for shallow foundations (Barker *et al*. 1991; Withiam *et al*. 1997)

		Method and soil condition	Resistance factor, ϕ	FS_{LRFD}	FS_{ASD}
Bearing capacity	Sand	Semi-empirical using SPT data	0.45	3.2	3.0
		Semi-empirical using CPT data	0.55	2.6	2.5
		Rational method, SPT	0.35	4.1	2.5
		Rational method, CPT	0.45	3.2	2.5
	Clay	Semi-empirical, CPT data	0.50	2.9	2.5
		Rational method, lab tests	0.60	2.4	2.5
		Rational method using field vane	0.60	2.4	2.5
		Rational method using CPT data	0.50	2.9	2.5
	Rock	Semi-empirical procedure	0.60	2.4	3.0
	Plate load test data		0.55	2.6	3.0
Sliding and passive pressure	Pre-cast	Sand, friction angle from SPT	0.90	1.6	1.5
	Cast-in-place	Sand, friction angle from SPT	0.90	1.6	1.5
		Sand, friction angle from SPT	0.85	1.8	1.5
		Sand, friction angle from SPT	0.85	1.8	1.5
		Cast on 15 cm sand sub-base over clay with Su < 0.5 normal pressure, estimated from:			
		Lab tests (UU triaxial)	0.85	1.7	1.5
		Field vane	0.85	1.7	1.5
		CPT data	0.80	1.8	1.5
		Cast on 15 cm sand sub-base over clay with Su > 0.5 normal pressure	0.85	1.7	1.5
		Passive earth pressure component of sliding resistance	0.50	2.9	1.5
Overall stability	Shallow foundations on or near slope evaluated for stability against deep-seated failure		0.85	1.6	1.5

reliability calibration approach, if sufficient data are available from which to estimate method biases (λ) and coefficients of variation.

18.3.2 Piles and drilled shafts

As for spread footings, piles and shafts are typically designed against six (6) limiting states, two pertaining to service limits, specifically, settlement and lateral deformations, and the remainder to strength limits (Table 18.9). Again, suggested or codified factors of

Table 18.9 Strength and service limit states for design of driven pile foundations. (After Withiam *et al*. 1998)

Performance limit	Strength limit state(s)	Service limit state
Settlement of pile group		x
Lateral displacement of pile/group		x
Bearing resistance of single pile/group	x	
Pile/group punching	x	
Tensile resistance of uplift-loaded piles	x	
Structural capacity of axially/laterally-loaded piles	x	

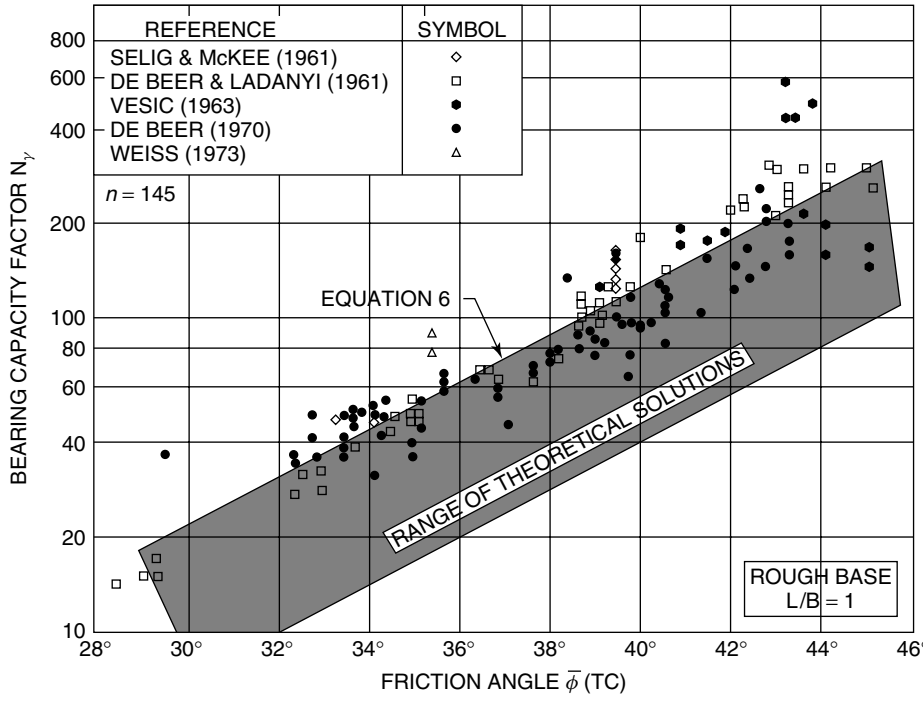


Figure 18.4 Experimental and theoretical bearing capacity results for strip footings (L/B = 6) on sand. (Ingra, T. S. and Baecher, G. B., 1983, 'Uncertainty in bearing capacity of sands,' *Journal of Geotechnical Engineering, ASCE*, Vol. 109, No. 7, pp. 899–914, reproduced by permission of the American Society of Civil Engineers).

safety in ASD typically are expressed as a range, depending on the type and extent of site characterization and the analytical method used in making performance predictions. Higher factors of safety have been recommended for use with empirical measures of soil properties; lower factors for use with measures based on the principles of mechanics or on laboratory measurements.

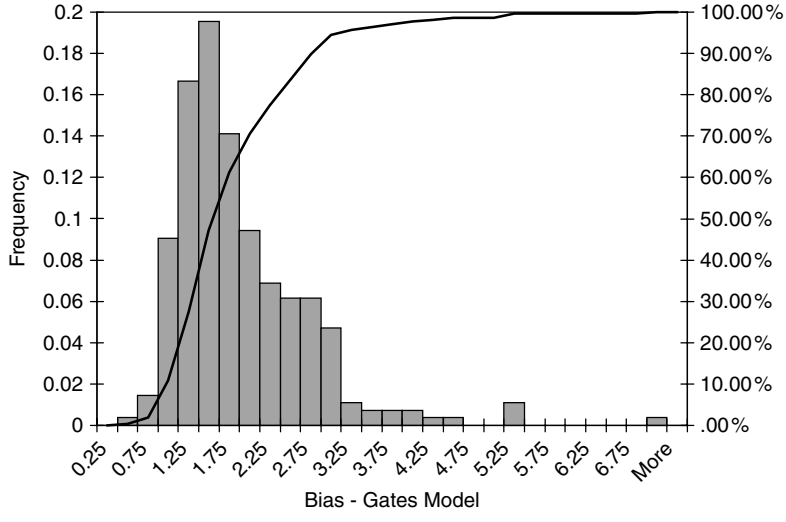


Figure 18.5 Empirical data for the Gates method of predicting the axial capacity of driven piles (Stenersen 2001).

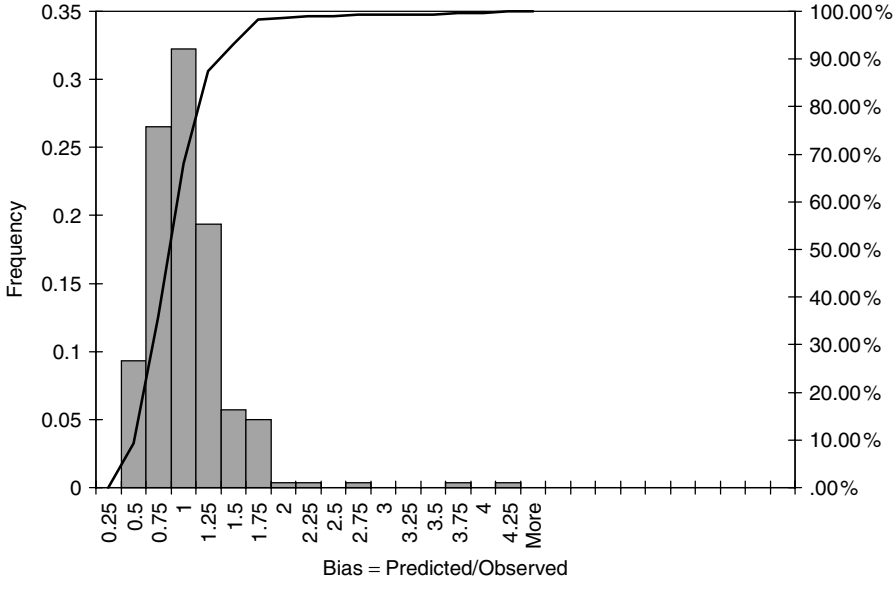


Figure 18.6 Empirical calibration data for the PDLT energy method of predicting the axial capacity of driven piles (Stenersen 2001).

Consider the data of Figure 18.5 from Stenersen (Stenersen 2001), which show empirical comparisons of observed *versus* predicted pile capacity using the Gates modification of the Engineering-News Record formula for predicting the capacity of driven piles (Bowles 1968). The abscissa shows the ratio of observed to predicted pile capacity at failure (method bias), with failure defined by Davisson's criterion. The abscissa shows number of observations. As can be seen, the data are more or less logNormally distributed, with a mean bias about 1.8 and a coefficient of variation about 0.48. The Gates formula is strongly negatively biased, on overage under-predicting the actual resistance of a single pile by almost a factor of 2. In contrast, the data of Figure 18.6 show empirical comparisons of observed *versus* predicted pile capacity using the PDLT Energy method. Here, the mean bias is 0.92, and the coefficient of variation of the bias is 0.46. The PDLT Energy method is close to unbiased, although it also has a large coefficient of variation. Note that since the coefficient of variation is the ratio of standard deviation to mean, the standard deviation of the bias for the Gates method is about half that for the PDLT Energy method.

Table 18.10 shows recommended resistance factors based on AASHTO (1997) and Barker *et al.* (1991) for a variety of methods for predicting ultimate axial geotechnical

Table 18.10 Resistance factors for geotechnical strength limit state for axially loaded piles based on FOSM analysis (modified after AASHTO, 1998)

Method/Soil/Condition		Resistance factor, ϕ_{FOSM}
Ultimate bearing resistance of single piles	Skin friction, clay	0.70
	α-method	
	β-method	0.50
	λ-method	0.55
	End bearing, clay and rock	
	Clay	0.70
	Rock	0.50
	Skin friction and end bearing, sand	
	SPT-method	0.45
	CPT-method	0.55
	Wave equation analysis with,	
	Assumed driving resistance	0.65
	Load test	0.80
Block failure	Clay	0.65
Uplift resistance of single piles	α-method	0.60
	β-method	0.40
	λ-method	0.45
	SPT-method	0.35
	CPT-method	0.45
	Load Test	0.80
Group uplift resistance	Sand	0.55
	Clay	0.55
Stress wave measurements on 10% to 70% of piles, capacity verified by simplified methods, e.g. the pile driving analyzer		1.00

Table 18.11 Resistance factors for geotechnical strength limit state for axially loaded piles based on FORM analysis (Paikowsky *et al*. 2002)

Pile type	Soil type	Design method	Resistance factor ϕ		ϕ/λ	
			Redundant	Non-redundant	Redundant	Non-redundant
Concrete Pile	Mixed	SPT97 mob	0.70	0.50	0.40	0.29
	Clay	α-API	0.50	0.40	0.67	0.55
		λ-Method			0.63	0.55
	Sand	β-Method			0.46	0.34
		SPT97 mob			0.42	0.31
	Mixed	FHWA CPT			0.60	0.48
		β-Method/Thurman	0.40	0.30	0.51	0.39
		αTomlinson/Nordlund/Thurman			0.41	0.30
	Sand	Nordlund			0.42	0.31
	Clay	α-Tomlinson	0.35	0.25	0.41	0.30
	Mixed	α-API/Nordlund/Thurman			0.41	0.30
	Sand	Meyerhof	0.20	0.15	0.32	0.22
Pipe Pile	Sand	SPT97 mob,	0.55	0.45	0.38	0.28
		Nordlund			0.38	0.27
	Mixed	SPT 97 mob	0.40	0.30	0.51	0.40
		α-API/Nordlund/Thurman	0.35	0.25	0.44	0.31
	Sand	β-Method			0.31	0.21
	Clay	α-API	0.30	0.20	0.36	0.26
	Sand	Meyerhof			0.33	0.23
	Mixed	αTomlinson/Nordlund/Thurman	0.25	0.15	0.32	0.23
		β-Method/Thurman			0.41	0.30
	Clay	α-Tomlinson			0.40	0.29
		λ-Method			0.36	0.25
H Piles	Mixed	SPT 97 mob	0.55	0.45	0.45	0.33
	Sand	SPT 97 mob			0.46	0.35
		Nordlund	0.45	0.35	0.49	0.37
		Meyerhof			0.51	0.39
	Clay	α-API			0.48	0.37
		α-Tomlinson	0.40	0.30	0.49	0.37
		λ-Method			0.50	0.39
	Mixed	α-API/Nordlund/Thurman	0.35	0.25	0.45	0.34
		αTomlinson/Nordlund/Thurman	0.30		0.51	0.39
	Sand	β-Method			0.39	0.28
	Mixed	β-Method/Thurman	0.20	0.15	0.42	0.31

Non-Redundant = Four or less piles under one pile cap ($\beta = 3.0$ $p_f = 0.1\%$)
Redundant = Five piles or more under one pile cap ($\beta = 2.33$ $p_f = 1.0\%$)
λ = bias = Mean K_{SX} = measured/predicted
ϕ/λ = efficiency factor, evaluating the relative economic performance of each method (the higher the better)
ϕ/λ values relate to the exact calculated ϕ and λ and not to the assigned ϕ values in the table

resistance. These can also be found, with comment, in Withiam *et al*. (1997). The resistance factors from the 1998 code are based on FOSM calibrations using coefficients of variation based on expert opinion and a review of the literature, combined with calibrations back-calculated to earlier ASD codes (Table 18.10). A more recent study by Paikowsky *et al*. (2002) has introduced FORM calibrations to a series of comprehensive databases of pile and drilled shaft load test data. These results are shown in Table 18.11. AASHTO

Table 18.12 Typical load and resistance factors for axially loaded piles in compression on land (modified after Meyerhof 1994)

Code factor	Eurocode 7 (1993)	Canadian Geotechnical Society (1992)	AASHTO (1994)
Dead load	1.1	1.25	1.3
Live load	1.5	1.5	2.17
Resistance (driven piles with load tests)	0.4–0.6	0.5–0.6	0.80
Resistance (driven piles from SPT tests)	–	0.33–0.5	0.45
Resistance (bored piles with load tests)	–	–	0.80
Resistance (bored piles from static analysis)	–	–	0.45–0.65

Table 18.13 Predictions of the dynamic methods for three case history piles (Paikowsky and Stenersen 2001)

		Q_{ult} (kN)		
		Friction piles		End bearing pile
Type of analysis	Method	Newbury TP#2	Newbury TP#2	Choctawhatchee Pier 5
CAPWAP analyses	CAPWAP EOD	418	738	–
	CAPWAP BOR	–	–	2527
	CAPWAP BORL	1112	1228	2598
Simplified methods	EA EOD	792	1228	–
	EA BOR	–	–	4484
	EA BORL	1432	4075	5187
Dynamic equations	ENR x/FS = 6, EOD	347	409	–
	ENR w/FS = 6, BOR	–	–	4857
	ENR w/FS = 6, BORL	1041	2527	5480
	Gates EOD	721	783	–
	Gates BOR	–	–	2304
	Gates BORL	1219	1708	2447
	FHWA EOD	1157	1299	–
	FHWA BOR	–	–	4680
	FHWA BORL	2251	3354	4982
Static load test results		$Q_{ult} = 658$ kN	$Q_{ult} = 872$ kN	$Q_{ult} = 5560$ kN

1994 load and resistance factor results are compared with Eurocode 7 (Orr and Farrell 1999) and Canadian equivalents (Canadian Geotechnical Society 1992) in Table 18.12.

Example

Paikowsky and Stenersen (2001) have presented a comparison example of three test piles designed using ASD and LRFD, the latter based on calibrations developed in the course of a research project under sponsorship of the National Cooperative Highway Research Program (NCHRP). The first two test piles from Newbury, Massachusetts, are friction piles, driven into a soil formation that is mostly clay with some interbedded sand and silt. The third test pile is an end-bearing pile from the Choctawhatchee River project in Florida, driven into sandy soil. Table 18.13 gives the predicted capacities computed by various techniques.

Table 18.14 summarizes design capacities based on ASD and LRFD with the reported load and resistance factors. Four methods of pile capacity prediction were used, with factors of safety taken from the existing AASHTO code (1998): CAPWAP ($FS = 2.25$), Gates (3.5), FHWA (3.5) and the Energy Approach (2.75). The LRFD capacities were determined using Equation (18.6) with dead and live load factors of 1.25 and 1.75, respectively, and the resistance factors shown in Table 18.11.

Table 18.14 Summary of design capacity comparisons for three case history piles (Paikowsky and Stenersen 2001)

		Q_{ult} (kN)								
		Friction piles						End bearing piles		
Type of		Newbury TP #2			Newbury TP #3			Choctawhatchee Pier 5		
analysis	Method	WSD	AASHTO	Ch 9 ϕ's	WSD	AASHTO	Ch 9 ϕ's	WSD	AASHTO	Ch 9 ϕ's
Wave matching	CAPWAP EOD	187		178	329		320	–		–
	CAPWAP BOR	–	667	–	–	943	–	1121	1246	1290
	CAPWAP BORL	489		569	543		623	1157		1326
Simplified methods	EA EOD	285		320	445		489	–		–
	EA BOR	–	605	–	–	854	–	1628	1121	1432
	EA BORL	525		463	1486		1308	1886		1664
Dynamic equations	ENR EOD	98		53	116		62	–		–
	ENR BOR	–		–	–		–	1388		730
	ENR BORL	294		160	721		383	1566		818
	Gates EOD	205	534	383	222	756	418	–	996	–
	Gates BOR	–		–	–		–	658		1219
	Gates BORL	347		649	489		907	703		1299
	FHWA EOD	329		302	374		338	–		–
	FHWA BOR	–		–	–		–	1334		1219
	FHWA BORL	641		587	961		872	1423		1299
Static load test results		$Q_{ult} = 658$ kN			$Q_{ult} = 872$ kN			$Q_{ult} = 5560$ kN		

18.4 Concluding Remarks

The geotechnical community in the United States has not yet adopted LRFD, although it is now part of practice in Europe and Canada. Federal agencies such as the Federal Highway Administration have introduced it into the practice of highway abutment foundation and retaining wall design in order to achieve consistency between structural and geotechnical practice. The technique is equally applicable to other areas of geotechnical engineering, such as slope stability evaluation, and Taylor's (1948) book shows that much of the basic intellectual framework for partial safety factors in geotechnical engineering has existed for a long time. There are several reasons that the profession has been slow to adopt LRFD, including reluctance to change current and generally successful methods of analysis and design, poor understanding of the LRFD methodology and underlying analysis, and inadequate calibration against actual data. The authors expect that current research and training programs will address these problems. In the mean time, LRFD remains one of the most powerful methods for implementing probabilistic methods into design practice. It was developed for situations involving repeated design, such as estimation of pile capacity or design of conventional retaining walls. It is not applicable for special projects in which a full reliability analysis is called for; Chapter 22 gives such an example.

19 Stochastic Finite Elements

Advances in computer science have made numerical methods increasingly important in engineering calculations. Geotechnical engineers have found the finite element method to be easily the most versatile of the advanced numerical methods, but some very widely used computer codes are based on finite difference approximations. In this chapter the term 'finite element method' is used to include both finite element and finite difference methods unless there is particular reason to distinguish the two.

19.1 Elementary Finite Element Issues

Many books describing the finite element method have appeared since its first introduction in the late 1950s. Applications have spread from the original use in structural analysis for the aircraft industry into almost every branch of engineering science, and each field has spawned its own literature with its own notation and terminology. The book by Zienkiewicz and Taylor (2000a, 2000b, 2000c) provides a comprehensive view from a civil engineering point of view. Hughes (1987) gives a treatment that is mathematically more rigorous. There are a few books that deal specifically with geotechnical engineering issues (Desai and Christian 1977; Zienkiewicz *et al.* 1999). Haldar and Mahadevan (2000) have described the use of stochastic finite element methods for structural reliability applications. Since the details of the finite element method have been widely disseminated in books like these, there is little point in repeating them here. We shall proceed with the assumption that the reader is familiar with the basic finite element methodology or can readily look it up.

19.1.1 The Basic Finite Element Method

The essence of the finite element method is to convert a problem described by partial differential equations over space and time into one described by algebraic equations. This is done by dividing the space-time continuum into a set of discrete elements and assuming

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

that the unknowns vary in a simple way over each element. Usually, this means that the unknowns vary linearly or quadratically over each element, but more complicated patterns are also used. Because the method originated among structural engineers, structural notation and nomenclature have clung to the method, and some explanation is often necessary when the method is employed in another area.

The simplest form of the finite element method is the application to the static deformation of a linearly elastic body undergoing infinitesimal strains. The body is divided into m elements – usually triangles or quadrilaterals in the plane case. Each element has a number of nodes, and the deformation of the element is determined by the displacements of the nodes. In the basic case the patterns of deformation are chosen so that continuity is maintained across the element interfaces. Displacements are prescribed at some nodes. At the remaining nodes, the displacements must satisfy the matrix equation

$$\mathbf{Ku} = \mathbf{p} \tag{19.1}$$

in which $\mathbf{p}$ is a vector of the n loads, $\mathbf{u}$ is the vector of the corresponding n unknown displacements, and $\mathbf{K}$ is the $n \times n$ stiffness matrix. The parameter n is the sum of the degrees of freedom for all the nodes. The global stiffness matrix $\mathbf{K}$ is assembled by placing in appropriate locations the terms from the stiffness matrices of the individual elements. The individual element stiffnesses are

$$\mathbf{k}_e = \int_V \mathbf{B}^T \mathbf{C} \mathbf{B} dV \tag{19.2}$$

The subscript e indicates that this is the stiffness of a particular element. The matrix $\mathbf{C}$ contains the elastic constants. For example, for a linearly elastic body in plane strain and with shear defined by the engineering shear strain

$$\mathbf{C} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & -\nu & 0 \\ -\nu & 1-\nu & 0 \\ 0 & 0 & \dfrac{1-2\nu}{2} \end{bmatrix} \tag{19.3}$$

E is Young's modulus and ν is Poisson's ratio. The matrix $\mathbf{B}$ relates the strains at any point in the element to the nodal displacements:

$$\boldsymbol{\varepsilon} = \mathbf{B}\mathbf{u}_e \tag{19.4}$$

Once again, the subscript e indicates the nodal displacements for this particular element.

The integrations prescribed by Equation (19.2), the assembly of the global stiffness matrix, solution of the matrix Equation (19.1), and back substitution to obtain the strains and stresses are all done automatically in contemporary computer codes. Several user-friendly computer codes are available for applications in geotechnical engineering.

The meanings of the terms change when the method is applied to a problem other than elastic stress distribution. For example, the most common application to steady-state flow through porous media replaces load with prescribed flow and displacement with pore pressure. Some coupled formulations require that the 'load' and 'displacement' vectors consist of two or more different types of variables. If the problem involves developments

over time as in the cases of transient flow or dynamic excitation, additional terms must be included. In some formulations of the dynamic problem, the vectors and matrices have complex terms. Non-linear material properties usually require incremental or iterative approaches. Nevertheless, despite the large number of possible variations and extensions, the basic method is that described by Equations (19.1)–(19.4).

19.1.2 Stochastic Applications using Deterministic Finite Elements

The simplest way to employ finite element methods in a reliability analysis is to treat the material properties as random variables that do not vary across their region of definition. This approach does not try to use the finite element methodology to represent the correlation structure of the random field of the variables. For example, the means and standard deviations define Young's modulus and Poisson's ratios of particular regions in the finite element analysis and the loads are similarly described. This information can then be used in any of the methods of Chapters 14–17. In effect, the finite element code is being used in the same way that the stability analysis code was used in Chapter 14. This approach is an easy way to get an estimate of the reliability of a system, but at the expense of ignoring the spatial variation of the properties.

Schweiger *et al.* (2001) describe this approach for three applications: a sheet-pile wall, tunnel excavations, and comparison of a single and double shell for tunnel support. In the case of the sheet-pile wall, the analysis assumed that the angle of friction and the reference modulus of elasticity were described by logNormal distributions defined by means and standard deviations. All the other parameters were deterministic. The excavation was simulated in 0.5 m intervals. The finite element program used in the analysis reduces the strength parameters until no equilibrium can be found and uses these values to compute a factor of safety. At each level of excavation, this capability combined with the point-estimate method gave estimates of the reliability. Figure 19.1 is a plot of reliability index versus excavation depth reported by Schweiger *et al*. The analysis also yielded the means and standard deviations of the displacements at each depth of excavation.

The second application, the tunnel excavation, started with twelve random variables defined by their means and standard deviations. The variables were the friction angle ϕ, cohesion c, Young's modulus E, and coefficient of earth pressure at rest K_0 for the soil (four variables), similar parameters for the rock (four variables), parameters defining the pre-relaxation for the top heading and bench (two variables), and the shear modulus for young and old concrete (two variables). The K_0's and the shear modulus of the concrete were assumed to be Normally distributed; all the other variables were logNormal. They performed a sensitivity study using FOSM, identifying seven variables that had significant impact on the results. Then the point-estimate method was used with these seven variables to develop statistics of safety with respect to a limit state function and surface settlement.

The study of the single and double shells used five stochastic variables: the friction angle ϕ, cohesion c, and Young's modulus E of the soft rock, and the normal stiffness EA and uniaxial compressive strength β_R of the shotcrete lining. Parameters ϕ and c had a correlation coefficient of -0.5. The uniaxial compressive strength enters only in the evaluation of the performance function, so four variables were used in the point-estimate method to determine the probability of failure as a function of the thickness of the second shotcrete layer. Figure 19.2 shows the results.

These results demonstrate that the deterministic finite element method – that is, the finite element method formulated with deterministic variables – can be used in conjunction

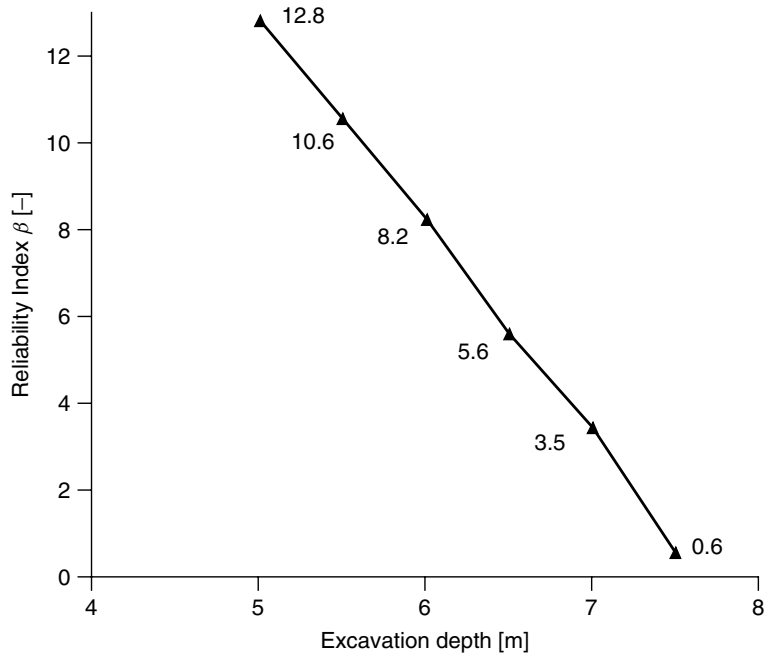


Figure 19.1 Reliability index versus excavation depth for sheet pile study. (Reprinted with permission from Schweiger, H. F., Turner, R. and Pottler, R. (2001). 'Reliability Analysis in Geotechnics with Deterministic Finite Elements.' *International Journal of Geomechanics*, Vol. 1, No. 4, pp. 389–413, © CRC Press, Boca Raton, Florida.)

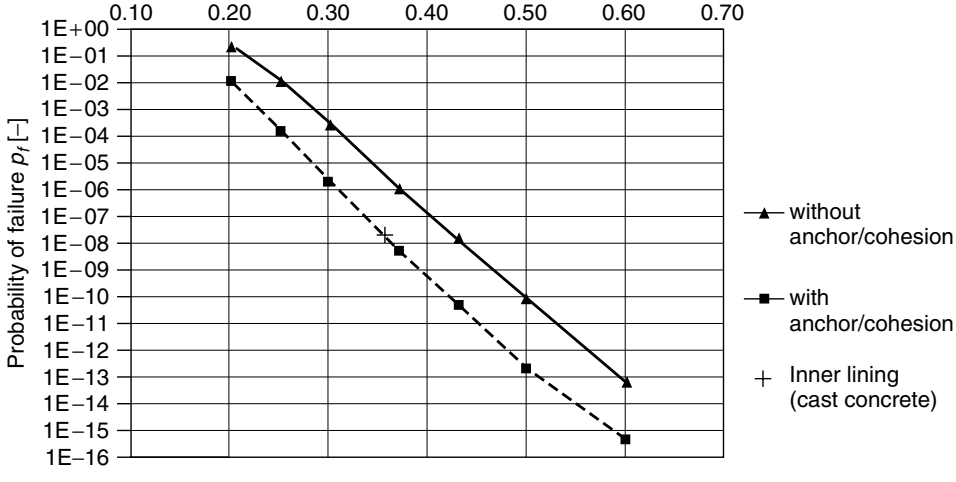


Figure 19.2 Probability of failure versus shotcrete thickness in tunnel section. (Reprinted with permission from Schweiger, H. F., Turner, R. and Pottler, R. (2001). 'Reliability Analysis in Geotechnics with Deterministic Finite Elements.' *International Journal of Geomechanics*, Vol. 1, No. 4, pp. 389–413, © CRC Press, Boca Raton, Florida.)

with means and standard distributions of the input variables to obtain reliability estimates. The results are particularly useful in design and sensitivity studies.

19.2 Correlated Properties

Although applications such as those described by Schweiger *et al*. demonstrate the power of the deterministic finite element method combined with reliability methods, they do not take advantage of the finite element method's power to model stochastic spatial variation. This requires modifying the finite element method to incorporate random fields and, in particular, the spatial correlation of the material properties. Most studies that model random fields are Monte Carlo simulations. That is, each analysis is a representation of the random field.

The simplest approach is simply to generate the material properties for each element randomly and independently. While we have seen this done in some quick studies, we do not recommend it because it ignores the correlation structure of the material properties. The deterministic approach of the previous section implicitly assumes that, within each material area, the properties are perfectly correlated, and this approach assumes they are perfectly uncorrelated. As will be pointed out below, the critical condition often lies between these extremes.

19.2.1 Turning Bands

Based on Matheron's (1973) original insight, Mantoglu and Wilson (1982) described the turning bands method as a way to generate correlated random fields. Matheron originated the name 'turning bands.' It is assumed that the field is defined by some trend and by some correlated, random, Normally distributed departure from that trend. First, the trend is removed; it can be added back in later. This leaves the problem of generating a random field with zero mean and correlated, Normally distributed, random error.

The method is illustrated in Figure 19.3 for the case of a two-dimensional field. The line labeled x is a coordinate axis. Line i is one of a set of *turning bands* lines, oriented by an angle θ_i with respect to the x axis. The parameter ζ_i describes distances along this line, and short increments along the line are $\delta\zeta$. A unit vector oriented along line i is $\mathbf{u}_i$. A one-dimensional, correlated random process is generated along each of the turning bands lines and defined in increments $\delta\zeta$. N is a point in the two-dimensional or three-dimensional region where the random field is to be generated; $\mathbf{x}_N$ is the vector from the origin to point N. The projection of N onto the line i is the point N_i, and the distance from the origin to N_i is the vector dot product $(\mathbf{x}_N \cdot \mathbf{u}_i)$. The value of the one-dimensional random process at N_i is $z_i(\zeta_{Ni}) = z_i(\mathbf{x}_N \cdot \mathbf{u}_i)$. Then, the value of the correlated random field at N is found by summing the contributions of all L lines:

$$z_s(\mathbf{x}_N) = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} z_i(\mathbf{x}_N \cdot \mathbf{u}_i) \tag{19.5}$$

The subscript s indicates that this is a simulated random field.

Mantoglu and Wilson report that the best results are obtained by distributing the L lines at equal angles to cover the region of the random field. They state that, in the

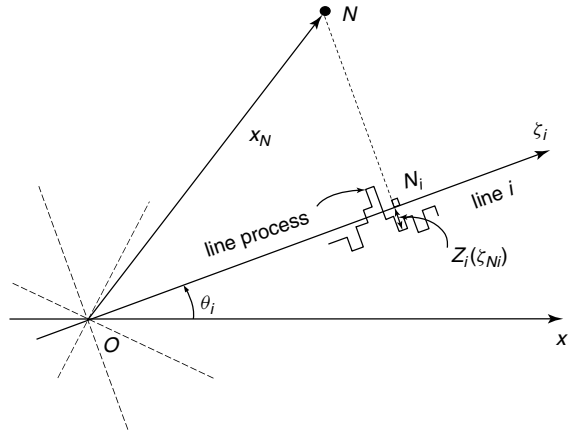


Figure 19.3 The Turning Bands Method. (Mantoglu, A. and Wilson, J. L., 1982, 'The Turning Bands Method for Simulation of Random Fields Using Line Generation by a Spectral Method,' *Water Resources Research*, Vol. 18, No. 5, pp. 1379–1394, © (1982) American Geophysical Union, Reproduced by permission of American Geophysical Union.)

two-dimensional case, 4–16 lines are adequate, more lines yielding finer results. In the three-dimensional case, they cite the finding of Journel and Huijbregts (1978) that accurate results are obtained with 15 lines connecting the opposite midpoints of the edges of a regular icosahedron (or, equivalently, the nodes of a regular dodecahedron) enveloping the region.

To generate the random process along the turning bands lines, it is necessary to develop a radial spectral density function $f(\omega)$ corresponding to the covariance function $C(r)$. Mantoglu and Wilson give the following pair for a simple exponential correlation

$$C(r) = \sigma^2 \exp(-br)$$
$$f(\omega) = \frac{\omega/b}{b[1 + (\omega/b)^2]^{3/2}} \tag{19.6}$$

and for a double exponential

$$C(r) = \sigma^2 \exp(-b^2 r^2)$$
$$f(\omega) = \frac{\omega/b}{2b} \exp\left[-\left(\frac{\omega}{2b}\right)^2\right] \tag{19.7}$$

They further show that the spectral density function of the one-dimensional process, $S_1(\omega)$, is related to the radial spectral density function by

$$S_1(\omega) = \frac{\sigma^2}{2} f(\omega) \tag{19.8}$$

The one-dimensional process along line i is generated by summing M harmonics

$$z_i(\zeta) = 2 \sum_{k=1}^{M} [S_1(\omega_k) \Delta\omega]^{1/2} \cos(\omega_k' \zeta + \phi_k) \tag{19.9}$$

in which ϕ_k are independent random angles uniformly distributed between 0 and 2π, $\omega_k = (k - 1/2)\delta\omega$, and $\omega'_k = \omega_k + \delta\omega$. The range of frequencies is $[-\Omega, +\Omega]$, and $\delta\omega$ is Ω/M. The term $\delta\omega$ is a small random frequency added to prevent periodicities. Mantoglu and Wilson give examples showing that, for a correlation distance b^{-1}, adequate accuracy results from using $\Omega = 40b$ and M = 50 to 100.

It should be noted that the points N at which the random field is simulated are not necessarily the centroids of simple elements. They could be Gauss integration points for more complicated elements. Therefore, this approach could be used in conjunction with sophisticated elements incorporating higher order expansions or coupled behavior.

The turning bands method has found its greatest application in the analysis of water resources and ground water hydrology. Gui *et al.* (2000) used the turning bands technique in successive Monte Carlo simulations of a homogeneous embankment. They assumed that the hydraulic conductivity was logNormally distributed about a constant mean value and that the spatial correlation was anisotropic. Therefore, the random field is defined by the mean $\sigma_{\ln K}$, the standard deviation $\sigma_{\ln K}$, and the spatial covariance function

$$C_P(\mathbf{v}) = \sigma_P^2 \exp\left\{-\left[\left(\frac{v_x}{I_x}\right)^2 + \left(\frac{v_y}{I_y}\right)^2\right]^{1/2}\right\} \tag{19.10}$$

where P is the function $\ln(K)$, I_x and I_y are the correlation lengths of P in the x and y directions, σ_P^2 is the variance of P, $\mathbf{v}$ is the offset vector with components v_x and v_y. For each realization of the random field, 50 simulations were run. The flow regime was evaluated at 50, 100, and 500 days, and the factor of safety and reliability index were computed. In this case, the other material properties were held constant. They concluded that the reliability index was very sensitive to the uncertainty in K. They also found that the factor of safety was adequately modeled as Normally or logNormally distributed when $\sigma_{\ln K} \leq 0.5$, but was not well represented by either distribution when $\sigma_{\ln K} > 0.5$.

19.2.2 Local Average Subdivision

The local average subdivision (LAS) method grew out of Fenton's doctoral work at Princeton University (Fenton 1990; Fenton and Vanmarcke 1991). The method works by successive division of the region over which the random field is to be defined. At each stage of the process the equal subdivisions must (1) have the correct variance according to local averaging theory, (2) be properly correlated with each other, and (3) average to the parent value.

Fenton, Griffiths, and their colleagues (Fenton and Griffiths 1996; Fenton *et al.* 1996; Paice *et al.* 1996; Griffiths and Fenton 1997; Griffiths and Fenton 1998) have used this technique to study a variety of problems, primarily in steady-state seepage through porous media. Griffiths and Fenton (1998) studied the classic problem of a sheet-pile wall penetrating into a single layer of soil. The particular geometry they used consisted of 3.2 m of soil and sheet piling penetrating 1.6 m. They performed 2000 realizations for each two-dimensional case and 1000 for each three-dimensional case. The variation in the exit gradient was considerably higher than that of other quantities of interest. They found that, for all values of the coefficient of variation of the hydraulic conductivity, the maximum values of the mean and standard deviation of the exit gradient corresponded to a correlation distance of 2–4 m. In other words, the most severe impact of randomness in the

hydraulic conductivity occurs when the correlation distance is nearly the same as the gap through which the seepage must flow. The probability of exceeding the estimate of the exit gradient based on mean values of the hydraulic conductivity could be as high as 50% but generally ranged near 40%. An interesting result from a design point of view was that a factor of safety of 5 could correspond to 10% probability of the exit gradient's exceeding unity when the coefficient of variation of the hydraulic conductivity was high. The authors concluded that, in such situations, factors of safety against initiating piping of 10 are not unreasonable.

In an earlier paper, Griffiths and Fenton (1997) studied the flow through a homogeneous dam. They observed that one effect of variability in hydraulic conductivity was to lower the position of the phreatic surface and the overall rate of flow. This seems to happen because the overall rate at which water moves through the dam is affected more by deceleration in areas of low conductivity than by acceleration in areas of high conductivity. They also found that the generally accepted design methods were indeed conservative for these cases and assumptions.

19.2.3 Cholesky Decomposition

Section 17.2.4, points out that a set of correlated Normally distributed random variables can be generated by a combination of the correlation matrix $\mathbf{K}$ and a vector $\mathbf{X}$ of uncorrelated random variables with Normal distribution and zero mean. If $\mathbf{K}$ is decomposed by the Cholesky algorithm:

$$\mathbf{S}^T \cdot \mathbf{S} = \mathbf{K} \tag{19.11}$$

then a vector $\mathbf{Y}$ of Normally distributed random variables with zero mean but with the correlation structure of $\mathbf{K}$ is

$$\mathbf{Y} = \mathbf{S}^T \cdot \mathbf{X} \tag{19.12}$$

This suggests another approach for generating a correlated set of values for finite element simulation. The analyst first generates randomly for all elements a set of Normally distributed but uncorrelated variables with zero mean and the desired coefficient of variation. Then the $\mathbf{K}$ matrix is constructed by using the midpoint distances between the elements to evaluate the off-diagonal terms. The row and column numbers in this matrix are element numbers, not node numbers. Many finite element codes already contain a Cholesky decomposition algorithm, so the decomposition is performed easily. Cholesky decomposition does not increase the bandwidth or the profile of the original matrix. Equation (19.12) yields the desired correlated values, to which the trend can be added.

While this technique is conceptually simple and easily accommodated into existing finite element software, it does have disadvantages. First, the construction of the correlation matrix may over-simplify the correlation structure between two finite areas, especially when they are closely spaced. The method is also more difficult to apply when values are desired not at the centroids of elements but at the Gauss integration points. Finally, Mantoglu and Wilson (1982) show that the computational cost of the algorithm is

$$C = \alpha \times NP^3 + \beta \times NS \times NP \tag{19.13}$$

in which NP is the number of points to be simulated, NS is the number of simulations, and α and β are proportionality parameters. By contrast, the cost of the turning bands method is

$$C = \gamma \times NS \times \sqrt{NP} \tag{19.14}$$

in which γ is another proportionality parameter. Although the exact comparison depends on the values of the parameters α, β, and γ, for large numbers of points the turning bands method is clearly more efficient.

19.3 Explicit Formulation

The approaches described previously in this chapter generate sets of random fields to be used in Monte Carlo simulation. This requires a large number of simulations if accurate results are desired. It can often be difficult to perform sensitivity analyses on the parameters in a Monte Carlo analysis. Therefore, a direct calculation of the effects of correlation becomes attractive. Unfortunately, the computational difficulties have stood in the way of many such efforts.

Baecher and Ingra (1981) describe a direct evaluation of correlation effects in the finite element analysis of settlements. Starting from Equation (19.1), they observe that a first-order estimate of the expected displacements is

$$E[\mathbf{u}] = (E^{-1}[\mathbf{K}])\mathbf{p} \tag{19.15}$$

where $(E^{-1}[\mathbf{K}])$ is the inverse of the matrix of expected stiffnesses. The first-order estimation of the variances and covariances requires partial differentiation of the displacement vector with respect to the element moduli:

$$\mathbf{K}\frac{\partial \mathbf{u}}{\partial \langle M \rangle_i} + \frac{\partial \mathbf{K}}{\partial \langle M \rangle_i}\mathbf{u} = \frac{\partial \mathbf{p}}{\partial \langle M \rangle_i} \tag{19.16}$$

in which $\langle M \rangle_i$ is the average modulus within element i. If the loads are independent of the element properties,

$$\mathbf{K}\frac{\partial \mathbf{u}}{\partial \langle M \rangle_i} = -\frac{\partial \mathbf{K}}{\partial \langle M \rangle_i}\mathbf{u} \tag{19.17}$$

and

$$\frac{\partial \mathbf{u}}{\partial \langle M \rangle_i} = -\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \langle M \rangle_i}\mathbf{K}^{-1}\mathbf{p} \tag{19.18}$$

Solution of Equation (19.18) for a system with n displacement components and m elements gives an $n \times m$ matrix:

$$\begin{bmatrix} \dfrac{\partial u_1}{\partial \langle M \rangle_1} & \dfrac{\partial u_1}{\partial \langle M \rangle_2} & \cdots & \dfrac{\partial u_1}{\partial \langle M \rangle_m} \\ \dfrac{\partial u_2}{\partial \langle M \rangle_1} & \dfrac{\partial u_2}{\partial \langle M \rangle_2} & \cdots & \dfrac{\partial u_2}{\partial \langle M \rangle_m} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial u_n}{\partial \langle M \rangle_1} & \dfrac{\partial u_n}{\partial \langle M \rangle_2} & \cdots & \dfrac{\partial u_n}{\partial \langle M \rangle_m} \end{bmatrix} \tag{19.19}$$

Then the first-order approximation for the covariance between displacements k and l is

$$C[u_k, u_l] \approx \sum_{i=1}^{m}\sum_{j=1}^{m} \left.\frac{\partial u_k}{\partial \langle M\rangle_i}\right|_{E[\langle M\rangle_i]} \left.\frac{\partial u_l}{\partial \langle M\rangle_j}\right|_{E[\langle M\rangle_j]} C[\langle M\rangle_i, \langle M\rangle_j] \tag{19.20}$$

In this expression $C[\langle M\rangle_i, \langle M\rangle_j]$ is the covariance of the material properties between two elements. Baecher and Ingra found that, for elements that are widely spaced with respect to the correlation distance, the covariance at element centroids can be used as an approximation. For more closely spaced elements, two approximations are suggested:

$$\begin{aligned} C[\langle M\rangle_i, \langle M\rangle_j] &\approx V[M]R(\xi) \\ C[\langle M\rangle_i, \langle M\rangle_j] &\approx V[\langle M\rangle]R(\xi) \end{aligned} \tag{19.21}$$

$R(\xi)$ is the distance between centroids. The first applies when the autocorrelation function is approximately linear between points in two elements, and the second when it is not.

Baecher and Ingra observed, "Two-dimensional settlement uncertainties depend on uncertainties in the properties of all the elements through the $\partial u_k/\partial\langle M\rangle_i$ terms. Correlation is introduced among nodal displacements due to the common dependence on the realizations of particular random variables. This correlation would exist even if the element properties themselves were mutually independent."

They introduced a further simplification by assuming that Poisson's ratio is deterministic so that the uncertain Young's modulus can be factored out of the element stiffnesses. Consequently the term at row s and column t of the stiffness matrix, to which w elements contribute, becomes

$$\begin{aligned} k_{s,t} &= \sum_{i=1}^{w} \langle M\rangle_i \alpha_i \beta_i \\ \frac{\partial k_{s,t}}{\partial \langle M\rangle_i} &= \sum_{i=1}^{w} \alpha_i \beta_i \end{aligned} \tag{19.21}$$

In these expressions α_i and β_i are constants that depend on the element geometry and Poisson's ratio. They ran analyses for the effects of a strip surface load of width B over a layer of thickness $4B$. The first-order approximation yielded estimates of the expected values of the displacements under various assumptions regarding the correlation function. The variance of the relative settlement is

$$V[\Delta u_r] = V[u_i - u_j] = V[u_i] + V[u_i] - 2C[u_i, u_j] \tag{19.23}$$

If it is assumed that the absolute differential settlements ($\Delta u_{abs} = |u_i - u_j|$) are Normally distributed, the moments of the absolute relative settlement become

$$E[\Delta u_{abs}] = \left(\frac{2}{\pi V[\Delta u_r]}\right)^{1/2}$$

$$V[\Delta u_{abs}] = \left(1 - \frac{2}{\pi}\right) V[\Delta u_r] \tag{19.24}$$

Figure 19.4 presents results for the effect of correlation distance on the implied differential settlement for several modeling assumptions. The 'one-dimensional' results refer to a simpler, one-dimensional model not discussed here. The two solid lines show that, in the two-dimensional finite element analyses for both exponential and double exponential isotropic correlation, the largest effect occurs when the correlation distance is near the width of the loaded strip. The figure also shows that the simplified one-dimensional model gives results very different from those obtained by the finite element analysis.

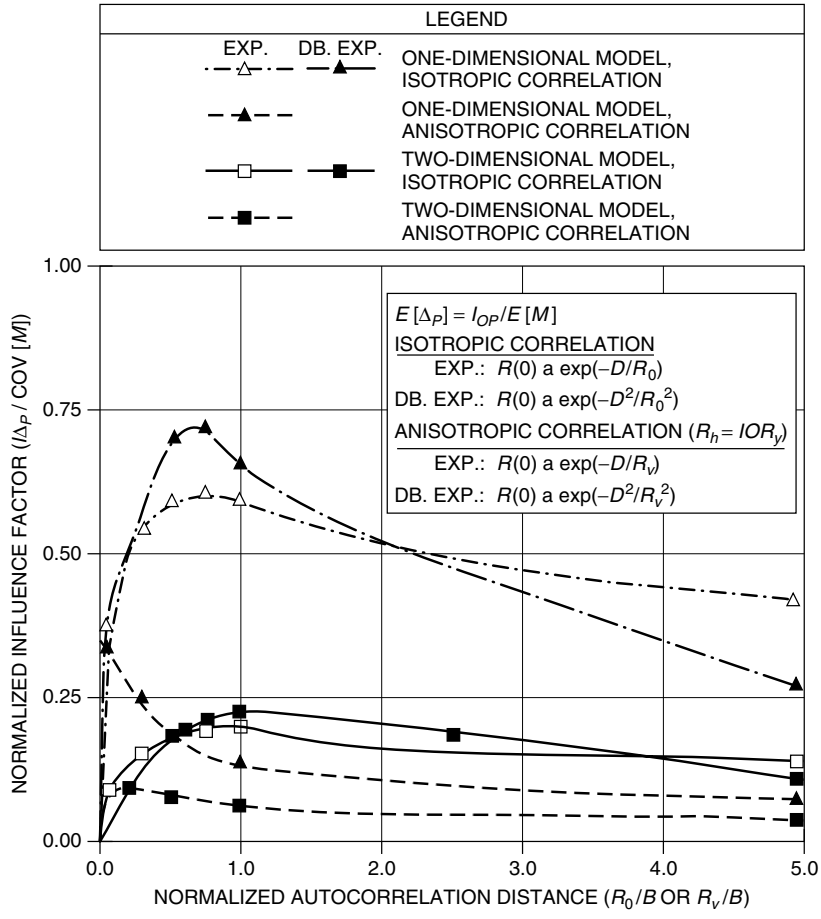


Figure 19.4 Absolute relative displacement under strip load in surface of stochastic layer versus autocorrelation. (Baecher, G. B. and Ingra, T. S., 1981, 'Stochastic FEM in Settlement Predictions,' *Journal of the Geotechnical Engineering Division, ASCE*, Vol. 107, No. GT4, pp. 449–463, reproduced by permission of the American Society of Civil Engineers.)

Because of computational burdens, Baecher and Ingra used a very coarse mesh of 99 nodes and 320 triangular elements. The recent increases in computing power and reductions in cost make it possible to apply the methodology to more complicated problems. However, this has not been done to date.

