Probability Theory and Stochastic Modelling 78

Pierre Brémaud

Discrete Probability Models and Methods

Probability on Graphs and Trees, Markov Chains and Random Fields, Entropy and Coding

Springer

Probability Theory and Stochastic Modelling

Volume 78

The **Probability Theory and Stochastic Modelling** series is a merger and continuation of Springer's two well established series Stochastic Modelling and Applied Probability and Probability and Its Applications series. It publishes research monographs that make a significant contribution to probability theory or an applications domain in which advanced probability methods are fundamental. Books in this series are expected to follow rigorous mathematical standards, while also displaying the expository quality necessary to make them useful and accessible to advanced students as well as researchers. The series covers all aspects of modern probability theory including

- Gaussian processes
- Markov processes
- Random fields, point processes and random sets
- Random matrices
- Statistical mechanics and random media
- Stochastic analysis

as well as applications that include (but are not restricted to):

- Branching processes and other models of population growth
- Communications and processing networks
- Computational methods in probability and stochastic processes, including simulation
- Genetics and other stochastic models in biology and the life sciences
- Information theory, signal processing, and image synthesis
- Mathematical economics and finance
- Statistical methods (e.g. empirical processes, MCMC)
- Statistics for stochastic processes
- Stochastic control
- Stochastic models in operations research and stochastic optimization
- Stochastic models in the physical sciences

More information about this series at http://www.springer.com/series/13205

Pierre Brémaud

Discrete Probability Models and Methods

Probability on Graphs and Trees, Markov Chains and Random Fields, Entropy and Coding

Springer

Pierre Brémaud
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne
Switzerland

ISSN 2199-3130 ISSN 2199-3149 (electronic)
Probability Theory and Stochastic Modelling
ISBN 978-3-319-43475-9 ISBN 978-3-319-43476-6 (eBook)
DOI 10.1007/978-3-319-43476-6

Library of Congress Control Number: 2016962040

Mathematics Subject Classification (2010): 60J10, 68Q87, 68W20, 68W40, 05C80, 05C81, 60G60, 60G42, 60C05, 60K05, 60J80, 60K15

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Pour Marion

Contents

Introduction

Discrete probability deals with random elements taking their values in a finite space, or an infinite yet denumerable space: integer-valued random variables, but also random elements with values in a complex space, for instance a graph, a tree, or a combinatorial stucture such as a set partition.

Many problems of a probabilistic nature arising in the applied sciences can be dealt with in the framework of discrete probability. This is especially true in the information, computing and communications sciences. For instance, the study of random graphs has relatively recently been revived with the advent of social networks and community marketing, and percolation graphs have become a popular model of connection in mobile communications. The link between randomness and computation is an area of investigation where discrete probability methods play a priviledged role. When does a logical equation admits a solution, when does there exist a graph with a given property? If the structure of the equation or of the graph is very complex, the probabilistic approach can be efficient. It also features random algorithms that efficiently solve a variety of problems, such as sorting a list of numbers in increasing order or deciding if a given (large) number is prime, and compete with the corresponding available deterministic algorithms. Also of interest to computer science is the Markov chain theory of sampling, exact or approximate, in view of evaluating the size of complex sets for instance. The theory of Markov chains also finds applications in the performance evaluation of communications systems as well as in signal processing.

Besides the information and communications sciences, discrete probability is of interest to qualitative physics. Phenomena such as percolation, phase transition, simulated annealing and thermodynamical irreversibility, can be explained by relatively simple discrete probability models. In the other direction, physics has been a source of inspiration for the information and computing sciences. For instance, Gibbs random fields find a role in image processing, and entropy turns out to be the central concept of information theory which has well-known applications in computer science, mainly for the efficient use of memory resources and the preservation of stored data integrity.

Four main themes with interactions can be distinguished:

Methods and tools. Although the examples and illustrations relate to the possible applications, mostly in the information, computing and communications sciences, but also subsidiarily in operations research and physics, this book is in the first instance concerned with theory. The emphasis is placed on universal methods (the probabilistic method, the coupling method, the Stein–Chen method, martingale methods) and tools (Chernoff's bound, Hoeffding's inequality, Holley's inequality) whose domains of application extend far beyond the present text.

Markov models. This includes Markov chains (Monte Carlo simulation, exact sampling, the electrical network analogy, the convergence rate theory for reversible Markov chains, the ergodicity theory of non-homogeneous Markov chains and its application to simulated annealing) and Markov fields (Gibbsian representation of random fields and their simulation).

Probability on trees and graphs. This refers to the classical random graphs such as the Galton–Watson branching tree, the Erdös–Rényi graphs and percolation graphs,

but it has a wider scope. In fact, a Markov chain can be viewed as a random walk on an oriented graph, Gibbs fields involve by essence a graph structure. Boltzmann samplers are intimately connected to graph theory in two ways: through the recursive procedures which can be assimilated to random walks on a graph, and because many examples of application concern the random generation of graph structures. The source coding issue of information theory gives rise to optimization problems on trees, and so do the algorithms for generating a random variable from a sequence of fair coin tosses.

Entropy and coding. This most important theme of applied discrete probability is connected to the last one, as we just mentioned, and to the first one because Shannon's coding theorem is perhaps the first and certainly the most spectacular application of the probabilistic method.

The book is self-contained. The mathematical level is that of a begining graduate course. The prerequisites consist of basic calculus (series) and linear algebra (matrices) and the reader is not assumed to be trained in probability. In fact, the first chapters constitute an introduction to discrete probability. I have avoided the "dead-end effect", the curse of introductory texts in probability that do not involve the measure-theoretical aspects of this subject. In this book, the terminology and notation are those of a standard theoretical course in probability and the proofs of many important results, such as the strong law of large numbers and the martingale convergence theorem, are the same as the corresponding ones in the general (non-discrete) case. Therefore, the time spent in absorbing the discrete theory will not be lost for the reader willing to pursue in the direction of more theory. In fact, most of the methods of discrete probability such as coupling, to name just one, can be easily adapted to the non-discrete case. Only, in the discrete case, they are more easily formulated and already find spectacular applications.

This book is merely an introduction to a few vast and flourishing domains of applied probability. However, since it reviews in detail the most important methods and tools of discrete probability, the reader will be ready for a direct access to the specialized and/or technical literature. The subsections entitled "Books for Further Information" at the end of each chapter provide a guide to both the advanced theory and its specific applications.

Practical issues. The index gives the page where a particular notation or abbreviation is explained. The position in the index of the corresponding item is the alphabetical one. For instance "$d_V(\alpha, \beta)$" appears is in the sublist for the letter "d". The index has several lines under the general heading "Example" which concern linked examples ("take 1", "take 2", etc.).

Acknowledgements

I wish to acknowledge my debt and express my sincere gratitude to Thomas Bonald, Anne Bouillard, Marc Lelarge, Arpan Mukhopadhyay, Thomas Nowak, Sibi Raj Pillai, Andrea Ridolfi, Christelle Rovetta, Justin Salez and Rémi Varloot for initiating me to some topics of the table of contents as well as for their comments, corrections and technical help in the preparation of the manuscript.

Paris, October 18, 2016

Chapter 1

Events and Probability

1.1 Events

1.1.1 The Sample Space

The study of random phenomena requires a clear and precise language that allows the neophyte to avoid the traps of fallacious intuition which pave the way. This section introduces the terminology and notation.

Probability theory features familiar mathematical objects, such as points, sets and functions, which however receive a particular interpretation: points are *outcomes* (of an experiment), sets are *events*, functions are *random numbers*. The meaning of these terms will be given just after we recall the notation concerning operations on sets, *union*, *intersection*, and *complementation*.