19.4 Monte Carlo Study of Differential Settlement

Fenton and Griffiths (2002) used Monte Carlo simulation of correlated random fields to study the differential settlement of two strip footings of equal size placed on the surface of an elastic medium for which Young's modulus was logNormally distributed and spatially correlated. They showed that the expected value of the differential settlement (μ_Δ) is zero, but its variance depends on the correlation structure, the size of the footings, and their spacing. They proposed that assuming the differential settlement δ is approximately Normal and that the settlements of the two footings (γ_1 and γ_2) are identically and log Normally distributed leads to

$$\mu_\Delta = 0, \quad \sigma_\Delta^2 = 2(1 - \rho_\delta)\sigma_\delta^2 \tag{19.25}$$

in which σ_γ is the standard deviation of each settlement and ρ_δ is the correlation coefficient between the settlements. When the correlation distance for the modulus ($\theta_{\ln E}$) approaches zero, the settlement variance σ_δ^2 does as well. When it becomes very large, the correlation between the settlements of the two footings approaches one. Thus, Fenton and Griffith observe that expected differential settlements would disappear for both very large and very small values of $\theta_{\ln E}$. This prediction is supported by the results depicted in Figure 19.5. In this figure W_f is the width of each footing, D is the distance between their centers,

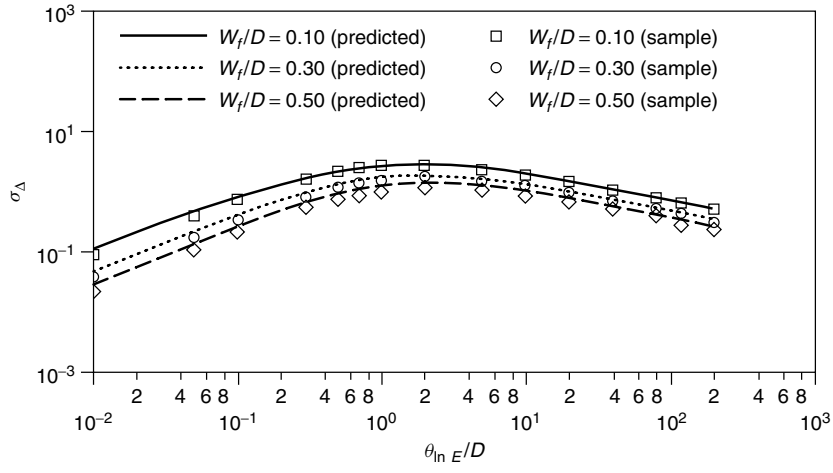


Figure 19.5 Predicted and sample standard deviations of differential settlement for two strip loads on elastic foundation material. (Fenton, G. A. and Griffiths, D. V., 2002, 'Probabilistic Foundation Settlement on Spatially Random Soil,' *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, Vol. 128, No. 5, pp. 381–390, reproduced by permission of the American Society of Civil Engineers).

and $\sigma_E/\mu_E = 1$. The points identified as 'sample' are the results of the Monte Carlo simulation. The maximum standard deviation of the differential settlement occurs when the separation distance has the same order of magnitude as the correlation distance for the modulus, a result consistent with other studies discussed above.

19.5 Summary and Conclusions

The finite element method is especially well suited to analyzing problems in which properties and geometries vary from place to place, but its combination with stochastic formulations presents some computational problems. Three approaches have been employed. First, the finite element method has been used as a deterministic computational tool with the uncertain properties varying uniformly in each region. This essentially assumes perfect correlation from point to point. Nevertheless, results useful for design studies have been obtained. Secondly, the random field of properties can be represented in a series of Monte Carlo simulations. In addition to the need for large computational times to obtain adequate results, this approach requires a method for developing the properties of the random field. The two most popular approaches are the Turning Bands Method (TBM) and Local Average Simulation (LAS). Finally, direct computation of the covariance matrices and solution for the variances by first-order second-moment methods has been described but not widely used.

The geotechnical literature contains relatively few descriptions of stochastic finite element analyses. Most of those deal with flow through porous media. Many of the results published to date tend to confirm the conservatism of conventional deterministic design approaches, in itself a welcome result.

A conventional way to deal with uncertainty deterministically is to bracket the results. When this is applied to problems in which spatial correlation is important, this approach would call for analysis with uncorrelated properties and with perfectly correlated properties. However, studies of exit gradients in flow under sheet piles and of differential settlements under strip loads demonstrate that the critical condition occurs when the correlation distance is approximately the same as the significant dimension of the problem being studied.

Part IV

20 Event Tree Analysis[1]

A variety of methods are available for analyzing engineering risks, but event trees have become the common approach for complex geotechnical systems. There are a number of reasons for this. Event trees provide an intuitive structure within which to organize issues about a particular site or structure. Because event trees typically progress from start to finish in chronological order, they follow a chain of events as it might unfold. They are also versatile in adapting to unique conditions at a particular site. All of this is comforting to practical engineers facing practical problems: Problems are decomposed into tractably small pieces and then brought back together to shed light on the prospect of system failures.

The place where event trees have had the most extensive use in geotechnical practice is the area of dam safety. This has been driven by work at the U.S. Bureau of Reclamation, British Columbia Hydro and Power Authority (BC Hydro), and other owners of large portfolios of hydropower and irrigation dams. This chapter discusses event trees primarily around this dam safety application. The application of event trees to another type of system safety, namely the seismic safety of a petrochemical tank farm, is discussed in Chapter 22.

20.1 Systems Failure

In the context of risk analysis, *system failure* means the cessation of proper functioning of what is expected of a site or structure as a whole. For example, consider a modern earth embankment dam, simplified in Figure 20.1. Since the function of a dam as a whole is to retain water (with some allowable seepage) functional failure occurs when the dam (i.e. the system) ceases to retain water. *System* means the group of interacting elements that form the complex whole of the site or structure, here, the dam, its foundation, and appurtenant mechanical equipment. The ultimate goal of Event Tree Analysis (ETA) is

[1] The authors are grateful to Karl Dise of the U.S. Bureau of Reclamation, Desmond Hartford of BC Hydro, and Andrew Zielinsky of Ontario Power Generation for their thoughts and involvement in the development of the materials in this chapter. Expanded discussion these issues appears in the Canadian Electricity Association's *Guide to Dam Safety Risk Management* (forthcoming).

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

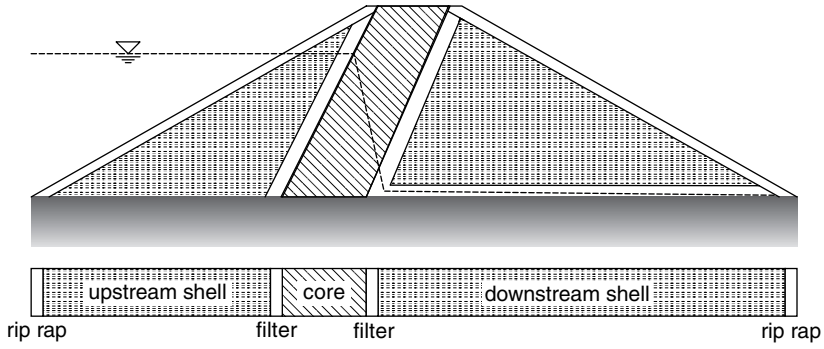


Figure 20.1 Schematic representation of earth dam system, showing typical discrete components for an event tree analysis: riprap, upstream and downstream shells, filters, and core.

to provide insight into the functioning of a system and into the associated uncertainties about the way the system functions. Along the way, this leads to a quantification of the probability that the system (i.e. the dam) may cease to provide its essential function. This is the *probability of system failure*. This probability reflects the aggregate uncertainty in knowledge about the functional performance of the dam and about the environmental loads and service conditions that the dam may face.

As in any modeling activity, assumptions and simplifications are made at each step in conceptualizing the dam as a system and creating an event tree. Different analysts have different ways of defining events, different ways of linking events together, and different ways of estimating parameters and assigning probabilities to events. An event tree reflects a belief structure about a system, about the natural environment within which the system resides, and about the natural and human processes that affect performance.

20.1.1 Event trees

An *event tree* is a graphical representation of the many chains of events that might result from some initiating event, a few of which, should they occur, would lead to system failure. As the number of events increases, the diagram fans out like the branches of a tree, suggesting the name (Figure 20.2).

An event tree begins with some accident-initiating event (Figure 20.3). This might be a flood, an earthquake, human agency, an internal flaw, or something else. The analysis attempts to generate all possible subsequent events, and correspondingly, events that might follow the subsequent events, and so on. The event outcomes are represented as branches issuing from the chance node representing a particular event. This process follows until many chains of events are generated, some of which lead to failure, but most of which do not.

A conditional probability is associated with each event, given all the events preceding in the tree. A joint probability of a chain of events is calculated by multiplying the conditional event probabilities along the chain. Summing the probabilities of all the chains of events that start from a single initiating event and lead to failure yields the total probability of failure due to that one initiating event. Summing the probabilities over all initiating events yields the total probability of system failure. Bury and Kreuzer (1985, 1986) and Kreuzer and Bury (1984) give examples of how event trees can be structured for a gravity dam.

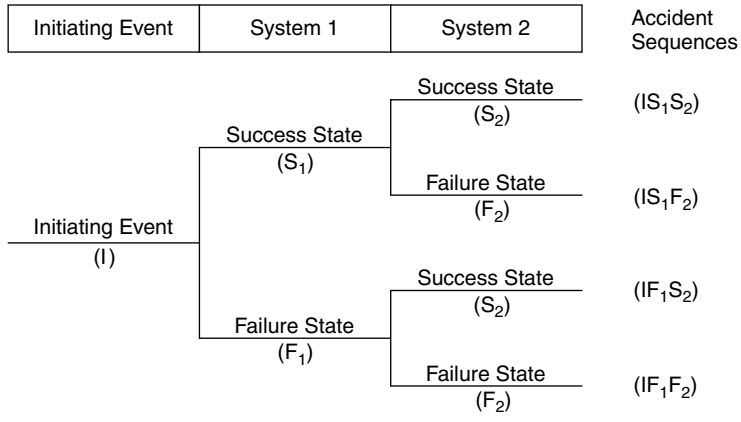


Figure 20.2 Illustration of a simple, generic event tree (US Nuclear Regulatory Commission 1975).

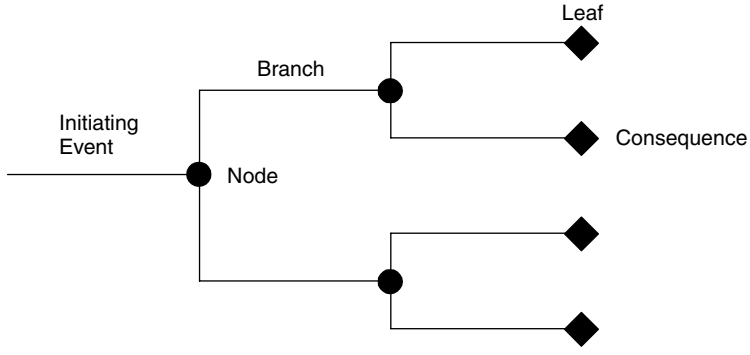


Figure 20.3 Event tree terminology.

In the risk analysis literature for power plants, aircraft, and other mechanical equipment, events within event trees are often limited to dichotomous outcomes (Leveson 1995; McCormick 1981). This is overly restrictive for most geotechnical applications. It is reasonable for events to have many possible discrete outcomes, or even to have continuous outcomes. For computational purposes, however, continuous outcomes are usually replaced by a discrete approximation. The only theoretical requirement on the outcome space for events within an event tree is that the outcomes be mutually exclusive and collectively exhaustive; that is, the outcome of a chance node follows exactly one branch.

The terminal node at the end of any chain of events through an event tree is referred to as a *leaf*. Each leaf in the event tree has associated consequences. The *consequences* are the costs or benefits accruing should the particular chain of events leading to that leaf obtain. In most cases, these consequences themselves may be complex and are analyzed by constructing a consequence tree. A consequence tree structures the set of considerations involved in estimating consequences of failures in the same way that an event tree structures chains of events leading to possible systems failure.

Event trees most likely originated in efforts to enumerate (also to visualize) multivariate outcomes in probability sample spaces. Event trees appear in the work of early writers on probability theory (Daston 1988; Hacking 1975). In recent times, event trees have formed the basis for statistical (Bayesian) decision analysis (Raiffa 1968) with its corresponding decision trees. Some writers have linked the origin of event trees to decision trees (Leveson 1995), but event trees seem to have a much earlier provenance.

20.1.2 Initiating events

Initiating events are the first nodes in an event tree, the events that precede or initiate subsequent chains of the events. For the example of dam safety, typical initiating events might include (1) storms leading to large reservoir inflows, (2) earthquakes creating ground shaking, (3) design or construction 'flaws,' (4) equipment failures, for example of spillway gates, or (5) human agency. Usually, initiating events and their consequences are considered in isolation from one another, that is, independent, but sometimes this is not a good assumption (Baecher 2002).

Initiating events external to the system, such as extreme storms or earthquakes, are usually treated as naturally varying phenomena, that is, as aleatory variables that occur randomly in time (Chapter 2). Thus, they might be modeled by a stochastic point process such as the Poisson or Negative Binomial. These initiating events are brought into an event tree much as they might be brought into a fragility curve. They are loadings on the system, while the event tree models system response to the loading (Figure 20.4). Presuming initiating events to be random implies annual probabilities of failures, as for example, in flood frequency curves or earthquake recurrence functions. In actuality, the uncertainties concerning initiating events may be caused in large part by limited knowledge, in which case the annualized probabilities of failure are an artifact of the modeling. Event trees for

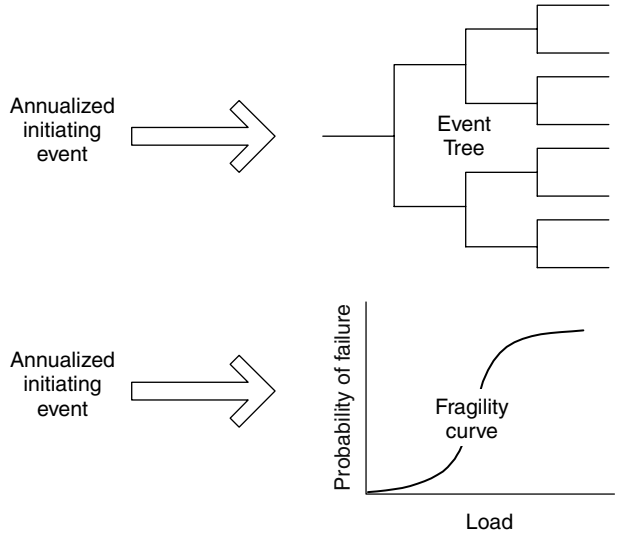


Figure 20.4 Annualized initiating events generate input to a system failure model that can be represented as an event tree, a fragility curve, or some other way.

dam safety typically focus on relatively few initiating events. Internal initiating events have also been called, 'failures under normal operating conditions' (Von Thun 1996).

20.1.3 Event trees and fault trees

Fault trees, which begin with the final system failure and reason backward to identify causes, are the principal way mechanical systems such as power plants or ships are analyzed (Bedford and Cooke 2001; McCormick 1981). Strictly speaking, there is no reason that dams could not be modeled by fault trees, and indeed, Parr and Cullen (1988), Vick and Bromwell (1989) and Bury and Kreuzer (1985; 1986) report risk analyses of dam safety using fault trees (Figure 20.5). McCann (2002) provides a pedagogical discussion of the use of fault trees for dam safety.

Event trees, at least in concept, start with initiating events or causes to the left-hand side of the drawing and progress toward ever more detailed consequences to the right-hand side. The ordering of events in an event tree can be rearranged so long as the relations among conditional probabilities are adjusted (more below), but in concept, the logical progression from cause to effect in a tree is an important if sometimes concealed principle of event tree analysis. In contrast, fault trees start with consequences (i.e. failures), typically at the top of the diagram, and progress backwards toward the ever more detailed causes, typically at the bottom of the diagram. Thus, the logical structure of fault trees is reverse to that of event trees, in that the logic moves from consequence to cause. A system failure mode is considered the 'top event' and a fault tree is developed in branches below this event showing causes. An event tree analyst asks, "what might happen if an initiating event occurs?" A fault trees analyst asks, "how can a particular outcome come to pass?" In concept, fault trees are photographs showing conditions at an instant in time. In contrast, event trees might be thought of as interaction models of the system components.

It is sometimes suggested that event trees and fault trees can be used interchangeably, but this is true only with difficulty. Event trees and fault trees evolved out of different needs and are adapted to different tasks (Bier 1997). A fault tree presents all possible ways of producing a single event, using binary logic from Boolean algebra and resembling a 'root system rising to a main stem' (McCann 2002). It is most suited to well-defined systems, such as mechanical equipment, for which all the components can be enumerated and their interrelationships specified. It is less well suited to poorly defined situations, such as earth structures or the geological subsurface, for which discrete components may not be differentiable or interrelationships may not be unequivocal. It is also poorly suited to non-Boolean conditions in which component performance may have multiple states.

Sometimes fault trees can be used as the reliability analyses that underlie the estimation of probabilities for event branches (Figure 20.6). For example, spillway gate structures are well-modeled using fault trees because they are mechanical systems. The probabilities associated with gate malfunction can be calculated using a fault tree just as they might be using some other structural reliability approach, and then input as the probability associated with a simple event branch, 'spillway gate fails to function,' in an event tree.

20.1.4 Conceptual basis of event trees

The definition of event as used in common speech has to do with outcomes and occurrences: with things that come to pass or come into being, with things that happen:

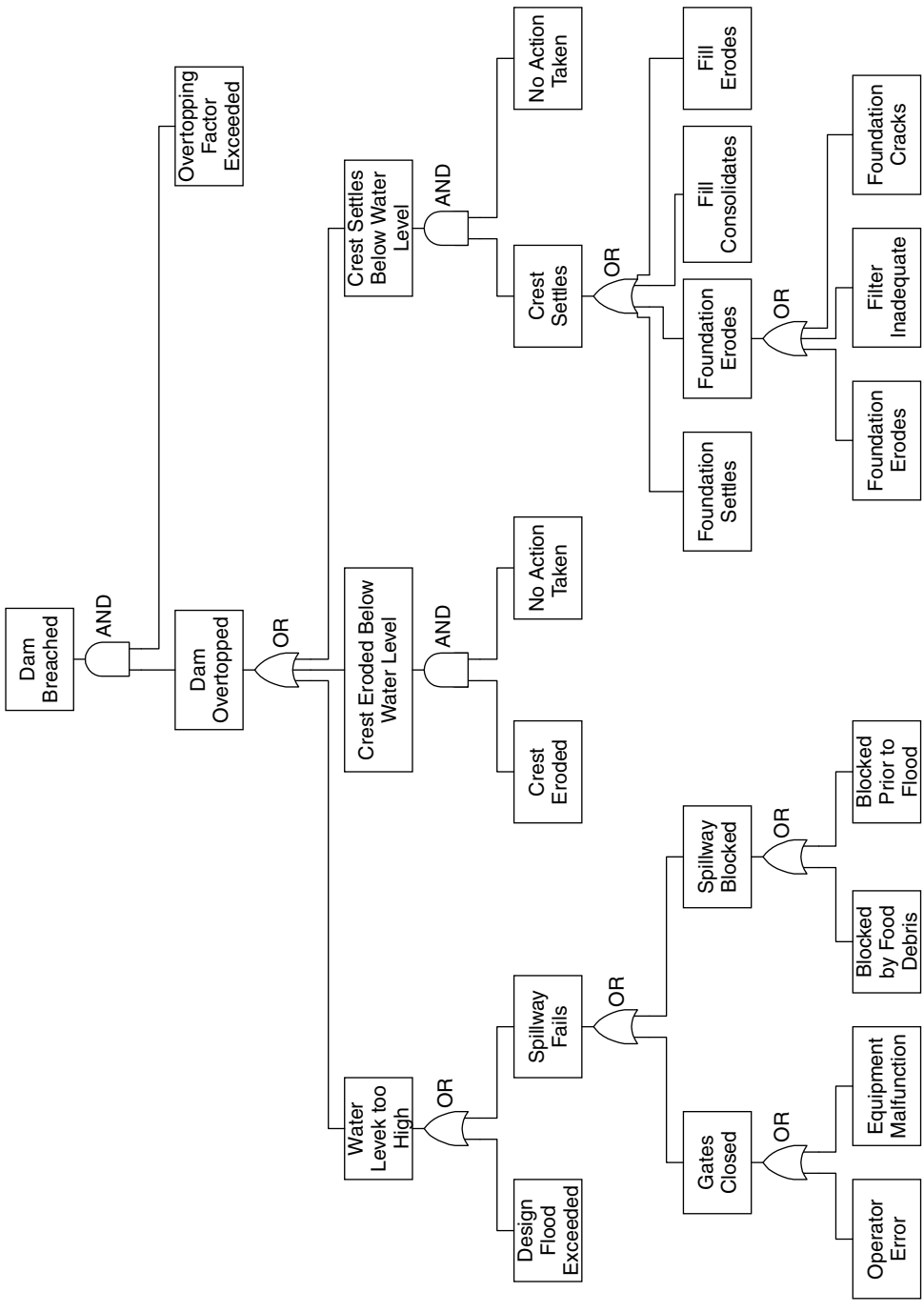


Figure 20.5 Example of a fault tree applied to the problem of dam failure (Parr and Cullen 1988).

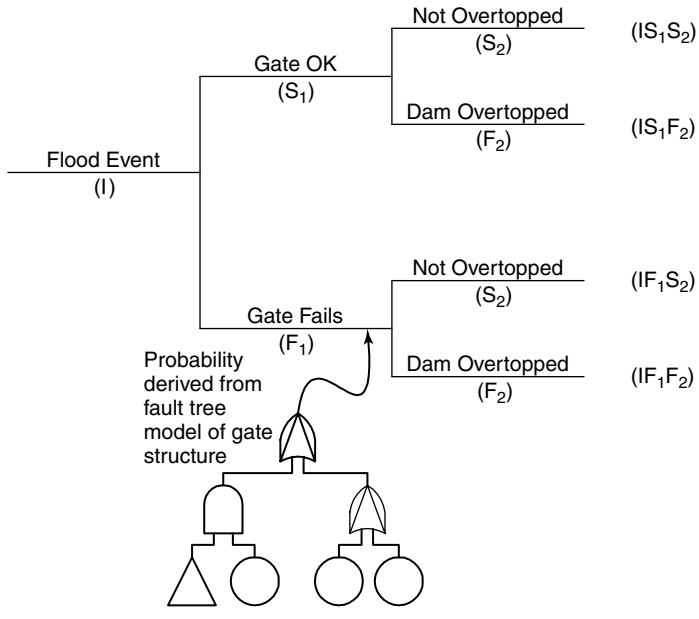


Figure 20.6 Event tree showing event probabilities calculated with associated fault trees for individual events (McCormick 1981).

> Event. 1 a: OUTCOME, something that follows as a result or consequence; b: the final outcome or determination of an action c: a postulated outcome, condition, or eventuality; 2 a: something that happens: OCCURRENCE, a noteworthy happening. (Merriam-Webster 1998).

This definition is less restrictive than, although similar to, those found in textbooks on probability:

> Event. 1 a: OUTCOME, or a collection of outcomes, of a defining experiment; b: a set of sample points in a sample space (Davenport 1970).

In probability theory, the notion of an *event* has to do with random experiments and sample spaces. This has led some people to think of event trees as the graphical expressions of sample spaces. As such, causality, time ordering, and other logical structures become moot. Events can be arbitrarily rearranged as long as conditional probabilities are adjusted accordingly. On the other hand, it may be helpful from an engineering view to retain the chronological ordering or the logical structure of the set of events within an event tree. Thus, different people may view the purpose of event trees differently.

What does an event tree intend to be? The most common concepts are: (1) a model of a physical system (e.g. a model of a particular dam); (2) a statement about joint probabilities of random variables; and (3) an accounting scheme for information and beliefs. These distinctions have to do with the way uncertainties are divided between natural variations and limited knowledge and which uncertainties are included in the tree. Some workers limit the events represented as nodes in the event tree to only those having to do with

aleatory uncertainty: things that happen in space or time. In this approach, epistemic uncertainties are represented in event tree as 'states of nature.' A *state of nature* is a fixed condition of the world, the value of which may be unknown. For example, a 'constant' in the sense of an engineering model is a fixed parameter, the value of which is known imprecisely.[2] As in all aspects of geotechnical risk and reliability, it is essential to draw a distinction in event trees between aleatory and epistemic uncertainties.[3]

If an event tree is viewed as a *model of a physical system*, then only events happening in time or space, and happening to or within the system, should be represented as nodes in the tree. States of nature that describe the external environment or parameters of engineering models should enter the tree only through reliability models used to assign branch probabilities. To the extent possible, events should be arranged in causal order.

A stricter interpretation of event-trees-as-systems-models is that only nodes that have to do with system states should be included in the tree. Using this interpretation, a node such as 'liquefaction slumping of crest greater than 3 m' would be included in the event tree, because it describes a physical state of the dam. An event such as, 'heroic effort made to protect eroding toe' would not be included because, even though it is an event that occurs in time, it is not the description of a physical state of the dam.

In contrast, existing but unknown conditions may be important uncertainties determining probabilities associated with event outcomes, but these uncertainties would belong in the reliability analysis leading to branch probabilities, not as event nodes in the tree. These conditions would be viewed as states of nature. Thus, for example, the suspected but uncertain existence of a low density fill in a dam foundation is important to the probability that the fill liquefies under seismic ground shaking, but it would not be an event node itself in the tree, although it would influence probabilities associated with events in the tree. Approaching an event tree as a model of a physical system greatly simplifies the tree structure, because condition variables that might appear as nodes are now subsumed within the reliability models leading to branch probabilities.

If an event tree is viewed as a *statement about the possible joint outcomes of a set of random variables*, then either (1) both events and condition variables co-exist in the tree, presuming a degree-of-belief view of probability is adopted, or (2) only events are represented, presuming a frequentist view is adopted. In either case, the ordering of the events and variables can be arbitrarily rearranged, and the tree is used mostly as a calculation engine.

Figure 20.7 shows an event tree for flood discharge and stage. The probabilities associated with the first branching are the marginal probabilities of discharge taken from the flood frequency curve. The probabilities associated with the second branching are the conditional probabilities of stage for known discharge. Reversing the order, putting stage as the first event and discharge as the second, changes only the branch probabilities. The probabilities at the first branching are the marginal probabilities of stage, while the

[2] The present usage is more restrictive than that common in the decision analysis literature (Raiffa 1968), where the meaning of states of nature also includes what are here called events.

[3] It is interesting to speculate whether these two analysts would end up with the same estimate final of risk. On the one hand, the total uncertainty should be unchanged by the assumption that one part thereof is epistemic and the other aleatory. On the other hand, we use the separation between aleatory and epistemic uncertainties to exploit differences in the power that statistical and probabilistic models bring to the analysis. In a specific case, there may well be an optimum way of dividing the two types of uncertainty, although it would be difficult to prove.

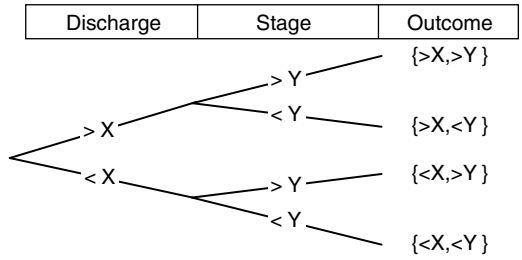


Figure 20.7 Simple event tree for discharge and stage of a river reach.

probabilities at the second branching are the conditional probabilities of discharge for known stage. The joint probabilities of the right-hand side outcomes are the same in both cases. The event tree is simply a convenient way of consistently representing the relationships among events and conditional probabilities.

Viewing event trees as statements of joint probability presumes no causality from parent node to child, and no temporal ordering. Events can logically appear in the tree in arbitrary chronology. Causality is irrelevant.

If an event tree is viewed as a *logical statement about information and beliefs*, then again both events and condition variables co-exist in the tree. This is the common approach in most geotechnical studies. The uncertainties associated with events and condition variables can be convolved, or they may be separated into two distinct parts: an event tree to summarize occurrences in time or space, along with a *logic tree* to summarize the state of knowledge about the dam and its environment. An event tree is thus a representation of what we know, not a model of a particular dam (Figure 20.8).

Evolving practice in seismic hazard, nuclear safety, and some other disciplines is to separate aleatory random variables from epistemic uncertainties. Two trees are then created: an event tree and a logic tree. The event tree contains only aleatory events; the logic tree contains only epistemic uncertainties. The logic tree is a representation of the possible realizations of the states of nature controlling probabilities within the event tree.

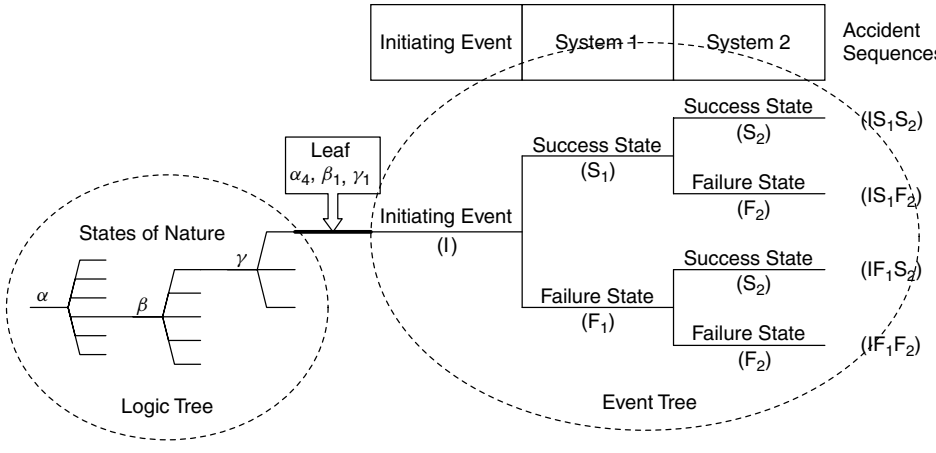


Figure 20.8 Logic tree describing fixed-but-unknown conditions (states) of nature, as conditioning point for event tree of system.

The event tree structure appends to individual 'leaves' of the logic tree, so that all calculations within the event tree are conditioned on the realized states of nature in the logic tree. This simultaneously accounts for dependencies anywhere in the event tree due to the common realization of states of nature in the logic tree. The final combined tree need be no more complex than before, since the number of combinations of all uncertain events, both aleatory and epistemic, remains the same.

20.2 Influence Diagrams

An *influence diagram* graphs relationships among initiating events, states of nature, conditions of the system, and consequences. Creating an influence diagram is, in essence, creating the basic model of the risk analysis from which event trees, reliability calculations, and other models arise. There is not a rigid set of steps or a recipe for creating an influence diagram, but only a structured procedure to help maintain logical consistency.

As in all modeling, the influence diagram should represent the logic of the system, influences upon the system, and uncertainties affecting system performance. To the extent possible, it should do so parsimoniously; it should seek an efficient representation. Obviously, this, too, is an art, requiring interpretation and judgment. The enterprise of creating an influence diagram seeks a relatively simple representation with comparatively few parameters, which, despite its simplicity, closely represents the behavior of the dam.

An influence diagram provides a visual model of the structure of a risk analysis problem. This includes the timing of events, relationships among risky outcomes, and uncertain events, quantities, or variables. The intent of an influence diagram is to help visualize how system components, uncertainties, and outcomes are interrelated; and especially to support the development of a systems risk model of the components, uncertainties, and outcomes. The influence diagram involves no mathematical model; it deals only with relationships among entities.

Two important advantages of influence diagrams over event or decision trees make them useful in the early stages of risk analysis. First, even a complicated problem can sometimes be represented in a simple influence diagram. An influence diagram represents a single uncertain quantity by a single node, and does not lead to the combinatory explosion of branches associated with event trees and fault trees. Secondly, influence diagrams explicitly show interrelationships among events. In an event tree there is usually no immediately obvious way to track the causal dependence of one uncertain quantity on another.

20.2.1 Influence diagrams and event trees

Influence diagrams and event trees are alternate representations for the same systems. It is often convenient first to structure a systems model as an influence diagram, and then to use the insight gained from the influence diagram to structure an event tree for the same system. Each tool provides a means of graphically representing events and the interrelationships among events, although these means differ between influence diagrams and event trees. Each tool can be used to quantify the effects of external risks on process activities, or on engineered or natural systems (Marshall and Oliver 1995).

An influence diagram uses event nodes representing the uncertain conditions surrounding a process, as well as activity nodes representing discrete process steps. Relationships among nodes are shown by directed paths, which sometimes are referred to as arcs or influence lines. The graphical display of events, process steps, and consequences, and their relationships to components of the system being analyzed is used to visualize the impacts of external events or uncertainties. An influence diagram is often used as an exploratory tool leading to the development of more formal event trees, and, as such, system states and uncertainties of various types may be mixed within the same diagram.

The quantification of uncertainty in an influence diagram is accomplished by building a compatible event tree in conjunction with the diagram. Such a tree consists of chance nodes corresponding to the risk and activity nodes in the influence diagram, connected by branches delineating their paths of influence (Clemen and Reilly 1991; Rowe 1977; Stedinger *et al.* 1996).

Relationships among events in event trees follow a hierarchy from independence, through causality, to networked causality, and to temporal orderings. These may be shown in a number of ways, as suggested in Figure 20.9. Many analysts make no distinction between statistical correlations and causal effects, in that the mathematics of probability theory can be made to account for inverted ordering of events, even if the events are causally related.

20.2.2 Elements of influence diagrams

Standard practice in constructing influence diagrams (Marshall and Oliver 1995), similar to standard practice for decision trees (Pratt *et al.* 1995), is to represent node types by circles or ellipses for uncertainties, squares for decisions, and diamonds for outcomes. Typically, influence diagrams in dam safety analyses do not include decision nodes, but that need not be the case. Some authors add a fourth, deterministic node, but this could be viewed as a degenerate case of an uncertain node (Bedford and Cooke 2001).

A directed arc is used to denote dependence or conditionality in influence diagrams. A directed arc joining node A to node B (Figure 20.10) may denote that, (1) event A causes

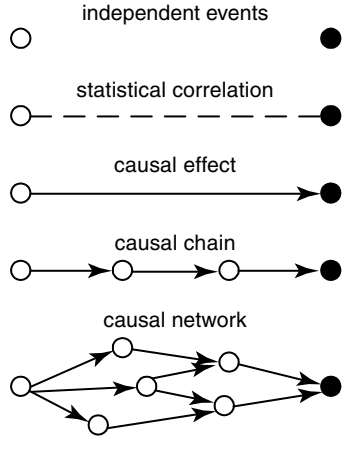


Figure 20.9 Relational structures in event trees.

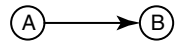


Figure 20.10 Directed arcs in influence diagrams.

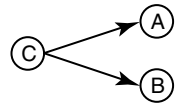


Figure 20.11 Two uncertain variables, A and B, each dependent on a third variable, C.

event B or influences the way B happens, or (2) knowing the value of A changes the probability or probability distribution of B (i.e. A and B are probabilistically dependent, and A or the knowledge of A occurs first in time).

Two events A and B can be probabilistically correlated and yet not be joined by a directed arc, for example, if they share common dependence upon some third variable C (Figure 20.11). Knowing the value of C, the uncertainties A and B are assumed probabilistically independent in this influence diagram (because they are not directly joined by an arc); but if C is unknown, A and B are probabilistically dependent, because each depends on the realization of C.

Two events A and B can also be *statistically correlated* and yet not joined by a directed arc, for example, if estimates of their values are based on the same set of experimental observations. Estimating the Mohr-Coulomb strength parameters c and ϕ from a single set of tests introduces correlation because the parameters are effectively regression coefficients.

20.2.3 Developing an influence diagram

The typical procedure for creating an influence diagram is:

1. Identify the events and uncertain quantities whose outcomes could be important to the safety of the system.
2. For each uncertain quantity, assign a name and a unit of measure; if decision variables are involved, define specific, quantitative alternatives.
3. Identify influences or dependencies among the uncertain quantities, and between each decision and each uncertain quantity; then represent each uncertain quantity and each decision by a node in a graph, aligning them in order of occurrence, if time is relevant.
4. Assign directed arcs among uncertainty nodes and between decision and uncertainty nodes, with the direction of the arcs indicating the direction of presumed influence.
5. In constructing the influence diagram, there should be no closed loops or cycles among the nodes. That is, there should be no connected path along the directed arcs that leads back to an earlier node.

We earlier introduced the system states and uncertainties associated with a simple floodway levee. These can be combined in an influence diagram to begin to show the interrelationship among the various elements and considerations in beginning a risk analysis of the levee. A first attempt at such an influence diagram is shown in Figure 20.12. This diagram consists both of system states and uncertainties affecting those states.

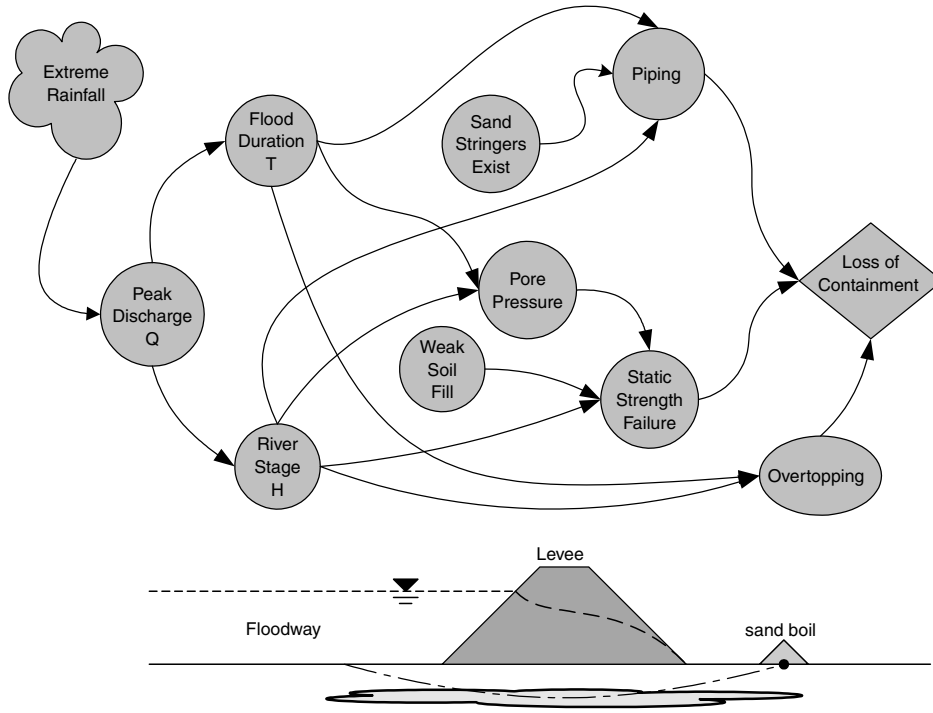


Figure 20.12 Influence diagram for levee failure. (Canadian Electricity Association, *Guide to Dam Safety Risk Management*, forthcoming.)

While it is not common for an influence diagram in dam safety studies to include decision nodes, it is also not unknown. Figure 20.13 shows the same influence diagram, now with the addition of an active decision node representing the option of releasing water upstream of the levee reach into a flood by-pass. The directed arc between flood duration and overtopping is neglected in this diagram as a simplification. If the operators of the floodway recognized that a potentially dangerous flood was heading into the levee reach, they now might make an active decision to intervene. As with the other nodes in the diagram, the behavior of the decision node may itself be affected by many variables and uncertainties, for example, ones associated with the availability of timely information and with human factors.

20.2.4 Markovian behavior in influence diagrams

The Markovian property, so prized for its practical usefulness in belief networks, (Pearl 1988) sometimes runs counter to good practice in influence diagrams involving decision nodes (belief networks typically do not include decision nodes). Consider a site characterization situation involving the potential for two steps of data gathering (akin to a 'second-opinion' decision problem in the clinical literature). Presume that the performance of the levee with respect to under-seepage depends on the possible, but uncertain, existence of sand lenses in the substrata. Geological information may lead to a prior probability of lenses existing (Figure 20.14), and based on this information, a decision can be made. The decision might be to choose a remedial option or to drill more borings to try to gain more

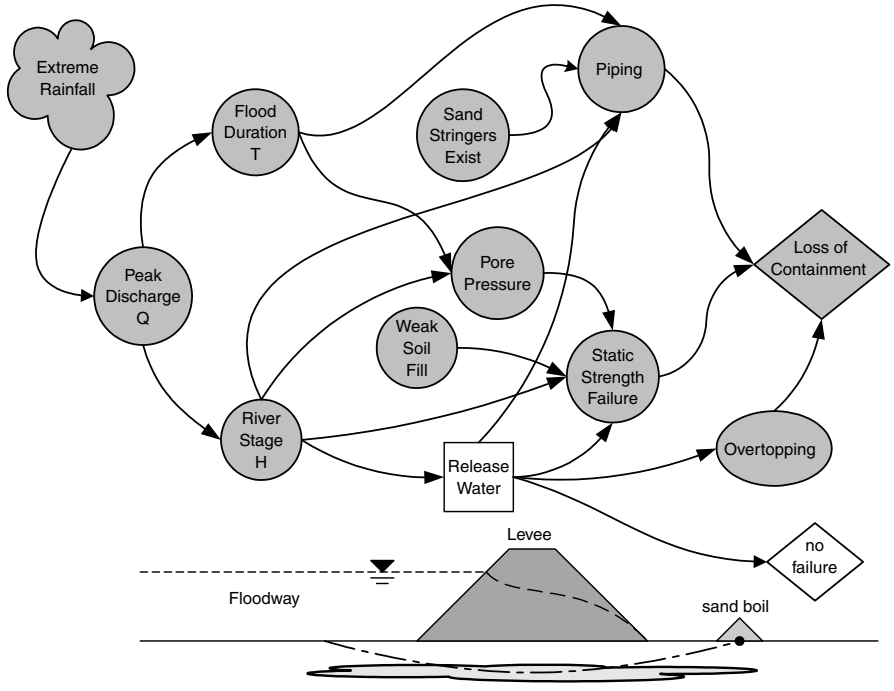


Figure 20.13 Influence diagram for levee failure, including active decision node. (Canadian Electricity Association, *Guide to Dam Safety Risk Management*, forthcoming.)

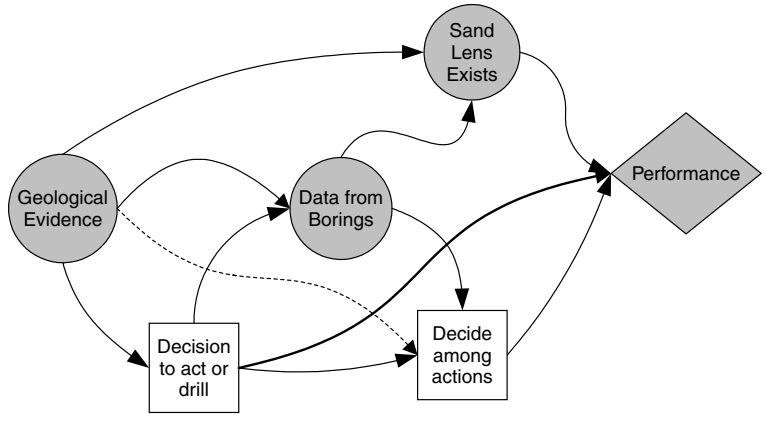


Figure 20.14 Influence diagram for the two-staged exploration decision. (Canadian Electricity Association, *Guide to Dam Safety Risk Management*, forthcoming.)

information about whether or not sand lenses exist. Once this additional information has been gathered, a new (posterior) probability of a lens existing is calculated through Bayes' Theorem, and a decision on remedial actions revisited. After a decision on remedial actions is taken, and a lens turns out either to exist or not to exist, some performance result obtains. Before the fact, of course, this performance result can only be described probabilistically.

The directed arc between the geological information and the second decision, denoted by a dotted line, must be there. The influence diagram cannot describe a meaningful decision problem without this arc. Absent the arc, the geological information known at the time of the first decision would be forgotten by the time of the second decision, and this is not realistic, since the geological information is pertinent to the second decision. A requirement for a proper influence diagram is that information known at an earlier decision is remembered at a later decision. The arc from the 'geological evidence' node to the second decision node is sometimes called a 'no-forgetting arc' (Marshall and Oliver 1995).

The nature of the geological evidence at the LHS initial node leads to a conditional probability of a sand lens existing via Bayes' Theorem,

$$\Pr(lens|geology\ evidence) \propto \Pr(lens)L(geology\ evidence|lens) \tag{20.1}$$

in which the left-hand side is the conditional probability of a lens existing given the geological evidence, and the right-hand side is the product of the marginal probability of a lens existing, by the likelihood (i.e. conditional probability) of the geological evidence given that a lens exists.

One could just as logically reason that the existence or non-existence of a lens influences the probabilities of different sorts of geological evidence, and the corresponding conditional probabilities are found by reversing Equation (20.1). Restructuring the influence diagram in this way (Figure 20.15), however, creates a disallowed cycle within the graph, illustrating that it is comparatively easy to create improper graphs. In a simple influence diagram with only a few nodes, it is relatively easy to detect this sort of inconsistency and reverse the offending arc(s). It is more difficult with large diagrams, and more formal methods are available to detect cycles in such cases (Marshall and Oliver 1995).

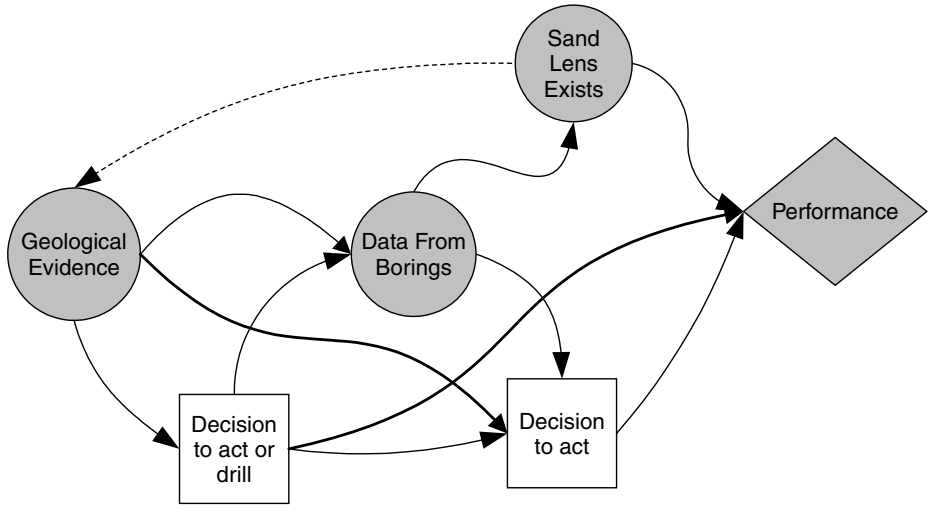


Figure 20.15 Arc reversal in the influence diagram for two-staged exploration. (Canadian Electricity Association, *Guide to Dam Safety Risk Management*, forthcoming.)

20.3 Constructing Event Trees

Consider the homogeneous, compacted soil flood levee shown in Figure 20.16. The purpose of the levee is to retain high water during floods. As a first approximation, this levee might 'fail' – and thus allow flood waters to escape – in one of three ways: (1) high river stage might overtop the levee, causing embankment erosion and consequent loss of containment; (2) the levee itself might fail due inadequate soil strength and high internal pore pressures; or (3) seepage in the soils beneath the levee might lead to piping and a deep sliding failure. How might we build an event tree with which to analyze the risk of failure of this levee?

20.3.1 Considerations in developing the event trees

Some of the considerations in analyzing the risk that the levee will fail are suggested by Table 20.1. First, at least four system states are relevant to a potential failure: the flood height relative to top of the levee, internal pore pressures in the levee and its foundation, the geotechnical strength and stability of the levee proper, and the presence of piping (internal erosion) in the foundation. For this example, piping in the levee itself is ignored. The probabilities associated with each system state depend on various uncertainties. Since some of these uncertainties may be common to different system states, the uncertainties about system states are likely not to be independent (National Research Council 2000). The separation of natural variations from knowledge uncertainties in Table 20.2 is, in fact, a modeling decision that could be made in different ways by different people, or made in different ways by the same person if the purpose of the modeling changes.

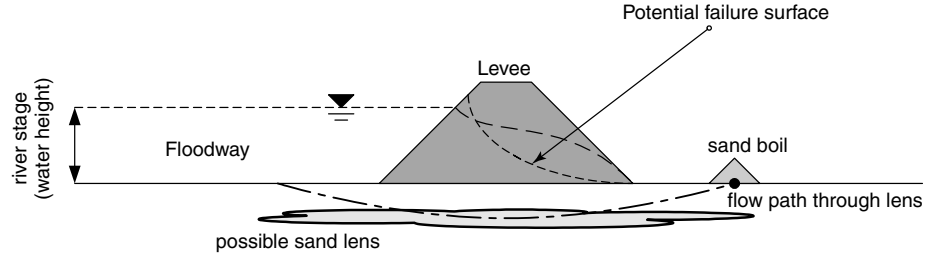


Figure 20.16 Example of levee subject to flood flow. (Canadian Electricity Association, *Guide to Dam Safety Risk Management*, forthcoming.)

Table 20.1 Important system states for levee example

Failure modes	System states	State variable
Overtopping of levee	Overtopping of levee	Relative water height
Structural or geotechnical instability in fill	Internal pore pressure in levee Subgrade water pressure difference Levee strength stability	Pressure on pertinent failure surface Differential water pressure Levee deformations
Erosion of the foundation	Subsurface piping	Rate of seepage in subsurface

Table 20.2 Division of important uncertainties for levee example onto those related to natural variability and those related to knowledge uncertainty

Naturally varying (aleatory) conditions	Knowledge (epistemic) uncertainties
Extreme storm in upstream basin	Average strength of the levee fill
Peak flood discharge (flood-frequency)	Presence of 'flaw' zones in the levee
Peak river stage for a given discharge	Existence of permeable lenses in foundation
Flood duration given peak flood	Critical duration of overtopping to cause failure
Spatial variation of soil properties	

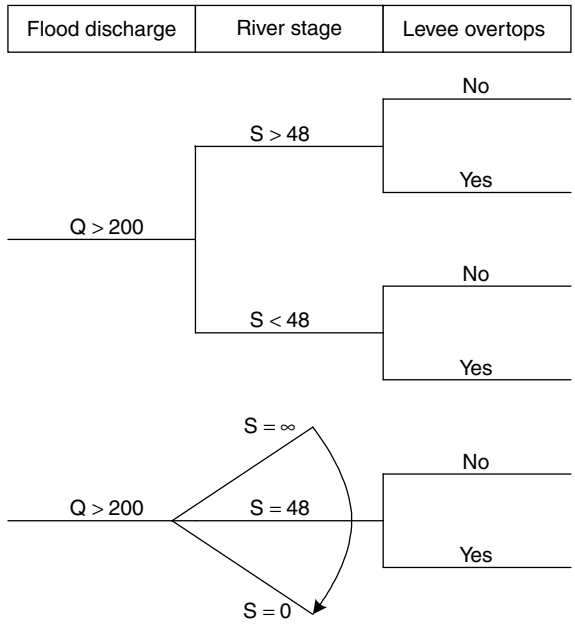


Figure 20.17 Alternative simple event trees for levee overtopping: Top showing dichotomous representation of outcome: bottom showing continuous representation of outcome.