If A and B are subsets of some set Ω, $A \cup B$ denotes their union and $A \cap B$ their intersection. In this book we shall denote by $\overline{A}$ the complement of A in Ω. The notation $A + B$ (the sum of A and B) implies by convention that A and B are disjoint, in which case it represents the union $A \cup B$. Similarly, the notation $\sum_{k=1}^{\infty} A_k$ used for $\cup_{k=1}^{\infty} A_k$ implies that the A_k's are pairwise disjoint. The notation $A - B$ implies that $B \subseteq A$, and it stands for $A \cap \overline{B}$. In particular, if $B \subseteq A$, then $A = B + (A - B)$. Recall *De Morgan's identities* for a sequence $\{A_n\}_{n \geq 1}$ of subsets of Ω:

$$\overline{\left(\bigcap_{n=1}^{\infty} A_n\right)} = \bigcup_{n=1}^{\infty} \overline{A}_n \text{ and } \overline{\left(\bigcup_{n=1}^{\infty} A_n\right)} = \bigcap_{n=1}^{\infty} \overline{A}_n \,.$$

The indicator function of the subset $A \subseteq \Omega$ is the function $1_A : \Omega \to \{0, 1\}$ defined by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \,, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Let $\mathcal{P}$ be a property that an element x of some set E may or may not satisfy. By definition

$$1_{\mathcal{P}}(x) = \begin{cases} 1 & \text{if } x \text{ satisfies } \mathcal{P} \,, \\ 0 & \text{if otherwise.} \end{cases}$$

The cardinality of a set A (the number of its elements in case it is finite or denumerable) will be denoted by $|A|$.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_1

Random phenomena are observed by means of experiments (performed either by man or nature). Each experiment results in an outcome. The collection of all possible outcomes ω is called the sample space Ω. Any subset A of the sample space Ω can be regarded as a representation of some event.

EXAMPLE 1.1.1: TOSSING A DIE, TAKE 1. The experiment consists in tossing a die once. The possible outcomes are $\omega = 1, 2, \ldots, 6$ and the sample space is the set $\Omega = \{1, 2, 3, 4, 5, 6\}$. The subset $A = \{1, 3, 5\}$ is the event "result is odd."

EXAMPLE 1.1.2: THROWING A DART. The experiment consists in throwing a dart at a wall. The sample space can be chosen to be the plane $\mathbb{R}^2$. An outcome is the position $\omega = (x, y)$ hit by the dart. The subset $A = \{(x, y); x^2 + y^2 > 1\}$ is an event that could be named "you missed the dartboard" if the dartboard is a disk centered at 0 and of radius 1.

EXAMPLE 1.1.3: HEADS AND TAILS, TAKE 1. The experiment is an infinite succession of coin tosses. One can take for sample space the collection of all sequences $\omega = \{x_n\}_{n\geq 1}$, where $x_n = 1$ or 0, depending on whether the n-th toss results in heads or tails. The subset $A = \{\omega;\ x_k = 1 \text{ for } k = 1 \text{ to } 1{,}000\}$ is a lucky event for anyone betting on heads!

1.1.2 The Language of Probabilists

The probabilists have their own dialect. They say that outcome ω realizes event A if $\omega \in A$. For instance, in the die model of Example 1.1.1, the outcome $\omega = 1$ realizes the event "result is odd", since $1 \in A = \{1, 3, 5\}$. Obviously, if ω does not realize A, it realizes $\overline{A}$. Event $A \cap B$ is realized by outcome ω if and only if ω realizes both A and B. Similarly, $A \cup B$ is realized by ω if and only if *at least* one event among A and B is realized (both can be realized). Two events A and B are called incompatible when $A \cap B = \varnothing$. In other words, event $A \cap B$ is impossible: no outcome ω can realize both A and B. For this reason one refers to the empty set $\varnothing$ as the impossible event. Naturally, Ω is called the certain event. Recall now that the notation $\sum_{k=1}^{\infty} A_k$ is used for $\cup_{k=1}^{\infty} A_k$ only when the subsets A_k are pairwise disjoint. In the terminology of sets, the sets $A_1, A_2, \ldots$ form a partition of Ω if