An event tree is built from a set of events, each represented by a node. The branches emanating from an event node represent possible outcomes. Since these are mutually exclusive and collectively exhaustive, the outcome follows exactly one branch. Figure 20.17 shows the initiating event of a flood discharge greater than some value. The subsequent event, 'river stage,' is summarized in the upper tree by a dichotomous outcome. In this case, the corresponding river stage is either greater than some value, or less than that value (by convention, the upper branch at each event usually corresponds to the associated event being 'true'). In the lower tree, the event is summarized by a continuous outcome between lower and upper bounds.

A probability distribution is associated with each event node. When the outcomes are defined as discrete, the probabilities are represented in a probability mass function (pmf). When the outcomes are continuous, the probabilities are represented in a probability density function (pdf). Calculations in the first case are made by multiplying branch

probabilities by their corresponding consequences, and in the latter case by integrating the consequences over the pdf.

An event tree quickly becomes large. Expanding a simple chain of events leading from an initiating event of high rainfall to strength failure of a levee could pass through at least the following five events, each of which might have many possible outcomes: (1) an extreme storm generates rainfall of a given amount; (2) discharge in the river exceeds some value; (3) stage rises above some critical elevation; (4) high water lasts for a critical length of time; and finally (5) the levee fails due to high loads and rising pore pressures. Simplifying to dichotomous events yields the tree of Figure 20.18, which is already large.

A more realistic model might include consideration of whether soft zones (flaws) exist within the embankment, how high the pore pressures rise, whether a levee experiencing structural instability might yet retain the flood flow, and so on. One problem with any of the simple logic tools used to model risks associated with system performance, such as influence diagrams or event trees, is that the tools are static analyses that decompose events into individual realizations. This simplifies the analysis, but may lose important features. Since the tree grows combinatorially with the number of events, care must be taken to keep event chains as short as can be reasonably accommodated.

Consider an issue suggested by Zielinsky (2001) of the height and duration of overtopping. Each is important for failure, and they act jointly. Shallow overtopping, that is, not much water volume passing over the levee, may cause failure, but only if the condition lasts for a long time; if the overtopping is brief, the levee may withstand potential erosion. On the other hand, deep overtopping even if brief may cause failure. Figure 20.19 shows two cases of overtopping. On the left, the overtopping is shallow but lasts for 20 hours; on the right, the overtopping is deep but lasts a much shorter time, perhaps three or four hours. Both are cases of overtopping, but the levee may respond differently to each. A simple influence diagram or event tree may not capture this combined dependence. Events may need to be defined by depth-duration pairs in order to model the behavior of the levee.

For convenience, event trees are usually separated by initiating event (Figure 20.20). Thus, in a dam safety study it is common to see three, four, or even more event trees, associated respectively with extreme storms, earthquakes, flaws in design or construction,

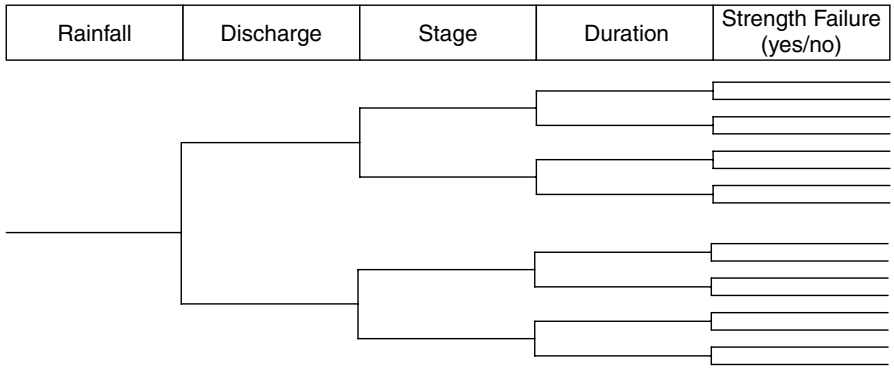


Figure 20.18 Simple event tree for structural/geotechnical strength instability of levee.

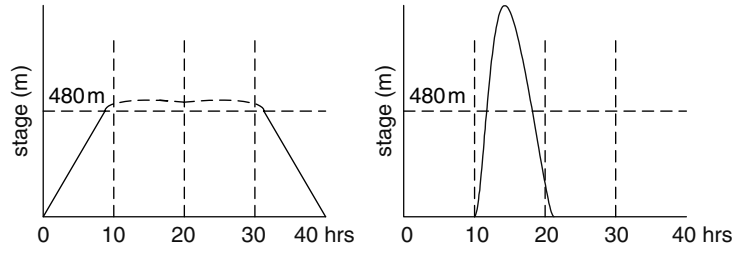


Figure 20.19 Height and duration of overtopping interact to affect levee performance.

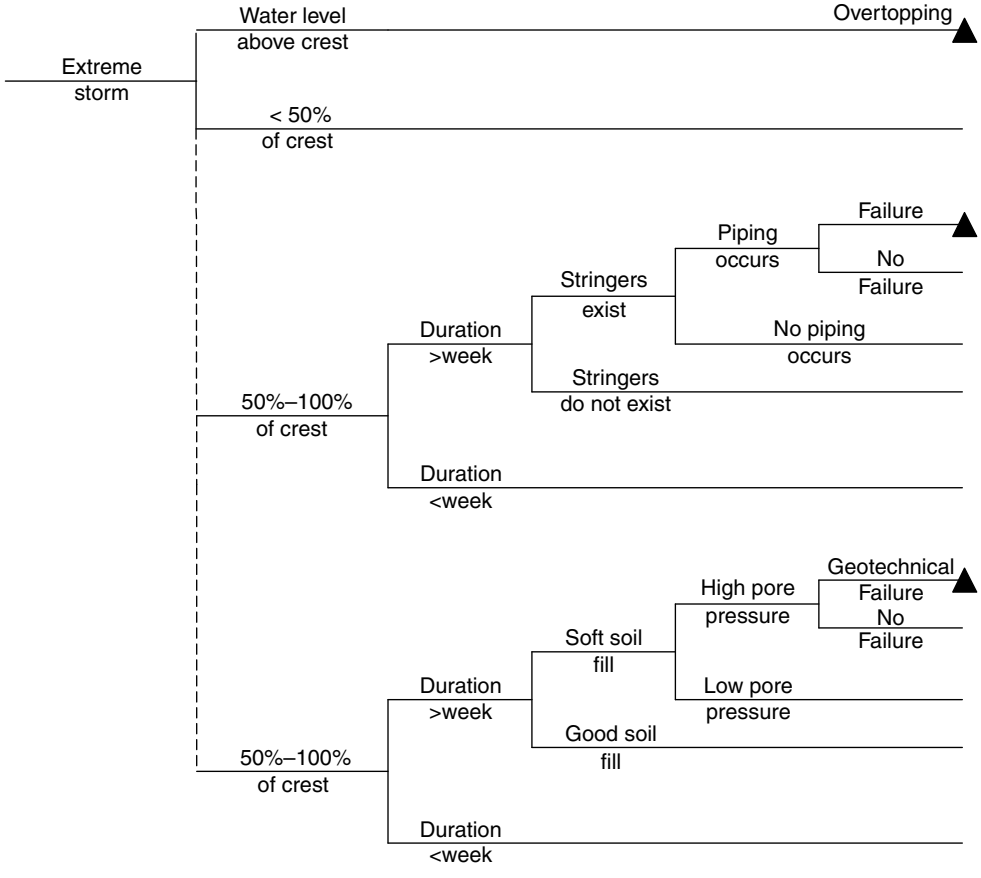


Figure 20.20 Hydrologic, piping, and strength failure parts of the levee failure event trees.

mechanical or electrical dysfunction of an appurtenant structure, and possibly more exotic things. Note that this tree combines states of nature ('soft soil fill exists') with system states of the levee ('high pore pressure in levee'). That is, it does not separate aleatory and epistemic uncertainties into a logic tree and event tree but combines them. The dashed vertical line indicates separate event trees leading to distinct failure modes but arising from the same initiating event. The triangle at a terminating end of a path through the

Table 20.3 A comparison of influence diagrams and decision trees (Marshall and Oliver 1995)

Model features	Influence diagrams	Event or decision trees
	Modeling	
Timing	Shows timing of all decisions and uncertain events.	Shows timing of all decisions and uncertain events.
Conditional independence	Shows dependence among uncertain events and decisions.	Dependence among uncertain events and decisions not shown.
Size	Number of nodes grows linearly with the number of variables.	Number of terminal nodes and paths grows exponentially with the number of variables.
Data	Identifies dependencies of variables without need for data.	Decision, probability and result data shown explicitly.
Variable type	Both continuous and discrete decisions and probabilities.	Both continuous and discrete decisions and probabilities.
Asymmetry	Scenarios with different event sequences not distinguished.	Shows asymmetric structure of problem.
Modeling usefulness	Most useful in initial stages of modeling; captures interaction between decision maker and analyst.	Useful in depicting detailed uncertain event outcomes and decisions, and model solution; difficult to display large problems.
	Solution process	
Bayes' theorem	Indicated by arc reversal, but calculation not shown.	Indicated by node reversal (a separate event tree may be used as an aid for calculation).
Solution method	Reduction by a set of reduction operations possible using advanced methods.	Uses simple rollback algorithm.

tree indicates a 'failure' consequence. Despite the common logical content of influence diagrams and event or decision trees, the practical uses of the two differ, as summarized in Table 20.3.

20.3.2 Consequences

The consequences that result from a dam incident or failure depend on a large number of factors beyond the behavior of the dam itself. For example, consequences depend on safety functions such as barriers, safety systems, operating procedures, operator actions, and so on, and on how they respond to an initiating event or an indication of dysfunction of the dam proper. Safety functions may include systems that respond automatically to an initiating event or to a failure, alarms that alert operators or other cognizant personnel that an event or failure has occurred, predefined operating procedures that follow an alarm, or barriers or other containment facilities intended to limit the effects of an initiating event or failure. Other considerations that influence the consequences resulting from initiating event or failure include time of day, meteorological conditions, downstream warning systems, and emergency response procedures to protect life and property.

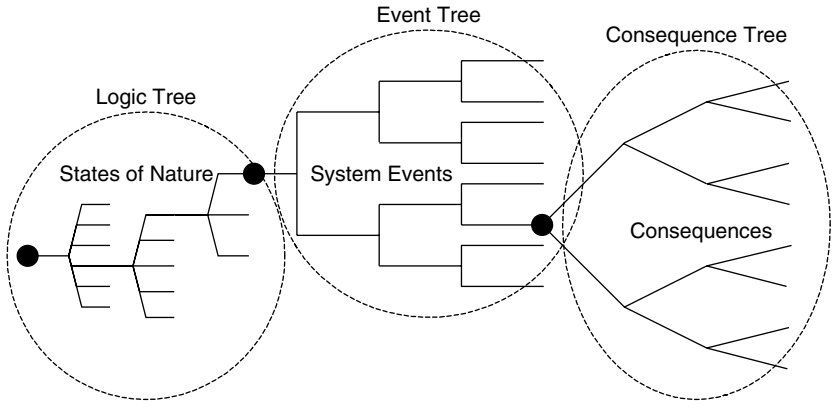


Figure 20.21 Consequence tree summarizing various events that may occur upon a particular outcome (leaf) of the event tree.

Typically, these many considerations are not of the dam system itself or are not part of the uncertainties related to limited knowledge about natural processes or the functioning of the dam. Thus they are separated into a consequence model or, more typically, into a separate consequence tree.

20.3.3 Consequence trees

As in probabilistic risk analyses (PRA) of nuclear power plants (McCormick 1981), a separate event tree is often used to model consequences of a failure; for example to provide an estimate of the consequences conditional on loss of pool or some other damage state of the dam (Figure 20.21). This sub-tree incorporates exposure cases and other downstream activities or events, such as the effectiveness of evacuations or other risk mitigation procedures. Usually, this downstream sub-tree can be treated independently of the event tree for the dam itself, but it is often no less complicated.

Event trees are evaluated by generating a set of probability-consequence pairs for each mutually exclusive end node in the tree. In principle, consequences could be expressed in utilities (Keeney and Raiffa 1976), rather than direct physical units (e.g. dollars or lives lost), but usually they are not. Multiplying probability by consequence to obtain an expected consequence implicitly treats high-probability low-consequence outcomes as equivalent to low-probability high-consequence outcomes, as long as the product is the same. For government projects this is usually taken to be reasonable, given governments' asset positions and the multiplicity of projects (Morgan and Henrion 1990). For a private owner, such risk neutrality may not be reasonable.

20.4 Branch Probabilities

An individual branch probability is typically assigned by one of several procedures: (1) statistical (i.e. empirical) estimates; (2) reliability (i.e. engineering) models; (3) fault tree analysis; or (4) expert opinion. Statistical estimates and reliability models are

discussed here. Fault trees are outside the present scope, but are discussed by Bedford and Cooke (2001). Expert opinion is discussed in Chapter 21.

There is a trade-off in assigning probabilities to event branches. On the one hand, an event tree can be large and finely detailed with many branches. The probabilities assigned to these branches might be estimated from simple data analysis or models. On the other hand, the tree can be compact, with relatively few event nodes and branches, but with probabilities estimated from complexly detailed sub-models. The choice along this spectrum is part of the craft of risk assessment.

20.4.1 Statistical estimates

Among the more common places that statistical estimates are used is in estimating probabilities of initiating events such as extreme floods or earthquakes. The occurrence of initiating events is typically modeled as a stationary Poisson process, with a single parameter, λ, describing the number of occurrences per unit time, usually a year. This model is

$$f_n(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{20.2}$$

in which n is the number of occurrences, and λ is the rate per unit time. Stationary means that λ is constant over time. The number of occurrences in time interval t is

$$f_n(n|\lambda, t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \tag{20.3}$$

for which the mean of n is $E[n] = \lambda t$ and the variance, $Var[n] = (\lambda t)^2$. Other point process models could also be used to represent the occurrence of initiating events, for example, if one wanted to introduce probabilistic dependence among the event occurrences (Lewis 1972).

The most important assumption of the Poisson distribution is that events occur independently. Thus, if a large storm occurs this year, the probability of a similar storm occurring next year is unaffected: In principle, one could have two $Pr = 0.001$ events (1000-year storms) back-to-back. The exception to independence in the Poisson model is that, for most natural hazards, the length of historical record is short. New information, such as the occurrence of an extreme storm or a large earthquake, affects the statistical estimate of the parameter λ. Thus, the occurrence of an extreme event changes the parameter of the model of the process itself, and consequently influences the estimated probabilities of future events. For example, in constructing Oroville Dam, two presumably 250-year ($p = 0.004$) storms and one 1000-year ($p = 0.001$) storm occurred during the few years of construction (Gilbert 2002). With the occurrence of each of these storms the rate parameter λ for the site changed, as did subsequent estimates of the annual exceedance curve.

Historical statistical records are usually the basis for estimating exceedance probabilities of external initiating events, but in most places these records are short. Estimates of the exceedance probability of rare events, such as the $p = 0.01$ ('100-year') event are prone to error. Estimates of yet lower probability events are prone to even more error. Common practice – presupposing a degree-of-belief view of probability – is to represent statistical uncertainty in a model parameter such as λ by a probability distribution, $f_\lambda(\lambda)$. Then a

predictive probability distribution can be taken on the number of events n in a unit length of time by integrating the model forecast over uncertainty in the parameter (Aitchison and Brown 1969)

$$f_n(n) = \int_\lambda f_n(n, \lambda) d\lambda = \int_\lambda f_n(n|\lambda) f_\lambda(\lambda)\, d\lambda \tag{20.4}$$

This predictive distribution is simply the marginal distribution on n after integrating out uncertainty on λ. Similar results obtain for the case when more than one uncertain parameter is involved, that is for vector $\boldsymbol{\lambda}$.

Note that the predictive distribution combines uncertainty due to natural variability, that is, the number of occurrences n given the parameter λ, with uncertainty due to limited knowledge, that is, the uncertainty in the model parameter itself. If more than one aspect of natural variability depends on the same uncertain parameter (e.g. number of occurrences in two separate periods, n_1 and n_2) then forming the predictive distribution of each in isolation from the other will mask the implied correlation caused by their shared dependence on the same uncertain realization of λ. This is one of the arguments for separating out natural variability from limited knowledge using logic trees.

The usual approach to quantifying the probability distribution for an uncertain model parameter, such as λ, based on statistical data is to start from some uniform or non-informative probability density function (pdf) on the parameter, presumed to exist prior to having observed the data, and then to update that probably distribution using Bayes' Theorem, as reviewed in Chapter 4. Modeling the occurrence of initiating events as a Poisson process typically presumes the uncertainty in those events to be due to natural variability in time or space. Given the average rate of occurrence specified by the parameter λ, the actual occurrence of initiating events is a purely stochastic process. The only place that knowledge uncertainty enters the formulation is in the statistical estimate of the parameter λ itself (which may be far from negligible). The assumption that initiating events are stochastic in time or space is convenient but not necessary. In the case of extreme storms, one could model storm events based on first principles of meteorology. Given this approach, the occurrence and magnitude of storm events are results of the modeling activity, and are not necessarily treated as random in time or space. Uncertainty enters this approach to modeling extreme events in the statistical estimates of the parameters of the meteorological model and in model error. This approach trades off the uncertainty due to natural variability for uncertainty due to limited knowledge.

Usually, an event tree analysis presumes stationarity for stochastic initiating events, but for events such as flood volumes or river stages the assumption may not be good. Changing upstream conditions in land use, channelization, and other factors may cause flood frequencies or rating curves to gradually change with time. This means that the probability distributions used to represent random variables, such as peak flood discharge, may change.

20.4.2 Reliability model estimates

Reliability models reason from first principles of mechanics or natural science to calculate uncertainties in the performance of specified variables (Chapter 13). For example, such models might start from uncertainties in soil engineering properties in an embankment

and calculate probabilities of excessive settlement of the crest of the embankment, or probabilities of strength failure. Conceptually, they are similar to the predictive distributions discussed in the previous section, in which statistical uncertainties in parameter values are integrated out to give an aggregate uncertainty in some random variable. With probabilistic engineering models, uncertainties in input parameters such as, say, undrained soil strength, are integrated over to yield implied uncertainties in model predictions that might otherwise have been deterministic.

Presume an engineering model, perhaps deterministic, that relates some dam performance variable x to a parameter or set of parameters θ through the equation(s)

$$x = g(\theta) \tag{20.5}$$

Presume also that uncertainty in θ can be expressed in a pdf $f_\theta(\theta)$. Then uncertainty in the performance variable x can be expressed as

$$f_x(x) = \int_\theta g(\theta) f_\theta(\theta)\, d\theta \tag{20.6}$$

This is akin to the predictive pdf of Equation (20.4), in that uncertainty in the parameter is integrated out. If the engineering model $g(\theta)$ is deterministic, then $f_x(x)$ is a simple probability density function reflecting the effect of uncertainty in the model parameter on uncertainty in the performance variable. If the engineering model is itself a stochastic relationship, then $f_x(x)$ will have a more complicated form, and some of the model parameters of $g(\theta)$ may themselves be expression of spatial or temporal variability.

Whether the prediction $f_x(x)$ reflects natural variation or limited information, or some combination of both, depends on the nature of the uncertainty in the model parameters, and on whether the model itself is deterministic or stochastic. If the model is deterministic and the uncertainty in θ is statistical, then $f_x(x)$ reflects limited knowledge. If the model is deterministic and the uncertainty in θ is spatial or temporal, then $f_x(x)$ reflects natural variation. If the model is stochastic, then $f_x(x)$ will typically reflect both natural variation and limited knowledge.

The other source of uncertainty in a prediction of the performance variable, in addition to that due to parameter uncertainty, is that due to uncertainty in the model formulation itself. This model uncertainty manifests in many ways. For example, a model may itself be statistical, as in the case of models based on regression analysis. Many current soil liquefaction models are of this variety. For such statistical models there may be well founded estimates of the uncertainty in model predictions made when using reasonably well identified input parameters. These uncertainties have to do with the set of calibrating data used to develop the original model and may reflect both natural variability and limited knowledge but usually in unknown proportions.

Another way model uncertainty manifests is in models that are approximations to complicated physical conditions. For example Mohr–Coulomb strength theory is a linearization to what is more commonly a nonlinear strength relationship between normal and shear stresses. Model uncertainties based on such approximations are due to limited knowledge and may be systematic, that is, they may tend consistently to bias predictions in one direction or another.

20.4.3 Probabilistic dependence

Dependence among branch probabilities associated with separate nodes in an event tree can be critical to obtaining proper numerical results in an event tree analysis, but at the same time both subtle to identify and difficult to estimate. It is imperative that correlations among probabilities be dealt with properly, as the following example shows.

Presume that one is concerned both about liquefaction during an earthquake and also about overtopping after a storm. The probability of liquefaction sometime during the life of the dam, accounting for the various earthquakes that might occur, but marginal of all other uncertainties, is estimated as 0.09. Similarly, the probability of overtopping sometime during the life of the dam, accounting for the various storms that might occur, but marginal of all other uncertainties, is estimated as 0.076. Then, were liquefaction and overtopping independent, the probability of one or the other, or both, occurring during the life of the dam would be

$$\begin{aligned}\Pr[\textit{liquefaction OR overtopping}] &= \Pr[\textit{liquefaction}] + \Pr[\textit{overtopping}] \\ &\quad - \Pr[\textit{liquefaction}]\Pr[\textit{overtopping}] \\ &= (0.09) + (0.076) - (0.09)(0.076) \\ &= 0.16\end{aligned} \tag{20.7}$$

Consider, however, that both liquefaction and overtopping – due in part to settlement of the embankment – are influenced by the possible existence of low-density soil lenses under the embankment. The respective conditional probabilities are given in Table 20.4, from which the marginal probabilities in Equation (20.7) can be verified. Given that the two events depend in common on the existence of low-density lenses, the actual probability of one or the other, or both, happening during the life of the dam is the weighted sum of the calculation of Equation (20.7), taken once conditioned on the zones occurring and a second time conditioned on the zones not occurring. This yields an actual probability of 0.11, about a third less than the calculation above:

$$\begin{aligned}&\Pr[\textit{liquefaction OR overtopping}] \\ &\quad = \Pr[\textit{zones}]\left\{\begin{array}{l}\Pr[\textit{liquefaction}|\textit{zones}] \\ +\Pr[\textit{overtopping}|\textit{zones}] \\ -\Pr[\textit{liquefaction}|\textit{zones}]\Pr[\textit{overtopping}|\textit{zones}]\end{array}\right\} \\ &\qquad + \Pr[\textit{no zones}]\left\{\begin{array}{l}\Pr[\textit{liquefaction}|\textit{no zones}] \\ +\Pr[\textit{overtopping}|\textit{no zones}] \\ -\Pr[\textit{liquefaction}|\textit{no zones}]\Pr[\textit{overtopping}|\textit{no zones}]\end{array}\right\} \\ &\quad = (0.1)[0.9 + 0.67 - (0.9)(0.67)] + (0.9)[0 + 0.01 - (0)(0.01)] \\ &\quad = 0.11\end{aligned} \tag{20.8}$$

Table 20.4 Conditional probabilities of liquefaction and overtopping given the existence or non-existence of low-density, soft lenses in an embankment

	Pr	Liquefaction	No liquefaction	Overtopping	No overtopping
Low-density lenses exits	0.1	0.9	0.1	0.67	0.33
No low-density lenses exist	0.9	0	1.0	0.01	0.99

Table 20.5 Branch probabilities and the bases for their calculation

Event node	Conditioning event	Event branch	Pr*	Method	Natural variability	Limited knowledge
Extreme storm		Yes	0.001	Flood frequency curve	Stochastic process	Distribution type, mean, variance, & skew
Water height	Extreme storm	Water level above crest	0.1	Rating curve	Bottom conditions, debris/ice, etc.	Statistical regression parameters
	Extreme storm	50–99% of crest	0.7	Rating curve	Bottom conditions, debris/ice, etc.	Statistical regression parameters
	Extreme storm	less than 50% of crest	0.2	Rating curve	Bottom conditions, debris/ice, etc.	Statistical regression parameters
Duration	Extreme storm	More than a week	0.25	Flood frequency curve	ditto	ditto
	Extreme storm	Less than a week	0.75	Flood frequency curve	ditto	ditto
Stringers	None	Yes	0.15	Poisson process	Ave number per reach and size	Chance exploration detected
Sand boils	Stringers exist	Yes	0.5	Seepage model	Spatial variability of permeability	Soil parameters and spatial model parameters
Failure	Sand boils	Yes	0.2	Slope failure model	Spatial variability of soil parameters	Statistics of soil properties, model uncertainty
Soft soil	None	Yes	0.1	Poisson process	Ave number per reach and size	Chance exploration detected
High pore pressure	Duration	Yes	0.5	Seepage model	Spatial variability of permeability	Soil parameters and spatial model parameters
GT Failure	Soft soil, pore pressure	Yes	0.2	Slope failure model	Spatial variability of soil parameters	Statistics of soil properties, model uncertainty

*Conditional probability given earlier events in tree.

The reason for this result is that the existence of the soft lenses makes both events simultaneously more likely, while the non-existence makes both simultaneously less likely. Since the two events are no longer independent, their marginal probabilities can no longer be simply summed to give the probability that either one or the other occurs. The correlation due to a common dependence reduces the risk to the structure in this case, but the outcome could be the reverse in other circumstances.

Branch probabilities associated with separate event nodes can be correlated through of any of the following:

1. *Causal dependence*, meaning that one event physically causes another. For example, liquefaction-induced settlement may directly lead to overtopping of an embankment, thus the liquefaction event and overtopping event would not be independent of one another. If the liquefaction settlement occurs, the probability of overtopping might be greatly enhanced.
2. *Probabilistic correlation*, meaning that two uncertainties may share a common dependence on a third uncertainty, as in the case of the low-density soil lenses in the example above. Whether the low-density soil lenses exist or not simultaneously changes the probability of liquefaction cracking and of overtopping.

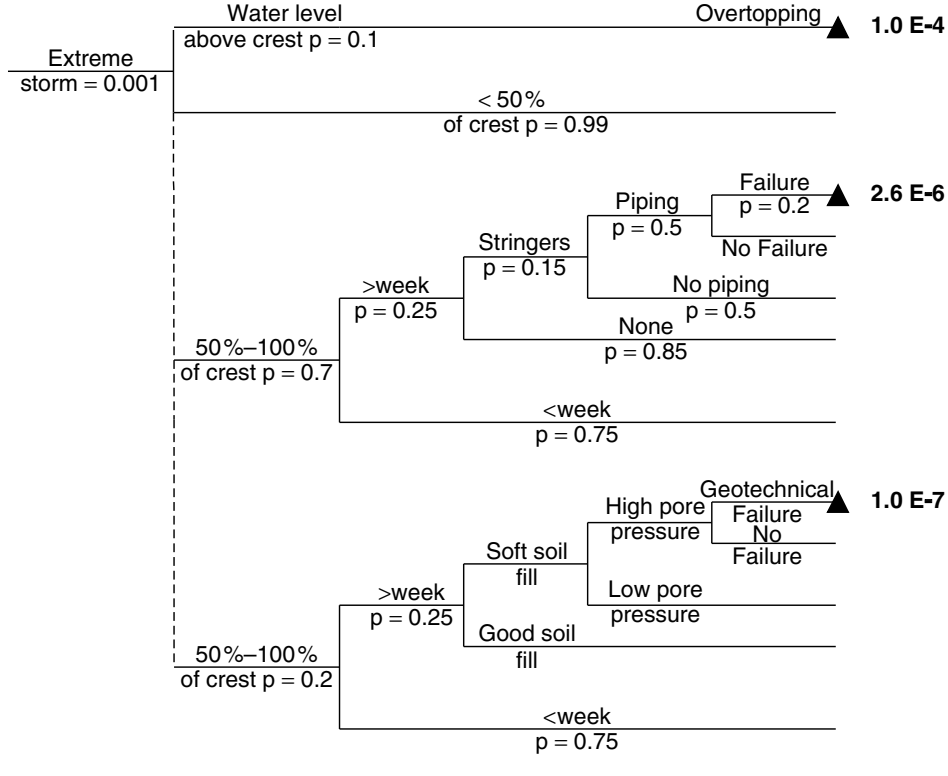


Figure 20.22 Event tree for levee failure during extreme storm, with estimated branch probabilities.

3. *Spatial or temporal autocorrelation*, meaning that two uncertainties depend on the spatial or temporal realization of some third uncertainty which itself exhibits stochastic dependence in space or time. The performances of two sections of a long levee may depend on soil engineering properties in the naturally occurring valley bottom, which, when modeled as a stochastic (aleatory) process, exhibit a long wave length of correlation in space; thus adjacent sections will exhibit similar settlements or factors of safety against slope instability.
4. *Statistical correlation*, meaning that two uncertainties are simultaneously estimated from a fixed set of data and therefore influenced by a common sampling variability error. In soil mechanics, a common – and almost always overlooked – statistical correlation is that between soil cohesion, c, and soil friction angle, ϕ, which, being regression parameters of the Mohr-Coulomb model, are negatively correlated, given a finite number of test data.

20.5 Levee Example Revisited

Figure 20.22 shows the event tree for levee failure during an extreme flood that was developed earlier, but now with branch probabilities. Table 20.5 shows the justification for each probability estimate, the estimation approach, and the sources of uncertainty due to natural variation and limited knowledge.

21 Expert Opinion

Many important uncertainties in risk analysis are not amenable to quantitative estimation from data. In some cases there are no data at all, only the judgment of experts. These uncertainties have traditionally been treated using expert opinion. The tacit knowledge of experts is based on intuition, unenumerated past experience, subjective theory, and other important but qualitative beliefs. How should we interpret expert testimony and include it in quantitative risk analysis? How do experts estimate probabilities associated with qualitative judgment? Can expert opinion be quantified in a way that is repeatable, and if so, how? Is there any way to test the external validity of expert opinion? This chapter surveys the growing field of expert elicitation of subjective probabilities, and summarizes the emerging understanding of the psychology of probability assignment.

21.1 Expert Opinion in Geotechnical Practice

In almost all risk analyses there are some uncertainties that are simply not amendable to quantitative estimation based on data and models. These may reflect unique situations that are not found in the historical record of experience, they may reflect uncertainties associated with poorly understood physical phenomena, they may reflect conditions for which data could, in principle, be collected but only at a prohibitive price, and so forth. Formally incorporating such uncertainties in a risk analysis relies on professional judgment (Gilbert *et al*. 1998). In most cases, this judgment has to do with tacit rather than explicit knowledge. It is based on intuition, qualitative theory, anecdotal experience, and other sources that are not easily amenable to mathematical representation. Yet, this judgment of experts is important information in analyzing risk.

Within the degree-of-belief school of probability, personal opinion is an admissible basis for estimating probability, as long as that opinion is quantified in a way that maintains consistency. That is, given certain assumptions, degrees of belief have been shown by Ramsey (1978) and others to satisfy the axioms of probability theory, and, thus, the mathematics of probability theory can be used to operate on them, and they can be formally incorporated in a quantitative risk analysis. Such degree-of-belief probabilities, which mostly have to do with epistemic uncertainties, may or may not be appropriately

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

convolved with probabilities associated with natural variability. Whether they are or are not depends upon the decisions at hand and not on matters of principle. In principle, within the probability-as-logic school of thought, quantitative probabilities inferred from data and associated with natural variation may indeed be mixed with degree-of-belief probabilities inferred by eliciting expert opinion.

21.2 How do People Estimate Subjective Probabilities?

While it is evident that subjective probability requires integrating information of various kinds within a consistent framework, it is less clear how people do this. Subjective probabilities should be concordant with probability theory (e.g. they should sum to 1.0), and one would prefer they be calibrated to observed frequencies in the physical world (i.e. they should have predictive value). These properties are called, *coherence* and *calibration*. A considerable body of behavioral research indicates that people are not well equipped for mentally processing uncertainty according to probability theory, or in harmony with observed frequencies (Hogarth 1975). The mathematics of probability theory describes how people ought to quantify uncertainties, not how they do.

In practice, it appears that people use simple mental strategies or rules of thumb to simplify the task of quantifying subjective probabilities. In the literature of cognitive psychology, these are called, *heuristics.* A large literature has grown up around heuristics and, as importantly, around the systematic errors to which heuristics lead. The literature calls the latter *cognitive biases*. The conclusion is that people use rules of thumb to simplify judgments about probability, and to the extent these fall short of normative standards, so do the assessed probabilities. A surprising number of engineers take subjective probability assessments for granted, treating them as though they were coherent and calibrated, when, in fact, they often are not. This *heuristics-and-biases* school of thought is most associated with the work of Edwards and Tversky (1967), Kahneman *et al*. (1982), and their colleagues. Criticism of the approach can be found in Gigerenzer (1991). Folayan *et al*. (1970), Baecher (1972) and Hynes and Vanmarcke (1976) have reported on its use for subjective probability assessment in geotechnical engineering, and some of their results are discussed below. The work on heuristics and biases has been used extensively in nuclear risk assessment (Meyer and Booker 1990, Mosleh *et al*. 1987), and in seismic hazard analysis (Budnitz *et al*. 1997). The following subsections describe common heuristics revealed by behavioral research. These have mostly been identified in controlled settings. In practical settings, heuristics are more difficult to distinguish.

21.2.1 Representativeness

Representativeness is more often illustrated than precisely defined (Kahneman *et al*. 1982). In general, it reflects subjective probability judgments based on the resemblance of particular conditions in one circumstance to those in another. In a classic experiment, subjects are provided with a detailed profile of the behavior and personality characteristics of a hypothetical person, then asked to estimate the probability that the person is a lawyer vs. an engineer. Subjects told that the individual was drawn from a group of 70 lawyers and 30 engineers produced the same estimates as those told that the group contained 30 lawyers and 70 engineers. The subjects' judgments are based on matching the description

to stereotypes of lawyers and engineers, ignoring other information – in this case prior frequencies.

In geotechnical practice, representativeness is often undermined by over-reliance on complex models, while discounting simple observations. Uncertainties remain disguised by embedded approximations, simplifications, and assumptions so that analysis results are taken as uniquely representative of field conditions with near-certainty, because 'it's the best analysis we have' (Vick 2002).

21.2.2 Anchoring and adjustment

Anchoring-and-adjustment is easily illustrated. When asked to estimate a quantity or an uncertainty, people often start with a 'best estimate,' and adjust up or down. Unfortunately, people tend to stick too close to the initial value, not adjusting sufficiently to reflect uncertainty. Asked to estimate the undrained shear strength of a clay and the uncertainty in that value, one's natural reaction is first to think about a typical value. What is the average shear strength for this type soil across the many sites I have dealt with? How does the present site differ? Do I think the soil here is stronger or weaker, stiffer or looser? How much should I adjust up or down? Might this soil be 10% stronger? How different could this site be, to suggest upper and lower bounds? These are the questions one asks oneself, but this chain of reasoning is exactly that which has been shown to lead to significant overconfidence in resulting estimates.

Quite a different result is obtained if one first states the largest value the strength might have, then the lowest, and only afterwards hones in on a central value. The latter yields a broader range of assessed uncertainty and a better calibration to the physical world. Folayan *et al*. (1970) present an example from geotechnical practice.

21.3 How Well do People Estimate Subjective Probabilities?

The heuristics of the last section deal with *how* people quantify probabilities, but they also influence *how well* people do so. The many fallacies that people – even technically trained people – exhibit at the gambling table should dissuade us from thinking that one's natural tendencies concerning probability are well calibrated to the physical world. People behave as if games of chance even out, or as if pulling the slot machine handle oneself improves the chance of winning, or as if small numbers of observations are highly representative of a random process. These things are all false, and, left to one's own devices, the probabilities we estimate are usually neither coherent nor consistent. In particular, people tend to over-confidence in their assessments, and mis-calibration seems to vary systematically with the difficulty of the assessment.

21.3.1 Overconfidence

Overconfidence is the most pervasive bias in assessing subjective probability (Lichtenstein *et al*. 1982). It manifests in probability estimates that are too extreme at both ends of the probability scale and estimated distributions having insufficient dispersion about the mean. People – even experts – rarely assess their uncertainty to be as large as it usually turns

out to be, sometimes to a shocking extent, as reported in a well-known study by Alpert and Raiffa (1982).

Figure 21.1 plots the results of experiments reported by Fischhoff *et al*. (1997) in which three groups of subjects provided answers to general-knowledge questions as well as estimated probabilities that their answers were correct. The estimated error probabilities were found to be reasonably well calibrated relative to the actual error frequencies only when the actual probabilities were no less than about 0.1. Their overconfidence, expressed as the difference between actual and judged error probabilities, increased dramatically at smaller values of actual error frequency. The subjects estimated a subjective probability of error as small as 10^{-6} when the actual error frequency was slightly less than 10^{-1}, a ratio of five orders of magnitude. Moreover, the subjects showed little ability to distinguish among varying degrees of extreme likelihood, with judged probabilities ranging from 10^{-2} to 10^{-6} despite actual error frequencies hovering near 10^{-1}. A related and surprising finding is that the harder the probability estimation task, the greater the associated overconfidence. For quite easy tasks sometimes under-confidence is displayed, although this effect is poorly understood (McClelland and Bolger 1994).

Neither experts in general nor geotechnical experts in particular seem immune from overconfidence. Hynes and Vanmarcke (Hynes and Vanmarcke 1976) reported on predictions of embankment failure height made by seven internationally-known geotechnical engineers for a test embankment on soft clay at the MIT I-95 test site. Figure 21.2 shows each expert's best estimate and 50% confidence interval for the amount of additional fill needed to fail the embankment, the average of the best estimates, and the actual amount required to cause failure. While the average of the seven best estimates is reasonably close to the outcome, no individual estimate had 50% error bounds large enough to encompass the actual outcome. Had the estimates been unbiased, half would have encompassed the actual failure height at the 50% confidence level, but none did so. Figure 21.3 is similar

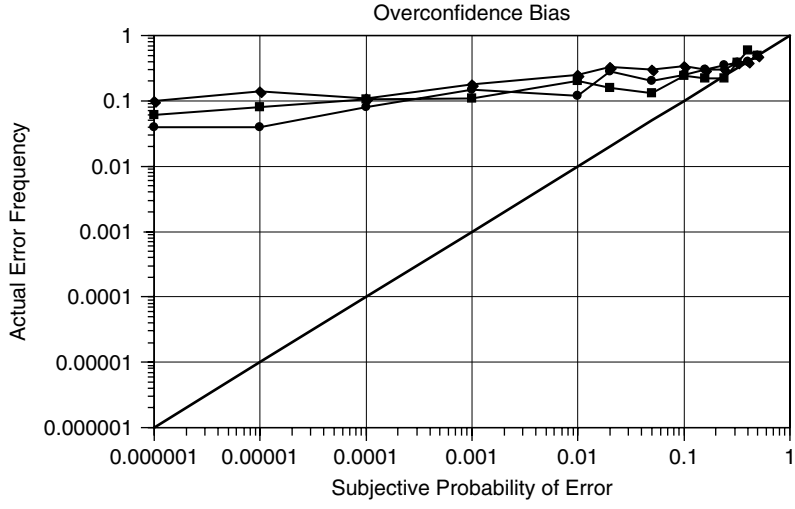


Figure 21.1 Subjectively estimated vs. actual probabilities. (Data from Fischhoff *et al*. (1997), after Vick, S. G., 1997, 'Dam safety risk assessment: new directions,' *Water Power and Dam Construction*, Vol. 49, No. 6, reproduced with permission of International Water Power and Dam Construction.)

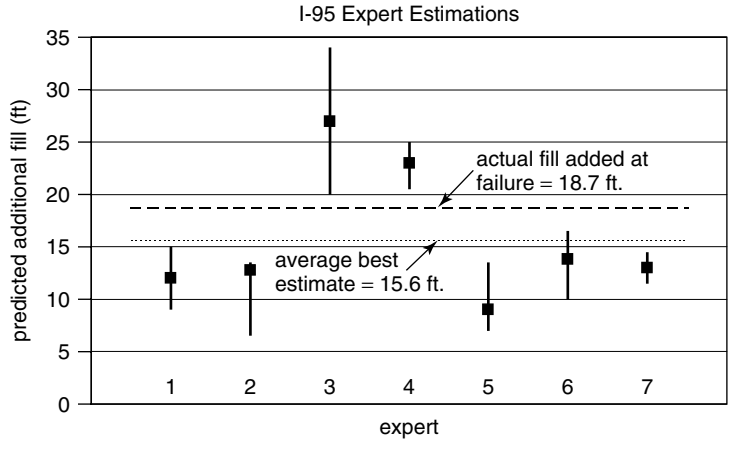


Figure 21.2 MIT I-95 embankment predictors' 50% confidence ranges of added height to cause failure (after Hynes, M. and Vanmarke, E., 1977, 'Reliability of Embankment Performance Predictions,' *Proceedings ASCE Engineering Mechanics Division Specialty Conference. Mechanics in Engineering*, University of Waterloo Press, pp. 367–384, reproduced with permission of University of Waterloo Press).

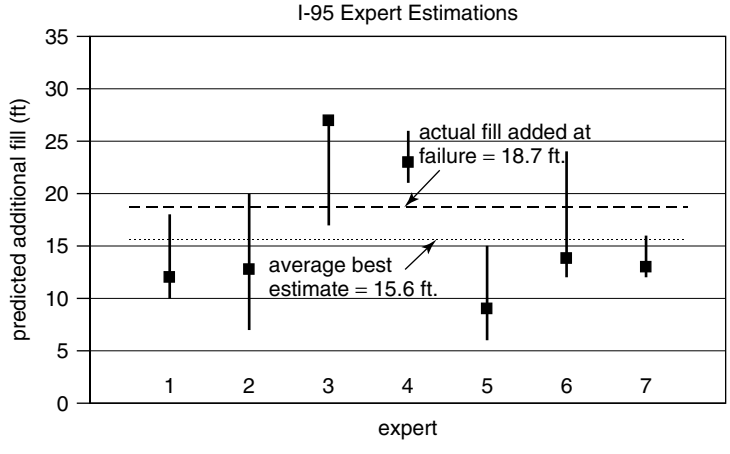


Figure 21.3 MIT I-95 embankment predictors' maximum and minimum ranges of added height to cause failure (after Hynes, M. and Vanmarke, E., 1977, 'Reliability of Embankment Performance Predictions,' *Proceedings ASCE Engineering Mechanics Division Specialty Conference. Mechanics in Engineering*, University of Waterloo Press, pp. 367–384, reproduced with permission of University of Waterloo Press).

to Figure 21.2, except that it shows the experts' estimates of the maximum and minimum additional needed fill. In some cases these estimates do encompass the actual event, but the values are clearly not consistent. Some maximum and minimum estimates actually fall within an expert's 50% confidence limits. Hynes and Vanmarcke state, "It is clear that there are wide differences among engineers in the way they interpret the terms 'minimum' and 'maximum.' These widely used terms are essentially meaningless unless related to relative likelihood or probability." It should also be noted that those in the audience at

the seminar were also invited to give their estimates and their 50% intervals "capture[d] the true value 62% of the time."

Slovic *et al*. (1982) and others have cited these and similar findings in suggesting that *substantive expertise*, or capability within one's specialized knowledge domain, has no necessary relationship to *normative expertise* or ability to provide coherent and unbiased probability judgments. On the other hand, at least some evidence with experts assessing probabilities about professional subjects with which they are familiar suggests that they may be more calibrated than are non-experts (*viz*., weather forecasters (Winkler and Murphy 1968) and auditors (Smith and Kida 1991)).

Similar over-confidence effects have been shown in subjective estimates of probability distributions provided by geotechnical engineers. Folayan *et al*. (1970) obtained estimated distributions for compressibility parameters of San Francisco bay mud from engineers with up to 17 years of experience. Baecher (1972) further analyzed these prior distributions in comparison to that obtained from subsequent laboratory tests. As shown in Figure 21.4, the estimated means were lower than that measured, but, more significantly, overconfidence produced distributions too narrow to encompass most of the measured data. The one exception was subject 5, a graduate student whose estimate showed gross under-confidence.

Why over-confidence? There are many suggestions in the literature. Keren (1994) suggests that anchoring and adjustment may be to blame. People may anchor on a probability estimate reflecting intermediate difficulty, say 75%, and adjust up or down – but not sufficiently – depending of the perceived difficulty of the estimation task. Ferrell and McGoey (1980) present a similar argument, but the literature contains other attempted explanations as well (McClelland and Bolger 1994).

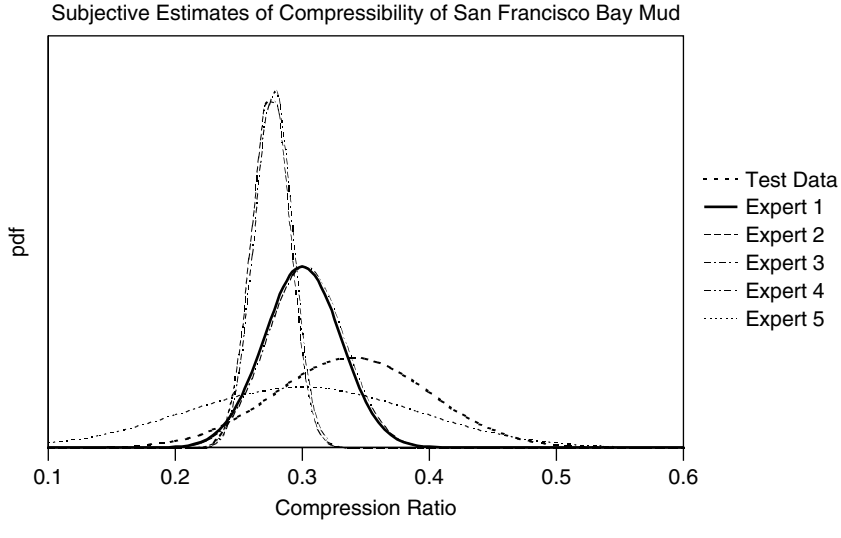


Figure 21.4 Subjective estimates of the compressibility of San Francisco Bay mud compared to test results for five experts (after Folayan, J., Höeg, K. and Benjamin, J., 1970, 'Decision Theory Applied to Settlement Predictions,' *Journal of the Soil Mechanics and Foundations Division, ASCE*, Vol. 96, No. SM4, pp. 1127–1141, reproduced by permission of the American Society of Civil Engineers).

21.3.2 Neglect of base rates

Neglect-of-base-rate bias reflects people's tendency to judge probability by the similarity of one circumstance to another. Typically, this manifests in ignoring the average frequency or *a priori* probability of an event in light of new information, even when that new information may have limited diagnostic strength. A typical result is that an event, known to be rare in the physical world, will be assigned a reasonably high probability based on fragmentary evidence. Forensic engineering (failure-cause analysis) is a rich breeding ground for examples of neglect-of-base-rate bias. The presumption is that people, even people with expertise in an area, focus on recent identifiable results as if they in themselves were wholly typical of some uncertain process or quantity. In so doing, people forget about – or at least under-rate – prior information or probabilities.

On a seismic hazard project undertaken by one of the authors some years ago, and described in more detail in the next chapter, great care had been invested in calculating liquefaction potential index (Yegian and Whitman 1978) for a particular industrial site in Tokyo Bay. This LPI was used in the analysis to predict ground failure. The result, expressed as a PDF, is shown in Figure 21.5. At about the time the analysis was done, a magnitude 6.5 earthquake occurred in the Tokyo area, causing 65 gal peak ground acceleration at the site and evidence of liquefaction within the facility perimeter. How should this affect the calculated probability distribution of LPI?

Subjective estimates by several engineers were made of the new most likely value of LPI. All of these were significantly in excess of LPI = 1. Figure 21.5 shows a calculation of the posterior PDF of LPI using Bayes' Theorem. The statistical calculation shows less updating and more influence of the prior PDF. Because many uncertainties are involved in calculating LPI (corrected below count, water level, effect of vertical stress, cyclic shear strength, etc.), the combined effect of small changes in each of the many uncertain quantities easily accounts for the observed liquefaction. Our intuition suggests that the updating should be greater than statistical theory requires.

A particularly malicious manifestation of base-rate neglect appears in what Tversky and Kahneman (1971) and Kahneman *et al*. (1982) have called, the 'law of small numbers.' Recall that the law of large numbers states that as the number of observations (i.e. sample

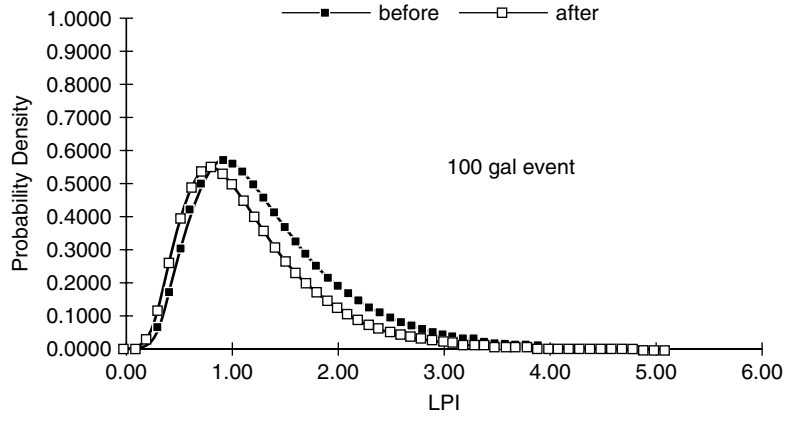


Figure 21.5 Liquefaction Potential Index (LPI) probability distribution for 100 gal event, before and after updating by observed earthquake performance.

size) becomes large, statistical patterns of the sample asymptotically resemble those of the sampled population. Tversky and Kahneman's law states that most people believe the same resemblance manifests in small samples. The enormous variability in statistical properties of small samples is largely ignored by engineers. Thus, we should not be surprised when subsequent measurements vary considerably from those taken initially, but we usually are. Interestingly, Gigerenzer (1991) argues that the neglect of base-rate effects can be reduced or eliminated by describing priors in frequency terms rather than as probability functions, but, while interesting, this is beyond the present scope.