$$\sum_{k=0}^{\infty} A_k = \Omega\,.$$

One then says that events $A_1, A_2, \ldots$ are mutually exclusive and exhaustive. They are exhaustive in the sense that any outcome ω realizes at least one among them. They are mutually exclusive in the sense that any two distinct events among them are incompatible. Therefore, any ω realizes *one and only one* of the events $A_1, \ldots, A_n$. In terms of indicator functions,

$$\sum_{k=0}^{\infty} 1_{A_k} = 1\,.$$

If $B \subseteq A$, event B is said to imply event A, because ω realizes A whenever it realizes B. In particular $1_B(\omega) \leq 1_A(\omega)$.

1.1.3 The Sigma-field of Events

Probability theory assigns to each event a number, the *probability* of the said event. The collection $\mathcal{F}$ of events to which a probability is assigned is not always identical to the collection of all subsets of Ω. The requirement on $\mathcal{F}$ is that it should be a sigma-field:

Definition 1.1.4 *Let $\mathcal{F}$ be a collection of subsets of Ω, such that*

(i) the certain event Ω is in $\mathcal{F}$,

(ii) if A belongs to $\mathcal{F}$, then so does its complement $\overline{A}$, and

(iii) if $A_1, A_2, \ldots$ belong to $\mathcal{F}$, then so does their union $\cup_{k=1}^{\infty} A_k$.

One then calls $\mathcal{F}$ a sigma-field on Ω, here the sigma-field of events.

Note that the impossible event $\varnothing$ being the complement of the certain event Ω is in $\mathcal{F}$. Note also that if $A_1, A_2, \ldots$ belong to $\mathcal{F}$, then so does their intersection $\cap_{k=1}^{\infty} A_k$ (Exercise 1.4.1).

The trivial sigma-field and the gross sigma-field are respectively the collection $\mathcal{P}(\Omega)$ of all subsets of Ω, and the sigma-field with only two members: $\{\Omega, \varnothing\}$.

If the sample space Ω is finite or countable, one usually (but not always and not necessarily) considers any subset of Ω to be an event. In other words, the sigma-field of events is the trivial one.

EXAMPLE 1.1.5: BOREL SIGMA-FIELD. The Borel sigma-field on $\mathbb{R}^n$, denoted $\mathcal{B}(\mathbb{R}^n)$, is by definition the smallest sigma-field on $\mathbb{R}^n$ that contains all rectangles, that is, all sets of the form $\prod_{j=1}^{n} I_j$, where the I_j's are arbitrary intervals of $\mathbb{R}$. The sets in this sigma-field are called Borel sets. The above definition is not constructive and therefore one may wonder if there exist sets that are not Borel sets. It turns out that such sets do exist, but they are in a sense "pathological". In practice, it is enough to know that all the sets for which you had been able to compute the n-volume in your earlier life are Borel sets.

EXAMPLE 1.1.6: HEADS AND TAILS, TAKE 2. Take $\mathcal{F}$ to be the smallest sigma-field that contains all the sets $\{\omega\,;\, x_k = 1\}$, $k \geq 1$. This sigma-field also contains the sets $\{\omega\,;\, x_k = 0\}$, $k \geq 1$ (pass to the complements), and therefore (take intersections) all the sets of the form $\{\omega\,;\, x_1 = a_1, \ldots, x_n = a_n\}$ for all $n \geq 1$, all $a_1, \ldots, a_n \in \{0, 1\}$.

1.2 Probability

1.2.1 The Axioms

The probability of an event measures the likeliness of its occurrence. As a function defined on the sigma-field of events, it is required to satisfy a few properties, called the axioms of probability.

Definition 1.2.1 *A probability on* $(\Omega, \mathcal{F})$ *is a mapping* $P : \mathcal{F} \to \mathbb{R}$ *such that*

(i) $0 \leq P(A) \leq 1$,

(ii) $P(\Omega) = 1$, *and*

(iii) $P(\sum_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$ *(sigma-additivity property).*

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space, or *probability model.*

Now is perhaps the best time to introduce a notation that will become standard starting from the next chapter. In this notation, commas replace intersection symbols, for instance $P(A, B) := P(A \cap B)$.

EXAMPLE 1.2.2: TOSSING A DIE, TAKE 2. Formula $P(A) = \frac{|A|}{6}$ defines a probability P on $\Omega = \{1, 2, 3, 4, 5, 6\}$.

EXAMPLE 1.2.3: HEADS AND TAILS, TAKE 3. Choose probability P such that for any event $A = \{x_1 = a_1, \ldots, x_n = a_n\}$, $P(A) = \frac{1}{2^n}$. Note that this does not define the probability of any event of $\mathcal{F}$. But the theory tells us that there does exist such a probability satisfying the above requirement and that this probability is unique.

EXAMPLE 1.2.4: RANDOM POINT IN THE SQUARE, TAKE 1. The following is a model of a random point in the unit square $\Omega = [0, 1]^2$: $\mathcal{F}$ is the collection of Borel sets of $\mathbb{R}^2$ contained in $[0, 1]^2$. Measure theory tells us that there exists one and only one probability P satisfying the requirement $P([a, b] \times [c, d]) = (b-a) \times (d-c)$, called the Lebesgue probability on $[0, 1]^2$, and that formalizes the intuitive notion of "area".

The probability of Example 1.2.2 suggests an unbiased die, where each outcome 1, 2, 3, 4, 5, or 6 has the same probability. As we shall soon see, the probability P of Example 1.2.3 implies an *unbiased* coin and *independent* tosses (the emphasized terms will be defined later).

The axioms of probability are motivated by the heuristic interpretation of probability as empirical frequency. If n "independent" experiments are performed, among which n_A result in the realization of A, then the empirical frequency

$$F(A) := \frac{n_A}{n}$$

should be close to $P(A)$ if n is "sufficiently large." (This statement has to be made precise. It is in fact a loose expression of the law of large numbers that will be given later.) Clearly, the empirical frequency satisfies the axioms.

The properties of probability stated below follow directly from the axioms:

Theorem 1.2.5 *For any event $A \in \mathcal{F}$*

$$P(\overline{A}) = 1 - P(A)\,, \tag{1.1}$$

and in particular $P(\varnothing) = 0$.

Proof. For (1.1), use additivity:

$$1 = P(\Omega) = P(A + \overline{A}) = P(A) + P(\overline{A})\,.$$

Applying (1.1) with $A = \Omega$ gives $P(\varnothing) = 0$. □

Theorem 1.2.6 *Probability is monotone, that is,*

$$A \subseteq B \Longrightarrow P(A) \leq P(B). \tag{1.2}$$

(Recall the interpretation of the set inclusion $A \subseteq B$: event A implies event B.)

Proof. When $A \subseteq B$, $B = A + (B - A)$, and therefore

$$P(B) = P(A) + P(B - A) \geq P(A).$$

□

Theorem 1.2.7 *Probability is sub-sigma-additive, that is:*

$$P(\cup_{k=1}^{\infty} A_k) \leq \textstyle\sum_{k=1}^{\infty} P(A_k). \tag{1.3}$$

Proof.

Observe that

$$\cup_{k=1}^{\infty} A_k = \sum_{k=1}^{\infty} A'_k,$$

where $A'_k := A_k \cap \left\{\overline{\cup_{i=1}^{k-1} A_i}\right\}$. Therefore,

$$P\left(\cup_{k=1}^{\infty} A_k\right) = P\left(\sum_{k=1}^{\infty} A'_k\right) = \sum_{k=1}^{\infty} P(A'_k).$$

But $A'_k \subseteq A_k$, and therefore $P(A'_k) \leq P(A_k)$. □

We now introduce a central notion of probability theory.