21.3.3 Misperceptions of independence

Over-confidence and neglect of base rates deal with single variables. In risk and reliability analysis a critical factor is often the probabilistic independence – or lack thereof – among variables. How good are people at quantifying conditional probabilities or correlations? Somewhat less work has been done on this topic.

Some evidence exists that people judge the probability of the combined occurrence of events as being higher than the probabilities of the constituent events, when the events are representative of a pattern. This has been called the conjunction fallacy by Tversky and Kahneman (1983). Consider the following question: "A small dam was built across a stream with similar geology at both abutments. The regional geology is flat lying sedimentary rocks. A spring flowing muddy water appears at the downstream toe of the dam, creating a small volcano of silty sand. Which of these two alternatives is more probable? (a) There is a geologic fault beneath the dam, or (b) there is a geologic fault beneath the dam allowing internal erosion of the embankment." While many experienced people judge the second alternative as more probable, clearly it cannot be, because the joint occurrence of two events is necessarily less likely than either event alone. Hacking (2001) presents a lucid discussion of this and similar fallacies.

Another place where misperceptions of independence arise is in combining statistical data with judgment. This is especially apparent in quantifying the uncertainties in soil engineering parameters, which may be statistically correlated. For example, soil cohesion and friction angle are in essence regression parameters. Larger values of c imply smaller values of ϕ, and *vice versa*, because a Mohr envelope is fit to given data. Thus, judgmental estimates of uncertainties in c and ϕ cannot be made separately, although common practice is to do so.

21.4 Can People Learn to be Well-calibrated?

A great deal of work has gone into finding out whether people can be trained to be better calibrated in their probability assessments. The common conclusion is: yes, but the results of training seem not to generalize well to new tasks (Ferrell and McGoey 1980; Keren 1994; McClelland and Bolger 1994). People seem to provide better calibrated probabilities when asked questions within their realm of expertise, although this is not without controversy.

A highly encouraging result comes from the realm of weather forecasting. In a series of studies, Murphy and Winkler established that National Weather Service forecasters are perhaps the world's best-calibrated probability assessors (Murphy and Winkler

1974,1977a,1977b; Winkler and Murphy 1973). Why should this be? The authors attribute the result to (i) practice (forecasts are made every day), (ii) immediate feedback (outcome on the following day), (iii) quantitative scoring of performance, and (iv) promotion and pay incentives for accuracy. The important suggestion for geotechnical practice is, the very act of risk assessment and quantifying subjective probabilities may improve practitioners' ability to assess their own personal uncertainties, whether or not a project calls for formal risk analysis.

21.5 Protocol for Assessing Subjective Probabilities

The preceding discussion has summarized empirical findings suggesting that people are not inherently adept at quantifying intuitive probability values, at least in the sense of providing numbers that are consistent, coherent, and well calibrated. This should not surprise us. People are not adept at all sorts of intuitive mathematical tasks, nor for that matter are they adept at consistently judging the loudness of sounds or other psychophysical phenomena (Stevens 1951). That is why we calculate probabilities and perform risk analysis in the first place. The question is, what do we learn from the preceding evidence that helps us structure a protocol for assessing subjective probabilities that supports useful outcomes?

People, even geotechnical engineers, do not enter a situation with a well-structured, mathematical conception of the probabilities of events pre-formed in their minds. The protocol of assessment must evoke such a structure. Current usage calls this process, *elicitation*. The protocol cannot simply ask a subject to guess a number for the probability for an event and expect that numbers so generated will be consistent, coherent, and well calibrated.

Even if the cognitive influences of heuristics and biases would seem to paint a bleak picture, people are nevertheless remarkably well adapted to dealing with uncertainty in everyday life, and successfully accommodating uncertainty has always been a hallmark of geotechnical practice. The problem is less one of expressing uncertainty judgments than in doing so in ways congruent with probability theory. The first requirement is external validity or *coherence* with basic probability axioms. For example, a set of subjectively-estimated probabilities for mutually exclusive and collectively exhaustive events must sum to 1.0.

Beyond this, internal consistency depends on how information may be presented or synthesized, and here is where heuristics and biases come into play. Defeating or at least reducing their effects is one important factor, but, unfortunately for practitioners, the heuristics-and-biases school is mostly silent on this topic. Elicitation procedures from related fields, however, provide a framework for subjective probability assessment that acknowledges behavioral effects.

21.6 Conducting a Process to Elicit Quantified Judgment

A common misconception in eliciting expert judgment is that people carry fully-formed probabilistic opinions around in their heads on almost any subject of interest and that the focus of an elicitation process is merely to access these pre-existing opinions. Actually,

people do not carry fully-formed constructs around in their heads but develop them during the process of elicitation. Thus, the elicitation process needs to help experts think about uncertainty, needs to instruct and clarify common errors in how people quantify uncertainty, and needs to lay out checks and balances to improve the consistency with which probabilities are assessed. A successful process of elicitation is one that helps experts construct carefully reasoned judgments. The process should never be approached as a 'cookbook' procedure, for the results will be unsatisfactory, as will the risk assessment in which the results are used.

The steps in using expert elicitation to quantify judgmental probabilities are the following:

1. Decide on the general uncertainties for which the probabilities need to be assessed.
2. Select a panel of experts displaying a balanced spectrum of expertise about the identified uncertainties.
3. Refine issues in discussions with the panel, and decide on the specific uncertainties for which the probabilities need to be assessed.
4. Expose the experts to a short training program on concepts, objectives, and methods of eliciting judgmental probability, and on common errors that people make when trying to quantify probability.
5. Elicit the judgmental probabilities of individual experts on issues pertinent to their individual expertise.
6. Allow the group of experts to interact, supported by a facilitator, to explore hypotheses, points of view, and quantified estimates of probability, toward the goal of aggregating probabilities and resolving the breadth of opinion.
7. Document the specific process used to elicit judgmental probabilities and communicate the results back to the panel of experts.

It is important for credibility – as well as defensibility – that the process be well documented, open to inspection, and methodologically transparent to peer review.

21.6.1 Choosing experts: who should be on the panel?

The choice of experts is the most important step in determining success or failure of the expert elicitation process. Depending on personality, experience, and technical background, individual experts may serve different functions. These individuals can be categorized into distinct types:

- *Resources* – individuals with special knowledge of a problem, technology, site, or project.
- *Proponents* – individuals with a particular and strongly argued point of view about a technology, issue, or project.
- *Evaluators* – individuals with open mind, willing to judge objectively the credibility of different hypotheses, points of view, or opinions.
- *Facilitators* – individuals versed in the technical issues who are also willing to lead discussions and manage interactions among other experts.

To be successful, an expert panel usually needs to include an evaluator and a facilitator. Resources and proponents are typically used to elucidate specific issues, to give depth in

specific areas; they are typically not used to give balance and breadth. It is important in considering potential experts that individuals be willing to be objective, willing to commit time, and willing to interact with others in a professional manner. Panel members who cannot be counted on to commit to these terms harm the process more than they help it.

21.6.2 Systematic process of elicitation

The process of eliciting expert opinion in judgmental probabilities has five phases:

1. Motivating phase
2. Training phase
3. Structuring (deterministic) phase
4. Assessing (probabilistic) phase
5. Documenting phase

The first two, motivating and training, set the stage for the whole process. The next two, structuring and assessing, comprise the central effort of the process, the former establishing a structure for the uncertainties to be assessed, and the latter performing the actual quantification. The process is complete with documentation, the purpose of which is to create a defensible set of results. To obtain consistent and defensible results, it is important that a systematic process like this be used, although individual organizations may tailor the details of the protocol to their own special needs.

21.6.3 Motivating phase

The motivating phase intends to develop rapport with the experts and to explain why and how judgmental probabilities will be elicited and how the results will be used in risk assessment. Experts are justifiably reluctant to participate in probability elicitation unless assured about how resulting assessments will be used in practice. During this phase, the basic philosophy of judgmental probability is reviewed, and an attempt is made to bring motivational biases within the expert panel out into the open. Motivational biases include those factors that might lead experts to provide assessments that do not accurately or completely reflect actual beliefs, whether generated consciously or unconsciously. Examples of motivational bias include the desire to appear knowledgeable and thus under-report uncertainty, the desire to influence a decision and thus prejudice answers, and the desire to advance a special cause and thus refuse to credit alternate points of view.

21.6.4 Training phase

The training phase has the purpose of making the experts aware of the processes and aids people typically use in quantifying judgmental uncertainties and how well calibrated judgmental probabilities are with respect to observed frequencies of assessed events in the world. The goal of this training is to encourage the experts to think critically about how they quantify judgment and to avoid the common biases encountered in quantifying judgmental probability.

A typical training phase involves having the experts explain how they proceed to make probabilistic judgments and how they use data and other information in arriving at

expressions of uncertainty about the technical issues to be faced in the elicitation. Common cognitive biases are discussed, and warm-up exercises involving probabilistic predictions are used to illustrate these biases. It is helpful if these exercises involve uncertain quantities of an engineering nature, such that the experts feel professionally engaged. A useful exercise is to have the experts explain retrospectively how unanticipated outcomes of an engineering project might have occurred; that is, "If you know that an earth embankment settled excessively on what would otherwise appear to be a stable and homogeneous sand foundation, how might you explain that observation?" The benefit of such thought experiments is that they open up the range of considerations the experts might consider and illustrate how easily one can become overconfident in assessing probabilities.

21.6.5 Deterministic phase

The structuring or deterministic phase has the goal of defining the specific uncertainties to be addressed, and the relationships among those uncertainties that allow individual probability assessments to be combined. The outcome of the deterministic phase should be clear, specific definitions of the uncertainties to be assessed, such that, were a new expert added to the panel, the descriptions of the uncertainties to be assessed would stand on their own without further clarification. The goal is to make unstated assumptions explicit and to disaggregate the technical problem into components with which experts can readily deal.

The deterministic phase begins with a review of the technical issues to be addressed in the elicitation. If time permits beforehand, a literature review and collection of documents should be made. A summary of this material should be prepared, along with graphs and tables comparing principal results or issues. Copies of the summary book should be distributed to the experts before the meeting, and time should be allocated at the beginning of the elicitation sessions to review the information. This helps the experts understand why the elicitation is approached the way that it is and establishes a sense that the elicitation process is being attempted in a serious, complete, and professional manner.

Two procedural issues affect the outcomes of a risk assessment: the way failure sequences are conceptualized, and away probabilities are assessed. The deterministic phase considers the way failure sequences are conceptualized and then decomposed into tractable parts.

One reason for decomposing failure sequences is that research suggests that people tend to be better at working with decomposed problems, and with estimating the probabilities associated with components than they are at working with entire systems and assessing probabilities for holistic processes (Gettys *et al.* 1973). By design, engineered systems have very low probabilities of adverse performance or failure. From empirical experience, people are not well able to estimate accurately these low probabilities. Thus, by decomposing a system into a series of events, the conditional probabilities associated with individual components are made larger, and people can more readily deal with them.

21.6.5.1 Asking questions about problem structure

The most common way of decomposing an engineering problem is by building an event tree or a fault tree, which were the subjects of the previous chapter. An event tree starts with some initiating event and then considers all possible chains of events, which could lead

from that first event to various performances by the system. Some of these chains of events lead to adverse outcomes or failures, some do not. For each event in the tree, a probability is assessed presuming the occurrence of all the events preceding it in the tree, that is, a conditional probability. The total probability for a particular chain of events or path through the tree is found by multiplying the sequences of conditional probabilities. Bury and Kreuzer (1986) describe how an event tree might be structured for a gravity dam.

A fault tree decomposes the problem in the reverse direction. The fault tree starts with some failure condition and then considers all possible chains of faults that could lead to that failure. Conditional probabilities for each fault are assessed in the same way as with an event tree, but the total probability is calculated by starting at the failure and moving backwards, rather than starting from an initiating event and moving forward, as in an event tree. The advantage of the event tree is that it comprehensively uncovers combinations of events leading to failures; the disadvantage is that event trees can become bushy messes. The advantage of a fault tree is that it focuses only on chains of events leading to failures; the disadvantage is that it may fail to uncover important combinations of events.

An event tree can be used to decompose a problem at different levels of detail. Usually, judgment is more easily applied to smaller components, and research suggests that more detailed decomposition, within reason, enhances the accuracy of calculated failure probabilities. One reason, presumably, is that the more detailed the event tree is, the less extreme the conditional probabilities which need to be estimated.

Both event trees and fault trees require a strict structuring of a problem into sequences. This is what allows probabilities to be decomposed into manageable pieces and provides the accounting scheme by which those probabilities are put together. In the process of decomposing a problem, however, it is sometimes convenient to start not with highly structured event or fault trees, but with an influence diagram. An influence diagram is a graphical device for exploring the interrelationships of events, processes, and uncertainties. Once the influence diagram has been constructed, it can be readily transformed into event or fault trees. Stedinger *et al*. (1996) show how influence diagrams can be used in assessing hydrologic risks of dam projects.

21.6.5.2 Finding an appropriate level of detail

Decomposition of a probability estimation problem relies on disaggregating failure sequences into component parts. Usually, these are the smallest sized pieces that can be defined realistically and analyzed using available models and procedures. Decomposition can be used for any failure mode that is reasonably well understood. Clearly, decomposition cannot be used for failure modes for which mechanistic understanding is lacking. Internal erosion leading to piping is arguably one such, poorly understood failure mode.

In most cases, the extent of decomposition, that is the size of the individual events into which a failure sequence is divided, is a decision left to the panel of experts. Most real problems can be analyzed at different levels of disaggregation. Considerations in arriving at an appropriate level of disaggregation include the availability of data pertinent to the components, the availability of models or analytical techniques for the components, the extent of intuitive familiarity experts have for the components, and the magnitude of probabilities associated with the components. Typically, best practice dictates disaggregating a failure sequence to the greatest degree possible, subject to the constraint of being able to assign probabilities to the individual components. Usually, it is a good practice to

disaggregate a problem such that the component probabilities that need to be assessed fall with the range 0.01 to 0.99 (see e.g. Vick 1997). If this range can be limited to 0.1 to 0.9, all the better. People have great difficulty accurately estimating judgmental probabilities outside these ranges.

21.6.6 Assessment (probabilistic) phase

Research in psychology has shown that people have limitations in their ability to assess consistently the numerical probabilities representing their beliefs. Therefore, procedures that aid probability assessment and help minimize common forms of assessment bias should be routinely used in expert elicitation. The elicitation of judgmental probability borders on experimental psychology, and as such it is strongly influenced by procedural details.

21.6.6.1 Associating numbers with descriptive statements

Probability theory is one way of measuring uncertainty. Probability theory expresses uncertainty as a number between 0 and 1. People also use descriptive phrases to express the notion of uncertainty. For example, people say that something is 'likely,' or 'probable.' People might say, 'there is a very good chance of that happening.' These verbal descriptions carry meaning but they are less precise than numbers. Thus, it is tempting to attach verbal descriptions to the numerical scale of probability, and then to use verbal descriptions as a means of measuring judgmental probability. Common experience suggests that, at least in the early stages of expert elicitation, people find verbal descriptions more intuitive than they do numbers. Such descriptions are sought for components within the event or fault tree. Then, using approximate transformations between verbal descriptions and quantitative judgmental probabilities approximations can be assigned to component events. The warning about using verbal descriptions is that the range of implied numerical probability different people associate with verbal descriptions can be wide and the number even an individual associates may change with the semantic context in which the verbal description is used.

Table 21.1, taken from Lichtenstein and Newman (1967) and Vick (1997), shows a simple list of verbal descriptions of uncertainty that have been used in the course of dam safety studies. More extensive studies have also been performed in the psychology literature to attempt to quantify the relationship between verbal descriptions of uncertainty and probability. The results show encouraging consistency from one person to another, but the ranges of responses are large; also, mirror-image pairs sometimes give asymmetric results. Thus, verbal descriptions of uncertainty are a useful tool, but they are only a starting point.

21.6.6.2 Avoiding intuitive or direct assignment

It is common for experts who have become comfortable using verbal descriptions to describe probability to want to begin directly assigning numerical values to those probabilities. This should be discouraged, at least initially. The opportunity for systematic error or bias in directly assigning numerical probabilities is too great (*viz.* Phillips and Edwards 1966; Luce and Suppes 1965). More experience with the process on the part of

Table 21.1 Empirical translations of verbal descriptions of uncertainty. (Source: Vick (1997), and Lichtenstein and Newman (1967))

Verbal description	Probability equivalent	Low	High
virtually impossible	0.01	0.00	0.05
very unlikely	0.10	0.02	0.15
unlikely	0.15	0.04	0.45
fairly unlikely, rather unlikely	0.25	0.02	0.75
fair chance, toss-up	0.50	0.25	0.85
usually, good chance, probable, likely	0.75	0.25	0.95
quite likely	0.80	0.030	0.99
very likely, very probably	0.90	0.75	0.99
virtually certain	0.99	0.90	1.00

the experts should be allowed to occur before directly assigning numbers. At this initial point, no more than order-of-magnitude bounds on the elicited numerical degrees belief are a realistic goal.

The direct numerical values generated by experts unfamiliar with judgmental probability elicitation tend to be unstable and often to violate basic axioms of probability. Because the judgmental probabilities elicited for component events are almost always conditional on the occurrence or non-occurrence of other events in the event schema, it is useful to use these numerical values to calculate total probabilities and then to review implications with the panel.

Intuitive approaches that simply ask people what they think a probability to be tend to make no attempt to foster careful weighing of judgmental opinions or to disassociate events from their consequences. One result, for example, is that people tend to overestimate the probability of events having a favorable consequence and to underestimate the probability of those with an unfavorable consequence. Intuitive approaches make use of specific numerical scales, and these may be suggestive of things not intended.

21.6.6.3 Action approach to elicitation

The degree-of-belief theory is based on the notion that judgmental probability and actions are inseparable: judgmental probabilities may be inferred from behavior, but they are not necessarily intuitive. As a consequence, simply asking an expert what he or she judges a probability to be will not necessarily result in an accurate assessment of his or her judgmental probability (i.e. as might be reflected in a situation requiring the expert to make a decision). Thus, to assess judgmental probability accurately one needs some form of 'action' approach in which probabilities are inferred from behavior in a controlled situation.

The action approach to assessment presents subjects with comparisons and decisions, as the well-known urn example asks a person to compare and thus gauge the uncertainty about a particular event with that about drawing colored balls from an urn. It places gambles involving the probabilities to be assessed. In the literature, the hypothetical gamble is called a 'reference lottery.' Judgmental probabilities are inferred from the resulting decisions. The expert is given the choice between two lotteries: one presents a

probability p of winning a significant cash prize, C, with a complementary probability $(1 - p)$ of wining nothing; the other presents the same cash prize C if a discrete event A occurs, and nothing if A does not occur. The expert is asked to adjust the value of p until he or she is indifferent between the two lotteries. The resulting p is assumed to be the same as the judgmental probability of A.

Consider a dam site at which one potential mode of troublesome performance involves the unlikely but possible existence of a fault or shear zone in the rock formation under the dam. If such a zone were there, problems of uplifting and potential internal erosion would be of increased concern. Some amount of site characterization has been carried out, but the results are not definitive, and the resulting uncertainty is sought as part of a risk assessment. That uncertainty depends in part on the prior experience of the expert making the judgment and in part on the inconclusive information from the site characterization program. How might this judgmental uncertainty of the fault existing be quantified?

Were no other information available (e.g. regional frequencies of shear features), the expert's uncertainty about the fault can be approximately measured by comparison to familiar events. For example, if the expert prefers to bet on the toss of a fair coin rather than on the existence of the fault, by implication his or her judgmental probability of the fault existing must be less than 1/2. Should he or she prefer to bet on the existence of the fault over the roll of a six-sided die, then the judgmental probability of the fault existing must be greater than 1/6, and so forth, leading to a bounding of the uncertainty. Changing the payoff odds on the gambles is another means for bounding the assessment.

Research on expert elicitation has addressed a number of methodological issues of how probability questions should be formulated. For example, should questions ask for probabilities, percentages, odds ratios, or log-odds ratios? In dealing with relatively probable events, probabilities or percentages are often intuitively convenient to experts; but in dealing with rare events, odds ratios (such as '100 to 1') may be easier because they avoid very small numbers. Also, do devices such as probability wheels – which spin like a carnival game and represent probability as a slice of the circle – help experts visualize probabilities? Definitive conclusions from research are lacking, and in the end, facilitators and experts must pick and choose a protocol that is comfortable to the individuals involved.

21.6.6.4 Quantifying judgmental probability distributions

Not all uncertain quantities involve simple probabilities of discrete events. Many of the parameters needed for engineering analysis have a realized value along a scale, and the issue facing expert elicitation is to assess the uncertainty about the parameter's value over that scale. For example, the base friction between a concrete mass and its foundation, measured as a drained friction angle, could in principle have a value anywhere between zero and ninety degrees. Uncertainty about the interval of this scale in which the actual value lies is described by a probability distribution. A probability distribution summarizes the relative uncertainty about the parameter's value lying within specific intervals of the scale. In expert elicitation it is often convenient to represent probability distributions as cumulative functions, which graph the scale of the parameter along the horizontal axis and the (judgmental) probability that the realized value of the parameter is less than specific values along the vertical axis.

The process starts by asking the expert to suggest extreme values for the uncertain quantity. It is useful to have the expert describe ways that values outside these extremes might

occur. Then, the expert is asked to assess probabilities that values outside the extremes occur. Starting with extreme values rather than best estimates is important in guarding against overconfidence and anchoring. Asking the expert to conceive extreme scenarios makes those scenarios 'available', and allows the expert to think about the extremes more readily. As numerical values are elicited, the facilitator should begin plotting these on graph paper; however, at this point the plot should not be showed to the expert, because it might bias future responses to conform to the previous ones. As ever more assessments are made, these are plotted on the graph to begin establishing bounds and to point out inconsistencies.

After establishing extreme values and their exceedance probabilities, the next step is to fill in values and their corresponding probabilities between the extremes. As in the beginning, it is good practice to avoid early assessment of the mean or mode and rather to focus on values without particular significance to the expert. The judgmental probability that the realized value of the uncertain quantity is less than a specific number can be elicited using comparative gambles, a probability wheel, or other aid, as discussed above. The plot of assessed values is used to identify gaps where more assessments are needed, and inconsistencies in the expert's responses where further assessments or clarification is called for. Toward the end of the process, the interval technique might be used to elicit the median and quartiles of the distribution.

In checking for consistency, it is useful to compare numerical results elicited as values with those elicited as probabilities. In the *fixed probability* approach, the expert is given a probability, and asked for a corresponding value of the uncertain quantity; or given a probability interval, and asked for corresponding ranges of the uncertain quantity. For example, "what value of friction angle do you think has a 1/3 chance of being exceeded?" "What values of friction angle do you think have a 50:50 chance of bounding the true value?" In the *fixed value* approach, the expert is given a value of the uncertain quantity and asked the probability that the true value is less than that value, or the expert is given a range of values and asked the probability that the true value lies within that range. In the *interval* approach, the expert is asked for the median of the uncertain quantity, then the half points (quartiles) of each resulting interval, then the next half points (octiles), and finally the 0.01 and 0.99 points.

Limited research suggests that, in general, fixed value procedures produce probability distributions that are more diffuse and usually better calibrated than do fixed probability or interval procedures. Furthermore, for reasons of avoiding overconfidence, the interval approach is discouraged (Baecher 1972).

21.6.6.5 Normalized frequency approach to estimating probabilities

The normalized frequency approach to assessing a judgmental probability starts with an observed, empirical frequency of similar events and adjusts those background rates either up or down to reflect local conditions.

A number of issues arise in using the normalized frequency approach. The first of these is identifying the relevant subcategory of events or processes in the inventory to which the present project relates. The collection of items in the inventory is heterogeneous and often of insufficient size from which to draw firm statistical conclusions. Unique aspects of the project at hand may make it difficult to select a sufficiently similar subset within the inventory from which to infer base rate frequencies. Secondly, failures or incidents

within the inventory seldom display simple cause and effect. Therefore, it is often difficult to isolate the empirical frequency of narrowly defined events and processes. Thirdly, the calculation procedures for adjusting base-rate frequencies to the peculiarities of the current project are themselves not without difficulties. On the other hand, the normalized frequency approach has intuitive appeal in that it begins with empirical frequencies.

Two approaches have been used for adjusting base-rate frequencies either up or down to account for local conditions. Each of these starts with a list of factors to differentiate the current project from the aggregate of projects in the inventory, that is, the factors that make a particular item either better worst then average. The question is how to combine these unique factors with the base-rate frequencies. The less formal approach is to make adjustments directly, through intuition and discussions among the expert panel. A significant hazard to this informal approach is that it exacerbates the anchoring bias discussed below. Nonetheless, experts often find direct, intuitive adjustments reasonable.

The more formal procedure is to use Bayes' Theorem to update the base frequency. Bayes' Theorem provides a calculation procedure for formally combining prior probabilities (that is, base rates) with other information that can be captured in likelihood functions (that is, conditional probability statements). To apply Bayes' Theorem, first features are identified which distinguish the current project from those in the catalog. Then, estimates are made of the conditional probability of these features being associated with a project performing adversely, and correspondingly the conditional probability of those features being associated with a project not performing adversely. The ratio of these two conditional probabilities, the likelihood ratio, is multiplied by the ratio of the base rate frequency and its complement (that is, one minus the base rate) to obtain an updated odds ratio. While, in principle, Bayes' Theorem provides a vehicle for quantitatively updating a base rate for project specific factors, in application it requires that the reliability of indicator factors be known in both a false-positive and false-negative way, which is often difficult to assess.

Consider again the uncertainty surrounding the possible existence of a fault of modest size in the foundation of a proposed dam site. Presume that in the experience of the expert panel, the incidence of such faulting is about 0.2 with projects in similar geology. That is, the base rate, as a first approximation, is $p(\text{fault}) = 0.2$. In the present project, however, a reasonable number of borings has failed to detect faulting. How should the base rate be adjusted? The base rate can be updated from Bayes' Theorem, expressed as odds (i.e. the ratio the probability in favor to the probability against, $p/(1-p)$):

$$\frac{p(\textit{faulting}|\textit{data})}{1-p(\textit{faulting}|\textit{data})} = \frac{p(\textit{faulting})}{1-p(\textit{faulting})}\,\frac{p(\textit{failed to find}|\textit{faulting})}{p(\textit{failed to find}|\textit{no faulting})} \tag{21.1}$$

The expression reads, the "odds ratio that faulting exits given the observed data, equals the product of the odds ratio before observing the data, times the ratio of the likelihoods (the likelihood ratio) of the data presuming that faulting were present compared to the likelihood if it were not. Let us say that the expert panel, using geometric arguments, estimates that the probability that the boring grid would have successfully detected existing faulting is 2/3. Conversely, if no faulting existed, the probability is by definition, zero. Thus,

$$\frac{p(\textit{faulting}|\textit{data})}{1-p(\textit{faulting}|\textit{data})} = \frac{0.2}{1-0.2}\,\frac{1-0.67}{1.0} = 0.0825 \tag{21.2}$$

So, the site characterization data reduces the odds from 1:4 to 1:12, and the corresponding probability from 0.2 to 0.076.

21.6.6.6 Reliability modeling for assessing probabilities

For some component events, engineering models are available for predicting behavior. In these cases, reliability analysis can be used to assess probabilities associated with the components. Reliability analysis propagates uncertainty in input parameters to uncertainties in predictions of performance. The assessment problem is changed from estimating probabilities of adverse performance directly to estimating probabilities for the input parameters. Once probabilities for the input parameters are assessed, any of a variety of simple mathematical techniques can be used to calculate probabilities associated with performance. Among these are first-order second-moment approximations, advanced second-moment techniques, point-estimate calculations, or Monte Carlo simulation. Sometimes, experts elect to assess an additional component of uncertainty in the reliability analysis to account for model error. While there are many ways to do this, the most common is to assign a simple, unit-mean multiplier to the model output, having a standard deviation estimated by the experts to reflect model uncertainty.

Experience with panels of experts suggests that model uncertainty is among the least tractable issues dealt with. The difficult questions about model uncertainty have to do with underlying assumptions, with conceptualizations of physical processes, and with phenomenological issues. Experts tend to have strongly held beliefs on such matters, so that discussions can become intense. Nonetheless, model uncertainty is a critical aspect of risk assessment. Most models engineers deal with in their daily work were developed for design purposes. They deal with incipient failure conditions and with assuring that loads and resistances remain within working ranges. Risk assessment deals with adverse performance and failures. Thus, models which were developed to prevent the precursors of failure are now used to model failure processes themselves. Failure processes involve strongly nonlinear behaviors in considerations of time rates and sequences. Traditional engineering models may require a good deal of manipulation to draw conclusions about failure processes, and this typically requires a good deal of qualitative reasoning by experts.

21.6.6.7 Correlations among uncertainties

Most experts find the notion of judgmental probability intuitively reasonable and, with practice, develop proficiency at assessing their own uncertainty about individual events or parameters. On the other hand, most people, whether experts or not, have difficulty thinking about the correlations among the uncertainties pertaining to different events or parameters (*viz.* Alloy and Tabachnik 1984). Correlation means that a person's uncertainty about one event or parameter is affected by knowing whether another event occurred or by knowing the value of another parameter.

People usually require significant analytical assistance when grappling with correlations, so it is good practice not to be overly aggressive in trying to assess correlations. The easiest way to assess probabilistic dependence between two uncertain quantities, x_1 and x_2, is first to assess the conditional probabilities for x_2, assuming various values for x_1, then

to assess the marginal probabilities for x_1 (i.e. irrespective of the value of x_2). The joint probabilities are found from the relationship

$$P(x_1, x_2) = P(x_2|x_1)P(x_1) \tag{21.3}$$

in which

$P(x_1, x_2)$ = the probability of x_1 and x_2 occurring together,
$P(x_2|x_1)$ = the probability of x_2, given the value of x_1, and
$P(x_1)$ = the probability of x_1 irrespective of the value of x_2.

This approach requires that multiple assessments of the conditional probabilities $P(x_2|x_1)$ be made for various values of x_1, but the advantage is that the expert does not have to grapple explicitly with the concept of correlation coefficients, which tend not to be intuitive. The reverse conditional probabilities, $P(x_1|x_2)$, which are often needed for the risk assessment, can be calculated using Bayes' Theorem.

In practice, it is better to attempt to restructure a problem rather than to assess correlations among uncertainties. This can be done, for example, when two uncertainties are correlated because they each depend on some third uncertainty, as in the case of downstream costs of flooding. Emergency mobilization costs and property damage costs in the future may be correlated because each depends on inflation. It would be more effective to assess the uncertainties in each cost conditioned on inflation, and then to combine the two independent assessments, rather than to attempt to assess the correlated behavior of the two uncertainties. Of course, not all correlated uncertainties can be handled in this convenient way.

21.6.6.8 Verifying assessed probabilities

Once a set of probabilities has been elicited, it is important to check for internal consistency, which statisticians sometimes call 'coherence.' Coherence means that the numerical probabilities obtained are consistent with probability theory. This can be done, first, by making sure that simple things, such as the probabilities of mutually exclusive and collectively exhaustive events adding up to 1.0, are confirmed. Secondly, it is also good practice to reword or restructure questions in logically equivalent ways to see if the resulting answers change. Also, redundant questions can be asked about events or parameters whose probabilities could be calculated from previous answers. Inconsistencies are resolved by discussion and reconsideration.

Beyond coherence, the implications of the elicited probabilities for risk estimates and for the ordering of one set of risks against other sets is also useful feedback to the experts. Sometimes, seeing unexpected or unanticipated implications of a set of assessments causes experts to reflect again upon the answers that have been given and the probabilities that have been elicited.

21.6.7 Documentation Phase

Documentation is an essential part of an exercise in expert elicitation. In the first place, a great deal of time and money has been invested in the elicitation, and the sponsoring organization needs to know what it has obtained. Perhaps more important is the fact that many expert elicitations take place in a contentious regulatory or political environment.

Documentation consists of records of everything that was done, from the earliest expert training to the final report. Much of this material can be saved in files that are accessible to interested parties, but it is important that the reports describe the process and the rationale behind major decisions so that they can be reviewed without searching the files. Budnitz *et al*. (1998) give a detailed description of documentation for expert elicitation for seismic hazard analyses for nuclear power plants.

21.7 Practical Suggestions and Techniques

Elicitation of subjective probabilities is best conducted with the assistance of probability analysts or facilitators versed in probability and decision theory, behavioral effects, and ideally geotechnical engineering as well. Their job is principally one of ensuring external validity and promoting internal consistency of the probabilities produced, much of which involves a process of prompting and querying the assessor. For example, probabilities of 0.4 and 0.8 for the presence of either soil or rock at a particular location in the ground would fail the requirements of external validity. Further questioning of the underlying reasoning would be necessary either to reconcile the discrepancy, or to reveal hidden assumptions (perhaps a perceived potential for residual weathering products at the soil/rock interface) which could then be made explicit by restructuring the problem with a revised set of possible outcomes.

Formal elicitation techniques customarily rely on education and training of assessors for reducing bias, under the precept that an assessor conscious of bias can and will compensate for its effects. General knowledge questionnaires are sometimes used to demonstrate overconfidence on a personal level. However, research findings on the effectiveness of such *debiasing* techniques are mixed, with some (Alpert and Raiffa, 1982) showing subjects to be nearly impervious to training even when offered rewards for unbiased results. Another perhaps more consistently effective debiasing technique for overconfidence is to request that the assessor specify and list the reasons why an elicited probability might be wrong, for instance by imagining reasons that might explain the occurrence of a less-likely outcome after the fact.

Encoding refers to the actual specification of numerical probability values, and various aids are useful in geotechnical applications. One such technique adopted for risk analysis in tunneling applications (Einstein *et al*. 1996) uses a 'probability wheel' device that separates a circle into adjustable sectors of different color. A visual reference is obtained by adjusting the sector size until the assessor is indifferent between its relative proportion of the circle and the likelihood for the outcome in question.

Like people in general, most assessors express uncertainties more readily verbally than they do numerically, and another encoding technique used extensively in dam safety risk analysis exploits this trait using transformations from verbal to numerical designations (Vick 1997). Table 21.1 shows conventions defined for mapping of verbal expressions of uncertainty to probability values. The table also provides median values and ranges for these and similar terms from a considerable body of behavioral research summarized by Reagan *et al*. (1989), who show that not only do people adopt these transformations in fairly consistent ways but that they also use them commutatively in converting both words to numbers and numbers to words.

The conventions help promote consistency among likelihood judgments expressed by a particular assessor and from one assessor to another. In practice, the verbal expressions

allow encoding to proceed along the lines of a multiple-choice exercise in first establishing a bounding range on verbal expressions by eliminating those which do not apply, then converging on the most appropriate one(s), followed by refinement of an interpolated probability value as needed. The defined conventions are purposefully truncated at probabilities of 0.01 and 0.99, since cognitive discrimination limitations usually produce little meaning for probabilities outside this range. At the same time, this maintains encoded values within or at least reasonably near the well-calibrated range and thereby provides some structural limitation on the effects of overconfidence bias. These constraints are relaxed in some circumstances, such as when indicated by further decomposition or by applicable base-rate frequency information.

21.8 Summary

It many practical instances the only way to obtain estimates of probability distributions is by expert elicitation. Although this seems at first sight to be a relatively simple matter, a large body of experience in a wide variety of disciplines has shown that experts are subject to many subtle pressures and make many errors. They tend to be overconfident. They tend to be influenced by earlier estimates. They often do not understand the statistical implications of their estimates. To mitigate these problems, a set of procedures has been developed, originally in the social sciences but now extending into many branches of engineering such as seismic hazard analysis and dam safety studies. Effective use of these measures requires the use of experienced facilitators.

22 System Reliability Assessment

Up to this point we have mostly considered failures of individual soil or rock structures, such as the instability of a slope or excessive settlement of a foundation. In many cases, we are interested not only in the performance of individual structures, but also in the way a group of structures performs as a whole system. For example, a tank farm for the storage of petroleum products comprises many tanks in which oil is stored. Some of these tanks are large, some are small; some are used to store oil, others are used to store other distillates. These tanks are typically grouped within patios surrounded by fire walls. The purpose of the patios is to trap oil that might be spilled should a tank fail, to separate different types of spilled products, and to contain a fire, should a spill ignite. The entire site may also be surrounded by an external fire wall, whose purpose is to contain oil and fires on site even if individual patios are breached. How safe is this system of facilities in the face of seismic ground shaking? Clearly, the safety assessment must consider more than the isolated behavior of the individual tanks, patios, and firewalls. Safety of the system depends on how all these structures work together.

22.1 Concepts of System Reliability

Most system reliability studies begin with an initiating event, that starts a sequence of subsequent events, that ultimately leads to failure of the overall system. Some examples of natural hazards that serve as initiating events are earthquakes, floods, and hurricanes. Things other than natural hazards may also be initiating events; for example, excessive settlement in a manufacturing facility may cause equipment failure and simultaneously disrupt utility services needed to deal with the equipment failure. Failures of the latter type are of considerable concern in assessing the safety of nuclear power plants.

The steps involved in a system reliability assessment are:

1. Identify initiating events and determine the probabilities of these events.
2. Quantitatively define the meaning of failure of the system.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

3. Develop quantitative models of the performance of individual components.
4. Identify mechanical interactions among component failures and failure modes.
5. Investigate statistical or probabilistic correlations, if any, among component failures and failure modes.
6. Integrate component performance models, interactions, and correlations within an overall system performance model.
7. Calculate numerical results for system reliability.

22.2 Dependencies Among Component Failures

The interdependencies of component failures or failure modes, whether caused by mechanical interaction as in step four or by correlation as in step five above, are extremely important. Consider the design of Figure 22.1, in which one tank in a patio is surrounded by a firewall to contain leaks. Presume that the annual probability of the tank failing and spilling its contents into the patio is $p_T = 0.01$ and that the overflow capacity of the patio is sufficient to retain the full volume of one tank. For oil to leak out of the patio, the tank must fail, and then the firewall must fail, too. Let the probability of the firewall failing given an oil load behind it be $p_F = 0.01$. The joint probability of both the tank and firewall failing, presuming the probabilities independent, is the product, $Pr\{\text{oil loss}\} = P_T P_F = 0.0001$, a fairly small number. However, what if liquefaction of the site caused by seismic ground shaking had an annual probability of occurring of 0.001, and should liquefaction occur, both the tank and firewall would fail? The probability of this failure is then 0.001. While the probability of liquefaction is a small contributor to the annual risk of tank failure alone, the probabilistic dependence it causes between tank and firewall failure increases the annual probability of loss of oil off the site (system failure) by a factor of ten.

Dependencies in component failure probabilities can arise in at least three ways:

1. Mechanical interaction among failure modes (e.g. the tank fails and in so doing uproots the soil under the firewall, and the wall then fails, too).
2. Probabilistic correlation (e.g. a common initiating event affects both the tank and firewall).

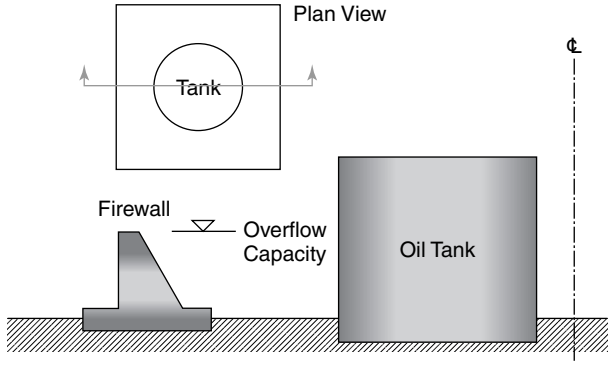


Figure 22.1 Typical tank farm geometry.

3. Statistical correlation (e.g. uncertainty about the consolidation coefficient of the foundation soils affects the performance of the tank and firewall in the same way; excessive settlement of each occurs together).

The last two categories are often called *common mode* failures.

22.3 Event Tree Representations

The most common way of decomposing a geotechnical risk assessment is by building an *event tree*, which we described in Chapter 20. An event tree starts with some initiating event, and then considers all possible chains of events that could lead from the first event. Each chain of events leads to some performance of the system. Some of these chains of events lead to adverse outcomes; some do not. For each event in the tree, a probability is assessed presuming the occurrence of all the events preceding it in the tree, that is, a conditional probability. The total probability for a particular chain of events or path through the tree is found by multiplying the sequences of conditional probabilities.

In this way, we can build an event tree for the oil storage patio of Figure 22.1. This event tree is shown in the upper part of Figure 22.2. The first event is loss of oil from the tank. This occurs with probability p. If the tank leaks, then either the fire wall retains the spilled oil or it does not. Let the probability that the fire wall fails to retain the oil be, q. Note, this probability q depends on whether the tank leaks or not. The pressure of

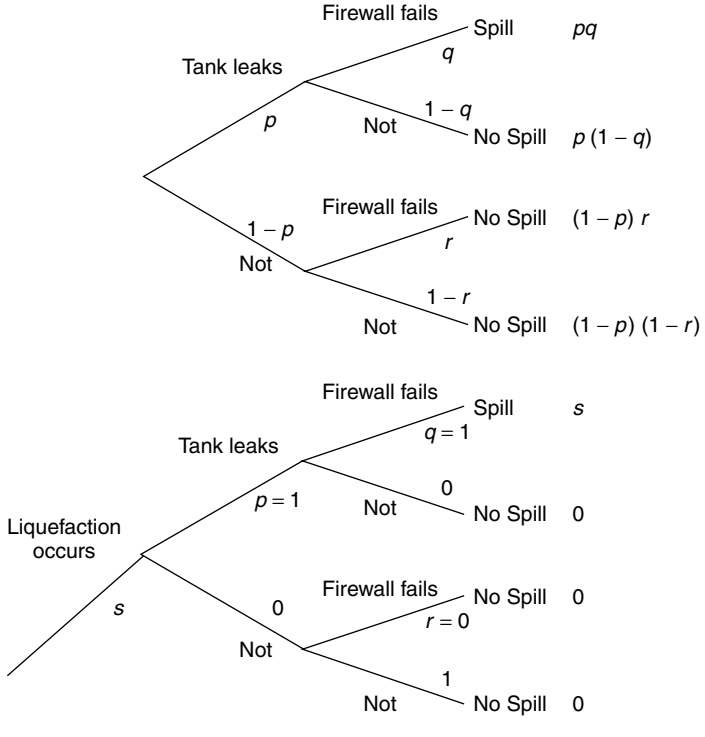


Figure 22.2 Event tree for tank farm failure.

ponded oil against the firewall presumably makes the wall more likely to fail, compared to the case without oil pressure. If system failure is defined as loss of oil off the site, then the only end node in the event tree that includes a failure is that for which both the tank has spilled its oil and the fire wall does not retain the oil. Thus, the probability of system failure is $p \times q$.

Now consider the initiating event that seismic ground shaking leads to liquefaction of the soils underlying the patio. This event tree is shown in the lower part of Figure 22.2. Once liquefaction occurs, the assumption is made that both the tank and the fire wall fail. So, in this case, rather than the two failures being probabilistically independent, they are correlated through the occurrence of a common event which causes each of them to fail at the same time. An event tree serves as a simple way of showing the interrelationship of events in a system failure.

An event tree can be used to decompose a problem at different levels of detail. Bury and Kreuzer (1986) and Vick (1997) describe in simple terms how event trees can be structured for gravity dams. Usually, analytical calculations or judgment are more easily applied to smaller components, and research suggests that more detailed decomposition, within reason, enhances the accuracy of calculated failure probabilities. One reason, presumably, is that the more detailed the event tree is, the less extreme the conditional probabilities that need to be calculated or estimated (Vick 2002).

Whitman (1984) describes a simple event tree that is part of the risk assessment for erosion in an earthen dam. The issue being addressed is scour in the channel downstream of the dam, caused by large releases over the concrete spillway. The spillway is capable of passing large flows, and the natural channel below may be eroded by these large discharges. Headward erosion of the channel may undermine the spillway basin and then possibly undermine the spillway itself. If the spillway fails, this may lead to breaching of the dam directly, or to erosion of an adjacent earth embankment that in turn could lead to breaching of the dam.

The initiating event in this case is a flood discharge of some specified range of magnitudes centered on the probable maximum flood, PMF. The subsequent events leading from this initiating event are shown on the event tree of Figure 22.3. In sequence, these potential events are:

1. The natural downstream channel erodes back to the stilling basin, causing scour holes of various depths.
2. The foundation of the stilling basin collapses as a result of a scour hole.
3. Collapse of the stilling basin leads to undermining of the spillway and consequent breaching of the earthen dam.
4. Collapse of the stilling basin leads to erosion of an adjacent earthen embankment and thus to breaching of the dam.

The probability of the initiating event, occurrence of the PMF flood, is established from hydrologic studies. In this case, the probability of the PMF within the design life of the dam was estimated to be 10^{-4}. Based on model hydraulic tests, it was concluded that the natural channel was certain to erode under the discharge of the PMF. Thus, the branch probability at the first node after the initiating event was taken to be 1.0. Using stability calculations and other hydraulic model tests, the various other branch probabilities were estimated and filled into the event tree. Each branch probability is conditional on

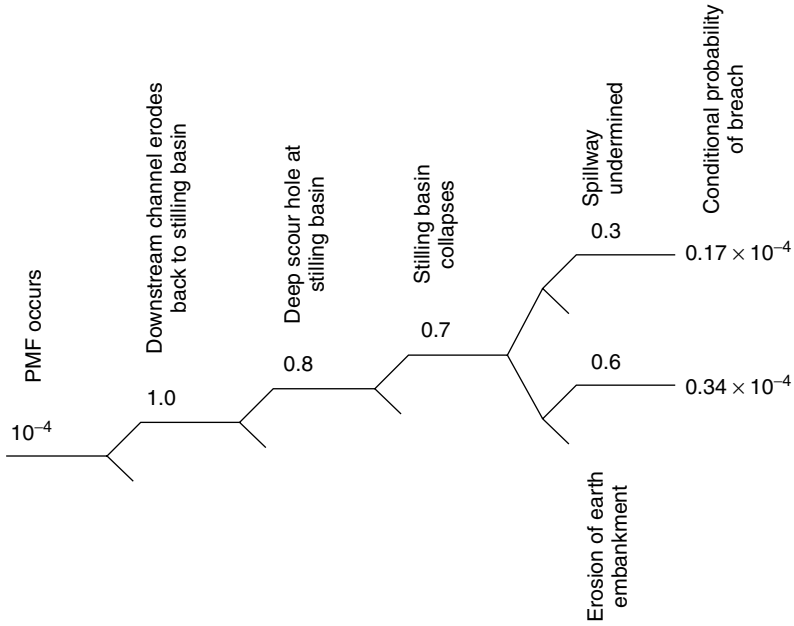


Figure 22.3 Event tree for dam failure. (Whitman, R. V., 1984, 'Evaluating the Calculated Risk in Geotechnical Engineering,' *Journal of Geotechnical Engineering, ASCE*, Vol. 110, No. (2), pp. 145–188, reproduced by permission of the American Society of Civil Engineers.)

the occurrence of events leading into its node. The probability of any path of branches through the tree is found by multiplying the individual branch probabilities. The final result is shown at the right hand side of the figure. The total probability of the earth dam failing by loss of containment is the sum of the probabilities of the two ways in which that failure could occur, or in this case, about 0.51×10^{-4}, not much less than the probability of the initiating event. In other words, if the initiating event occurs, it is reasonably likely (about a 50:50 chance) that the dam will fail.

The event tree provides a convenient way for decomposing a system reliability problem into smaller pieces that are easier to analyze and then provides a vehicle with which to recombine the results obtained for the smaller pieces in a logically coherent way to obtain the reliability of the system itself. Other examples of relatively simple event trees used in geotechnical practice are provided by Vick and Bromwell (1989) for dam failure caused by the collapse of a sinkhole, and by Wu *et al*. (1989) for liquefaction of a sand caused by seismic ground shaking. More complex event trees for existing dams to assess reliability are given by Vick and Stewart (1996) for Terzaghi and Duncan Dams in British Columbia and by Von Thun (1996) for Nambe Falls Dam in New Mexico.

One sub-tree from the US Bureau of Reclamation's event tree for Nambe Falls Dam is shown in Figure 22.4. This particular sub-tree is associated with hydrologic initiating events. The tree is separated into panels. The first simply identifies the loading case. The second panel describes the loading conditions with associated probabilities, and continuing through events that affect the dam. The third panel enumerates the ways in which the dam may react to the various loading conditions (e.g. failure conditions, damage states, and no-failure events). Finally, the fourth and fifth panels enumerate the potential

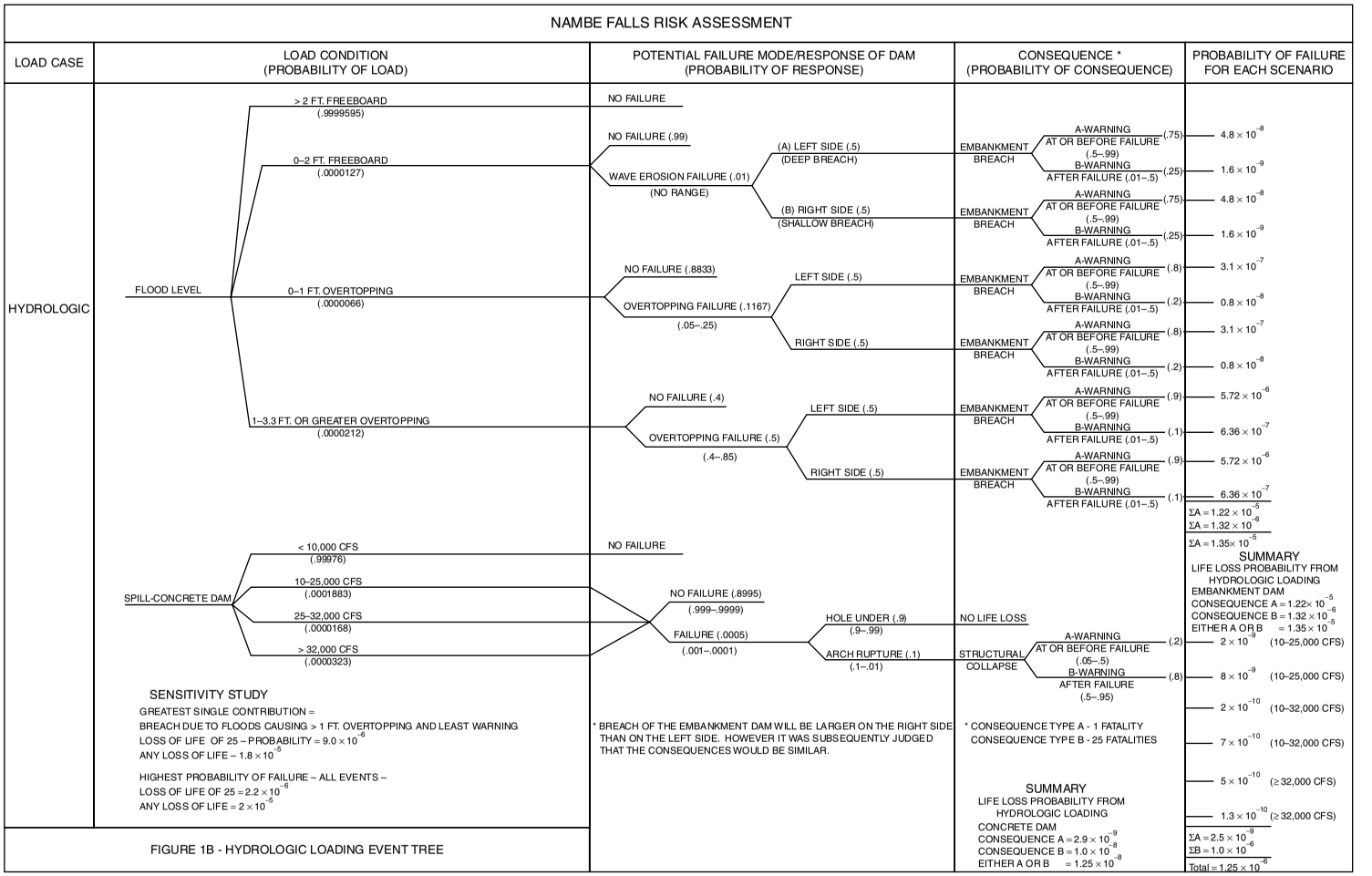


FIGURE 1B - HYDROLOGIC LOADING EVENT TREE

Figure 22.4 Nambe Falls event tree. (Von Thun, J. L., 1996, 'Risk Assessment of Nambe Falls Dam,' *Uncertainty in the Geologic Environment*, Madison, WI, pp. 604–635, reproduced by permission of the American Society of Civil Engineers.)

consequences of each chain events and show the calculated probability associated with each chain.