Definition 1.2.8 *A set $N \subset \Omega$ is called P-negligible if it is contained in an event $A \in \mathcal{F}$ of probability $P(A) = 0$.*

Theorem 1.2.9 *A countable union of negligible sets is a negligible set.*

Proof. Let N_k, $k \geq 1$, be P-negligible sets. By definition there exists a sequence A_k, $k \geq 1$, of events of null probability such that $N_k \subseteq A_k$, $k \geq 1$. We have

$$N := \cup_{k\geq 1} N_k \subseteq A := \cup_{k\geq 1} A_k,$$

and by sub-sigma-additivity of probability, $P(A) = 0$. □

EXAMPLE 1.2.10: RANDOM POINT IN THE SQUARE, TAKE 2. Each rational point therein has null area and therefore null probability. Therefore, in this model, the (countable) set of rational points of the square has null probability. In other words, the probability of drawing a rational point is null.

Definition 1.2.11 *A property $\mathcal{P}$ relative to the samples $\omega \in \Omega$ is said to hold P–almost surely ("P–a.s.") if*

$$P(\{\omega;\ \omega \text{ verifies property } \mathcal{P}\}) = 1.$$

If there is no ambiguity as to the underlying probability P, one usually abbreviates "P–almost surely" in "almost surely".

1.2.2 The Borel–Cantelli Lemma

The following property comes close to being a tautology and is of great use.

Theorem 1.2.12 *Let $\{A_n\}_{n\geq 1}$ be a non-decreasing sequence of events (that is, for all $n \geq 1$, $A_{n+1} \supseteq A_n$). Then*

$$P(\cup_{n=1}^{\infty} A_n) = \lim_{n\uparrow\infty} P(A_n)\,. \tag{1.4}$$

Proof. Write

$$A_n = A_1 + (A_2 - A_1) + \cdots + (A_n - A_{n-1})$$

and

$$\cup_{k=1}^{\infty} A_k = A_1 + (A_2 - A_1) + (A_3 - A_2) + \cdots\,.$$

Therefore,

$$\begin{aligned} P(\cup_{k=1}^{\infty} A_k) &= P(A_1) + \sum_{j=2}^{\infty} P(A_j - A_{j-1}) \\ &= \lim_{n\uparrow\infty} \left\{ P(A_1) + \sum_{j=2}^{n} P(A_j - A_{j-1}) \right\} = \lim_{n\uparrow\infty} P(A_n). \end{aligned}$$

□

Corollary 1.2.13 *Let $\{B_n\}_{n\geq 1}$ be a non-increasing sequence of events (that is, for all $n \geq 1$, $B_{n+1} \subseteq B_n$). Then*

$$P(\cap_{n=1}^{\infty} B_n) = \lim_{n\uparrow\infty} P(B_n)\,. \tag{1.5}$$

Proof. Using De Morgan's identity and applying (1.4) with $A_n = \overline{B}_n$, observing that $\{\overline{B}_n\}_{n\geq 1}$ is a non-decreasing sequence of events:

$$\begin{aligned} P(\cap_{n=1}^{\infty} B_n) &= 1 - P\left(\overline{\cap_{n=1}^{\infty} B_n}\right) = 1 - P(\cup_{n=1}^{\infty} \overline{B}_n) \\ &= 1 - \lim_{n\uparrow\infty} P(\overline{B}_n) = \lim_{n\uparrow\infty}(1 - P(\overline{B}_n)) = \lim_{n\uparrow\infty} P(B_n)\,. \end{aligned}$$

□

EXAMPLE 1.2.14: EXTINCTION OF THE BLUEPINKOS. Let A_n be the event that in the year n, the population of bluepinkos (a rare species of australian birds) is not null. The event that the bluepinkos eventually become extinct is $\mathcal{E} = \cup_{n\geq 1} A_n$. Obviously the sequence $\{A_n\}_{n\geq 1}$ is non-increasing (if there is no bluepinko left at time n, then there will be no bluepinko at all subsequent years). Therefore, according to Theorem 1.2.12, $P(\mathcal{E}) = \lim_{n\uparrow\infty} P(A_n)$.