In most cases, the activity of constructing event trees for a system is in itself instructive, whether or not the resulting probabilities are used in a quantitative way. The exercise requires project engineers to identify chains of events that could potentially lead to a failure of one sort or another. This explicit activity, especially when carried out with a group of people, often leads to insights that might not otherwise have been obvious and therefore that might have been overlooked. In some cases, the event tree becomes a 'living' document that follows the progress of design; it is changed or updated as new information becomes available or as design decisions are changed.

22.4 Fault Tree Representations

Consider again the simple case of the oil storage patio of Figure 22.1. In building an event tree of this system, we started with a spill of oil from the tank, and then considered the subsequent event that the fire wall fails to contain the oil. In building a fault tree of the same system, we start with the system failure, 'oil spilled off site,' and then ask, how this might have happened. For oil to be spilled off site, both the fire wall must fail to contain any spilled oil existing on the site, and the tank must somehow fail to hold the oil that was contained within it.

Common practice is to draw a fault tree from top to bottom as in Figure 22.5. At the top is the system failure condition, 'oil spilled off site.' Beneath are the two faults that need to occur to enable this system failure, namely, 'tank fails to contain oil,' and 'fire wall fails to contain oil.' Since both of these things must occur for the system failure to occur, they are connected in the tree by an 'AND' node, denoted by a bullet shaped symbol with a multiplication sign. This shows that the probabilities are multiplied to obtain the probability of the next higher fault

$$p_0 = p_1 p_2 \tag{22.1}$$

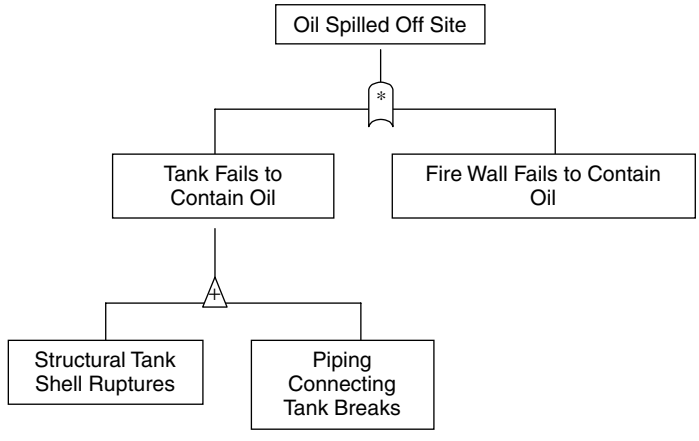


Figure 22.5 Fault tree for tank farm failure.

in which $p_0 = Pr\{\text{oil spilled off site}\}$, $p_1 = Pr\{\text{tank fails to contain oil}\}$, and $p_2 = Pr\{\text{fire wall fails to contain oil}\}$.

One continues to decompose the fault at each level into the contributing faults that would cause it to occur. For the fault, 'tank fails to contain oil,' to occur, one could assume that either of two other faults might have to occur, namely, 'structural tank shell ruptures,' or 'piping connecting tank breaks.' Since only one or the other (or both) of these need occur for containment to be lost by the tank, they are connected by an 'OR' node, denoted with a plus sign. This shows that the probabilities are added to obtain the probability of the next higher fault

$$\begin{aligned} (1 - p_0) &= (1 - p_1)(1 - p_2) \\ p_0 &= p_1 + p_2 - p_1 p_2 \end{aligned} \tag{22.2}$$

in which $p_0 = Pr\{\text{tank fails to contain oil}\}$, $p_1 = Pr\{\text{structural tank shell ruptures}\}$, and $p_2 = Pr\{\text{piping connecting tank breaks}\}$.

The advantage of a fault tree over an event tree is that it focuses only on chains leading to failures; the disadvantage is that it may fail to uncover important combinations of events. For example, it is not clear that the fault tree of Figure 22.5 can identify the hazard posed by liquefaction induced failure of both the tanks and the fire walls. Thus, the advantage of the event tree is that it comprehensively uncovers combinations of events leading to failures, but the disadvantage is that event trees quickly become bushy messes.

Fault trees have been comparatively less used in geotechnical risk and reliability than have event trees. One geotechnical application in which fault trees have been widely used, however, is safety assessment for nuclear waste disposal facilities (e.g. Arthur D. Little 1978; Battelle Pacific Northwest Laboratories, Inc. 1981). Presumably, the close connection of this enterprise to the nuclear industry, where fault tree analysis is common, explains the phenomenon.

Van Zyl *et al*. (1996) used fault tree analysis to assess the risk associated with tailings impoundment in a region of Ireland underlain by geologically old karst formations. The study defines system failure as 'uncontrolled release of tailings from the impoundment as a result of a karst cavity,' and developed the fault tree show in Figure 22.6 to structure the analysis. At the top level, the tree is separated by an OR node between the fault, 'tailings embankment failure,' and 'liner system failure.' The tree in the figure shows just the first half, associated with embankment failure. The boxed faults are those for which a further decomposition is made; the circled faults are those for which a probability is calculated or assessed directly, and they are called *basic faults* (i.e. they are not further structured).

The rationale for estimating probabilities associated with a partial list of the basic faults in the analysis is presented in Table 22.1. Analytical models were used to calculate probabilities associated with some basic faults, and subjective estimation was used to estimate others. For example, the probability of a solution cavity of sufficient size to cause structural failure occurring under the embankment was calculated by first using reasoning based on soil mechanics to arrive at a critically sized cavity and second assuming that uncertain cavity locations within the karst formation can be modeled as a two-step process. The size distribution of cavities and intact rock units among cavities were modeled with exponential probability density functions (Figure 22.7); then the number of cavities occurring within a given area was modeled with a binomial process (Appendix A). On the other hand, the probability of cavity collapse caused by ground water lowering was based

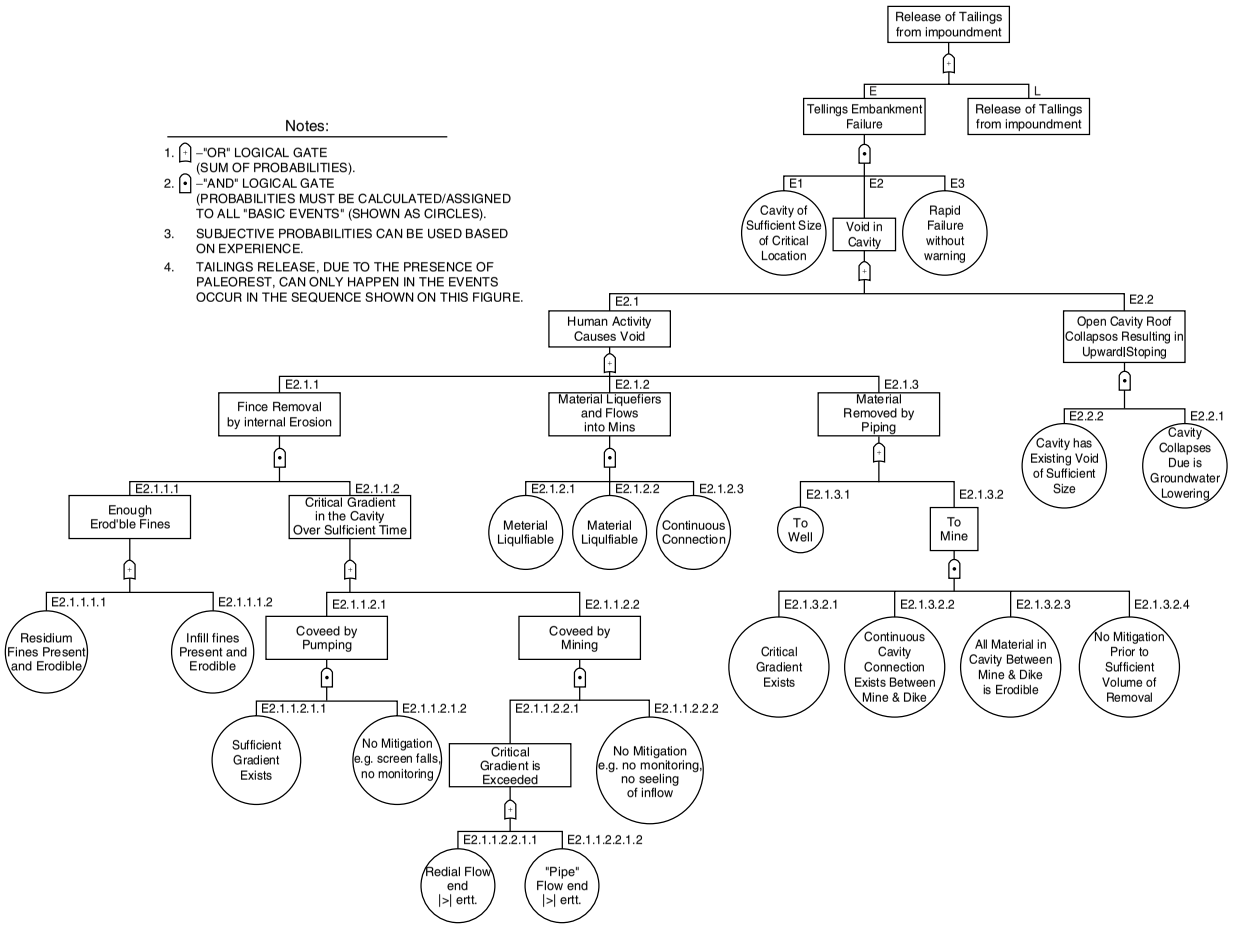


Figure 22.6 Fault tree for Karst. (Van Zyl, D., Miller, I., Milligan, V. and Tilson, W. J., 1996, 'Probabilistic Risk Assessment for Tailings Impoundment Founded on Paleokarst,' *Uncertainty in the Geologic Environment*, Madison, WI, pp. 563–585, reproduced by permission of the American Society of Civil Engineers).

Table 22.1 Summary of basic events and their probabilities for tailings embankment failure (after Van Zyl, *et al*. 1996)

Basic event	Description	Remarks	Probability
E1	Cavity of sufficient size at critical location	The probability of a cavity at the critical location was evaluated using eight to state Markov process. This analysis was based on data from the exploratory drilling program.	0.048
E2.1.1.1.1	Residuum fines present and erodible	Confirmation drilling indicates that residuum fines are present. Grain size analysis of the residuals indicate clay and sand. The sand material provides a filter for the fine clay material.	0.099
E2.1.1.1.2	Infill fines present and erodible	No evidence of karstic features at Gelmoy. Infill material would consist of peat and glacial till with low erodibility.	0.0001
E2.1.1.2.1.1	Sufficient gradient exists from pumping	Higher flow gradient exists close to the wells. Gradient at a distance from the wells is less than critical.	0.001
E2.1.1.2.1.2	No mitigation e.g. screen fails, no monitoring	Wells will be constructed to prevent removal of fines. Discharge of dewatering will be monitored for turbidity.	0.001
.1.1.2.2.1.1	Radial flow and i > icrit	Average gradient is about all 0.04 using a 'capacity-demand' model for the critical gradient for fine sand a probability is calculated.	0.0007
E2.1.1.2.2.1.2	Pipe flow and i > icrit	Unlikely that a continuous pipe exists therefore, the gradient is taken as the effective gradient of a 0.09.	0.0005
E2.1.1.2.2.2	No mitigation, e.g., no monitoring, and sealing of inflow	Mine inflows will be mitigated through regular monitoring during operations.	0.25
E2.1.2.1	Material liquefiable	Requires a 10% size of a 0.01 to 0.25 mm and a relative density of less than 5%. It is highly unlikely that continuous zones of such material exists.	0.001
E2.1.2.2	material liquefies	Requires stress conditions including rapid strain or dynamic loading such as blasting. Infill material would have to be submerged. Low probability of occurrence as blast loading attenuates rapidly.	0.001
E2.1.2.3	Continuous connection	Continuous connection would require mature continuous karstic cavity in Waulsortian formation. the Gelmoy area does not contain mature karst.	0.001

Table 22.1 (*continued*)

Basic event	Description	Remarks	Probability
E2.1.3.1	Material removed by piping from well	Well pack and screens designed to prevent removal of fines. Much local experience, negligible erodibility	0.001
E2.1.3.2.1	Critical gradient exists	A high gradient would have to exist throughout the entire length of the continuous cavity. For a continuous infilled connection the gradient would decrease with distance from the exit point.	0.001
E2.1.3.2.2	Continuous cavity connection exists between mine and dike	Continuous connection would require mature continuous karstic cavity in Waulsortian formation. The Gelmoy area does not contain mature karst.	0.001
E2.1.3.2.3	All material in cavity between mine and dike is erodible	Would require a continuous connecting of erodible materials. Various materials ()clay, sand, gravel, and peat) would have varying degrees of erodibility.	0.001
E2.1.3.2.4	No mitigation prior to sufficient volume of removal	For sufficient volume of material to be removed, considerable fill material would be deposited within the drift. Monitoring in the underground working would detect this condition. Difficult to initiate large opening.	0.25
E2.2.2	Cavity has existing void of sufficient size	Upward stopping could result if a cavity of critical volume is located beneath the embankment. Confirmation drilling indicated a low occurrence of cavities filled with void apace.	0.001
E2.2.1	Cavity collapse due to groundwater lowering	Collapse of cavity at the bedrock surface would most likely occur during lowering of the groundwater table prior to construction. Failure at a later time is less likely	0.01
E3	Rapid failure without warning	Settlement beneath compacted earthen structures typically result in signs of distress. Daily monitoring will be conducted to detect signs of distress.	0.01

on the subjective expert opinion of a group of geologists, using qualitative observations of the rate of cavity collapse in paleokarst, and reasoning from rock mechanics principles.

The study concluded that the probability of tailings release due to embankment failure is of the order 5×10^{-7}, and due to liner failure, of the order 2.5×10^{-6}. Thus, the incremental probability of failure of the tailings embankment due to karst is small compared

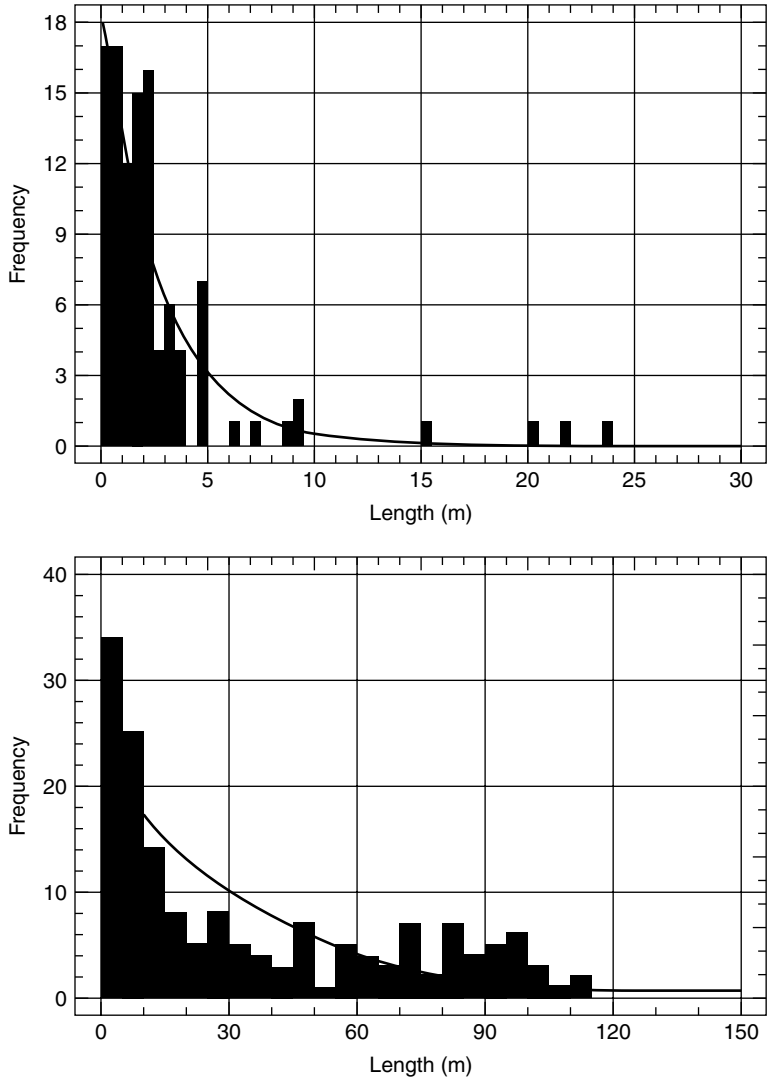


Figure 22.7 Exponential distribution for Karst analysis. (Van Zyl, D., Miller, I., Milligan, V. and Tilson, W. J., 1996, 'Probabilistic Risk Assessment for Tailings Impoundment Founded on Paleokarst,' *Uncertainty in the Geologic Environment*, Madison, WI, pp. 563–585, reproduced by permission of the American Society of Civil Engineers.)

with commonly accepted probabilities of failure of other water retaining structures such as earth dams (i.e. 10^{-4}).

22.5 Simulation Approach to System Reliability

Event and fault trees approach system reliability by developing a logical structural model of the system built around events. A geotechnical structure such as a dam, foundation,

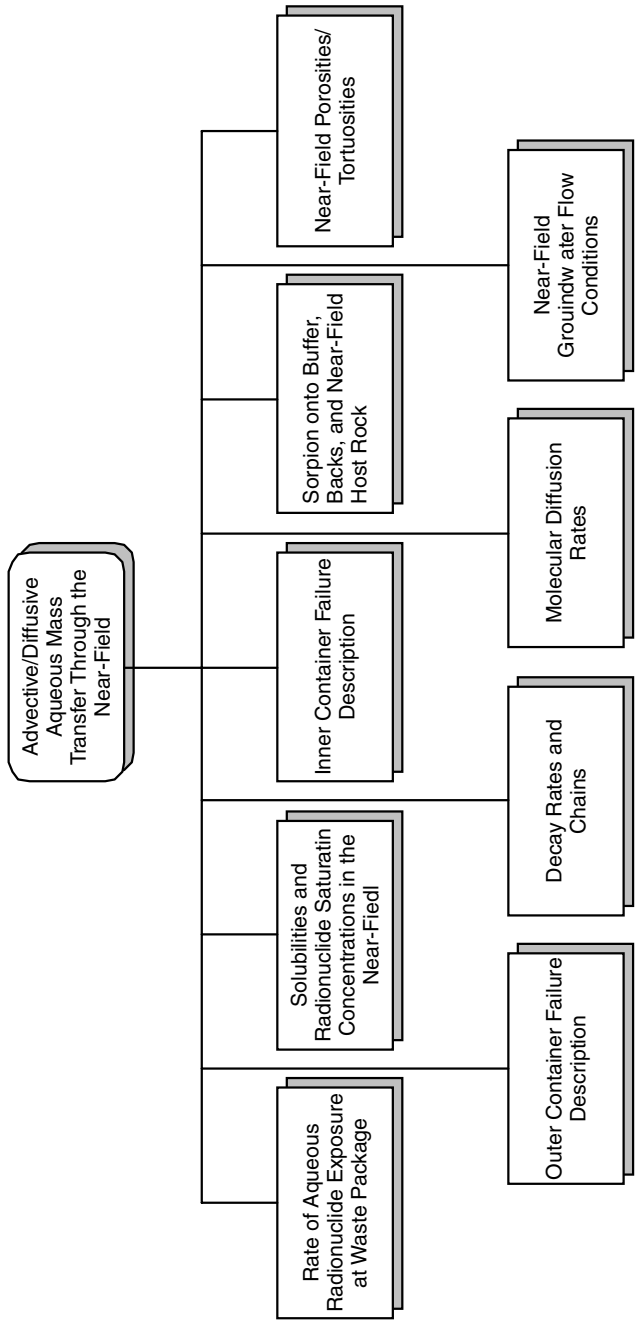


Figure 22.8 Factors contributing to repository simulation. (Miller, I. and Kossik, R., 1996, 'Probabilistic Simulation of Geologic Waste Disposal Facilities Using the Repository Integration Program (RIP),' *Uncertainty in the Geologic Environment*, Madison, WI, pp. 944–964, reproduced by permission of the American Society of Civil Engineers.) (Also Miller *et al.* 1992.)

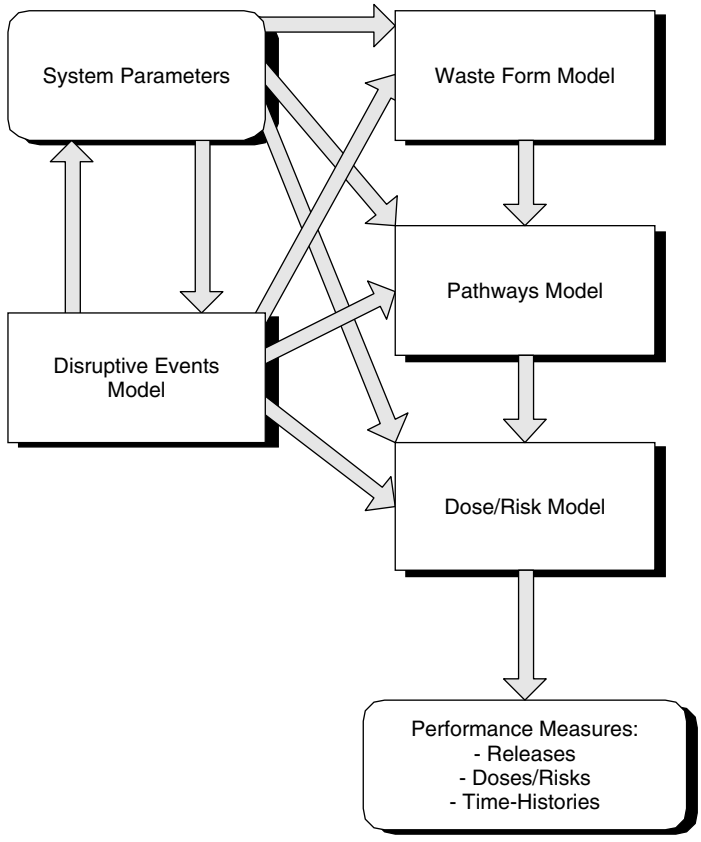


Figure 22.9 Interaction diagram for repository simulation. (Miller, I. and Kossik, R., 1996, 'Probabilistic Simulation of Geologic Waste Disposal Facilities Using the Repository Integration Program (RIP),' *Uncertainty in the Geologic Environment*, Madison, WI, pp. 944–964, reproduced by permission of the American Society of Civil Engineers.) (Also Miller *et al*. 1992.)

or tunnel is represented not through models of engineering mechanics but in an abstract relationship of events. Probabilities are assigned to these events in one manner or another, and then a calculation is made combining these probabilities with the logic-tree structure assumed to apply to the events (i.e. the event or fault tree), and an analytical probability of system failure is calculated. A limitation of this approach is that, for complex systems, the logic tree becomes large and messy. Similarly, the analyst may overlook important interrelationships or functional dependencies among events not obvious by inspection, even to the trained observer.

Simulation approaches to system reliability use a mathematical model of the "physics" (in our case, engineering mechanics) of the system as a vehicle for experimentation. The simulation is subject to a wide range of parameter values and boundary and initial conditions and is run many times to assess system performance. Incorporating randomness in the parameter values and boundary and initial conditions leads to variability in system response. A statistical sampling approach is used to draw conclusions about the reliability of the system. Often, uncertainties in the models used to calculate performance (i.e. model

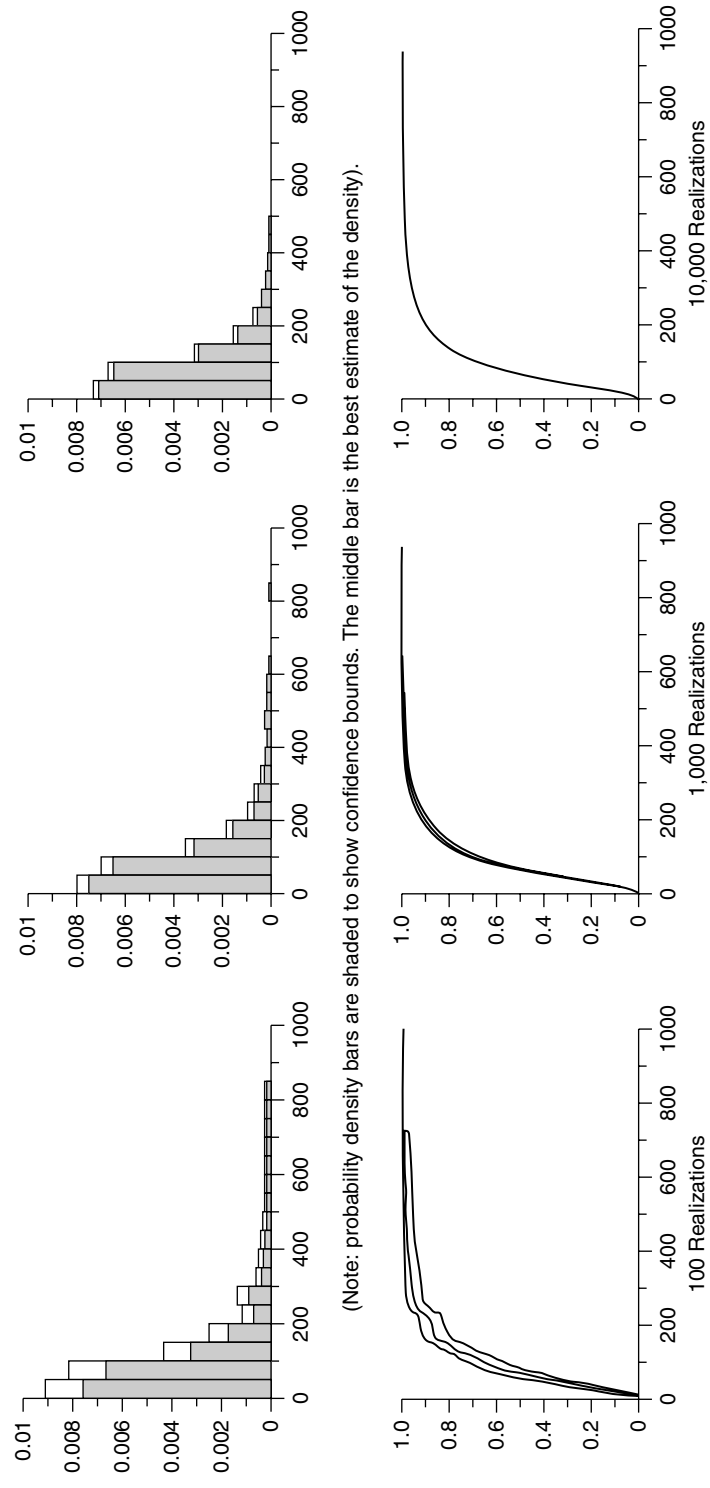


Figure 22.10 Output of repository simulation. (Miller, I. and Kossik, R., 1996, 'Probabilistic Simulation of Geologic Waste Disposal Facilities Using the Repository Integration Program (RIP),' *Uncertainty in the Geologic Environment*, Madison, WI, pp. 944–964, reproduced by permission of the American Society of Civil Engineers.) (Also Miller *et al.* 1992.)

uncertainty) are also included in the calculations. The simulation approach to system reliability is an extension of the Monte Carlo methods discussed in Chapter 17 to more complex applications.

In the simulation approach, a large number of equally likely runs (or *trials*) are made. These are viewed as potential realizations of the future performance of the system. Using sampling theory and statistical inference, the uncertain future performance is summarized in probabilistic descriptions, and investigations can be made of the influence of individual uncertainties or combinations of uncertainties on that future performance. It has become customary to display the outcomes of simulation studies as Cumulative Density Functions (CDF) or Complementary Cumulative Density Functions (CCDF) of performance variables. The CDF expresses the probability of an outcome being less than or equal to some value (Appendix A). The CCDF is the complement of the CDF, and expresses the probability of an outcome being greater than some value.

Miller *et al*. (1992, 1996) performed a large system simulation of a geologic disposal facility for high-level radioactive waste. In this modeling effort, major components of the repository system – such as the container system isolating the radioactive materials, the response of excavated openings to thermo-mechanical loadings, or the deep groundwater flow regime about the repository – are modeled using deterministic procedures, but with randomly generated input parameter values, and sometimes with randomly generated initial conditions. The interrelationships among the various models in the overall simulation of the repository are shown in Figures 22.8 and 22.9. The random variables used to characterize uncertainty in model parameters relate to the degree of uncertainty about soil or rock mass conditions or about other processes important to repository performance (e.g. waste form properties).

The system model has four principal components: disruptive events model, waste form model, pathways model, and dose/risk model. The outputs of the simulation are releases of radiation over time and human or ecosystem doses and toxicological risk over time. The results of the realizations are summarized statistically, as shown in Figure 22.10, in histograms and CDFs. The bounds show 5% and 95% statistical confidence limits on the inferences, in the relative frequentist sense (Chapter 3).

It is sometimes the case that the inputs and outputs of individual models used in a simulation relate to the inputs and outputs of the other models. For example, advective flow of ground water around and through the repository can be modeled using finite element procedures, but the boundary conditions of the model depend on the thermo-mechanical response of the rock mass to in situ stresses and to heat generated by the waste canisters. Thus, the outputs of the stress model and of the waste form model may be required before executing the flow model, and conversely if the stress or heat calculations depend on the fluid flow. In such cases an iterative approach may be needed, further increasing the computational burden of simulation.

22.6 Combined Approaches

The TONEN refinery and tank farm sit on reclaimed land in Tokyo Bay near the city of Kawasaki, south of Tokyo. The region is seismically active, and signs of liquefaction in the bay fill have been observed during moderate earthquakes in the area. The refinery and tank farm is a major installation, storing large volumes of oil and petroleum products.

Should foundation failures occur during an earthquake, leading to the loss of stored products off site into Tokyo Bay, both environmental and financial costs could be substantial. The owners of the facility, desiring to manage these risks consistently with other corporate activities and to identify remedial actions that might be taken to reduce risk, needed to quantify the magnitude and character of the risks the facility posed. The risk assessment performed for the site combined event and fault tree analysis with simulation to obtain estimates.

22.6.1 The site

The facility is divided into four principal areas, as shown in Figure 22.11. The present discussion focuses on the KSS site, to the northwest and inland of the larger Sites 200 and 400. Risk assessments were performed on each site, but the KSS analysis is representative. The analysis was carried under the direction of T. W. Lambe & Associates (1989).

The KSS oil storage area is located on 8 m of loose sand and silty sand. The area consists of six large tanks (greater than 20,000 kl capacity) and sixty smaller tanks (Figure 22.12). These tanks are separated into patios and surrounded by fire walls, the purpose of which is to contain any spilled oil. The spill capacity of the patios is sized for individual tank failures, since in normal operations the probability is small that more than one tank within a patio fails or leaks simultaneously. Should more than one tank within a patio

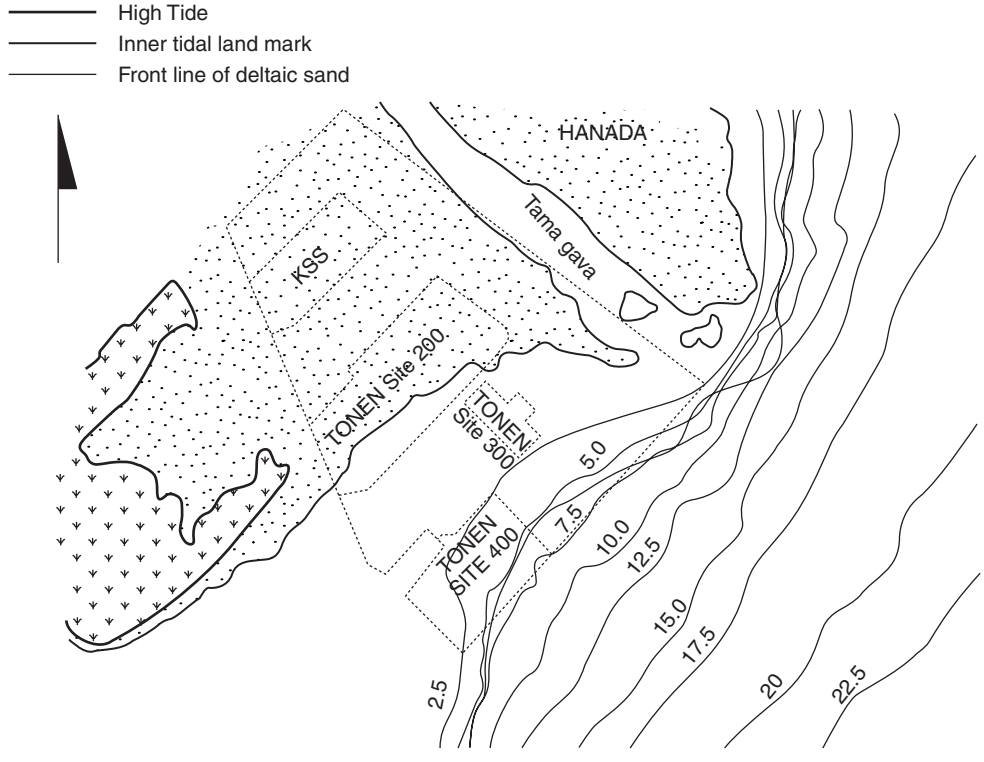


Figure 22.11 The TONEN site. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

simultaneously spill their full volume of oil, the fire walls would be overtopped, resulting in oil released outside the patio.

Previous risk assessments of Sites 200 and 400 showed unacceptably high levels of risk (T.W. Lambe & Associates, 1982). In many ways the KSS site is similar to the 200 and 400 sites, but with three important differences: (1) the SPT blow count value for the KSS site, and thus the imputed cyclic shear strength of the foundations, are lower than at the other sites; (2) the fines content is higher; and (3) most of the tank foundations were treated by vibroflotation.

The risk assessment of Sites 200 and 400 identified the principal source of risk as an off-site spill caused by rupture of multiple tanks due to foundation failure. Seismic ground shaking causes the foundation soils to deform or liquefy. This deformation results in large differential settlements of the tank bottoms, resulting in rupture of the tank and loss of contents. Multiple tanks failing within a patio cause the fire walls to be overtopped, or the same soil deformations that cause the tanks to rupture also cause the fire wall to fail. In either case, oil is spilled off site. Proximity to Tokyo Bay and the enormous costs associated with an oil spill into the Bay, make an off-site spill the dominating risk. Earlier experience by other oil companies in Alaska and Japan indicated that clean-up and legal costs attending a significant off-site spill could be well in excess of US$ 10^9 (*ca.* 1980 dollars). Structural failure was deemed unlikely, given the highly conservative dynamic load assumptions used in design. Losses from fire were not addressed.

The KSS site, comprising about 300,000 m^2, is located in an industrial area on reclaimed land developed by hydraulic filling the in late 1950s and early 1960s. Approximately 4 m of fill was placed, the first part hydraulically, and the last 1–3 m by dumping landslide debris and miscellaneous fill, and by spreading with earthmoving equipment. A typical profile is shown in Figure 22.13. The bay bottom sands and clays are recent (Holocene) alluvial deposits. The hydraulic fill was dredged from the sea floor and shoreline.

The clay is weak and compressible and the potential source of meters of long-term settlement. Concern existed that it was a possible source of foundation failure by sliding during an earthquake. The bay bottom sands are relatively clean sands mixed with silty

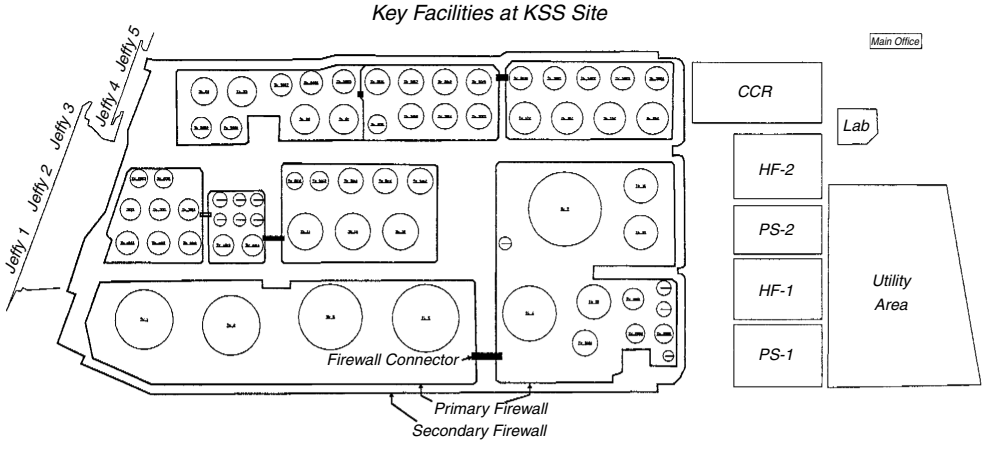


Figure 22.12 KSS region of the TONEN site. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

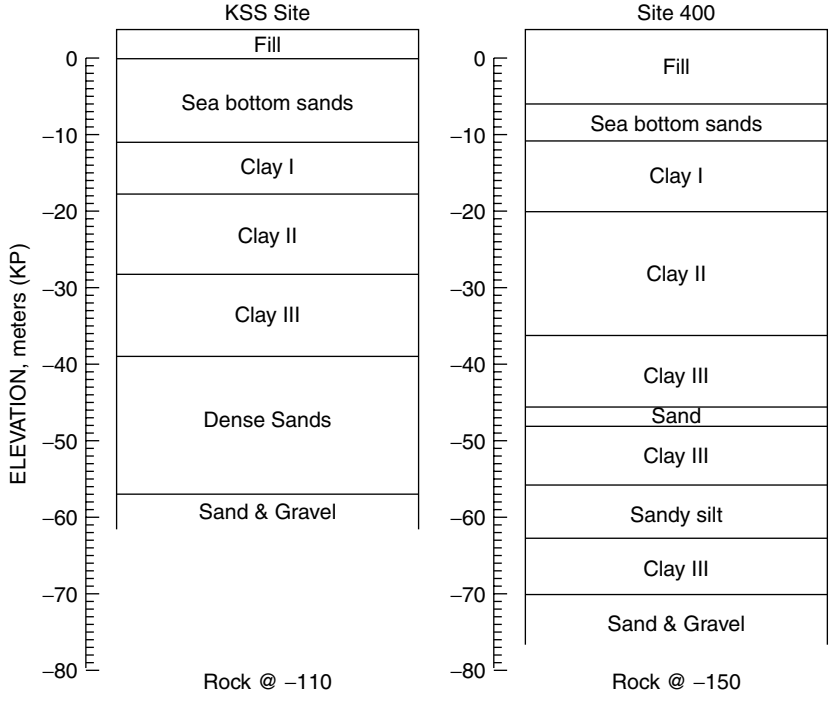


Figure 22.13 KSS soil profile. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

sand in places. The lower portions of the sand are relatively dense, with blow counts greater than 20 BPF. The upper 3–5 m contain more silt and have lower blow counts, in the range 1–9 BPF. These upper sands could potentially liquefy during an earthquake. The surficial fill consists of sands, silty sands, and sandy silts, with very low blow counts (0–8 BPF). The ground water level (gwl) is at about 1-2 m depth. The sands and gravels below −40 m elevation are overconsolidated and thought to pose no risk during an earthquake.

The low densities (manifest in low blow counts) of the bay bottom sands and fill, combined with their particle size distributions, suggests they may cause problems during earthquakes. In 1987, a magnitude 6.7 earthquake 65 km from the site caused liquefaction near one of the tanks at the KSS site. Measurements at site 400 indicated 85 gal (0.085 g) peak ground surface acceleration. A ground surface crack was observed 20 m from the tank, surrounded by a mound of fine sand. No tanks were damaged.

Vertical effective stresses at the site were calculated from the expression

$$\overline{\sigma}_v = \gamma_t H - (H - D_w)\gamma_w \tag{22.3}$$

in which

$\overline{\sigma}_v$ = effective vertical stress
γ_t = total unit weight of the soil
H = depth

D_w = depth to ground water surface
γ_w = unit weight of water

Since γ_w was presumed to be known, the variance in $\overline{\sigma}_v$ as a function of depth was calculated by a first-order approximation as

$$Var[\overline{\sigma}_v] = H^2\, Var[\gamma_t] + {\gamma_w}^2\, Var[\mathrm{D_w}] \tag{22.4}$$

From actual measurements, the mean and standard deviation of γ_t were 1.85 t/m^3 and 0.093 t/m^3. There was a systematic variation (standard deviation) of the depth to gwl of 0.45 m, on top of which was seasonal variation (standard deviation) of 0.6 m. Combining these two (by summing the variances) led to a standard deviation of depth to gwl of 0.75 m. Given the large coefficient of variation of depth to gwl (Ω_{Dw} = standard deviation/mean), the variable D_w was modeled as logNormally distributed to preclude negative values of depth in the analysis. This distributional assumption also seemed to fit the empirical histogram of observations made at the site.

A total of 48 SPT borings were made at the site between 1959 and 1988, at locations shown schematically in Figure 22.14. Values of blow count (N) greater than 50 were ignored as spurious. Figure 22.15 summarizes the uncorrected N values with depth for the entire site, separated into values before and after vibroflotation and shows means and standard deviations for one-meter depth intervals. The average N in the upper 5 m before treatments is about 4; after treatment, about 8. The mean value of 12 to 13 BPF between depths 5 m and 8 m was unchanged by treatment. The mean below 8 m was consistently greater than 20 BPF

For use in the liquefaction analysis, the blow counts were corrected by the Gibbs and Holtz (1957) factor

$$N_C = \frac{50N}{\overline{\sigma}_v + 10} \tag{22.5}$$

in which

N_C = corrected blow count
N = measured blow count
$\overline{\sigma}_v$ = effective vertical stress

The average corrected blow count in the top 5 m was 14 and 30, for treated and untreated areas, respectively. The spatial variance of N_c across the site was 8.4 BPF^2. In addition, the limited number of measurements and the regression fitting with depth to estimate a mean trend in N_c and to estimate $\overline{\sigma}_v$ created a systematic error variance in N_c of about 0.2 BPF^2.

22.6.2 Tank, fire wall and revetment reliability

The probability of tank and fire wall failures was based on Yegian and Whitman's (1978) Liquefaction Potential Index (LPI). LPI is a ratio of applied shear stress due to ground shaking to effective shear resistance

$$LPI = \frac{He^{0.M}}{\overline{\sigma}_v(R+16)\overline{S}_c} \tag{22.6}$$

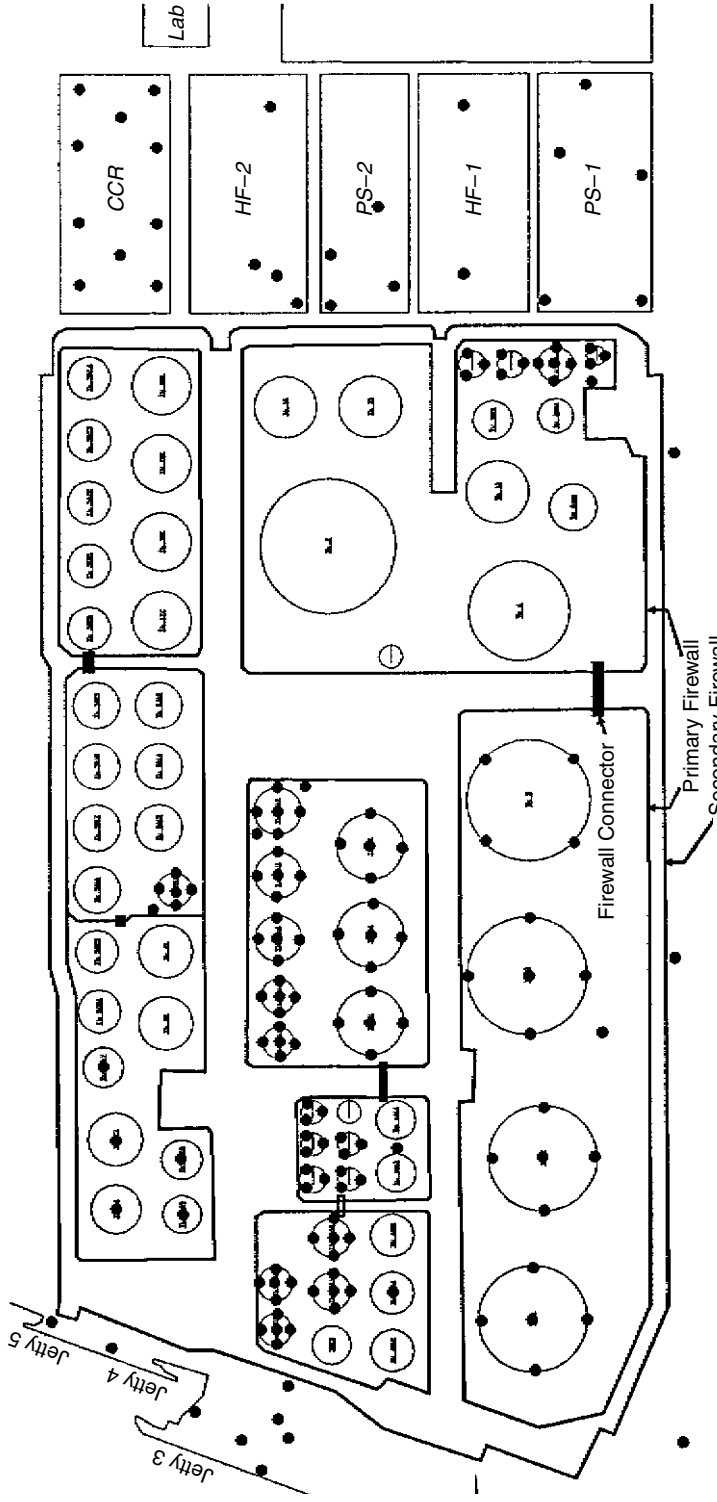


Figure 22.14 Boring locations at KSS. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

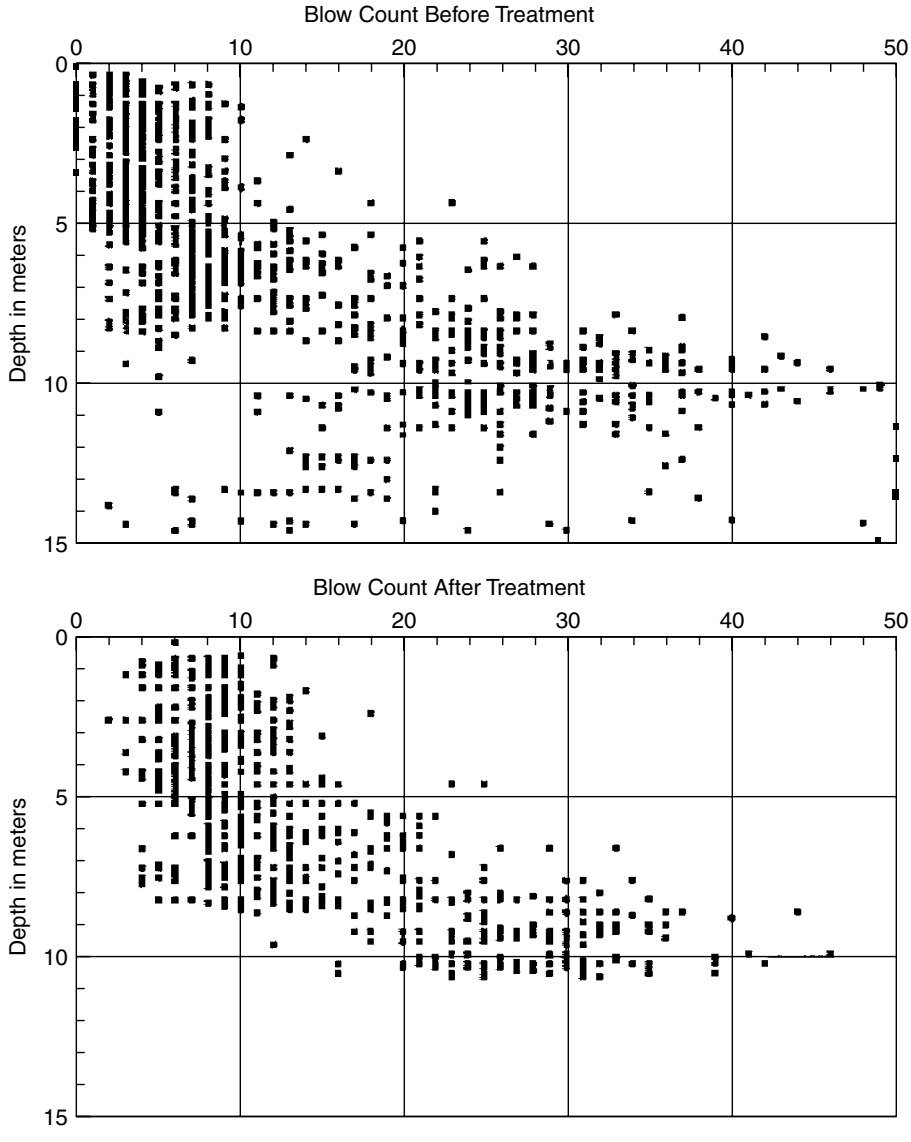


Figure 22.15 Uncorrected SPT N values at KSS site. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

in which

H = *depth*
M = *magnitude*
$\overline{\sigma}_v$ = *effective vertical stress in psi*
R = *hypocentral dist. in miles*
$\overline{S}_c$ = *empirical cyclic strength*

Using regression analysis of worldwide data, Yegian and Whitman developed an empirical relationship between cyclic strength and corrected blow count, N_c,

$$\overline{S}_c = \frac{N_c^{0.839}}{8.12} \exp\{0.5\text{Var}[\ln S_c/N_c]\} = 0.1336N_c^{0.839} \tag{22.7}$$

in which the variance term in the exponential comes from the regression fitting, and equals 0.163, leading to the shorter expression on the right-hand side.

Tank settlements were correlated to LPI based on finite element analyses of representative conditions at the site, and the probability of tank failure as a function of settlement was approximated by the curve shown in Figure 22.16, based on work by Marr *et al.* (1982). For each acceleration level of Table 22.2, a mean and variance of LPI were calculated, and a logNormal probability density function (pdf) applied to LPI. The pdf of LPI was propagated through the settlement vs. LPI curve to obtain a pdf of settlement, and this, in turn, was propagated through the settlement vs. probability of failure curve to obtain p_f for the tank.

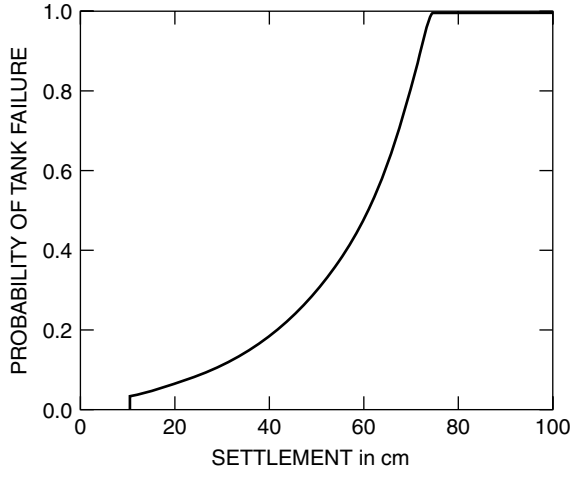


Figure 22.16 Probability of tank failure as a function of settlement. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

Table 22.2 Discrete approximation to seismic hazard at the KSS site

Case	Surface acceleration (cm/sec^2) value (range)	Base level velocity (cm/sec) value (range)	Average return period (years)	Probability (in 20 yrs)
I	50 (30–75)	2 (1–3)	2.5	≈1.0
II	100 (75–150)	5 (3–7)	16	0.71
III	200 (150–250)	10 (7–15)	100	0.12
IV	300 (>250)	20 (>15)	300	0.06

Two special concerns arose in performing the analysis just described. The first was the effect of spatial correlation of soil properties, and thus LPI, on the probabilities of failure calculated for tanks, fire wall sections, and revetment sections. The second was the lack of independence among probabilities of failure for tanks, wall sections, and revetment sections adjacent to one another.

The performance of the tank foundations was indexed to average LPI beneath the tank. This means that the variability of the soil properties entering the LPI equation was averaged over a volume of soil. For the large tanks, this volume was larger than for the small tanks, and as a consequence the variance of the average LPI under the larger tanks was less than under the small tanks. Thus, the probability of failure for the small tanks was larger than for the large tanks, although the consequences of failure measured in the volume of oil spilled, would be less. A variance reduction factor based on the methodology of Chapter 10 was applied to each tank, depending on the volume of soil over which the soil properties were averaged.

Spatial variation of soil properties affected the probabilities of failure of the fire walls and revetment in a related way. Both the fire walls and the revetment are long, linear features. Failure at any location along the length of the wall or revetment means failure to contain a spill. Thus, the concern in assessing the probability of failure of a wall or revetment is one of 'first crossing' or 'worst condition,' in the sense of the exceedances of random processes. To simplify calculations, the walls and revetments were divided into a number of 'equivalent independent sections,' each of length $1/\delta_o$, in which/δ_o is the autocorrelation distance. The probability of failure of an individual section was calculated, based on LPI and spatial averaging, and the probability of at least one such section failing in a wall containing n such sections was calculated by the relation

$$\Pr\{\text{wall failure}\} = 1 - (1 - \Pr\{\text{section failure}\})^n \tag{22.8}$$

Special considerations were made for corners in the walls, or other places where wall geometry deviated from a straight line. These considerations were made by numerically solving the geometrical averaging of spatial variation for the particular geometries.

Chapter 8 describes the importance of spatial and systematic uncertainty. Spatial correlations and systematic uncertainty in soil properties also affected the degree of correlation among the failures of multiple tanks, fire wall segments, and revetment segments. First, the uncertainty in soil properties has both a systematic and a spatial component. The systematic component comprises estimation error in the mean and bias error in the geotechnical model used to interpret test results. This affects all the tanks, walls, etc. the same way. If the mean is in error under one tank, it is in error under all. Thus, to the extent that the probabilities of failure of the tanks, walls, etc. depend upon systematic uncertainty, they are perfectly correlated. At the same time, to the extent that the probabilities of failure depend on spatial variation, they are independent, but only if the individual facilities are widely separated in space. Adjacent tanks rest on soils which, given the autocovariance of the soil properties, have average properties that are to some degree correlated, and thus the probabilities of failure of the adjacent tanks must be to some degree correlated. Therefore, in calculating probabilities of failure, both the effects of systematic error and spatial correlation must be accommodated in inferring the degree of correlation among the failure probabilities of tanks, wall segments, and revetment segments.

22.6.3 Event tree analysis

The risk assessment consisted of five steps:

1. Identify the factors influencing seismic performance of the storage tanks: frequency of earthquakes, variability of soil properties, behavior of the tanks during earthquakes, performance of the foundation soils during earthquakes, performance of the fire walls and revetments during earthquakes, performance of other site facilities designed to prevent spills, and so forth.
2. Analyze earthquake source models to establish seismic hazard at the site and establish levels of ground shaking and their respective probabilities of being exceeded.
3. Develop an event tree for a typical patio; the event tree is used to illustrate how various facilities in a tank patio interact to lead to system failure.
4. Build a numerical (Monte Carlo) simulation of the entire site to analyze random behavior of each key component; run the model through many realizations to develop a statistical sample of the behavior of the site and facility.
5. Compare the predicted probability of an off-site spill and the corresponding consequences to other risk; use the comparison and the annual expected monetary loss to judge the acceptability of the risk faced at this site.

The seismic hazard at the site was developed using methods similar to those based on Cornell's (1968) work. The outcome of this hazard assessment is shown in Figure 22.17 as a complementary cumulative distribution (CCDF) of peak surface ground acceleration. For the numerical computation, this curve was approximated by the four discrete levels shown in Table 22.2.

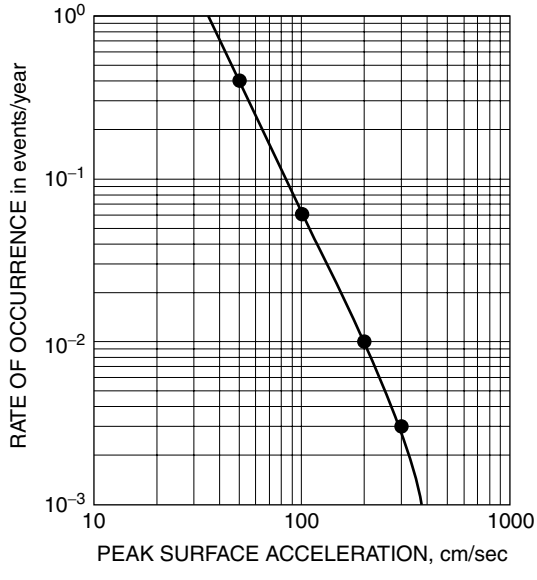


Figure 22.17 Seismic hazard results for TONEN region. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

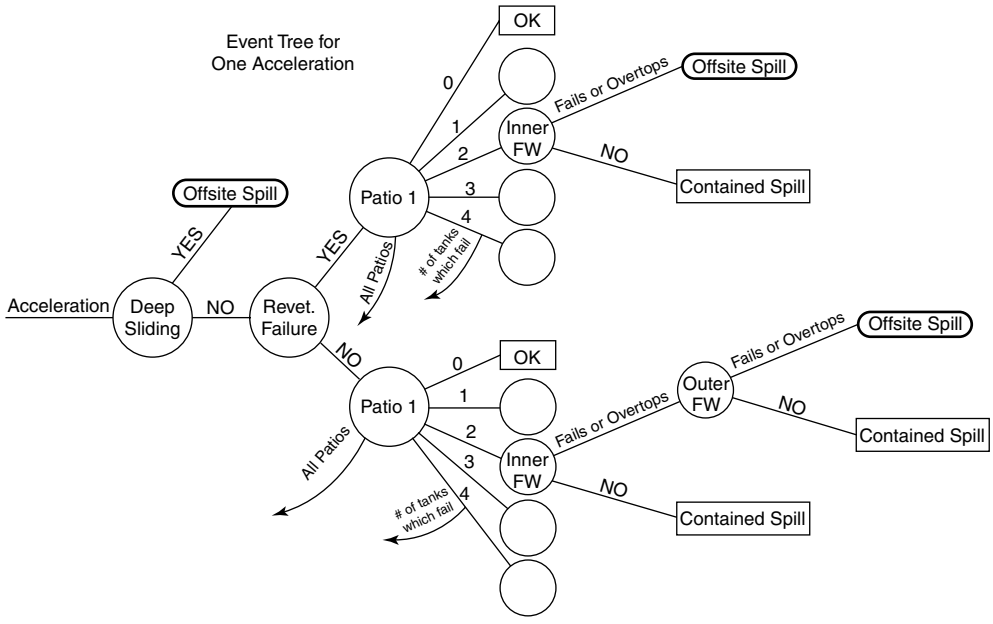


Figure 22.18 Event tree for Patio1. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

An event tree was developed for Patio 1, the large patio at the south of the site, comprising four 20,000kl tanks. This event tree is shown in Figure 22.18 for one particular level of acceleration. The complete event tree for Patio 1 has four branches, each identical in form to that of the figure, corresponding to the four levels of acceleration listed in Table 22.2.

The event tree begins from the left with the occurrence of a given level of seismic ground shaking. Given this ground shaking, the first event checked is deep sliding within the clays. If sliding of the entire site within the clays occurs, complete disruption of the site is assumed, leading to an off-site spill, and no further analysis is required.

If deep sliding does not occur, the next event checked is failure of the revetment surrounding the entire site. The revetment along the waters edge, with its associated outer fire wall, is the last bulwark against an off-site spill. Even if oil is released from a patio, it may still be trapped behind the revetment, but if the revetment fails, this is no longer the case. Also, because certain patio walls and even tanks are within the zone of influence of a revetment failure, depending where the revetment failure occurs, it may mechanically affect the stability of the firewalls or tanks.

Within the patio are four tanks. One or more of these may fail, or none may fail. If one or more fail, some amount of oil is released into the patio, depending on how full the tanks are at the time of failure. Throughout the year, the volume of product stored in the tanks varies, and the volume in some tanks varies more than that in others. Since the four tanks are near one another in space, their respective probabilities of failure tend to be dependent. The failure of one tank may mechanically affect another tank, and the soil properties underlying the tanks are spatially correlated. If no tanks fail, then there is no off-site spill.

If one or more tanks fail, then next event checked is breaching (failure or overtopping) of the firewall. In Figure 22.18 the subsequent branches of the tree are only shown for two tanks failing, but the same subsequent branch emanates from each tank failure event. If the volume of spilled oil within the patio exceeds the storage capacity of the fire walls (which by fire code needs only equal the volume of one full tank), then the fire wall is overtopped and some volume of off-site spill occurs. If one or more tanks fail and the fire wall also fails, then off-site spill occurs. As for multiple tank failures, fire wall failure is correlated with the failures of tanks, revetment, and possibly other facilities at the site.

A complete event tree for the entire KSS site becomes the bushy mess mentioned at the start of this section. It leads to a few thousand end nodes, because the combination of facility failures and non-failures grows exponentially. Nonetheless, a rough approximation of the system failure probability for the entire site can be calculated starting with the probability of a spill from Patio 1, and making the highly simplifying assumption that the entire site consists merely of some number, k, of 'equivalent, independent' such patios. Then, the probability of system failure becomes the probability of an off-site spill from at least one of these equivalent, independent patios

$$\Pr\{\textit{system failure}\} = 1 - (1 - \Pr\{\textit{patio failure}\})^k \tag{22.9}$$

Among the conclusions drawn from an analysis of the Patio 1 event tree were (i) the logical relationship among failure elements, (ii) the importance of both mechanical and probabilistic correlation among facility (tank, walls, revetment, piping, etc.) failures, (iii) the influence of specific combinations of failure conditions on the volume of oil spilled off site, and (iv) a rough estimate of the overall probability of an off-site spill.

22.6.4 Simulation analysis

The event tree analysis provided an understanding of the importance of different risk elements at the site, and a rough estimate of the probability of an off-site spill. To develop a more precise estimate of system failure probability, a simulation model of the tank farm was built.

The occurrence of an off-site spill of oil during an earthquake depends on the simultaneous occurrence of many interdependent elements within the oil storage areas. The many possible combinations of failure of revetments, outer fire walls, inner fire walls, overflow of unfailed fire walls, and tank failures from foundation settlement, caused by an earthquake the size of which is, itself, uncertain, cannot easily be resolved analytically. In the simulation model, the principal components affecting the likelihood of an off-site spill were considered. For each component, random numbers were generated to represent the component's response to a particular seismic event. The random numbers were chosen to fit probability distributions whose means and variances reflected the inferred uncertainty in the behavior of each component. Systematic uncertainties and uncertainties due to spatial variation were analyzed separately, since realizations of the systematic uncertainties apply to every component in the same way, while those of the spatial variation differ from one component to another.

This simulation model started with the plan of the facility shown in Figure 22.11, and reliability models of the individual components of the site were developed to predict probabilities of component failure given different levels of peak ground acceleration. In most

cases, these were the same reliability models used to calculate branch probabilities for the event tree described above. From these reliability representations, failure probabilities and correlations among failure probabilities were derived for each facility component, and for each of the four identified levels of ground acceleration. Then, runs were made in which a random number generator was used to assign failure or non-failure conditions to each component in a calculated realization, volumes of oil spilled in each realization were summed, and a storage volume accounting was made to determine whether oil was spilled off site in each realization. A large number of realizations was run, and the results analyzed statistically to infer exceedance probabilities of the volume of oil spilled off site in a given time period.

At the end of each run, an accounting is made to determine how much, if any, oil is spilled. Each tank is polled to see whether it failed. If it did fail, a random number ranging between 0 and 100% is generated to reflect the volume of oil stored in the tank at the instant of failure. The statistical parameters of this random variable are taken from production records at the tank farm, and differ from one tank to another depending on the product stored and the size and location of the tank. Typical site records for three patios are shown in Figure 22.19.

For each patio, an accounting of the total volume of oil spilled by tanks within is made. If one or more segments of the patio fire wall has also failed, the total volume of oil spilled within the patio is considered released onto the facility. If no wall segments have failed, a check is made to see if the fire wall has been overtopped, and, if so, the balance of oil volume larger than the patio capacity is considered to be released. Finally, a summation of oil spilled out of all the patios onto the facilities is made, and the revetment (and its associated outer fire wall) is checked to see whether failure occurred, in which case all the oil spilled onto the facility is considered to be released off site. If the revetment has not failed, a check is made for overtopping, in which case the balance of oil greater than the site capacity is considered to have been spilled off site. At the end, a histogram of oil

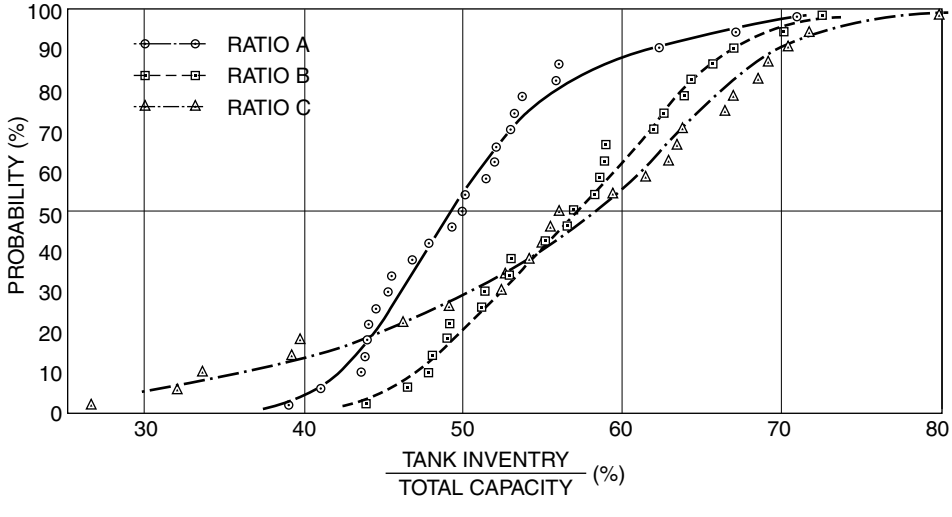


Figure 22.19 Site production records expressed as probability of proportion of tank inventory in use. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

spilled off site in the individual realizations is plotted, and the frequency with which a non-zero volume was spilled is used as an estimator of the probability of system failure for the particular acceleration interval.

The entire simulation was carried out four times, once for each interval of peak ground acceleration in Table 22.2, using 2000 realizations each. Typical results are shown in Figure 22.20 for levels of dewatering from 0 m to 7 m. The summary of results for the three facility locations at Kawasaki is shown in Table 22.3.

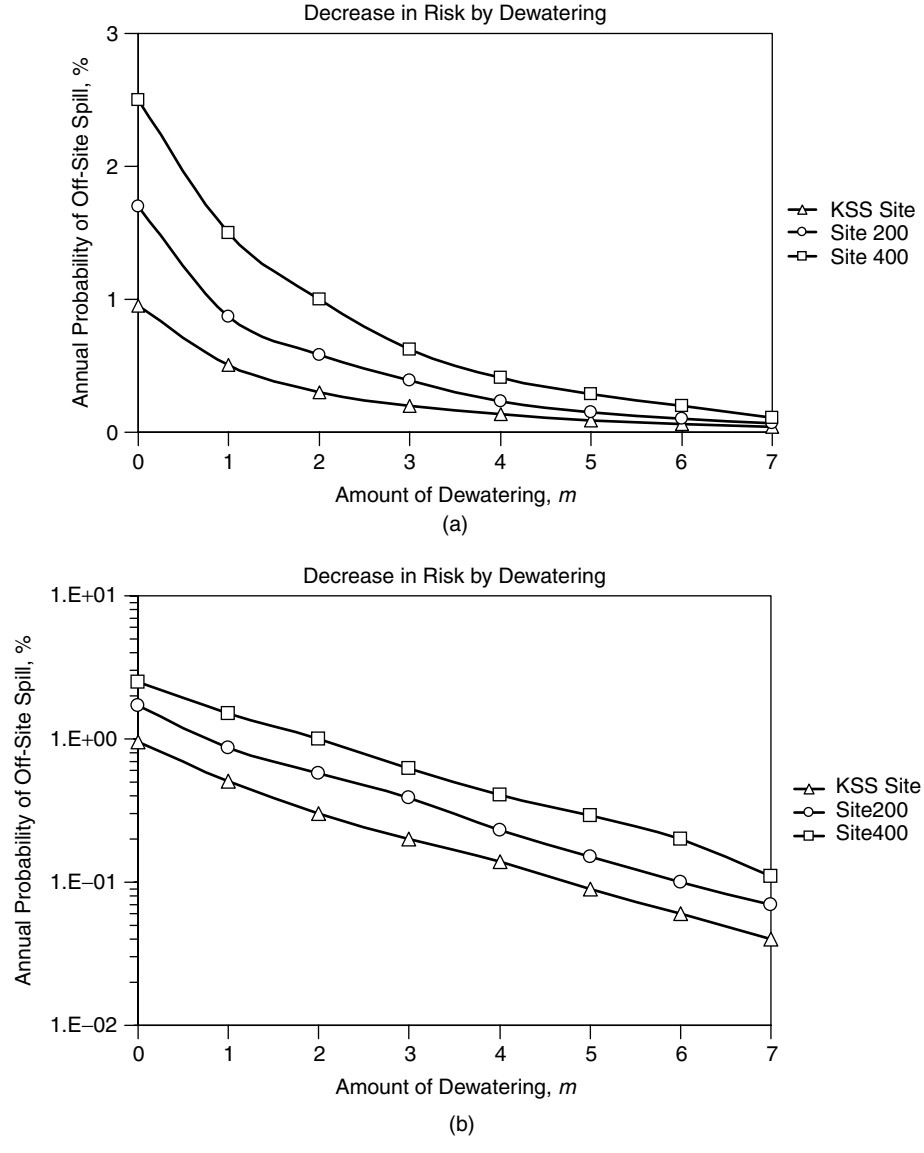


Figure 22.20 Results of simulation for three sites and lowering water table by 0 to 7 m. (a) linear probability scale, (b) logarithmic probability scale (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

Table 22.3 Calculated probability of off-site spill at KSS and TONEN 200 and 400 sites

Event (gal)	Probability (annual)	KSS Site (%)	Site 200 (%)	Site 400 (%)
50	0.400	0.002	0.008	0.013
100	0.625	0.006	0.011	0.018
200	0.010	0.340	0.450	0.590
300	0.003	0.567	0.667	0.800
Total	–	0.010	0.017	0.025

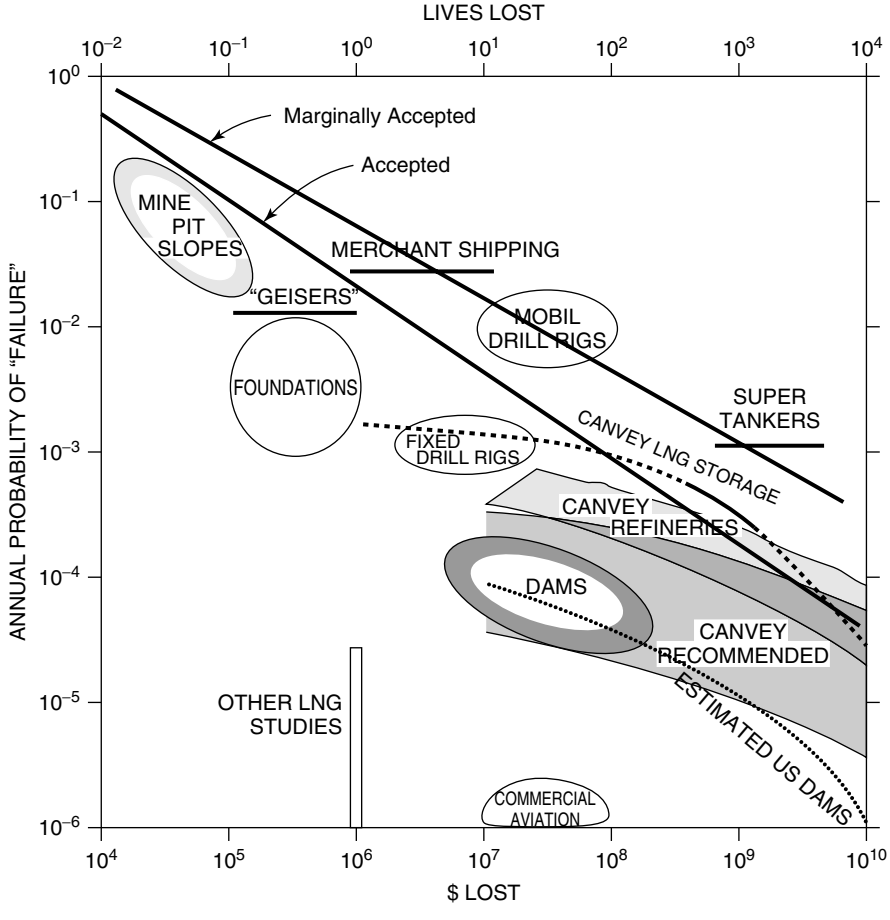


Figure 22.21 Acceptable and marginally acceptable levels of risk potted on F-N chart. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

22.6.5 Comparative risk at Kawasaki

To understand the risk faced by the tank farm owners, the calculated exceedance probabilities and corresponding consequences were plotted in a probability-consequence or F-N

chart. This chart (Figure 22.21) shows annual probability of occurrence on the vertical axis with the corresponding consequence(s) of failure – either in monetary cost, lives lost, or both – on the horizontal axis. The envelopes marked, 'accepted' and 'marginally accepted' reflect risks inferred for other civil works. Because the monetary cost associated with an off-site spill is of an order greater than US10^9, clearly the risk faced at the tank farm (0.01 to 0.025) is too high.

One way to reduce the annual probability of an off-site spill is to lower the ground water level beneath the facility. This has the effect of reducing the pore water pressure, increasing the effective vertical stress in the soil, and thus decreasing the probability of liquefaction given seismic ground shaking. The simulation model allowed a direct calculation of the effectiveness of dewatering on risk reduction, since the effect of dewatering on LPI could be calculated in Equation (22.6), and new statistical parameters for LPI for each interval of peak ground acceleration used in the simulation model. The results of these re-calculations are shown in Figures 22.20 and 22.22, in which the reduction in annual probability of an off-site spill as a function of ground water lowering is shown.

The summary of the results presented in the form of Figures 22.20–22.22 enabled the corporate management to make a rational decision about the course of action it should undertake. By comparing the annualized cost of failure due to earthquakes with the annual cost of preventive measures, the management decided that continuous lowering of the groundwater level was justified economically. An impervious slurry wall was built around the site, pumps were installed, and a continuous program of pumping and monitoring of groundwater levels was instituted. This remains in place at the time of writing, and similar solutions have been implemented at other sites in the area.

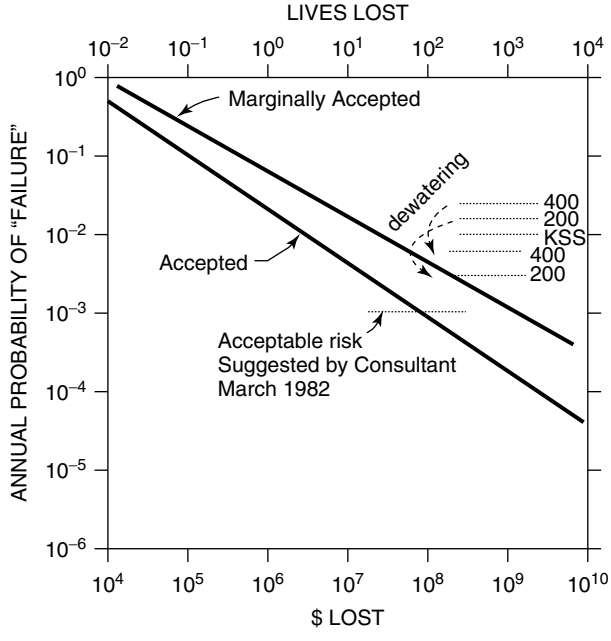


Figure 22.22 Effects of dewatering compared to acceptable and marginally acceptable risks on F-N chart. (T. W. Lambe & Associates 1982, 1989, reproduced with permission.)

22.7 Summary

This chapter describes some of the procedures for evaluating the reliability of a system rather than a single component. The example of the tank farm in the seismically active region shows how this is carried out in practice. Similar efforts have been undertaken for evaluating the safety of dams. Several points will have become obvious to the reader but should be emphasized in any case. First, all the tools and insights of the preceding chapters must be brought to bear on such an analysis, including logic trees, expert elicitation, detailed reliability assessment, and simulation. Second, some components of the system are inevitably better understood than others, and judgment must be applied to include them in the analysis. Third, the major limitation on such studies is not a limitation in reliability theory or methodology or restrictions in computer facilities but the investment of time and effort required for interaction between personnel who understand the system to be studied and those experienced in reliability studies.

Finally, effective use of reliability analysis of a system requires commitment by management, as typified in the TONEN case just described. Without such commitment, the reliability analysis is a sterile exercise. Of course, the same comment applies to the use of the observational method. It is to be hoped that the future will see broader use of both reliability and observational approaches to deal with complicated geotechnical systems and the two approaches will be combined more closely than in the past.

Appendix A: A Primer on Probability Theory

A.1 Notation and Axioms

Probability theory is a branch of algebra with its own axioms and notation. While one can construct an algebra with almost any set of self-consistent axioms, some are more useful than others, and the particular axioms and theorems of probability theory have proven very useful over the years. The notation is sometimes opaque to the casual user; in this appendix we set out the forms most widely used in civil engineering applications.

The first notation is the expression for the probability of A, where A could be an event, a condition, a state of nature, or almost anything else. We write that the probability of A is $P[A]$. Obviously, we can replace 'A' with any appropriate expression or phrase such as '$X \leq x$' or 'soil liquefies.'

All the theorems of probability theory can be derived from three axioms:

1. $P[A] \geq 0$. (That is, a probability must be equal to or greater than zero.)
2. $P[A] = 1$ means that A is certain.
3. $P[A \cup B] = P[A] + P[B]$ if A and B are mutually exclusive. The notation '$\cup$' stand for the 'union' of A and B or that A or B or both occur.

The Venn diagram, illustrated in Figure A.1, is a great help in visualizing probability concepts. The rectangular box represents the universe, that is all possible events. The circle labeled A represents event A, and that labeled B represents event B. For axiom 3 these events are mutually exclusive, so their circles do not intersect. Venn diagrams will be used below to elucidate more complicated situations.

A.2 Elementary Results

By convention $\overline{A}$ or $\tilde{A}$ means 'not A.' In words, if A means 'the soil liquefies,' $\overline{A}$ means 'the soil does not liquefy.' From the axioms

$$P[A] + P[\overline{A}] = 1 \tag{A.1}$$

or, in words, the soil either liquefies or does not, but not both.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian
 ISBN: 0-471-49833-5

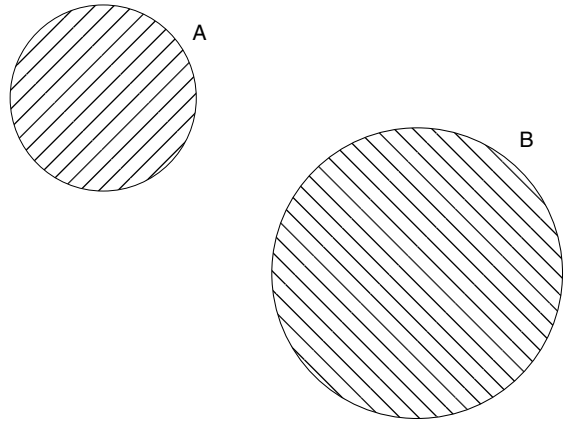


Figure A.1 Venn diagram for two mutually exclusive events.

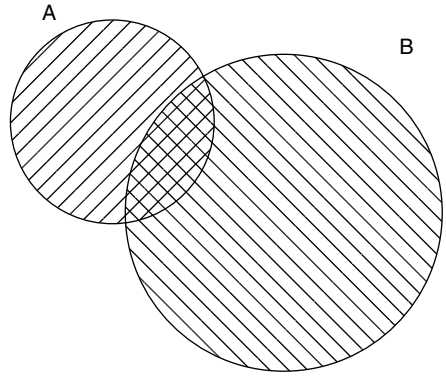


Figure A.2 Venn diagram for two events that are not mutually exclusive.

When both A and B occur, the event is said to be their intersection, denoted by $A \cap B$ or simply AB. Figure A.2 shows a situation in which A and B are not mutually exclusive; that is, they can both occur. It is easily verified that

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \tag{A.2}$$

By extension, and by reference to Figure A.3, when there are three events

$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[AB] - P[AC] - P[BC] + P[ABC] \tag{A.3}$$

This can be extended to any number of events, with the general rule that items with an odd number of intersecting events have positive signs and those with even numbers have negative signs.

Manipulation of combinations of events is helped by de Morgan's rule:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \tag{A.4}$$

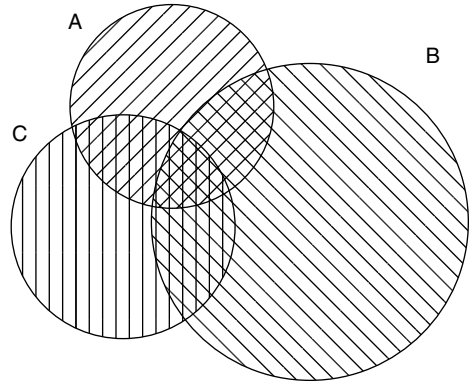


Figure A.3 Venn diagram for three events that are not mutually exclusive.

In many cases, we want to express the probability that A occurs, given that B also occurs. This is called the conditional probability. It is expressed as $P[A|B]$. Furthermore:

$$P[A|B] = \frac{P[AB]}{P[B]} \tag{A.5}$$

and

$$P[AB] = P[A|B] \cdot P[B] = P[B|A] \cdot P[A] \tag{A.6}$$

This can be extended to more variables:

$$P[ABC] = P[A|BC] \cdot P[B|C] \cdot P[C] \tag{A.7}$$

A.3 Total Probability and Bayes' Theorem

A very useful result, analogous to the chain rule for partial differentiation is that, if $E_1, E_2, \ldots, E_n$ is a set of mutually exclusive and exhaustive events (that is, they include all possible events E_i) and A is some event conditional on one or more of the E_i, then

$$\begin{aligned} P[A] &= P[A|E_1] \cdot P[E_1] + P[A|E_2] \cdot P[E_2] + \cdots + P[A|E_n] \cdot P[E_n] \\ &= \sum_{i=1}^{n} P[A|E_i] \cdot P[E_i] \end{aligned} \tag{A.8}$$

This is known as the Total Probability Theorem.

Suppose that there is an exhaustive series of events with estimated probabilities E_i. These are called prior probabilities. Now, suppose some event A occurs. (This might be the result of some measurements or some happening such as a failure.) We want to know how this affects the estimate of the probability of E_i. From the above results,

$$P[AE_i] = P[A|E_i] \cdot P[E_i] = P[E_i|A] \cdot P[A] \tag{A.9}$$

$$\therefore\ P[E_i|A] = \frac{P[A|E_i] \cdot P[E_i]}{P[A]} = \frac{P[A|E_i] \cdot P[E_i]}{\sum P[A|E_i] \cdot P[E_i]} \tag{A.10}$$

In other words, Equation (A.10) describes how to revise the estimate of the probability of E_i after we have learned that A has occurred. To do this we must know the prior estimates of the probabilities of the E_is and the conditional probabilities of A. This theorem is known as Bayes' Theorem; it forms the basis of much modern reliability work.

A.4 Discrete Distributions

Several well-known functions can be used to describe the uncertainty in a variable. These have become widely used because they describe situations that arise in practice, but many variables do not conform to any of the standard functions. The distributions are usually divided into two categories: discrete and continuous. Discrete distributions describe cases in which the uncertain variable can have only discrete values, such as the value of pips on a die or the number of floods in a given period of time. Continuous distributions describe cases in which the uncertain variable can have any value within a given range, such as the amount of rainfall in a year or the shear strength of a soil. This section describes discrete distributions, and the next deals with continuous distributions. It is customary to refer to the variable by an upper case letter (e.g. X) and to specific values that the variable assumes by a lower case latter (e.g. x).

Evans *et al.* (1993) have presented, for most of the commonly encountered distributions, a compendium of parameters and related functions, including many that are not discussed here. Most of the material included in the following two sections and in Tables A.1, A.2 and A.3 can be found in their book. Some of the other material was taken from standard references such as Burington and May (1970), or independently derived by the authors.

A discrete variable can have n distinct values. Each of these values has a probability of occurrence, which we will designate p_i. The collection of all the p_i is called the probability mass function or pmf. It is usually written as $p_X(x)$ or p_i. Obviously, since the n distinct values exhaust al the possibilities, $\sum_{i=1}^{n} p_i = 1$.

The probability that the variable has the ith value is p_i, and the probability that the variable has a value equal to or less than the ith value is $\sum_{j=1}^{i} p_j$. This is called the cumulative distribution functio*n* or CDF, written $F_X(x)$. If the ith value of the variable is x_i, we can define moments of the distribution, m_k:

$$m_k = \sum_{i=1}^{n} p_i \cdot x_i^k \tag{A.11}$$

The first moment, for $k = 1$, is the mean or expected value of the variable. It is denoted by μ_X or $E[X]$. It is convenient to center the other moments about μ_X instead of about the origin. The second moment taken about μ_X is the variance, denoted by $Var[X]$, and its square root is the standard deviation, σ_X. Algebraic manipulation gives

$$Var[X] = \sigma_X^2 = \sum_{i=1}^{n} p_i \cdot (x_i - \mu_X)^2 = \sum_{i=1}^{n} p_i \cdot x_i^2 - \mu_X^2 \tag{A.12}$$

Table A.1 Parameters for common discrete distributions

Name	Parameters	Symbol	pmf ($= p_X(x)$)	Mean	Variance	Skewness coefficient
Uniform	a, n, h	$\mathbf{D}(a, n, h)$	$\frac{1}{n+1}$	$a + \frac{nh}{2}$	$\frac{h^2 n(n+2)}{12}$	0
Uniform (counting)	$0, n$	$\mathbf{D}(0, n)$	$\frac{1}{n+1}$	$\frac{n}{2}$	$\frac{n(n+2)}{12}$	0
Binomial	n, p	$\mathbf{B}(n, p)$	$\binom{n}{x} p^x q^{n-x}$	np	npq	$\frac{q-p}{(npq)^{1/2}}$
Geometric	p	$\mathbf{G}(p)$	pq^n	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{1+q}{q^{1/2}}$
Negative Binomial	y, p	$\mathbf{NB}(y, p)$	$\frac{\Gamma(x+y)}{\Gamma(y)x!} p^{\mathrm{y}} q^{\mathrm{x}}$	$\frac{\mathrm{yq}}{p}$	$\frac{\mathrm{yq}}{p^2}$	$\frac{1+q}{(yq)^{1/2}}$
Pascal (N. B. with integer parameters)	y, p	$\mathbf{NB}(y, p)$	$\binom{x+y-1}{y-1} p^{\mathrm{y}} q^{\mathrm{x}}$	$\frac{yq}{p}$	$\frac{yq}{p^2}$	$\frac{1+q}{(yq)^{1/2}}$
Poisson	λ	$\mathbf{P}(\lambda)$	$\lambda^{\mathrm{x}} \exp(-\lambda)/x!$	λ	λ	$\lambda^{-1/2}$

Notes: 1. $q = 1 - p$.

2. $\mathrm{m}\binom{n}{m} = \frac{n!}{(n-m)!m!}$ = number of combinations of n things taken m at a time.

3. $\Gamma(x)$ is the gamma function of x and is $(x-1)!$ when x is an integer.

4. In the Poisson distribution λ is often replaced by νt.

The coefficient of variation, Ω, is the ratio of the standard deviation and the mean:

$$\Omega = \frac{\sigma}{\mu} \tag{A.13}$$

Higher moments are used less often, but the skewness, ν_X, is occasionally employed:

$$\nu_X = \frac{\sum_{i=1}^{n} p_i \cdot (x_i - \mu_X)^3}{\sigma_X^3} \tag{A.14}$$

Table A.1 lists the parameters, the conventional symbol, the pmf, the mean, the variance, and the skewness coefficient for commonly encountered discrete variables. The Uniform distributions describe conditions when the variable has equal probability at $n + 1$ equally spaced points. The counting special case occurs when the points are the integers from 0 to n.

A Bernoulli process consists of n trials, each of which has a probability p of success. Drawing balls from and urn containing red and white balls is an example if the ball is

Table A.2 Parameters for common continuous distributions

Name	Parameters	Symbol	pmf(= $f_X(x)$)	Mean	Variance	Skewness coefficient
Uniform	a, b	$\mathbf{R}(a, b)$	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	0
Triangular	a, b, c		$\dfrac{2(x-a)}{(b-a)(c-a)}$ if $a \le x \le c$ $\dfrac{2(b-x)}{(b-a)(b-c)}$ if $c \le x \le b$	$\dfrac{a+b+c}{3}$	$(a^2+b^2+c^2$ $-ab-ac-bc)/18$	$[(a^3+b^3+c^3)/135$ $-(a^2b+ab^2+a^2c$ $+ac^2+b^2c$ $+bc^2)/90$ $+2abc/45]/\sigma^3$
Normal	μ, σ	$\mathbf{N}(\mu, \sigma)$	$\dfrac{1}{\sigma(2\pi)^{1/2}} \exp\left[\dfrac{-(x-\mu)^2}{2\sigma^2}\right]$	μ	σ^2	0
logNormal	λ, ζ	$\mathbf{L}(\lambda, \zeta)$	$\dfrac{1}{\zeta x(2\pi)^{1/2}} \exp\left[\dfrac{-(\ln x-\lambda)^2}{2\zeta^2}\right]$	$\exp\left(\lambda+\dfrac{1}{2}\zeta^2\right)$	$[E(X)]^2[\exp(\zeta^2)-1]$	$[\exp(\zeta^2)+2]$ $\times[\exp(\zeta^2)-1]^{1/2}$
Exponential	λ	$\mathbf{E}(\lambda)$	$\lambda\exp(-\lambda x)$	$1/\lambda$	$1/\lambda^2$	2
Gamma	b, c	$\boldsymbol{\gamma}(b, c)$	$\dfrac{(x/b)^{c-1}\exp(-x/b)}{b\Gamma(c)}$	bc	b^2c	$2c^{-1/2}$
Beta	p, q	$\boldsymbol{\beta}(p, q)$	$\dfrac{x^{p-1}(1-x)^{q-1}}{B(p,q)}$	$\dfrac{p}{q+p}$	$\dfrac{pq}{(p+q)^2(p+q+1)}$	$\dfrac{2(q-p)(p+q+1)^{1/2}}{(p+q+2)(pq)^{1/2}}$
Chi-squared	ν	$\boldsymbol{\chi}^2(\nu)$	$\dfrac{x^{(\nu-2)/2}\exp(-x/2)}{2^{\nu/2}\Gamma(\nu/2)}$	ν	2ν	$2^{3/2}\nu^{-1/2}$
Student's t	ν	$\mathbf{t}(\nu)$	$\dfrac{1}{\sqrt{\nu\pi}}\dfrac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\left(1+\dfrac{x^2}{\nu}\right)^{-(\nu+1)/2}$	0	$\dfrac{\nu}{\nu-2}$	0
von Mises	a, b		$\dfrac{\exp[b\cos(x-a)]}{2\pi I_0(b)}$	a	$\dfrac{1-I_1(b)}{I_0(b)}$	0

Notes: 1. In the logNormal distribution λ and ζ are the mean and standard deviation of $\ln X$, and μ and σ are the mean and standard deviation of X.
5. $\Gamma(x)$ is the gamma function of x and is $(x-1)!$ when x is an integer.
6. $\mathrm{B}(p, q)$ is the beta function $= \int_0^1 u^{p-1}(1-u)^{q-1}du$.
7. The variable x in von Mises distribution ranges from 0 to 2π, the mean is the mean direction, and the variance is a circular variance.

Table A.3 Parameters for extreme value distributions

Name	Parameters	pmf ($= f_X(x)$)	Mean	Variance	Skewness coefficient or third moment
Type I (Gumbel) largest value	u, α	$\alpha \exp[-\alpha(x-u)] \times \exp\{-\exp[-\alpha(x-u)]\}$	$u + \frac{\gamma}{\alpha} \approx u + \frac{0.57721}{\alpha}$	$\frac{\pi}{\sqrt{6}\alpha}$	1.1396 note 3
Type I (Gumbel) smallest value	u, α	$\alpha \exp[\alpha(x-u)] \times \exp\{-\exp[\alpha(x-u)]\}$	$u - \frac{\gamma}{\alpha} \approx u - \frac{0.57721}{\alpha}$	$\frac{\pi}{\sqrt{6}\alpha}$	-1.1396 note 3
Type II	u, k	$\frac{k}{u}\left(\frac{u}{x}\right)^{k+1} \exp\left[-\left(\frac{u}{x}\right)^k\right]$	$u\Gamma\left(1-\frac{1}{k}\right)$	$u^2\left[\Gamma\left(1-\frac{2}{k}\right) - \Gamma^2\left(1-\frac{1}{k}\right)\right]$	$E[X^3] = u^3\Gamma\left(1-\frac{3}{k}\right)$ note 4
Type III	u, k, ε	$\frac{k}{u-\varepsilon}\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k-1} \times \exp\left[-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^k\right] \quad x \geq \varepsilon$	$\varepsilon + (u-\varepsilon)\Gamma\left(1+\frac{1}{k}\right)$	$(u-\varepsilon)^2\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]$	$E[(x-\varepsilon)^3] = (u-\varepsilon)^3\Gamma\left(1+\frac{3}{k}\right)$ note 4
Type III, $\varepsilon = 0$ (Weibull)	u, k	$\frac{kx^{k-1}}{u^k} \exp\left[-\left(\frac{x}{u}\right)^k\right]$	$u\Gamma\left(1+\frac{1}{k}\right)$	$u^2\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]$	$E[x^3] = u^3\Gamma\left(1+\frac{3}{k}\right)$ note 4

Notes: 1. $\Gamma(x)$ is the gamma function of x and is $(x-1)!$ when x is an integer.
2. γ is Euler's constant $= 0.57721\ldots$
3. Skewness coefficient.
4. Third moment.

replaced after each drawing. The probability of having exactly x successes in n trials is described by the Binomial distribution. The Geometric distribution describes the number of the trial at which the first success occurs. This is also the number of trials necessary to obtain the next success after a particular success. The Negative Binomial distribution describes the number of the trial at which the kth success occurs. The Pascal distribution is the special case of the Negative Binomial distribution with integer values of X and Y.

A Poisson process is one in which an event can occur at any time but each event is independent of the previous events. In other words, the process has no memory. The Poisson distribution gives the number of events in a time period $\nu t (= \lambda)$. The Exponential distribution also arises in the Poisson process, but it is described under continuous distributions.

A.5 Continuous Distributions

A continuous distribution has a CDF that is defined as the probability that the value of X is less than or equal to x, i.e. $P[X \leq x]$. This is identified by the notation $F_X(x)$. The derivative of the CDF is called the probability density function, pdf, and describes how the probability is distributed over the range of X. It is also identified by the notation $f_X(x)$. Thus,

$$f_X(x) = \frac{dF_X(x)}{dx} \quad F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi \tag{A.15}$$

We can also talk about the probability that x lies between two values of X; it is simply the integral of $f_X(x)$ between those limits. The interval of integration must be finite if the result is to be a probability. For this reason $f_X(x)$ is called a density function. From the basic axioms of probability theory,

$$F_X(\infty) = \int_{-\infty}^{+\infty} f_X(\xi)\, d\xi = 1 \tag{A.16}$$

Continuous distributions have moments defined by

$$m_K = \int_{-\infty}^{+\infty} \xi^k \cdot f_X(\xi)\, d\xi \tag{A.17}$$

The first moment of a continuous function is the mean or expected value of the variable. The second moment taken about μ_X is the variance, and so forth. Algebraic manipulation gives

$$Var[X] = \sigma_X^2 = \int_{-\infty}^{+\infty} (\xi - \mu_X)^2 \cdot f_X(\xi)\, d\xi = \int_{-\infty}^{+\infty} \xi^2 \cdot f_X(\xi)\, d\xi - \mu_X^2 \tag{A.18}$$

The skewness coefficient is based on the third moment:

$$\nu_X = \frac{\displaystyle\int_{-\infty}^{+\infty} (\xi - \mu_X)^3 \cdot f_X(\xi)\, d\xi}{\sigma_X^3} \tag{A.19}$$

Higher moments are seldom used in reliability work, but the fourth moment is used to define the kurtosis (or 'flatness') coefficient:

$$\kappa_X = \frac{\int_{-\infty}^{+\infty} (\xi - \mu_X)^4 \cdot f_X(\xi)\, d\xi}{\sigma_X^4} \tag{A.20}$$

Table A.2 lists the parameters, the conventional symbol, the pdf, the mean, the variance, and the skewness coefficient or third moment for commonly encountered continuous variables. The Uniform distribution describes conditions when the variable X has equal probability density from a to b. The Triangular distribution has a pdf that rises linearly from zero at a to a peak at c and falls linearly back to zero at b.

The Normal distribution describes a large number of random processes. The Central Limit Theorem states that, in the limit, the sum of most random processes will be a Normal distribution. The logNormal distribution describes the product of many random processes. The logarithm of the variable X is Normally distributed.

The Exponential distribution arises from the Poisson process. It describes the time to the first occurrence in the Poisson process. The Gamma distribution describes the time until the cth occurrence in a Poisson process. The parameter b is equal to $1/\gamma$ from the Poisson process.

The Beta distribution is used to model variables that are bounded between 0 and 1. The bounds can be extended to a and b by simple algebra. Many different shapes can be achieved by judicious choice of p and q.

The Chi-squared and Student's t distributions are used in statistical tests. The von Mises distribution is a circular version of the normal distribution that is useful is studying the statistics of orientation of features such as joint sets. The Fisher (1953) distribution is a spherical extension of the Normal distribution used to describe the orientation of joint sets.

Table A.3 lists the parameters for the best known Extreme Value distributions. Type I distributions describe the limiting distributions of the largest or smallest values when the variable is unlimited in the upper and lower directions and the tails are approximately exponential. An example is the Normal distribution. Type II distributions are limited at one end but tail off at the other. They relate to the Type I distributions as the logNormal distribution relates to the Normal distribution. Type III distributions describe the smallest values when the underlying distributions are limited in the tail.