EXAMPLE 1.2.15: YOU CANNOT ALWAYS WIN. An event B can be logically impossible, that is $B = \varnothing$. It can also be negligible, that is $P(B) = 0$. Of course, a logically impossible event is a fortiori negligible. Probabilistic computations seldom lead to the conclusion that an event is impossible, but will tell that it has a null probability, which is for all practical purposes sufficient. For instance, in an infinite sequence of coin tosses with an unbiased coin, the event B that one always obtain heads is not logically impossible, but it has a null probability. In fact, the probability of the event B_n that the n first tosses give heads is $\frac{1}{2^n}$, and therefore, by sequential continuity (B_n is non-increasing and $B = \cap_{n\geq 1} B_n$), $P(B) = 0$.

Consider a sequence of events $\{A_n\}_{n\geq 1}$ where the index n may be interpreted as time. We are interested in the probability that A_n occurs infinitely often, that is, the probability of the event

$$\{\omega; \omega \in A_n \text{ for an infinity of indices } n\},$$

denoted by $\{A_n\, i.o.\}$, where *i.o.* is an abbreviation for "infinitely often". We have the direct Borel–Cantelli lemma:

Theorem 1.2.16 *For any sequence of events $\{A_n\}_{n\geq 1}$,*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P(A_n\, i.o.) = 0.$$

Proof. We first observe that

$$\{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k.$$

(Indeed, if ω belongs to the set on the right-hand side, then for *all* $n \geq 1$, ω belongs to at least one among $A_n, A_{n+1}, \ldots$, which implies that ω is in A_n for an infinite number of indices n. Conversely, if ω is in A_n for an infinite number of indices n, it is for *all* $n \geq 1$ in at least one of the sets $A_n, A_{n+1}, \ldots$)

The set $\cup_{k \geq n} A_k$ decreases as n increases, so that by the sequential continuity property of probability,

$$P(A_n \text{ i.o.}) = \lim_{n \uparrow \infty} P\left(\bigcup_{k \geq n} A_k \right). \tag{1.6}$$

But by sub-σ-additivity,

$$P\left(\bigcup_{k \geq n} A_k \right) \leq \sum_{k \geq n} P(A_k),$$

and by the summability assumption, the right-hand side of this inequality goes to 0 as $n \uparrow \infty$. □

Counting Models

A number of problems in Probability reduce to counting the elements in a finite set. The general setting is the following. The set Ω of all possible outcomes is finite, and for some reason (of symmetry for instance) we are led to believe that all the outcomes ω have the same probability. Since the probabilities sum up to one, each outcome has probability $\frac{1}{|\Omega|}$. Since the probability of an event A is the sum of the probabilities of all outcomes $\omega \in A$, we have

$$P(A) = \frac{|A|}{|\Omega|}. \tag{$\star$}$$

Thus, computing $P(A)$ requires *counting* the elements in the sets A and Ω.

Recall the two basic facts of combinatorics (the art of counting): (a) the number of permutations of a set with n elements is $n!$, and (b) in a set of n elements, the number of subsets of k elements is $\binom{n}{k} := \frac{n!}{k!(n-k)!}$. Also recall Stirling's equivalence $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$. A more precise formulation is: For all positive integers n,

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

A simpler form is

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq 2\sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \tag{1.7}$$

The following useful bound is a direct consequence of the above: For all positive integers $k < n$,

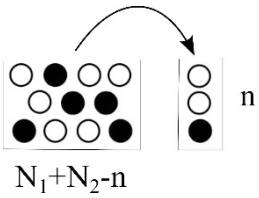
$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k . \tag{1.8}$$