A.6 Multiple Variables

When several random variables exist, they can be represented by a joint pmf if they are discrete or a joint pdf if they are continuous. For example, if there are two discrete variables, X and Y, $p_{X,Y}(x, y)$ is the probability that X has the value x and Y has the value y, or

$$p_{X,Y}(x, y) \equiv P[(X = x) \cap (Y = y)] \tag{A.21}$$

The joint CDF is then

$$F_{X,Y} \equiv P[(X \leq x) \cap (Y \leq y)] = \sum_{x_i \leq x} \sum_{y_i \leq y} p_{X,Y}(x_i, y_i) \tag{A.22}$$

By analogy, the continuous pdf and CDF also exist. For the case of two continuous variables X and Y

$$P[(x_0 \le X \le x_i) \cap (y_0 \le Y \le y_i)] = \int_{x_0}^{x_i} \int_{y_0}^{y_i} f_{X,Y}(\xi, \eta)\, d\xi\, d\eta \tag{A.23}$$

The joint CDF is then

$$F_{X,Y} = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(\xi, \eta)\, d\xi\, d\eta \tag{A.24}$$

and the pdf is

$$f_{X,Y} = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} \tag{A.25}$$

These equations can be extended for more than two variables.

The marginal distribution is the distribution of a variable for all values of the other variables. For example, for two discrete variables, the marginal pmf of X is

$$p_X(x) = P[X = x] = \sum_{all\ y_i} p_{X,Y}(x, y_i) \tag{A.26}$$

and, for two continuous variables, the marginal pdf is

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, \eta)\, d\eta \tag{A.27}$$

The marginal CDF is

$$F_X(x) \equiv P[X \le x] = \sum_{x_j \le x} \sum_{all\ y_i} p_{X,Y}(x_j, y_i) \text{ or } \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{X,Y}(\xi, \eta)\, d\xi\, d\eta \tag{A.28}$$

The conditional pmf or pdf represents the distribution of a variable when all the other variables are held fixed. Thus, in the case of two variables, they are the distributions of X when Y has some fixed value. The pmf is

$$p_{X|Y}(x, y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \tag{A.29}$$

and the pdf is

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{A.30}$$

These relations can be elaborated for more variables.

The moments of multivariate distributions are found by extension from the definitions for single variables. Let Z be some function g of X and Y, i.e. $Z = g(X, Y)$. Then the expected value or mean of Z is

$$E[Z] \equiv \mu_Z = E[g(X, Y)] = \sum_{all x_i} \sum_{all y_i} g(x_i, y_i) p_{X,Y}(x_i, y_i) \tag{A.31}$$

or

$$E[Z] \equiv \mu_Z = E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(\xi, \eta)\, d\xi\, d\eta \tag{A.32}$$

There are three central second moments of $g(X, Y)$. The first is

$$\begin{aligned} E[(X - \mu_X)^2] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi - \mu_X)^2 f_{X,Y}(\xi, \eta)\, d\xi\, d\eta \\ &= \int_{-\infty}^{+\infty} (\xi - \mu_X)^2 f_X(\xi)\, d\xi = Var[X] \end{aligned} \tag{A.33}$$

Similarly, the second is

$$E[(Y - \mu_Y)^2] = Var[Y] \tag{A.34}$$

The third is

$$\begin{aligned} Cov[X, Y] = \sigma_{X,Y} &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi - \mu_X)(\eta - \mu_Y) f_{X,Y}(\xi, \eta)\, d\xi\, d\eta \end{aligned} \tag{A.35}$$

This is called the covariance of the two variables. The notation 'Cov' should not be confused with the abbreviation 'COV' for the coefficient of variation. To reduce confusion, we have usually used the symbol Ω for the latter. A further problem is that the units of the covariance are the same as those for the variance, which is σ^2. It is unfortunate that the symbol σ with appropriate subscripts is used for both the standard deviation and covariance, even though they have different units, but that is the convention that has grown up over the yeas.

The dimensionless expression of the degree of correlation between two variables is expressed by the correlation coefficient:

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{A.36}$$

When there are two or more variables, their variance and correlation structure can be expressed by the covariance matrix, $\mathbf{C}$, which is also denoted by $\mathbf{\Sigma}$:

$$\mathbf{C} \equiv \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1,X_2} & \cdots & \sigma_{X_1,X_n} \\ \sigma_{X_1,X_2} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2,X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_1,X_n} & \sigma_{X_2,X_n} & \cdots & \sigma_{X_n}^2 \end{bmatrix} \tag{A.37}$$

or by the correlation matrix, $\mathbf{K}$:

$$\mathbf{K} \equiv \begin{bmatrix} 1 & \rho_{X_1,X_2} & \cdots & \rho_{X_1,X_n} \\ \rho_{X_1,X_2} & 1 & \cdots & \rho_{X_2,X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{X_1,X_n} & \rho_{X_2,X_n} & \cdots & 1 \end{bmatrix} \tag{A.38}$$

It follows that

$$\mathbf{C} = \begin{bmatrix} \sigma_{X_1} & 0 & \cdots & 0 \\ 0 & \sigma_{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{X_n} \end{bmatrix} \cdot \mathbf{K} \cdot \begin{bmatrix} \sigma_{X_1} & 0 & \cdots & 0 \\ 0 & \sigma_{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{X_n} \end{bmatrix} \tag{A.39}$$

A.7 Functions of Random Variables

Much work in probability theory involves manipulating functions of random variables. If some set of random variables has known distributions, it is desired to find the distribution or the parameters of the distribution of a function of the random variables. Some simple results having wide application are as follows (X and Y are random variables; a and b are constants, $g(X)$ is a function of X):

$$E[aX] = aE[X] \quad Var[X] = a^2\, Var[X] \tag{A.40}$$

$$E[a + bX] = a + bE[X] \quad Var[a + bX] = b^2\, Var[X] \tag{A.41}$$

$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)] \tag{A.42}$$

The variance of the sum of two functions is not so simple, as the variance is not a linear function.

The linear combination of several random variables yields the following simple results. If

$$Y = \sum_{i=1}^{n} a_i X_i \tag{A.43}$$

then

$$E[Y] = \sum_{i=1}^{n} a_i E[X_i] \tag{A.44}$$

$$Var[Y] = \sum_{i=1}^{n} a_i^2\, Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_i a_j\, Cov[X_i, X_j] \tag{A.45}$$

When the variables are uncorrelated, the last equation simplifies to

$$Var[Y] = \sum_{i=1}^{n} a_i^2\, Var[X_i] \tag{A.46}$$

When variables are multiplied, the expected value of the result is

$$E[XY] = E[X]E[Y] + Cov[X, Y] \tag{A.47}$$

However, only when the variables are uncorrelated is there a simple expression for the variance:

$$Var[XY] = \mu_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2 + \sigma_X^2\sigma_Y^2 \tag{A.48}$$

and

$$\Omega_{XY}^2 = \Omega_X^2 + \Omega_Y^2 + \Omega_X^2\,\Omega_Y^2 \tag{A.49}$$

Many applications of functions of random variables involve integrating the probability density functions. The most common examples are calculations of the moments of a joint probability distribution. From calculus we know that, if there is a set of variables x_i that can be expressed in terms of another set y_i and we want to integrate some function $g(x_i)$, it may be more convenient to express the integral in terms of the y_i variables. However, it is then necessary to transform the dx_i into the dy_i, and this requires using the Jacobian determinant. The equation for two variables in each set is

$$\iint g(x_i, x_2)\,dx_1dx_2 = \iint g[x_1(y_1, y_2), x_2(y_1, y_2)]\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}dy_1dy_2 \tag{A.50}$$

The Jacobian determinant is

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} \tag{A.51}$$

For example, suppose that X_1 and X_2 are independent and uniformly distributed between 0 and 1. Define

$$\begin{aligned} Y_1 &= \sqrt{-2\ln X_1}\cos 2\pi X_2 \\ Y_2 &= \sqrt{-2\ln X_1}\sin 2\pi X_2 \end{aligned} \tag{A.52}$$

It follows that

$$\begin{aligned} X_1 &= \exp\left[-\frac{1}{2}(Y_1 + Y_2)^2\right] \\ X_2 &= \frac{1}{2\pi}\arctan\frac{Y_2}{Y_1} \end{aligned} \tag{A.53}$$

If we evaluate the partial derivatives of the X's and substitute them into Equation (A.51), we find

$$\frac{\partial(X_1, X_2)}{\partial(Y_1, Y_2)} = -\left[\frac{1}{\sqrt{2\pi}}e^{-Y_1^2/2}\right]\left[\frac{1}{\sqrt{2\pi}}e^{-Y_2^2/2}\right] \tag{A.54}$$

Thus,

$$dX_1dX_2 = -\left[\frac{1}{\sqrt{2\pi}}e^{-Y_1^2/2}\right]\left[\frac{1}{\sqrt{2\pi}}e^{-Y_2^2/2}\right]dY_1dY2 \tag{A.55}$$

which demonstrates that Equations (A.52) transform the uniform variables into independent standard Normal variables $\mathbf{N}(0.1)$. This result is useful in generating random Normal variables for Monte Carlo simulation. Extension of the Jacobian determinant for more than two variables is straightforward.

References

AASHTO (1994). *LRFD Bridge Design Specifications, SI Units*. Washington, DC, American Association of State Highway and Transportation Officials.

AASHTO (1997). "1997 Interims to LRFD Highway Bridge Design Specifications, SI Units, First Edition (1997 LRFD Interims)." Washington, DC, American Association of State Highway and Transportation Officials.

AASHTO (1998). *Standard Specifications for Highway Bridges*. Washington, DC, American Association of State Highway and Transportation Officials.

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*. New York, Dover Publications.

ACI (2002). *Building Code Requirements for Structural Concrete and Commentary (318-02 Building Code)*. Farmington Hills, MI, American Concrete Institute.

Adler, R. J. (1981). *The Geometry of Random Fields*. Chichester, New York, John Wiley & Sons.

Agterberg, F. P. (1974). *Geomathematics. Mathematical Background and Geo-Science Applications*. Amsterdam, New York, Elsevier.

AISC (1994). "Load and Resistance Factor Design." *Manual of Steel Construction*. Chicago, American Institute of Steel Construction.

Aitchison, J. (1970). *Choice against Chance: An Introduction to Statistical Decision Theory*. Reading, MA, Addison-Wesley.

Aitchison, J. and Brown, J. A. C. (1969). *The Lognormal Distribution, with Special Reference to Its Uses in Economics*. Cambridge, Cambridge University Press.

Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge, Cambridge University Press.

Alaszewski, A., Harrison, L. and Manthorpe, J. (1998). *Risk, Health, and Welfare: Policies, Strategies, and Practice*. Philadelphia, PA, Open University Press.

Allais, M. (1957). "Methods of appraising economic prospects of mining exploration over large territories." *Management Science* **3**: 285–347.

Allen, D. E. (1994). "The history and future of limit states design." *Journal of Thermal Insulation and Building Envelopes* **18**: 3–20.

Alloy, L. B. and Tabachnik, N. (1984). "Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information." *Psychological Review* **91**: 112–149.

Alpert, M. and Raiffa, H. (1982). "A progress report on the training of probability assessors." *Judgment Under Uncertainty, Heuristics and Biases*. Kahneman, D., Slovic, P. and Tversky, A., eds., Cambridge, Cambridge University Press: 294–306.

ANCOLD (1994). "*Guidelines on Risk Assessment 1994*." Australia-New Zealand Committee on Large Dams.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York, John Wiley & Sons.

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian

ISBN: 0-471-49833-5

Ang, A. H.-S. and Tang, W. H. (1975). *Probability Concepts in Engineering Planning and design, Vol I*., New York, John Wiley & Sons.

Ang, A. H.-S. and Tang, W. H. (1990). *Probability Concepts in Engineering Planning and Design, Vol. II*, Republication of book originally published by John Wiley & Sons.

Angulo, S. M. and Tang, W. H. (1996). "Groundwater monitoring system design using a probabilistic observational method for site characterization." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 797–812.

API (1989). "Draft recommended practice for planning, design, and constructing fixed offshore platforms–load and resistance factor design." Dallas, TX, American Petroleum Institute.

Arthur D. Little, Inc. (1978). *Safety assessment of nuclear waste burial*. Cambridge, Massachusetts, U. S. Environmental Protection Agency.

ASCE (1993). *Minimum Design Loads for Buildings and Other Structures (formerly ANSI A58.1)*. Reston, VA, American Society of Civil Engineers.

ASTM (1997). *Standards Related to Environmental Site Characterization*. Philadelphia, PA. American Society for Testing and Materials.

Atterberg, A. (1913). "Die Plastizitaet und Bindigkeit liefernde Bestandteile der Tone." *Int. Mitteil. Bodenkunde* **3**.

Auff, A. "The variability in dimensions of 'as constructed' arterial roads with particular reference to pavement levels." *Ninth Conference on the Australian Road Research Board*, Brisbane: 91–110.

Ayyub, B. and McCuen, R. (1997). *Probability, Statistics, and Reliability in Civil Engineering*. Palm Beach, FL, CRC Press.

Ayyub, B. M., Assakkaf, I., Atua, K. I., Melton, W. and Hess, P. (1997). "LRFD rules for naval surface ship structures: reliability-based load and resistance factor design rules." Washington, U.S. Navy, Naval Sea Systems Command.

Ayyub, B. M., Beach, J. and Packard, T. (1995). "Methodology for the development of reliability-based design criteria for surface ship structures." *Naval Engineers Journal, ASNE* **107**(1): 45–61.

Azzouz, A., Baligh, M. M. and Ladd, C. C. (1983). "Corrected field vane strength for embankment design." *Journal of Geotechnical Engineering, ASCE* **109**(3): 730–734.

Baecher, G. B. (1972). "Site Exploration: A probabilistic approach." PhD thesis, Civil Engineering, Massachusetts Institute of Technology.

Baecher, G. B. (1982a). "Simplified geotechnical data analysis." *Reliability Theory and Its Application in Structural and Soil Engineering*. Thoft-Christensen, P., ed., Dordrecht, Reidal Publishing.

Baecher, G. B. (1982b). "Statistical methods in site characterization." *Updating Subsurface Samplings of Soils and Rocks and their In-Situ Testing*, Santa Barbara, Engineering Foundation: 463–492.

Baecher, G. B. (1983a). "Professional judgment and prior probabilities in engineering risk assessment." *Proceedings, 4th International Conference on Applications of Statistics and Probability to Structural and Geotechnical Engineering*, Florence, Pitagora Editrice, Bologna: 635–650.

Baecher, G. B. (1987a). "Statistical Analysis of Geotechnical Data." Vicksburg, US Army Corps of Engineers, Waterways Experiment Station.

Baecher, G. B. (1987b). "Statistical quality control for engineered fills." U.S. Army Corps of Engineers, Waterways Experiment Station.

Baecher, G. B. (1987c). "Geotechnical Risk Assessment User's Guide." McLean, VA, Federal Highway Administration.

Baecher, G. B. (2002). "Event Tree Analysis: Theory." *Guide to Risk Analysis for Dam Safety*. Hartford, D., eds., Vancouver, Dam Safety Interest Group, Canadian Electricity Association. III (3.3.2).

Baecher, G. B. and Ingra, T. S. (1981). "Stochastic FEM in settlement predictions." *Journal of the Geotechnical Engineering Division, ASCE* **107**(GT4): 449–463.

Baecher, G. B. and Ladd, C. C. (1997). " Formal observational approach to staged loading." *Transportation Research Record* (Dec).

Baecher, G. B., Marr, W. A. and Consla, J. (1980). "Statistical properties of mine tailiings." Denver, CO, U.S. Bureau of Mines.

Baecher, G. B., Marr, W. A., Lin, J. S. and Consla, J. (1983). "Critical Parameters for Mine Tailings Embankments." Denver, CO, U. S. Bureau of Mines.

Baecher, G. B., Pate, E. M. and de Neufville, R. (1980). "Risk of dam failure in benefit/cost analysis." *Water Resources Research*, **16**(3): 449–456.

Barker, R. M., Duncan, J. M., Rojiani, K. B., Ooi, P. S. K., Tan, C. K. and Kim, S. G. (1991). "Manuals for the Design of Bridge Foundations, Shallow Foundations, Driven Piles, Retaining Walls and Abutments, Drilled Shafts, Estimation Tolerable Movements, Load Factor Design Specifications, and Commentary." Washington, DC, Transportation Research Board.

Barnett, V. (1982). *Comparative Statistical Inference*. Chichester, New York, John Wiley & Sons.

Barron, J. M. and Peterson, R. L. (1976). *Consumer Choice and Search Theory*. West Lafayette, IN, Institute for Research in the Behavioral Economic and Managerial Sciences, Krannert Graduate School of Management, Purdue University.

Baston, V. J. and Garnaev, A. Y. (2000). "A search game with a protector." *Naval Research Logistics Quarterly* **47**(2): 85–96.

Battelle Pacific Northwest Laboratories, Inc. (1981). "Risk assessment studies for high level waste disposal." Hanford, WA, US Department of Energy.

Bayes, T. (1763). "An essay toward solving a problem in the doctrine of chances." *Philosophical Transaction of the Royal Society (London)* **53**: 370–418.

Bea, R. G. (1998). "Reliability characteristics of a platform in the Mississippi River Delta." *Journal of Geotechnical and Geoenvironmental Engineering ASCE*, **124**(8): 729–738.

Becker, D. E. (1996). "Limit states design for foundations. Part I. An overview of the foundation design process." *Canadian Geotechnical Journal* **32**: 956–983.

Bedford, T. and Cooke, R. (2001). *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge, Cambridge University Press.

Benjamin, J. R. and Cornell, C. A. (1970). *Probability, Statistics, and Decision for Civil Engineers* New York, McGraw-Hill.

Benkoski, S. J., Monticino, M. G. and Weisinger, J. R. (1991). "A survey of search theory literature." *Naval Research Logistics Quarterly* **38**: 469–494.

Benson, C. H. (1993). "Probability Distributions for Hydraulic Conductivity of Compacted Soil Liners." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **119**(3): 471–486.

Berger, J. O. (1993). *Statistical Decision Theory and Bayesian Analysis*. New York, Springer-Verlag.

Berger, J. O., De Oliveira, V. and Sanso, B. (2001). "Objective Bayesian analysis of spatially correlated data." *Journal of the American Statistical Association* **96**(456): 1361–1374.

Bernstein, P. L. (1996). *Against the Gods: The Remarkable Story of Risk*. New York, John Wiley & Sons.

Bertrand, J. (1907). *Calcul des probabilités*. Paris, Gauthier-Villars.

Bier, V. M. (1997). "An overview of probabilistic risk analysis for complex engineered systems." *Fundamentals of Risk Analysis and Risk Management*. Molak, V. (ed.), Boca Raton, Lewis Publishers.

Bjerrum, L. (1960). "Some notes on Terzaghi's method of working." *From Theory to Practice in Soil Mechanics: Selections from the writing of Karl Terzaghi*. Bjerrum, L., Casagrande, A., Peck, R. B. and Skempton, A. W., eds., New York, John Wiley & Sons: 22–25.

Bjerrum, L. (1972). "Embankments on soft ground: State-of-the-art report." *Specialty Conference on Performance of Earth and Earth-Supported Structures*, *ASCE*, Lafayette, Indiana: 1–54.

Bjerrum, L. (1973). "Problems of soil mechanics and construction on soft clays." *Eighth International Conference on Soil Mechanics and Foundation Engineering*, Moscow: 111–159.

Black, W. L. (1962). "Sequential Search." PhD thesis, Electrical Engineering, Massachusetts Institute of Technology.

Blodgett, C. F., Jakubauskas, M. E., Price, K. P., and Martinko, E. A. "Remote sensing-based geostatistical modeling of forest canopy structure." *American Society for Photogrammetry and Remote Sensing (ASPRS) Annual Meeting*.

Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Berkeley, CA, University of California Press.

Box, G. E. P. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. New York, John Wiley & Sons.

Box, J. F. (1978). *R. A. Fisher, the Life of a Scientist*. New York, John Wiley & Sons.

Breitung, K. W. (1984). "Asymptotic approximations for multinormal integrals." *Journal of Engineering Mechanics, ASCE* **110**(3): 357–366.

Breitung, K. W. (1994). *Asymptotic Approximations for Probability Integrals*. Berlin, Springer-Verlag.

Broemeling, L. D. (1985). *Bayesian analysis of linear models*. New York, M. Dekker.

Brown, A. A. (1960). "Search theory and problems of exploration boring." *Industry Experiment Station Bulletin* **72**: 33–37.

Brown, L. (2002). *The New Shorter Oxford English Dictionary*. Oxford, Oxford University Press.

Brunswik, E. (1969). *The Conceptual Framework of Psychology*. Chicago, University of Chicago Press.

Buckle, H. T. (1858). *History of civilization in England*. New York, Appleton.

Budnitz, R. H., Apostolakis, G., Boore, D. M., Culff, L. S., Coppersmith, K. J., Cornell, C. A. and Morris, P. A. (1997). "Recommendations for Probabilistic Seismic Hazard analysis: Guidance on Uncertainty and Use of Experts, NUREG/CR-6372." Report of the Senior Seismic Hazard Analysis Committee (SSHAC), Washington, DC, U. S. Nuclear Regulatory Commission.

Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. Berlin, Heidelberg, New York, Springer-Verlag.

Burington, R. S. and May, D. C., Jr. (1970). *Handbook of Probability and Statistics*. New York, McGraw-Hill Book Company.

Bury, K. V. and Kreuzer, H. (1985). "The assessment of the failure probability for a gravity dam." *Water Power and Dam Construction* **37**.

Bury, K. V. and Kreuzer, H. (1986). "The assessment of risk for a gravity dam." *Water Power and Dam Construction* **38**(2).

Canadian Geotechnical Society (1992) *Canadian Foundation Engineering Manual*, 3*rd Edition*, BiTech Publishers Ltd., Richmond, BC.

Carnap, R. (1936). "Testability and meaning." *Philosophy of science* **3**: 420.

Casagrande, A. (1932). "Research on the Atterberg limits of soils." *Public Roads* (Oct.).

Casagrande, A. (1936). "The determination of the pre-consolidation load and its practical significance." *International Conference on Soil Mechanics and Foundation Engineering*, Cambridge, MA.

Casagrande, A. (1965). "The role of the 'calculated risk' in earthwork and foundation engineering." *Journal of the Soil Mechanics and Foundations Division, ASCE* **Vol. 91**(SM4): 1–40.

Cetin, K. O., Der Kiureghian, A. and Seed, R. B. (2002). "Probabilistic models for the initiation of seismic soil liquefaction." *Structural Safety* **24**(1): 67–82.

Charnes, A. and Cooper, W. W. (1958). "The theory of search: Optimum distribution of search effort." *Management Science* **5**: 44–50.

Chatillon, G. (1984). "The balloon rules for a rough estimate of the correlation coefficient." *The American Statistician* **38**(1): 58–60.

Chew, M. G., Jr. (1967). "A sequential search procedure." *Annals of Mathematical Statistics* **38**: 494–502.

Chiasson, P., Lafleur, J., Soulie, M. and Law, K. T. (1995). "Characterizing spatial variability of a clay by geostatistics." *Canadian Geotechnical Journal* **32**: 1–10.

Chow, V. T., Maidment, D. R. and Mays, L. W. (1988). *Applied Hydrology*. New York, McGraw-Hill.

Christakos, G. (1992). *Random Field Models in Earth Sciences*. San Diego, Academic Press.

Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*. New York, Oxford University Press.

Christakos, G. and Hristopulos, D. T. (1998). *Spatiotemporal Environmental Health Modeling: A Tractatus Stochasticus*. Boston, MA, Kluwer Academic Publishers.

Christian, J. T., and Baecher, G. B. (1999) "Point-estimate method as numerical quadrature," *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, **125**(9): 779–787.

Christian, J. T. and Baecher, G. B. (2002) "The point-estimate method with large numbers of variables," *International Journal of Numerical and Analytical Methods in Geomechanics*, **26**(15): 1515–1529.

Christian, J. T., Ladd, C. C. and Baecher, G. B. (1994). "Reliability applied to slope stability analysis." *Journal of Geotechnical Engineering, ASCE* **120**(12): 2180–2207.

Christian, J. T. and Swiger, W. F. (1975). "Statistics of liquefaction and SPT results." *Journal of the Geotechnical Engineering Division, ASCE* **101**(GT11): 1135–1150.

Christian, J. T. and Swiger, W. F. (1976). "Closure to 'Statistics of liquefaction and SPT results"' *Journal of the Geotechnical Engineering Division, ASCE* **102**(GT12): 1279–1281.

Chudnovsky, D. and Chudnovsky, G. (1989). *Search Theory: Some Recent Developments*. New York, Marcel Dekker.

CIRIA (1977). "Rationalization of safety and serviceability factors in structural codes." London, Construction Industries Research and Information Association.
Clemen, R. T. and Reilly, T. (2001). *Making Hard Decisions with Decision Tools*. Pacific Grove, CA, Duxbury/Thomson Learning.
Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models & Applications*. London, Pion.
Cochran, W. G. (1977). *Sampling Techniques*. New York, John Wiley & Sons.
Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York, Oxford University Press.
Cooksey, R. W. (1996). *Judgment Analysis: Theory, Methods, and Applications*. San Diego, Academic Press.
Cornell, C. A. (1968). "Engineering seismic risk assessment." *Bulletin of the Seismological Society of America* **58**: 1583–1606.
Corotis, R. B., Azzouz, A. S. and Krizek, R. (1975). "Statistical evaluation of soil index properties and constrained modulus." *Second International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Aachen: 273–294.
Cramér, H. and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. New York, John Wiley & Sons.
Cranley, E. C. (1969). "Variations in bituminous construction." *Public Records* **35**.
Cressie, N. A. C. (1991). *Statistics for Spatial Data*. New York, John Wiley & Sons.
Danskin, J. M. (1962). "A theory of reconnaissance I." *Operations Research* **10**: 285–299.
Daston, L. (1988). *Classical Probability in the Enlightenment*. Princeton, NJ, Princeton University Press.
Davenport, T. H., and Prusak, L. (1998). "Working knowledge how organizations manage what they know." Boston, MA, Harvard Business School Press.
Davenport, W. B. (1970). *Probability and random processes; an introduction for applied scientists and engineers*. New York, McGraw-Hill.
David, F. N. (1962). *Games, Gods and Gambling: The Origins and History of Probability and Statistical Ideas from the Earliest Times to the Newtonian Era*. New York, Hafner.
David, M. (1977). *Geostatistical Ore Reserve Estimation*. Amsterdam; Amsterdam, Elsevier.
Davis, J. C. (1986). *Statistics and Data Analysis in Geology*. New York, John Wiley & Sons.
Davisson, M. T. (1972) "High capacity piles," *Soil Mechanics Lecture Series on Innovations in Foundation Construction*, Chicago, Illinois: 81–112.
De Beer, E. E. (1970). "Experimental determination of the shape factors and bearing capacity factors of sand." *Geotechnique*, **23**(4): 387–411.
De Beer, E. E., and Ladanyi, B., (1961) "Étude expérimetale de la capacité portante du sable sous des fondations circulaires établies en surface." *Fifth International Conference on Soil Mechanics and Foundation Engineering*, Paris.
De Finetti, B. (1937). "Foresight: its logical laws, its subjective sources." *Studies in Subjective Probability*. Kyburg, H. and Smokler, H., eds., New York, John Wiley & Sons: 93–158.
De Finetti, B. (1972). *Probability, Induction and Statistics: The Art of Guessing*. London, New York, John Wiley & Sons.
De Finetti, B. (1974). *Theory of Probability: A Critical Introductory Treatment*. London, New York, John Wiley & Sons.
De Geoffroy, J. G. and Wignall, T. K. (1985). *Designing optimal strategies for mineral exploration*. New York, Plenum Press.
De Groot, M. H. (1970). *Optimal Statistical Decisions*. New York, McGraw-Hill.
de Guenin, J. (1961). "Optimal distribution of effort: an extension of the Koopman basic theory." *Operations Research* **9**: 1–7.
DeGroot, D. J. (1985). "Maximum Likelihood Estimation of Spatially Correlated Soil Properties." MS thesis, Civil Engineering, Massachusetts Institute of Technology.
DeGroot, D. J. (1996). "Analyzing spatial variability of in situ soil properties." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 210–238.
DeGroot, D. J. and Baecher, G. B. (1993). " Estimating autocovariance of in situ soil properties." *Journal of the Geotechnical Engineering Division, ASCE* **119**(GT1): 147–166.
de Morgan, A. (1845). "Theory of probabilities." *Encyclopaedia Metropolitana*, B. Fellowes, ed., London: 393–490.

Der Kiureghian, A. and De Stefano, M. (1991). "Efficient algorithm for second-order reliability analysis." *Journal of Engineering Mechanics, ASCE* **117**(12): 2904–2923.

Der Kiureghian, A., Lin, H.-Z. and Hwang, S.-J. (1987). "Second-order reliability approximations." *Journal of Engineering Mechanics, ASCE* **113**(8): 1208–1225.

Dershowitz, W., Wallman, P. and Kindred, S. (1991). "Discrete fracture modelling for the Stripa site characterization and validation drift inflow predictions." Seattle, WA, Golder Associates.

Desai, C. S. and Christian, J. T. (1977). *Numerical Methods in Geotechnical Engineering*. New York, McGraw-Hill Book Co.

Dhillon, G. S. (1961). "The settlement, tilt, and bearing capacity of footings under central and eccentric loads." *Journal National Buildings Organization*, **6**.

Dillon, W. R. and Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. New York, John Wiley & Sons.

Dingman, S. L. (2002). *Physical Hydrology*. Upper Saddle River, NJ, Prentice-Hall.

Ditlevsen, O. (1997). "Structural reliability codes for probabilistic design – a debate paper based on elementary reliability and decision analysis concepts." *Structural Safety*, **19**(3): 253–270.

Dobbie, J. M. (1968). "A survey of search theory." *Operations Research* **16**: 525–537.

Dobry, R. and Alvarez, L. (1967). "Seismic failures of Chilean tailings dams." *Journal of the Soil Mechanics and Foundations Division, ASCE* **93**(6): 237–260.

Douglas, M. and Wildavsky, A. B. (1982). *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers*. Berkeley, University of California Press.

Dowding, C. H. (1979). "Perspectives and challenges of site characterization." *Site Characterization and Exploration*, *ASCE*, Evanston, IL: 10–38.

Dowding, C. H., ed., (1979). *Site Characterization and Exploration*, *ASCE*, Evanston, IL.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. New York, John Wiley & Sons.

Drew, J. L. (1966). "Grid drilling exploration and its application to the search for petroleum." PhD thesis, Geology, Pennsylvania State University.

Duda, R. O., Hart, P. E. and Stork, D. G. (2001). *Pattern Classification*. New York, John Wiley & Sons.

Duncan, J. M. (1999). "The use of back analysis to reduce slope failure risk: The Seventh Annual Arthur Casagrande Memorial Lecture." *Civil Engineering Practice, Journal of the Boston Society of Civil Engineers Section, ASCE* **14**(1): 75–91.

Duncan, J. M. (2000). "Factors of safety and reliability in geotechnical engineering." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **126**(4): 307–316.

Duncan, J. M., Tan, C. K., Barker, R. M. and Rojiani, K. B. (1989). "Load and resistance factor design of bridge structures." *Proceedings of the Symposium on Limit States Design in Foundation Engineering*, Toronto, Canadian Geotechnical Society, Southern Ontario Section. 47–63.

Edwards, W. and Tversky, A. (1967). *Decision Making: Selected Readings*. Harmondsworth, Penguin.

Einstein, H. H. and Baecher, G. B. (1983). "Probabilistic and statistical methods in engineering geology, Part one: Exploration." *Rock Mechanics and Rock Engineering* **16**(1): 39–72.

Einstein, H. H., Halabe, V. B., Dudt, J.-P. and Descoeudres, F. (1996). "Geologic uncertainties in tunneling." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 239–253.

Einstein, H. H., Labreche, D. A., Markow, J. J. and Baecher, G. B. (1978). "Decision analysis applied to rock exploration" *Engineering Geology* **12**(2): 143–161.

Elishakoff, I. (1999). *Probabilistic Theory of Structures*. Mineola, NY, Dover Publications.

Ellingwood, B., Galambos, T. V., MacGregor, J. G. and Cornell, C. A. (1980). "Development of a probability-based load criterion for American National Standard A58." Washington, DC, National Bureau of Standards.

Engle, J. H. (1957). "Use of clustering in mineralogical and other surveys." *First International Conference on Operations Research*, Oxford, The English Universities Press: 176–192.

Enslow, P. H., Jr. (1966). "A bibliography of search theory and reconnaissance theory literature." *Naval Research Logistics Quarterly* **13**: 177–202.

European, Committee for Standardization (1993). "Eurocode 7." Copenhagen, Danish Geotechnical Institute.

Evans, A. W. and Verlander, N. Q. (1997). "What is wrong with criterion FN-lines for judging the tolerability of risk?" *Risk Analysis* **17**: 157–167.

Evans, M., Hastings, N. and Peacock, B. (1993). *Statistical Distributions*. New York, John Wiley & Sons, Inc.

Everitt, B. and Dunn, G. (2001). *Applied Multivariate Data Analysis*. London & New York, Oxford University Press.

Fellenius, B. H. (1994). "Limit states design for deep foundations." *Proceedings, U.S. DOT International Conference on Deep Foundations*, Orlando, Federal Highway Commission.

Feller, W. (1967). *An introduction to probability theory and its applications, v. 1*, 3^{d} Ed., Wiley, New York.

Feller, W. (1971). *An introduction to probability theory and its applications, v. 2*, 2nd Ed., Wiley, New York.

Fenton, G. A. (1990). "Simulation and Analysis of Random Fields." PhD Thesis, Civil Engineering and Operations Research, Princeton University.

Fenton, G. A. and Griffiths, D. V. (1996). "Statistics of free surface flow through stochastic earth dam." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **122**(6): 427–436.

Fenton, G. A. and Griffiths, D. V. (2002). "Probabilistic foundation settlement on spatially random soil." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **128**(5): 381–390.

Fenton, G. A., Paice, G. M. and Griffiths, D. V. (1996). "Probabilistic analysis of foundation settlements." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 651–665.

Fenton, G. A. and Vanmarcke, E. H. (1991). "Simulation of random fields via local average simulation." *Journal of Engineering Mechanics, ASCE* **116**(8): 1733–1749.

Ferrell, W. R. and McGoey, P. J. (1980). "A model of calibration for subjective probabilities." *Organizational Behavior and Human Performance* **26**: 32–53.

Fishburn, P. C. (1964). *Decision and value theory*, Wiley, New York.

Fishburn, P. C. (1970). *Utility theory for decision making*, Wiley, New York.

Fischhoff, B., Slovic, P. and Lichtenstein, S. (1997). "Knowing with certainty: The appropriateness of extreme confidence." *Journal of Experimental Psychology* **3**(4): 552–564.

Fisher, R. A. (1921). "On the mathematical foundations of theoretical statistics." *Philosophical Transactions of the Royal Society, Series A* **222**: 309.

Fisher, R. A. (1925). "Theory of statistical estimation." *Proceedings of the Cambridge Philosophical Society* **26**: 528.

Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh, Oliver and Boyd.

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." *Annals of Eugenics* **7**: 179–188.

Fisher, R. A. (1953). "Dispersion on a sphere." *Proceedings of the Royal Society, Series A* **217**: 295–305.

Fisher, R. A. and Bennett, J. H. (1971). *Collected papers of R. A. Fisher*. Adelaide, University of Adelaide.

Fisher, R. A. and Yates, F. (1938). *Statistical tables for biological, agricultural and medical research*. London etc., Oliver and Boyd.

Fishman, G. S. (1995). *Monte Carlo: Concepts, Algorithms, and Applications*. New York, Springer-Verlag.

Folayan, J., Höeg, K. and Benjamin, J. (1970). "Decision theory applied to settlement predictions." *Journal of the Soil Mechanics and Foundations Division* **96**(SM4): 1127–1141.

Fredlund, D. G. and Dahlman, A. E. "Statistical geotechnical properties of glacial Lake Edmonton sediments." *First International Conference on Applications of Statistics and Probability to Soil and Structural Engineering*, Hong Kong: 203–228.

French, S. and Raios Insua, D. (2000). *Statistical Decision Theory*. London, Arnold.

Freudenthal, A. M. (1951). "Planning and interpretation of fatigue tests." *Symposium on Statistical Aspects of Fatigue*, ASTM Special Technical Publication.

Freudenthal, A. M. and Gumbel, E. J. (1956). "Physical and statistical aspects of fatigue." *Advances in Applied Mechanics* **4**.

Freudenthal, A. M., Garrelts, J. M. and Shinozuka, M. (1966). "The analysis of structural safety." *Journal of the Structural Division, ASCE* **92**(ST1): 267–325.

Frost, J. R. (2000). "Principles of search theory." International Search and Rescue Society, Ottawa, Ont.

Fu, K. S. (1968a). "Relationship among various learning techniques in pattern recognition systems." *Pattern Recognition*, L. N. Kanal, ed., Thompson Book Co., Washington, DC, 399–408.

Fu, K. S. (1968b). *Sequential Methods in Pattern Recognition and Machine Learning*, Academic Press, NY.

Fu, K. S. (1970). "Statistical pattern recognition." *Adaptive Learning and Pattern Recognition Systems*, J. M. Mendal and K. S. Fu, eds., Academic Press, New York, NY.

Galambos, T. V. and Ravindra, M. K. (1978). "Properties of steel for use in LRFD." *Journal of Structural Engineering, ASCE* **84**(9): 1459–1468.

Gelb, A. and The Analytical Sciences Corporation Technical Staff (1974). *Applied Optimal Estimation*. M. I. T. Press, Cambridge, MA.

Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology*. Englewood Cliffs, N.J., Prentice-Hall.

Gettys, C. M., Steiger, J. H., Kelly, C. W. and Peterson, C. R. (1973). "Multiple-stage probabilistic information Processing." *Organizational Behavior and Human Performance* **10**.

Gibbs, H. J. and Holtz, W. G. (1957). "Research on determining the density of sands by spoon penetration testing." *Proceedings, 4th International Conference on Soil Mechanics and Foundation Engineering*, London.

Gigerenzer, G. (1991). "How to make cognitive illusions disappear: Beyond 'heuristics and biases'." *European Review of Social Psychology* **2**: 83–115.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. and Kruger, L. (1989). *The Empire of Chance: How Probability Changed Science and Everyday Life*. Cambridge, Cambridge University Press.

Gilbert, P. (2002). Personal communication.

Gilbert, R., Wright, S. and Liedtke, E. (1998). "Uncertainty in Back Analysis of Slopes: Kettleman Hills Case History." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **124**(12): 1167–1176.

Gilbert, R. B. and McGrath, T. C. (1997a). "Design of site investigation programs for geotechnical engineering." *Uncertainty Modeling and Analysis in Civil Engineering*, CRC Press, Boca Raton, FL. 447–484.

Gilbert, R. B. and McGrath, T. C. (1997b). "Site investigation design for geotechnical engineering." *Uncertainty Modeling and Analysis in Civil Engineering*. Ft. Lauderdale, CRC Press.

Gilbert, R. B. and McGrath, T. C. (1999). "Analytical method for designing and analyzing 1D search programs." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **125**(12): 1043–1056.

Gilbert, R. B. and Tang, W. H. (1989). "*Progressive failure probability of soil slopes containing geologic anomalies*." *International Conference on Structural Safety and Reliability*, San Francisco: 255–262.

Gillham, N. W. (2001). *A Life of Sir Francis Galton: From African Exploration to the Birth of Eugenics*. London, Oxford University Press.

Gleick, J. (1988). *Chaos: Making of a New Science*. New York, Penguin.

Gluss, B. (1959). "An optimum policy for detecting a fault in a complex system." *Operations Research* **7**: 468–477.

Goble, G. G. (1998). "Geotechnical Related Development and Implementation Load and Resistance Factor Design (LRFD) Methods." Washington, DC, National Cooperative Highway Research Program.

Goldstein, M. and Dillon, W. R. (1978). *Discrete Discriminant Analysis*. New York, John Wiley & Sons.

Grayson, C. J. (1960). *Decisions under Uncertainty: Drilling Decisions by Oil and Gas Operators*. Cambridge, Harvard University Division of Research Graduate School of Business Administration.

Griffiths, D. V. and Fenton, G. A. (1997). "Three-dimensional seepage through spatially Random soil." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **123**(2): 153–160.

Griffiths, D. V. and Fenton, G. A. (1998). "Probabilistic analysis of exit gradients due to steady seepage." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **124**(9): 789–797.

Gruska, G. F., Mirkhani, K. and Lamberson, L. R. (1973). *Point estimation in non normal samples: using Pearson frequency curves, and Burr cumulative functions*. Published by the authors.

Gui, S., Zhang, R., Turner, J. P. and Xue, X. (2000). "Probabilistic slope stability analysis with stochastic soil hydraulic conductivity." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **126**(1): 1–9.

Gumbel, E. J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications: A Series of Lectures*. Washington, DC, U. S. Government Printing Office.
Gumbel, E. J. (1958). *Statistics of extremes*. New York, Columbia University Press.
Hacking, I. (1965). *Logic of Statistical Inference*. Cambridge, Cambridge University Press.
Hacking, I. (1975). *The Emergence of Probability*. Cambridge, Cambridge University Press.
Hacking, I. (1990). *The Taming of Chance*. Cambridge, New York, Cambridge University Press.
Hacking, I. (1999). *The social construction of what?* Harvard University Press, Cambridge.
Hacking, I. (2001). *An Introduction to Probability and Inductive Logic*. Cambridge, Cambridge University Press.
Hahn, G. J. and Shapiro, S. S. (1967). *Statistical Models in Engineering*. New York, John Wiley & Sons.
Haldane, J. B. S. (1932). "A mathematical theory of natural and artificial selection, part 9." *Proceedings of the Cambridge Philosophical Society* **28**: 58.
Haldar, A. and Mahadevan, S. (2000). *Reliability Assessment Using Stochastic Finite Elements*. New York, John Wiley & Sons.
Haldar, A. and Tang, W. H. (1979). "Probabilistic evaluation of liquefaction potential." *Journal of the Geotechnical Division, ASCE* **105**(2): 145–163.
Haley, K. B. and Stone, L. D. (1980). *Search Theory and Applications*. New York, Plenum Press.
Halim, I. and Tang, W. (1993). "Site exploration strategy for geologic anomaly characterization." *Journal of Geotechnical Engineering, ASCE*, **119**(2): 185–213.
Halim, I. S. (1991). "Reliability of Geotechnical System Considering Geological Anomaly." PhD thesis, Civil Engineering, University of Illinois.
Halim, I. S. and Tang, W. H. (1990). "Bayesian method for characterization of geological anomaly." *Proceedings of the First International Symposium On Uncertainty Modeling and Analysis*, College Park, MD, IEEE Los Alamitos, California: 585–594.
Halim, I. S. and Tang, W. H. (1991). "Reliability of undrained clay slope considering geologic anomaly." *International Conference on Applications of Statistics and Probability to Structural and Geotechnical Engineering*, Mexico City: 776–783.
Halim, I. S., Tang, W. H. and Garrett, J. H. (1991). "Knowledge-assisted interactive probabilistic site characterization." *Geotechnical Engineering Congress, GSP 27, ASCE*: 264–275.
Hamming, R. W. (1962). *Numerical Methods for Scientists and Engineers*. New York, Dover Publications.
Hammitt, G. M. (1966). "Statistical Analysis of Data from a Comparative Laboratory Test Program Sponsored by ACIL." Vicksburg, MS, U. S. Army Waterways Experiment Station.
Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York, Oxford University Press.
Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*, MIT Press, Cambridge, MA.
Haralick, R. M. and Dinstein, I. (1971). "An iterative clustering procedure." *IEEE Transactions* **SMC-1**: 275–289.
Harbaugh, J. W., Davis, J. C. and Wendebourg, J. (1995). *Computing Risk for Oil Prospects: Principles and Programs*. Oxford, Pergamon.
Harbaugh, J. W., Doveton, J. H. and Davis, J. C. (1977). *Probability Methods in Oil Exploration*. New York, John Wiley & Sons.
Harr, M. E. (1987). *Reliability-Based Design in Civil Engineering*. New York, McGraw-Hill.
Harr, M. E. (1989). "Probabilistic estimates for multivariate analyses." *Applied Mathematical Modelling* **13**(No. 5): pp. 313–318.
Hartford, D. N. D. (2000) "Judged values and value judgements in dam risk assessment: A personal perspective." *ANCOLD Bulletin* **114**: 78–86.
Hasofer, A. M. and Lind, N. C. (1974). "An exact and invariant first-order reliability format." *Journal of the Engineering Mechanics Division, ASCE* **100**(EM1): 111–121.
Hasofer, A. M. and Wang, J. Z. (1993). "Determining failure probability by importance sampling based on higher order statistics." *Probabilistic Methods in Geotechnical Engineering*, Canberra, Australia, A. A. Balkema: 131–134.
Haymsfield, E. (1971). "Probabilistic approach for determination of liquefaction potential." *Journal of Energy Engineering, ASCE* **125**(1): 18–33.

Hazen, A. (1911). "Discussion of 'Dams on sand foundations' by A.C. Koenig." *Transactions of the American Society of Civil Engineers* **73**: 199.
Hildebrand, F. B. (1956). *Introduction to Numerical Analysis*. New York, McGraw-Hill.
Hildebrand, F. B. (1976). *Advanced Calculus for Applications*. Englewood-Cliffs, NJ, Prentice-Hall.
Hilldale(-Cunningham), C. (1971). "A Probabilistic Approach to Estimating Differential Settlement." MS Thesis, Department of Civil Engineering, Massachusetts Institute of Technology.
Hillel, D. (1998). *Environmental Soil Physics*. San Diego, CA, Academic Press.
Ho, C. L. and Kavazanjian, E., Jr. (1986). "Probabilistic study of SPT liquefaction analysis." *Use of In situ Test in Geotechnical Engineering*, Blacksburg, VA, ASCE. 602–616.
Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (2000). *Understanding Robust and Exploratory Data Analysis*. New York, John Wiley & Sons.
Hoede-Keyser, J. (1970). "Design, specification, control and performance survey: a quality entity." *Symposium on Quality Control of Road Works*, Centre d'Etudes Techniques de l'Equipement.
Hoeg, K. and Murarka, R. P. (1974). "Probabilistic analysis of a retaining wall." *Journal of the Geotechnical Engineering Division, ASCE* **100**(GT3): 349–370.
Hogarth, R. (1975). "Cognitive processes and the assessment of wubjective probability distributions." *Journal of the American Statistical Association* **70**(35): 271–294.
Holling, C. S. (1980) *Adaptive Environmental Assessment and Management*, International Series on Applied Systems Analysis, Toronto, John Wiley & Sons.
Holtz, R. D. and Krizek, R. J. (1971). "Statistical evaluation of soils test data." *First International Conference on Applications of Statistics and Probability to Soil and Structural Engineering*, Hong Kong, Hong Kong University Press: 229–266.
Hong, H. P. (1996). "Point-estimate moment-based reliability analysis." *Civil Engineering Systems* **13**(4): 281–294.
Hong, H. P. (1998). "An efficient point estimate method for probabilistic analysis." *Reliability Engineering and System Safety* **59**(No. 3): pp. 261–267.
Hong Kong Government Planning Department (1994). "Hong Kong Planning Standards and Guidelings, Chapter 11, Potentially Hazardous Installations." Hong Kong.
HSE (1982). "A Guide to the Notification of Installations Handling Hazardous Substances Regulation." London, HSMO.
HSE (1991). "Major Hazard Aspects of the Transportation of Dangerous Goods." London, HMSO.
HSE (1992). "Safety Assessment Principles for Nuclear Plants." London, HMSO.
Huberty, C. J. (1994). *Applied discriminant analysis*, Wiley, New York, N. Y.
Hughes, T. J. R. (1987). *The Finite Element Method*. Englewood Cliffs, N. J., Prentice-Hall.
Hvorslev, M. J., (1949) *Subsurface exploration and sampling of soils for civil engineering purposes: report on a research project of the Committee on Sampling and Testing, Soil Mechanics and Foundations Division, American Society of Civil Engineers*. New York, Reprinted by Engineering Foundation.
Hynes, M. and Vanmarcke, E. (1977). "Reliability of Embankment performance predictions." *Proceedings ASCE Engineering Mechanics Division Specialty Conference. Mechanics in Engineering*, Univ. of Waterloo Press: 367–384.
Ingles, O. G. and Metcalf, J. B. (1973). *Soil Stabilization: Principles and Practice*. New York, John Wiley & Sons.
Ingles, O. G. and Noble, C. S. (1975). "The evaluation of base course materials." *Soil Mechanics-Recent Developments*. Valliappan, S., Hain, S. J. and Lee, I. K. Sydney, Unisearch Ltd.: 399–410.
Ingra, T. S. and Baecher, G. B. (1983). "Uncertainty in bearing capacity of sands." *Journal of Geotechnical Engineering, ASCE* **109**(7): 899–914.
Isaaks, E. H. and Srivastava, R. M. (1989). *Applied Geostatistics*. New York, Oxford University Press.
Jaksa, M. B., Brooker, P. I. and Kaggwa, W. S. (1997). "Inaccuracies associated with estimating random measurement errors." *Journal of the Geotechnical Engineering Division, ASCE* **123**(5): 393–401.
James, W. (1907). *Pragmatism, a New Name for Some Old Ways of Thinking*. New York, Longmans Green & Co.
Javete, D. F. (1983). "A simple statistical approach to differential settlements on clay." PhD thesis, Civil Engineering, University of California, Berkeley.

Jaynes, E. T. (1996). "Probability theory: the logic of science." Fragmental publication on internet site, omega.albany.edu:8008/JaynesBook.html.
Jeffreys, H. (1983). *Theory of Probability*. Oxford, New York, Oxford University Press.
Jensen, J. L. (1997). *Statistics for Petroleum Engineers and Geoscientists*. Upper Saddle River, NJ, Prentice Hall.
Johnson, J. (1960). *Econometric Methods*. New York, McGraw-Hill.
Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*. Boston, Houghton Mifflin.
Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions*. New York, Houghton Mifflin.
Journel, A. G. and Huijbregts, C. (1978). *Mining Geostatistics*. London, New York, Academic Press.
Jowett, G. H. (1952). "The accuracy of systematic sampling from conveyer belts." *Applied Statistics* **1**: 50–59.
Juang, C. H., Juang, T. and Andrus, R. D. (2002). "Assessing probability-based methods for liquefaction potential evaluation." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **128**(7): 580–589.
Kaderabek, T. J. and Reynolds, R. T. (1981). "Miami limestone foundation design and construction." *Journal of the Geotechnical Engineering Division, ASCE* **107**(GT7): 859–872.
Kahaner, D., Moler, C. and Nash, S. (1989). *Numerical Methods and Software*. Englewood Cliffs, N.J., Prentice-Hall.
Kahneman, D., Slovic, P. and Tversky, A., Eds. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge., Cambridge Univ. Press.
Kahneman, D. and Tversky, A. (1982a). "On the psychology of prediction judgment under uncertainty: heuristics and biases." *Judgment under Uncertainty: Heuristics and Biases*. Kahneman, D., Slovic, P. and Tversky, A., eds., New York, Cambridge Univ. Press: 48–68.
Kahneman, D. and Tversky, A. (1982b). "Subjective probability: A judgment of representativeness." *Judgment under Uncertainty: Heuristics and Biases*. Kahneman, D., Slovic, P. and Tversky, A. New York, Cambridge Univ. Press: 32–47.
Kahneman, D. and Tversky, A. (1982c). "Variants of uncertainty." *Cognition* **11**: 145–157.
Kaplan, S. and Garrick, B. J. (1981). "On the quantitative definition of risk." *Risk Analysis* **1**(1): 11–27.
Kaufman, G. M. (1963). *Statistical Decision and Related Techniques in Oil and Gas Exploration*. Englewood Cliffs, NJ, Prentice-Hall.
Keeney, R. L. and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York, John Wiley & Sons.
Kendall, M. G. and Moran, P. A. P. (1963). *Geometrical Probability [by] M.G. Kendall and P.A.P. Moran*. London, C. Griffin.
Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics*. London, C. Griffin.
Kendall, M. G., Stuart, A., Ord, J. K. and O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics*. London, New York, Edward Arnold, Halsted Press.
Kennedy, T. W. (1978). "Practical use of the indirect tensile test for the characterization of pavement materials." *Australian Road Research Board Conference*: 8–17.
Kennedy, T. W. and Hudson, W. R. (1968). "Applications of the indirect tensile test to stabilized materials." *Highway Research Record* **235**: 36–48.
Kent, S. (1949) *Strategic Intelligence for American World Policy*, Princeton, New Jersey, Princeton University Press.
Keren, G. (1994). "The rationality of gambling: Gamblers' conceptions of probability, chance and luck." *Subjective Probability*. Wright, G. and Ayton, P. New York, John Wiley & Sons.
Kitanidis, P. K. (1985). "Parameter uncertainty in estimation of spatial functions: Bayesian analysis." *Water Resources Research* **22**(4): 499–507.
Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications to Hydrogeology*. New York, Cambridge University Press.
Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston, New York, Houghton Mifflin.
Knuth, D. E. (1997). *The Art of Computer Programming: Vol. 1, Fundamental Algorithms*. Reading, MA, Addison-Wesley.
Koch, G. S. and Link, R. F. (2002). *Statistical Analysis of Geological Data*. Mineola, NY, Dover Publications.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*, Chelsea Pub. Co., New York.
Kolmogorov, A. N. (1941). "The local structure of turbulence in an incompressible fluid at very large Reynolds number." *Doklady Adademii Nauk SSSR* **30**: 301–305.
Koopman, B. O. (1946). "Search and screening." The Summary Reports Group of the Columbia University Division of War Research, Washington, DC.
Koopman, B. O. (1956a). "The theory of search I: Kinematic bases." *Operations Research* **4**: 324–346.
Koopman, B. O. (1956b). "The theory of search II: Target detection." *Operations Research* **4**: 503–531.
Koopman, B. O. (1957). "The theory of search III: The optimal distribution of searching effort." *Operations Research* **5**: 613–626.
Koopman, B. O. (1980). *Search and Screening: General Principles with Historical Applications*. Elmsford, N.Y., Pergamon Press.
Kotzias, P. C., Stamatopoulos, A. C. and Kountouris, P. J. (1993). "Field quality control on earth-dam: Statistical graphics for gauging." *Journal Geotechnical Engineering, ASCE* **119**(5): 957–964.
Kreuzer, H. and Bury, K. V. (1984). "A probability based evaluation of the safety and risk of existing dams." *International Conference on Safety of Dams*, Coimbra, Portugal.
Krumbein, W. C. and Graybill, F. A. (1965). *An Introduction to Statistical Models in Geology*. New York, McGraw-Hill.
Krumbein, W. C., and Sloss, L. L. (1963). *Stratigraphy and sedimentation*, W. H. Freeman, San Francisco, CA.
Kuhn, S. H. (1971). "Quality control in highway construction." *First International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Hong Kong, Honk Kong University Press: 287–312.
Kulhawy, F. H. and Mayne, P. W. (1990). "Manual on Estimating Soil Properties for Foundation Design, EL-6800 Final Report." Palo Alto, Electric Power Research Institute.
Kulhawy, F. H. and Phoon, K. K. (1996). "Engineering judgment in the evolution from deterministic to reliability-based foundation design." *Uncertainty in the Geological Environment*, Madison, WI, ASCE: 29–49.
Kulhawy, F. H. and Trautmann, C. H. (1996). "Estimation of in-situ test uncertainty." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 269–286.
Lacasse, S. and Nadim, F. (1996). "Uncertainties in characterizing soil properties." *Uncertainty in the Geologic Environment*, Madison, ASCE: 49–75.
Ladd, C. C. and Foott, R. (1974). "New design procedure for stability of soft clays." *Journal of the Geotechnical Engineering Division, ASCE* **100**(7): 763–786.
Ladd, C. C. (1991). "Stability evaluation during staged construction." *Journal of Geotechnical Engineering, ASCE* **117**(4): 540–615.
Ladd, C. C. (1996). Personal communication.
Ladd, C. C. and Da Re, G. (2001). "Discussion of 'Factors of safety and reliability in geotechnical engineering' by J. M. Duncan." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **127**(8): 710–714.
Ladd, C. C., Dascal, O., Law, K. T., Lefebvre, G., Lessard, G., Mesri, G. and Tavenas, F. (1983). "Report of the subcommittee on embankment stability – annexe II, Committe of specialists on Sensitive Clays on the NBR Complex." Montreal, Societe d'Energie de la Baie James.
Ladd, C. C. and Foott, R. (1974). "New design procedure for stability of soft clays." *Journal of the Geotechnical Engineering Division, ASCE* **100**(GT7): 763–786.
Ladd, C. C., Foott, R., Ishihara, K., Schlosser, F. and Poulos, H. G. (1977). "Stress-deformation and strength characteristics: State-of-the-art report." *Ninth International Conference on Soil Mechanics and Foundation Engineering*, Tokyo, Japan: 421–494.
Lambe, T. W. (1951). *Soil Testing for Engineers*. New York, John Wiley & Sons.
Lambe, T. W. (1985). "Amuay landslides." *XIth International Conference on Soil Mechanics and Foundation Engineering*, San Francisco, CA, A. A. Balkema: 137–158.
Lambe, T. W., Silva, F. and Lambe, P. C. (1988). "Expressing the level of stability of a slope." *The Art and Science of Geotechnical Engineering at the Dawn of the Twenty-First Century: A Volume Honoring Ralph B. Peck*, Prentice-Hall: 558–588.
Lambe, T. W. and Whitman, R. V. (1969). *Soil Mechanics*. New York, John Wiley & Sons.

Laplace, P. S. (1814). *Philosophical Essay on Probabilities*. Translated by Dale, A. I. New York, Dover Publications, Inc.

Leach, R. D. (1976). "Allocation of quality control resources." *Eighth Conference on the Australian Road Research Board*, Perth: 14–22

Lee, I. K., White, W. and Ingles, O. G. (1983). *Geotechnical Engineering*. Boston, Pitman.

Lefebvre, G., Ladd, C. C. and Pare, J. -J. (1988). "Comparison of field vane and laboratory undrained shear strength in soft sensitive clays." *Vane Shear Strength Testing in Soils: Field and Laboratory Studies; ASTM STP 1014*, Philadelphia, PA, ASTM: 233–246.

Legget, R. F. and Hatheway, A. W. (1988). *Geology and engineering*. New York, McGraw-Hill.

Leonards, G. A. (1975). "Investigation of failures." *Journal of the Geotechnical Engineering Division, ASCE* **Vol. 108**(No. GT2): pp. 187–246.

Lerche, I. (1992). *Oil exploration: basin analysis and economics*, Academic Press, San Diego, CA.

Lerche, I. (1997). *Geological risk and uncertainty in oil exploration*, Academic Press, San Diego, CA.

LeRoy, L. W. (1951). *Subsurface Geology: Petroleum, Mining, Construction*. Golden, CO, Colorado School of Mines.

Leveson, N. (1995). *SafeWare: System Safety and Computers*. Reading, Mass., Addison-Wesley.

Lewis, P. A. W. (1972). *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*. New York, Wiley-Interscience.

Li, K. S. (1992). "Point-estimate method for calculating statistical moments." *Journal of Engineering Mechanics, ASCE* **Vol. 118**(No. 7): pp. 1506–1511.

Liao, S. S. C., Druss, D. L., Neff, T. L. and Brenner, B. R. (1996). "Just one more boring, and we'll know for sure." *Uncertainty in the Geological Environment*, Madison, WI, ASCE: 119–133.

Liao, S. S. C., Veneziano, D. and Whitman, R. V. (1988). "Regression models for evaluating liquefaction potential of sandy soils." *Journal of Geotechnical Engineering, ASCE* **114**(4): 389–411.

Lichtenstein, S., Fischhoff, B. and Phillips, L. (1982). "Calibration of probabilities: The state of the art to 1980." *Judgment under Uncertainty: Heuristics and Biases*." Kahneman, D., Slovic, P. and Tversky, A., eds., New York, Cambridge Univ. Press: 306–334.

Lichtenstein, S. and Newman, J. r. (1967). "Empirical scaling of common verbal phrases associated with numerical probabilities." *Psychological Science* **9**: 563–564.

Lin, P. -L. and Der Kiureghian, A. (1991). "Optimization algorithms for structural reliability." *Structural Safety* **9**(3): 161–177.

Lind, N. C. (1972). *Theory of Codified Structural Design*. Waterloo, Ont., Solid Mechanics Division University of Waterloo.

Lind, N. C. (1983). "Modelling uncertainty in discrete dynamical systems." *Applied Mathematical Modelling* **7**(3): 146–152.

Lindblom, C. E. (1959) "The science of muddling through." *Public Administration Review*, **19**: 79–88.

Lindblom, C. E. (1979) "Still muddling, not yet enough." *Public Administration Review*, **39**: 222–233.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge, Cambridge University Press.

Lindley, D. V. (1971). *Bayesian Statistics: A Review*. Philadelphia, PA, Society for Industrial and Applied Mathematics.

Lippman, M. J. (1974). "Two-dimensional Stochastic Model of heterogeneous Geologic System." PhD thesis, Civil Engineering, University of California, Berkeley.

Loh, C. H., Cheng, C. R. and Wen, Y. K. (1995). "Probabilistic analysis of deposit liquefaction." *Lifeline Earthquake Engineering*, San Francisco.

Low, B. K. (1996). "Practical probabilistic approach using spreadsheet." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE.: 1284–1302.

Low, B. K. (1997). "Reliability analysis of rock wedges." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **123**(6): 498–505.

Low, B. K., Gilbert, R. B. and Wright, S. G. (1998). "Slope reliability analysis using Generalized method of slices." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **124**(4): 350–362.

Low, B. K. and Tang, W. H. (1997a). "Efficient reliability evaluation using spreadsheets." *Journal of Engineering Mechanics, ASCE* **123**(7): 749–752.

Low, B. K. and Tang, W. H. (1997b). "Reliability analysis of reinforced embankments on soft ground." *Canadian Geotechnical Journal* **34**(5): 672–685.
Luce, R. D. and Suppes, P. (1965). *Foundations of Measurement*. New York, Academic Press.
Lumb, P. (1966). "The variability of natural soils." *Canadian Geotechnical Journal* **3**: 74–97.
Lumb, P. (1974). " Application of statistics in soil mechanics." *Soil Mechanics: New Horizons*. Lee, I. K., ed., London, Newnes-Butterworth: 44–112, 221–239.
Maistrov, L. E. (1974). *Probability Theory: A Historical Sketch*. New York, Academic Press.
Mantoglu, A. and Wilson, J. L. (1982). "The turning bands method for simulation of random fields using line generation by a spectral method." *Water Resources Research* **18**(5): 1379–1394.
Mardia, K. V. and Marshall, R. J. (1984). "Maximum likelihood estimation of models for residual covariance in spatial regression." *Biometrika* **71**(1): 135–146.
Margolis, H. (1987). *Patterns, Thinking, and Cognition: A Theory of Judgment*. Chicago, University of Chicago Press.
Marr, W. A. (2000). Personal communication.
Marr, W. A., Ramos, J. A., and Lambe, T. W. (1982) "Criteria for differential settlement of oil storage tanks." *Journal of the Geotechnical Engineering Division, ASCE* **108**(8): 1017–1039.
Marshall, K. T. and Oliver, R. M. (1995). *Decision Making and Forecasting, with Emphasis on Model Building and Policy Analysis*. New York, McGraw-Hill.
Marsily, G. de (1986). *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*. Orlando, FL, Academic Press.
Matérn, B. (1947). "Methods of estimating the accuracy of line and sample plot surveys." *Meddelanden fran Statens Skogsforskningsinstitut* **36**: 1–138.
Matérn, B. (1960). "Spatial Variation." *Meddelanden fran Statens Skogsforskningsinstitut* **49**(5).
Matérn, B. (1986). *Spatial variation*. Berlin, New York, Springer-Verlag.
Matheron, G. (1971). *The Theory of Regionalized Variables and Its Application–Spatial Variabilities of Soil and Landforms*. Fontainebleau, Les Cahiers du Centre de Morphologie Mathematique 8.
Matheron, G. (1973). "The intrinsic random functions and their applications." *Advances in Applied Probability* **5**: 439–468.
Matheron, G. (1989). *Estimating and Choosing: An Essay on Probability in Practice*. Berlin, New York, Springer-Verlag.
Mattson, R. L., and Dammann, J. E. (1965) 'A technique for determining and coding subclasses in pattern recognition problems.' *IBM Journal of R & D* **9**: 294–302.
McCann, M. (2002). "Fault tree analysis." *Guide to Risk Analysis for Dam Safety*. Hartford, D. Vancouver, Dam Safety Interest Group, Canadian Electricity Association. **II (3.3.2)**.
McCann, M. W. (1999). Personal communication.
McCartney, K. (1988). "A computer program for the three-dimensional measurement of silicoflagellate skeletons." *Computers and Geosciences*, **14**(1): 99–111.
McClelland, A. G. R. and Bolger, F. (1994). "The calibration of subjective probabilities: Theories and models 1980–1994." *Judgmental Forecasting*. Wright, G. and Ayton, P., eds., Chichester, New York, John Wiley & Sons: 453–482.
McCormick, N. J. (1981). *Reliability and Risk Analysis: Methods and Nuclear Power Applications*. New York, Academic Press.
McCuen, R. H., and Snyder, W. M. (1986). *Hydrologic Modeling: Statistical Methods and Applications*, Prentice-Hall, Englewood Cliffs, N.J.
McGrath, T. C. and Gilbert, R. B. (1999). "Analytical method for designing and analyzing I-D search programs." *Journal of Geotechnical and Geoenvironmental Engineering* **125**(12): 1043–1056.
McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York, John Wiley & Sons.
Mela, D. F. (1961). "Information theory and search theory as special cases of decision theory." *Operations Research* **9**: 907–909.
Melchers, R. E. (1987). *Structural Reliability: Analysis and Prediction*. Chichester, New York, John Wiley & Sons.
Melchers, R. E. (1993). "Modern computational techniques for reliability." *Probabilistic Methods in Geotechnical Engineering*, Canberra, Australia, A. A. Balkema
Merriam-Webster, I. (1998). *Merriam-Webster's Collegiate Dictionary*. Springfield, Mass., Merriam-Webster.

Meyer, M. A. and Booker, J. M. (1990). "Eliciting and Analyzing Expert Judgment: A Practical Guide." Washington, DC, US Nuclear Regulatory Commission.

Meyerhof, G. G. (1994). "Evolution of Safety Factors and Geotechnical Limit State Design." *Spencer J. Buchanan Lecture*, College Station, Texas, Texas A and M University.

Miller, A. C., III and Rice, T. R. (1983). "Discrete approximations of probability distributions." *Management Science* **Vol. 29**(No. 3): 352–362.

Miller, I. and Kossik, R. (1996). "Probabilistic simulation of geologic waste disposal facilities using the repository integration model." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 944–964.

Miller, I., Kossik, R. and Cunnane, M. (1992). "A new methodology for repository site suitability evaluation." *Proceedings of the Third International Conference for High Level Radioactive Waste Management*, Las Vegas.

Miller, R. L., and Kahn, J. S. (1962). *Statistical analysis in the geological sciences*, Wiley, New York, N. Y..

Minty, E. J., Smith, R. B. and Pratt, D. N. (1979). "Interlaboratory testing variability assessed for a wide range of NSW soil types." *Third International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Sydney: 221–235.

Mitchell, J. K. (1993). *Fundamentals of Soil Behavior*. New York, John Wiley & Sons.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, New York, Cambridge University Press.

Morse, P. M. (1970). "On browsing: the use of search theory in the search for information." Cambridge, Operations Research Center, Massachusetts Institute of Technology.

Morse, R. K. (1971). "The importance of proper soil units for statistical analysis." *First International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Hong Kong, Hong Kong University Press: 347–355.

Moses, F. (1985). "Implementation of a reliability-based API RP2A format." Dallas, American Petroleum Institute.

Moses, F. and Lind, N. C. (1970). *Structural Reliability and Codified Design: eight special lectures presented at the University of Waterloo from October 6 to December 15, 1969*. Waterloo, Ont., Solid Mechanics Division University of Waterloo.

Mosleh, A., Bier, V. M., and Apostolakis, G. (1987) "Methods for the elicitation and use of expert opinion in risk assessment: Phase I, a critical evaluation and directions for future research." Washington, D. C., U. S. Nuclear Regulatory Commission.

Murphy, H. W. and Grahame, R. E. (1976). "Pavement construction standards and tolerances." *Eighth Conference of the Australian Road Research Board*, Perth: 7–13.

Murphy, A. H. and Winkler, R. L. (1974). "Probability forecasts: A survey of National Weather Service forecasters." *Bulletin of the American Meteorological Society* **55**: 1449–14453.

Murphy, A. H. and Winkler, R. L. (1977a). "Can weather forecasters formulate reliable forecasts of precipitation and temperature?" *National Weather Digest* **2**: 2–9.

Murphy, A. H. and Winkler, R. L. (1977b). "The use of credible intervals in temperature forecasting: some experimental results." *Decision Making and Change in Human Affairs*. Jungerman, H. and de Zeeuw, G., eds., Dordrecht, Reidel.

National Research Council (2000). *Risk analysis and uncertainty in flood damage reduction studies*. Washington, DC, National Academies Press.

National Research Council of Canada (1977). "National Building Code of Canada." Ottawa.

Newendrop, P. D. (1975). *Decision Analysis for Petroleum Exploration*. Tulsa, OK, Petroleum Publishing Company.

Newman, J. R. (1966). *The world of mathematics: a small library of the literature of mathematics from A*h-mosé the scribe to Albert Einstein*. New York, Simon and Schuster.

Neyman, J. and Pearson, E. S. (1933). "On the problem of the most efficient tests of statistical hypotheses." *Philosophical Transactions of the Royal Society*, ***A*231**: 289–337.

Noiray, L. (1982). "Predicted and Measured Performance of a Soft Clay Foundation under Stage Loading." MS thesis, Civil Engineering, Massachusetts Institute of Technology.

Nowak, A. S. (1995). "Calibration of LRFD bridge code." *Journal of Structural Engineering, ASCE* **121**(8): 1245–1251.

Nozick, R. (1993). *The Nature of Rationality*. Princeton, NJ, Princeton University Press.

NSRC (2000). "United States National Search and Rescue Supplement to the International Aeronautical and Maritime Search and Rescue Manual." Washington, DC, National Search and Rescue Committee.

Nunn, L. H. (1981). "An Introduction to the Literature of Search Theory." Alexandria, VA, Operations Evaluation Group Center for Naval Analyses.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics*. London, Edward Arnold.

Ohlson, M., and Okland, R. (1998). "Spatial variation in rates of carbon and nitrogen accumulation in a boreal bog." *Ecology*, **79**: 2745–2758.

Okabe, A. (2000). *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams*. New York, John Wiley & Sons.

Olea, R. A. (1984). "Systematic sampling spatial functions." Lawrence, Kansas Geological Survey.

O'Neill, M. W. (1995). "LRFD factors for deep foundations through direct experimentation." *US/Taiwan Geotechnical Engineering collaboration workshop*, Taipei. 100–114

Ord, J. K. (1972). *Families of Frequency Distributions*. New York, Hafner.

Orr, T. L. L. and Farrell, E. R. (1999). *Geotechnical Design to Eurocode 7*. Berlin, Springer-Verlag.

Otte, E. (1978). "A Structural Design Procedure for Cement Treated Layers in Pavements." DSc (Eng), Civil Engineering, University of Pretoria.

Otway, H. J. and Cohen, J. J. (1975). "Revealed Preferences: Comments on the Starr Benefit-Risk Relationships." Laxenburg, International Institute for Applied Systems Analysis.

Paice, G. M., Griffiths, D. V. and Fenton, G. A. (1996). "Finite element modeling of settlements on spatially random soil." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **122**(9): 777–779.

Paikowsky, S. G., Baecher, G. B., Ayyub, B., McVay, M., Birgisson, B. and Kuo, C. (2002). "NCHRP 24-17 LRFD foundation design." Lowell, MA, University of Massachusetts.

Paikowsky, S. G. and Stenersen, K. L. (2001). "Load and resistance factor design (LRFD) for dynamic analyses of driven piles." Lowell, MA, University of Massachusetts.

Papoulis, A. (2002). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, N. Y.

Papoulis, A., and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*, McGraw-Hill, New York, N, Y..

Park, S. K. and Miller, K. W. (1969). "Random number generators: Good ones are hard to find." *Communications of the ACM* **31**: 1192–1201.

Parkin, J. (2000). *Engineering Judgement and Risk*. London, T. Telford.

Parr, N. M. and Cullen, N. (1988). "Risk management and reservoir maintenance." *Journal of the Institute of Water and Environmental Management (IWEM)* **2**.

Parzen, E. (1964). *Stochastic Processes*. San Francisco, Holden-Day.

Parzen, E. (1992). *Modern Probability Theory and Its Applications*. New York, John Wiley & Sons.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, Calif., Morgan Kaufmann Publishers.

Pearson, K. and Pearson, E. S. (1978). *The history of statistics in the 17th and 18th centuries against the changing background of intellectual, scientific, and religious thought: lectures by Karl Pearson given at University College, London, during the academic sessions, 1921–1933*. London, C. Griffin.

Peck, R. B. (1969). " Ninth Rankine Lecture: Advantages and limitations of the observational method in applied soil mechanics." *Geotechnique* **19**(2): 171–187.

Peck, R. B. (1980). " Where has all the judgment gone?" *Canadian Geotechnical Journal* **17**: 584–590.

Peck, R. B., and Dunnicliff, J., and Deere, D. U., eds., (1984) *Judgment in Geotechnical Engineering: The Professional Legacy of Ralph B. Peck*, John Wiley & Sons, New York, N. Y.

Peirce, C. S., Cohen, M. R. and Dewey, J. (1998). *Chance, Love, and Logic: Philosophical Essays*. Lincoln, University of Nebraska Press.

PEMA (2000). "Summary of the commonwealth hazard vulnerability analysis." Pennsylvania Emergency Management Agency. 2002.

Pennington, N. and Hastie, R. (1993). "A theory of explanation-based decision making." *Decision Making in Action: models and methods*. Klein, G. A. *et al*., eds., Norwood, NJ, Ablex Publishing.

Pettijohn, F. J., Potter, P. E. and Siever, R. (1987). *Sand and Sandstone*. New York, Springer-Verlag.

Phillips, L. D. and Edwards, W. (1966). "Conservatism in a simple probability inference task." *Journal of Experimental Psychology* **72**: 346–357.

Phoon, K. K. and Kulhawy, F. H. (1996). "On quantifying inherent soil variability." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 326–340.

Phoon, K. K. and Kulhawy, F. H. (1999). "Characterization of geotechnical variability." *Canadian Geotechnical Journal* **36**(4): 612–624.

Pólya, G. (1954). *Mathematics and Plausible Reasoning*. Princeton, N.J., Princeton University Press.

Popper, K. R. (1968). *The Logic of Scientific Discovery*. New York, Harper & Row.

Porter, T. M. (1986). *The Rise of Statistical Thinking, 1820-1900*. Princeton, NJ, Princeton University Press.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ, Princeton University Press.

Posner, E. C. (1963). "Optimal search procedures." *IEEE Transactions on Information Theory* **IT-0**: 157–160.

Potter, P. E. and Siever, R. (1953). "A comparative study of Upper Chester and Lower Pennsylvanian stratigraphic variability." *Journal of Geology* **63**: 429–451.

Pratt, J. W., Raiffa, H. and Schlaifer, R. (1995). *Introduction to Statistical Decision Theory*. Cambridge, MA, MIT Press.

Press, S. J. and Wilson, S. (1978). "Choosing between Logistic Regression and Discriminant Analysis." *Journal of the American Statistical Association* **73**: 699–705.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computation*. Cambridge, England, Cambridge University Press.

Price, N. J. (1966). *Fault and Joint Development in Brittle and Semi-Brittle Rock*. New York, Pergamon Press.

Priest, S. D. and Hudson, J. A. (1976). "Discontinuity spacings in rock." *International Journal of Rock Mechanics and Mining Science* **13**: 135–148.

Quetelet, A. (1827). "Recherches sur la population, les naissances, les décès, les prisons, les dépôts de medicité, etc., dans le royaume des Pays-Bas." *Nouveaux mémoires de l'Académie royale ds sciences et belles-lettres de Bruxelles*, **4**: 117–192.

Rabcewitz, L. (1964a). "The New Austrian Tunneling Method: Part I." *Water Power* **Vol. 16**(No. 11): pp. 453–457.

Rabcewitz, L. (1964b). "The New Austrian Tunneling Method: Part II." *Water Power* **16**(12): 511–515.

Rabcewitz, L. (1965). "The New Austrian Tunneling Method: Part III." *Water Power* **17**(1): 19–24.

Rackwitz, R. (1976). "Practical Probabilistic Approach to Design, Bulletin 112." Paris, Comite Europeen du Beton.

Rackwitz, R. and Fiessler, B. (1978). "Structural reliability under combined load sequences." *Computers and Structures* **9**: 489–494.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. New York, Random House.

Raiffa, H. and Schlaifer, R. (1968). *Applied Statistical Decision Theory*. Cambridge, MA, MIT Press.

Raiffa, H. and Schlaifer, R. (2000). *Applied Statistical Decision Theory*. New York, John Wiley & Sons.

Ramsey, F. P. (1978). *Foundations: Essays in Philosophy, Logic, Mathematics, and Economics*. Atlantic Highlands, NJ, Humanities Press.

Ramsey, F. P. and Braithwaite, R. B. (1931). *The Foundations of Mathematics and Other Logical Essays*. London, K. Paul Trench Trubner & Co. Ltd., Harcourt Brace & Co.

Rao, C. R. and Toutenberg, H. (1999). *Linear Models: Least Squares and Alternatives*. New York, Springer-Verlag.

Ravindra, M. K. and Galambos, T. V. (1978). "Load and resistance factor design for steel." *Journal of the Structural Engineering Division, ASCE* **104**(9): 1337–1353.

Reagan, R., Mosteller, F. and Youtz, C. (1989). "Quantitative meanings of verbal probability expressions." *Journal of Applied Psychology* **74**(3): 433–442.

Richter, C. F. (1958). *Elementary Seismology*. San Francisco, W. H. Freeman.

Riela, J., Urzua, A., Christian, J. T., Karzulovic, A. and Flores, G. (1999). "Sliding rock wedge reliability analysis of Chuquicamata mine slopes." *XI Pan-American Conference on Soil Mechanics and Geotechnical Engineering*, Foz do Iguaçu, Brazil.

Ripley, B. D. (1981). *Spatial Statistics* New York, John Wiley & Sons.

Robertson, P. K., and Mayne, P. W., eds., (1998) *Geotechnical Site Characterization, Proceedings of the First International Conference on Site Characterization, ISC'98*, Atalanta, Georgia, A. A. Balkema.

Rosenblatt, M. (1952). "Remarks on a multivariate transformation." *Annals of Mathematical Statistics* **23**(3): 470–472.

Rosenblueth, E. (1975). "Point estimates for probability moments." *Proceedings, National Academy of Science* **72**(10): 3812–3814.

Rosenblueth, E. (1981). "Two-point estimates in probabilities." *Applied Mathematical Modelling* **Vol. 5**(No. 2): pp. 329–335.

Rosenblueth, E. and Esteva, L. (1972). "Reliability basis for some Mexican Codes." Detroit, American Concrete Institute.

Rowe, P. W. (1972). "Twelfth Rankine Lecture: The relevance of soil fabric to site investigation practice." *Geotechnique* **22**(2).

Rowe, W. D. (1977). *An Anatomy of Risk*. New York, John Wiley & Sons.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*. New York, John Wiley & Sons.

Salmon, G. M. and Hartford, D. N. D. (1995a). "Lessons from the application of risk assessment to dam safety." *ANCOLD/NXSOLD Conference on Dams*, Christchurch, ANCOLD Bulletin 101: 54–67.

Salmon, G. M. and Hartford, D. N. D. (1995b). "Risk analysis for dam safety." *Water Power & Dam Construction* **3** and **4**.

Salmon, G. M. and von Ehn, R. (1993). "Consequence based dam safety criteria for floods and earthquakes." *Workshop on Dam Safety Evaluations,* Grindelwald, Switzerland.

Salmon, W. C. (1998). *Causality and Explanation*. New York, Oxford University Press.

Salzer, H. E., Zucker, R. and Capuano, R. (1952). "Table of the zeros and weight factors of the first twenty Hermite polynomials." *Journal of Research, National Bureau of Standards* **48**(RP2294): 111–116.

Savage, L. (1954). *The Foundations of Statistics*. New York, John Wiley & Sons.

Savinskii, I. D. (1965). *Probability Tables for Locating Elliptical Underground Masses with a Rectangular Grid*. New York, Consultants Bureau.

Saxena, S. K., ed., (1982). *Updating Subsurface Samplings of Soil and Rocks and Their In-Situ Testing*. ASCE Conference, Santa Barbara, CA.

Schilling, M. F. (1984). "Some remarks on quick estimation of the correlation coefficient." *The American Statistician* **38**(4): 330–331.

Schlaifer, R. (1978). *Analysis of Decisions under Uncertainty*. New York, John Wiley & Sons.

Schultze, E. (1975). "Some aspects concerning the application of statistics and probability to foundation structures." *Second International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Aachen: 457–494.

Schweiger, H. F., Turner, R. and Pottler, R. (2001). "Reliability analysis in geotechnics with deterministic finite elements." *International Journal of Geomechanics* **1**(4): 389–413.

Seed, H. B. and Idriss, I. M. (1967). "Analysis of soil liquefaction: Niigata earthquake." *Journal of the Soil Mechanics and Foundations Division, ASCE* **93**(SM3): 83–108.

Seed, H. B. and Idriss, I. M. (1971). "Simplified procedure for soil liquefaction potential." *Journal of the Soil Mechanics and Foundations Division, ASCE* **97**(SM9): 1249–1274.

Seed, H. B. and Lee, K. L. (1966). "Liquefaction of saturated sands during cyclic loading." *Journal of the Soil Mechanics and Foundations Division, ASCE* **90**(SM6): 105–134.

Seed, H. B. and Peacock, W. H. (1971). "Test procedure for measuring soil liquefaction characteristics." *Journal of the Soil Mechanics and Foundations Division, ASCE* **97**(SM8): 1099–1119.

Selig, E. T., and McKee, K. E. (1961). "Static and dynamic behavior of small footings." *Journal of the Soil Mechanics and Foundation Division, ASCE*, **87**(SM6).

Shackelford, C. D., Nelson, P. P., and Roth, M. J. S., eds., (1996) *Uncertainty in the Geologic Environment: From Theory to Practice*, ASCE Conference, 2 Vols., Madison, WI.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL.

Sherwood, P. T. (1970). "An examination of the reproducibility of soil classification and compaction tests." *Symposium on Quality Control of Road Works*, Centre d'Etudes Techniques de l'Equipement

Simon, H. A. (1957). *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in a Social Setting*. New York, John Wiley & Sons.

Simpson, B., Pappin, J. W. and Croft, D. D. (1981). "An approach to limit state calculations in geotechnics." *Ground Engineering* **14**(6): 21–28.

Singh, A. (1971). "How reliable is the factor of safety in foundation engineering?" *First International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Hong Kong, Hong Kong University Press: 389–424.

Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial*. Oxford, Oxford University Press.

Slichter, L. B. (1955). "Geophysics applied to prospecting for ores." *Economic Geology* 50th Anniversary Volume: 885–915.

Slovic, P., Fischhoff, G. and Lichtenstein, S. (1977). "Behavioral decision theory." *Annual Review of Psychology* **28**: 1–39.

Smith, J. F. and Kida, T. (1991). "Heuristics and biases: Expertise and task realism in auditing." *Psychological Bulletin* **109**: 472–485.

Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods*. Ames, IA, Iowa State University Press.

Soulie, M. and Favre, M. (1983). "Analyse geostatistique d'un noyau de barrage tel que construit." *Canadian Geotechnical Journal* **20**: 453–467.

Soulie, M., Montes, P. and Silvestri, V. (1990). "Modeling spatial variability of soil parameters." *Canadian Geotechnical Journal* **27**: 617–630.

Sowers, G. F. and Sowers, G. B. (1979). *Introductory Soil Mechanics and Foundations: Geotechnical Engineering*. New York, Macmillan.

SPE (2001). "Hydrocarbon Economics and Evaluation Symposium." Dallas, TX.

Speltzer, C. S. and Stael von Holstein, C. -A. S. (1975). "Probability encoding in decision analysis." *Management Science* **22**(3).

Stamatopoulos, A. C. and Kotzias, P. C. (1975). "The relative value of increasing number of observations." *Second International Conference on Applications of Statistics and Probability in Soil and Structural Engineering*, Aachen. 495–510.

Starr, C. (1969). "Social benefit versus technological risk." *Science* **165**: 1232–1238.

Starr, C. and Whipple, C. (1980). "Risks of risk decisions." *Science* **208**(4448): 1114–1119.

Stedinger, J. R., Heath, D. C. and Thompson, K. (1996). "Risk assessment for dam safety evaluation: Hydrologic risk." Washington, DC, USACE Institute for Water Resources.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York.

Stenersen, K. (2001). "Load and Resistance Factor Design Method for Piles." MS Thesis, Department of Civil and Environmental Engineering, University of Massachusetts, Lowell, MA.

Stern, P. C., Fineberg, H. V. and National Research Council (U.S.). Committee on Risk Characterization (1996). Understanding Risk: Informing Decisions in a Democratic Society. Washington, D.C., National Academy Press.

Stevens, S. S. (1951). "Mathematics, measurements, and psychophysics." *Handbook of Experimental Psychology*. Stevens, S. S., ed., New York, John Wiley & Sons: 1–49.

Stigler, S. M. (1986). *The History of Statistics*. Cambridge, Massachusetts, Harvard University Press.

Stigler, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA, Harvard University Press.

Stone, L. D. (1975). *Theory of Optimal Search*. New York, Academic Press.

Stone, L. D. (1983). "The process of search planning: current approaches and continuing problems." *Operations Research* **31**(March-April): 207–233.

Stone, L. D. (1989). *Theory of Optimal Search*. Arlington, VA, Operations Research Society of America.

Sui, W. W. C., Parimi, R. R. and Lind, N. C. (1975). "Practical approach to code calibration." *Journal of the Structural Division, ASCE* **101**(ST7): 1469–1480.

Switzer, P. (1967). "Reconstructing patterns from sample data." *Annals of Mathematical Statistics* **38**: 138–154.

Switzer, P. (1995). "Spatial interpolation errors for monitoring data." *Journal of the American Statistical Society* **90**(431): 853–861.

T. W. Lambe & Associates (1982). "Earthquake risk to patio 4 and site 400." Longboat Key, Fl.

T.W. Lambe & Associates (1989). "Earthquake risk analysis for KSS tank area." Longboat Key, FL.

Tanaka, H., Loat, J., Shibuya, S., Soon, T. T. and Shiwakoti, D. (2001). "Characterization of Singapore, Bangkok, and Ariake clays." *Canadian Geotechnical Journal* **38**: 378–400

Tang, W. H. (1979). "Probabilistic evaluation of penetration resistance." *Journal Geotechnical Engineering Division, ASCE* **105**(10): 1173–1191.

Tang, W. H. (1987). "Updating anomaly statistics – single anomaly case." *Journal of Structural Safety* **4**(2): 1.

Tang, W. H. (1990). "Reliability of geotechnical performances considering geologic anomalies." *Workshop on Dynamic Effects of Structures and Earthquake Engineering*, Taiwan: 13–14.

Tang, W. H. and Angulo, M. (1996). "Bayesian liquefaction resistance analysis." *Uncertainty in the Geologic Environment: From Theory to Practice*, Madison, WI, ASCE: 1195–1209.

Tang, W. H. and Halim, I. (1988). "Updating anomaly statistics – multiple anomaly pieces." *Journal of Engineering Mechanics, ASCE*(6).

Tang, W. H., Halim, I. and Gilbert, R. B. (1988). "Reliability of geotechnical systems considering geologic anomaly." *Proceedings, ASCE Specialty Conference on Probabilistic Methods*, Blacksburg: 136–139.

Tang, W. H., and Quek, S. T. (1986). "Statistical model of boulder size and fraction." *Journal of Geotechnical Engineering*, **112**(1): 79–90.

Tang, W. H. and Saadeghvaziri, A. (1983). "Updating distribution of anomaly size and fraction." *Recent Advances in Engineering Mechanics and Their Impact on Civil Engineering Practice*. Chan, W. F. and Lewis, A. D. M, eds., **2:** 895–898.

Tang, W. H., ed., (1995). *Probabilistic Methods in Geotechnical Engineering*. Washington, DC, National Academy Press.

Taylor, D. W. (1948). *Fundamentals of Soil Mechanics*. New York, John Wiley & Sons.

Taylor, G. I. (1921). "Diffusion by continuous movements." *Proceedings London Mathematical Society (2)* **20**: 196–211.

Taylor, J. R. (1997). *An Introduction to Error Analysis*. Sausalito, California, University Science Books.

Terzaghi, K. (1925). *Erdbaumechanik auf bodenphysikalischer Grundlage*. Vienna, Franz Deuticke.

Terzaghi, K. (1929). "The effect of minor geological details on the stability of dams." *American Institute of Mining and Metallurgical Engineers*. Technical Publication 215: 31–44.

Terzaghi, K. (1943). *Theoretical Soil Mechanics*. New York, John Wiley & Sons.

Terzaghi, K. (1960). *From Theory to Practice in Soil Mechanics: Selections from the Writings of Karl Terzaghi, with Bibliography and Contributions on his Life and Achievements*. New York, John Wiley & Sons.

Terzaghi, K. (1961). "Past and future of applied soil mechanics." *Journal of the Boston Society of Civil Engineers* **68**: 110–139.

Terzaghi, K. and Peck, R. B. (1948). *Soil Mechanics in Engineering Practice*. New York, John Wiley & Sons.

Terzaghi, K., Peck, R. B. and Mesri, G. (1996). *Soil Mechanics in Engineering Practice*. 3rd ed., New York, John Wiley & Sons.

Thevanayagam, S., Kavazanjian, E., Jacob, A. and Nesarajah, S. (1996). "Deterministic and probabilistic analysis of preload design at a hydraulic fill site." *Uncertainty in the Geological Environment*, Madison, WI, ASCE: 1417–1431.

Thompson, S. K. (2002). *Sampling*. New York, John Wiley & Sons.

Tolstoy, I. (1981). *James Clerk Maxwell: a biography*. Edinburgh, Canongate.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, Mass., Addison-Wesley.

Turkstra, C. J., and Lind, N. C. (1991). *Human error and organization factors in bridge design and construction*. Downsview, Ont., Research and Development Branch, Ontario Ministry of Transportation.

Tversky, A. and Kahneman, D. (1971). "Belief in the 'law of small numbers'." *Psychological Bulletin* **76**: 105–110.

Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: heuristics and biases." *Science* **185**: 1124–1131.

U.S. Army Corps of Engineers (1999). "Risk-based analysis in geotechnical engineering for support of planning studies." Washington, DC, Headquarters, U.S. Army Corps of Engineers.

Underwood, L. B. (1974). "Exploration and geologic prediction for underground works." *Specialty Conference on Subsurface Exploration for Underground Excavation and Heavy Construction*, Henniker, NH.

USEPA (2000). "Data Quality Objectives Process for Hazardous Waste Site Investigations EPA QA/G-4HW." Washington, DC, US Environmental Protection Agency.

US Nuclear Regulatory Commission (1975). "Reactor Safety Study: An assessment of accident risks in U.S. commercial nuclear power plants." Washington, DC, U.S. Nuclear Regulatory Commission (WASH-1400).

US Nuclear Regulatory Commission (1999). "Framework for Risk-Informed Regulation in the Office of Nuclear Material Safety and Regulation. SECY-99-100," Washington, DC, U.S. Nuclear Regulatory Commission.

Van Zyl, D., Miller, I., Milligan, V. and Tilson, W. J. (1996). "Probabilistic risk assessment for tailings impoundment founded on paleokarst." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 563–585.

Vanmarcke, E. H. (1983). *Random Fields: Analysis and Synthesis*. Cambridge, MIT Press.

Veneziano, D. (1976). "Inductive Approaches to Structural Safety." *Journal of Professional Activities, ASCE*, **102**(1): 81–97.

Veneziano, D. (1979). "Probabilistic model of joints in fractured rock." Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Veneziano, D. and Liao, S. S. C. (1984). "Statistical analysis of liquefaction data." *ASCE Conference on Probabilistic Analysis of Liquefaction Data*: 206–209.

Venn, J. (1866). *The logic of chance. An essay on the foundations and province of the theory of probability, with especial reference to its application to moral and social science*. London, Cambridge, Macmillan.

Versteeg, M. (1987). "External safety policy in the Netherlands: An approach to risk management." *Journal of Hazardous Materials* **17**: 215–221.

Vesic, A. (1963). "Bearing capacity for deep foundations in sand." *Highway Research Board*, **39**.

Vick, S. G. (1983). *Planning, Design, and Analysis of Tailings Dams*. John Wiley & Sons, New York, N. Y.

Vick, S. G. (1997). "Dam safety risk assessment: New directions." *Water Power and Dam Construction* **49**(6).

Vick, S. G. (2002). *Degrees of Belief: Subjective Probability and Engineering Judgment*, ASCE Press, American Society of Civil Engineers, Reston, VA.

Vick, S. G. and Bromwell, L. F. (1989). "Risk analysis for dam design in Karst." *Journal of the Geotechnical Engineering Division, ASCE* **115**(6): 819–835.

Vick, S. G. and Stewart, R. A. (1996). "Risk analysis in dam safety practice." *Uncertainty in the Geologic Environment*, Madison, ASCE: 586–603.

von Mises, R., Neyman, J., Sholl, D. and Rabinowitch, E. I. (1939). *Probability, Statistics and Truth*. New York, The Macmillan Company.

von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton, NJ, Princeton University Press.

Von Thun, J. L. (1996). "Risk assessment of Nambe Falls Dam." *Uncertainty in the Geologic Environment*, Madison, ASCE: 604–635.

Vrijling, J. K., Hengel, W. v., and Houben, R. J. (1995). "A framework for risk evaluation." *Journal of Hazardous Materials*, **43**: 245–261.

Wagner, D. H. (1979). "Search theory in the exploration for and mining of polymetallic nodules on the ocean floor." *Search Theory and Applications*. Haley, K. B. and Stone, L. D., eds., New York, Plenum Press: 195–198.

Wald, A. (1944). "On a statistical problem arising in the classification of an individual into one of two groups." *Annals of Mathematical Statistics* **15**: 146–162.

Wald, A. (1950). *Statistical Decision Functions*. New York, John Wiley & Sons.

Wald, A. (1971). *Statistical Decision Functions*. Bronx, N.Y., Chelsea Pub. Co.

Wald, A. and Institute of Mathematical Statistics (1957). *Selected papers in statistics and probability*. Stanford, CA, Stanford University Press.

Wallace, W. (1995) "A New Framework for Hazardous Waste Remediation." *Geoenvironment 2000*: 1630–1645.

Washburn, A. R. (1989). *Search and Detection*. Arlington, VA, Operations Research Society of America.

Weinstock, H. (1963). "The description of stationary random rate processes." Cambridge, Massachusetts Institute of Technology Instrumentation Laboratory Report.

Weiss, K. (1973). "Die Formbeiwerte in der Grundbruchgleichung fuer nichtbindige Boden." *Degebo Heft*(29).

Whitman, R. V. (1984). "Evaluating the calculated risk in geotechnical engineering." *Journal of the Geotechnical Engineering Division, ASCE* **110**(2): 145–188.

Wignall, T. K. and De Geoffroy, J. G. (1985). *Designing Optimal Strategies for Mineral Exploration*. Cambridge, Perseus Publishing.

Wignall, T. K. and De Geoffroy, J. G. (1987). *Statistical Models for Optimizing Mineral Exploration*. New York, Plenum Press.

Winkler, R. L. and Murphy, A. H. (1968). "'Good' probability assessors." *Journal of Applied Meteorology* **7**: 751–758.

Winkler, W. F. and Murphy, A. H. (1973). "Experiments in the laboratory and the real world." *Organizational Behavior and Human Performance* **10**: 252–270.

Withiam, J. L., Voytko, E. P., Baker, R. M., Duncan, J. M., Kelly, B. C., Musser, S. C. and Elias, V. (1997). "Load and Resistance Factor Design (LRFD) for Highway Bridge Substructures (Participant Workbook)." Washington, DC, Federal Highway Administration.

Withiam, J. L., Voytko, E. P., Barker, R. M., Duncan, M. J., Kelly, B. C., Musser, C. and Elias, V. (1998). "Load and Resistance Factor Design (LRFD) of Highway Bridge Substructures." Washington, DC, Federal Highway Administration.

Wolff, T. F. (1996). "Probabilistic slope stability in theory and practice." *Uncertainty in the Geologic Environment*, Madison, WI, ASCE: 419–433.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media, Inc., Champaign.

Wong, F. S. (1985). "First-order, second-moment methods." *Computers and Structures* **20** (4): pp. 779–791.

Wu, T. H. (1974). "Uncertainty, Safety, and Decision in Soil Engineering." *Journal Geotechnical Engineering Division, ASCE* **100**(3): 329–348.

Wu, T. H. and Kraft, L. M. (1967). "The probability of foundation safety." *Journal of the Soil Mechanics and Foundations Division, ASCE* **93**(SM5): 213–231.

Wu, T. H., Abdel-Latif, M. A., Nuhfer, M. A. and Curry, B. B. (1996). "Use of geologic information in site characterization." *Uncertainty in the Geological Environment*, Madison, ASCE: 76–90.

Wu, T. H., Tang, W. H., Sangrey, D. A. and Baecher, G. B. (1989). "Reliability of offshore foundations: State of the art." *Journal of the Geotechnical Engineering Division, ASCE* **115**(2): 157–178.

Wu, T. H. and Wong, K. (1981). "Probabilistic soil exploration: A case history." *Journal of the Geotechnical Engineering Division, ASCE* **107**: 1693–1712.

Yaglom, A. M. (1962). *An Introduction to the Theory of Random Functions*. Englewood Cliffs, NJ, Prentice-Hall.

Yates, R. (1948). "Systematic sampling." *Philosophical Transactions of the Royal Society, Series A* **2441**: 345–377.

Yegian, M. K. and Whitman, R. V. (1978). "Risk analysis for ground failure by liquefaction." *Journal of the Geotechnical Engineering Division, ASCE* **104**(7): 921–928.

Youd, T. L., Idriss, I. M., Andrus, R. D., Arango, I., Castro, G., Christian, J. T., Dobry, R., Finn, W. D. L., Harder, L. F., Jr., Hynes, M. E., Ishihara, K., Koester, J. P., Liao, S. S. C., Marcuson, W. F., III, Martin, G. R., Mitchell, J. K., Moriwaki, Y., Power, M. S., Robertson, P. K., Seed, R. B. and Stokoe, K. H., II (2001). "Liquefaction resistance of soils: Summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils." *Journal of Geotechnical and Geoenvironmental Engineering, ASCE* **127**(10): 817–833.

Zielinsky, A. (2001). Personal communication.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York, John Wiley & Sons.

Zhu, G., Yin, J. -H. and Graham, J. (2001). "Consolidation modelling of soils under the test embankment at Clek Lap Kok International Airport in Hong Kong using a simplified finite element model." *Canadian Geotechnical Journal* **38**: 349–363.

Zienkiewicz, O. C., Chan, A. H. C., Pastor, M., Schrefler, B. A. and Shiomi, T. (1999). *Computational Geomechanics*. Chichester, John Wiley & Sons.

Zienkiewicz, O. C. and Taylor, R. L. (2000a). *The Finite Element Method, Vol. 1, The Basis*. Chichester, John Wiley & Sons.

Zienkiewicz, O. C. and Taylor, R. L. (2000b). *The Finite Element Method, Vol. 2*, *Solid Mechanics*. Chichester, John Wiley & Sons.

Zienkiewicz, O. C. and Taylor, R. L. (2000c). *The Finite Element Method, Vol. 3, Fluid Dynamics*. Chichester, John Wiley & Sons.

Index

Reliability and Statistics in Geotechnical Engineering Gregory B. Baecher and John T. Christian

ISBN: 0-471-49833-5