Indeed,

$$\begin{aligned}\binom{n}{k} &= \frac{n(n-1)\cdots(n-k+1)}{k!} \\ &\leq \frac{n^k}{k!} \leq \frac{n^k}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k} \leq \frac{1}{\sqrt{2\pi k}}\left(\frac{en}{k}\right)^k \leq \left(\frac{en}{k}\right)^k .\end{aligned}$$

EXAMPLE 1.2.17: URN. There is an urn containing N_1 black balls and N_2 red balls. You draw successively without replacement and at random n balls from the urn ($n \leq N_1 + N_2$). The probability of having drawn k black balls ($0 \leq k \leq \inf(N_1, n)$) is:

$$p_k = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N_1+N_2}{n}} .$$



Proof. The set of outcomes Ω is the family of all subsets ω of n balls among the $N_1 + N_2$ balls in the urn. Therefore,

$$|\Omega| = \binom{N_1 + N_2}{n}.$$

It is reasonable to suppose that all the outcomes are equiprobable. In this case, formula $(\star)$ applies. One must therefore count the subsets ω with k black balls and $n - k$ red balls. To form such a set, a set of k black balls among the N_1 black balls is formed, and there are $\binom{N_1}{k}$ such sets. To each such subset of k black balls, one must associate a subset of $n - k$ red balls. This multiplies the possibilities by $\binom{N_2}{n-k}$. Thus, if A is the number of subsets of n balls among the $N_1 + N_2$ balls in the urn which consist of k black balls and $n - k$ red balls

$$|A| = \binom{N_1}{k}\binom{N_2}{n-k},$$

and therefore $p_k = \frac{|A|}{|\Omega|}$ is as announced above. □

1.3 Independence and Conditioning

1.3.1 Independent Events

In the frequency interpretation of probability, a situation where $n_{A\cap B}/n \approx (n_A/n) \times (n_B/n)$, or

$$\frac{n_{A\cap B}}{n_B} \approx \frac{n_A}{n}$$

(here $\approx$ is a non-mathematical symbol meaning "approximately equal") suggests some kind of "independence" of A and B, in the sense that statistics relative to A do not vary when passing from a neutral sample of population to a selected sample characterized by the property B. For example, the proportion of people with a family name begining with H is the same among a large population with the usual mix of men and women as it would be among a large all-male population. This prompts us to give the following formal definition of independence, the single most important concept of probability theory.

Definition 1.3.1 *Two events A and B are called independent if*

$$P(A \cap B) = P(A)P(B)\,. \tag{1.9}$$

One should be aware that incompatibility is different from independence. As a matter of fact, two incompatible events A and B are independent if and only if at least one of them has null probability. Indeed, if A and B are incompatible, $P(A \cap B) = P(\varnothing) = 0$, and therefore (1.9) holds if and only if $P(A)P(B) = 0$.

The notion of independence carries over to families of events.

Definition 1.3.2 *A family $\{A_n\}_{n\in\mathbb{N}}$ of events is called independent if for all finite indices $i_1, \ldots, i_r \in \mathbb{N}$,*

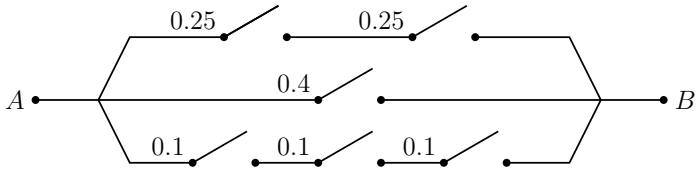
$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}\right) = P(A_{i_1}) \times P(A_{i_2}) \times \cdots \times P(A_{i_r})\,.$$

One also says that the A_n's are jointly independent.

Theorem 1.3.3 *Suppose that the family of events $\{A_n\}_{n\in\mathbb{N}}$ is independent. Then, so is the family $\{\tilde{A}_n\}_{n\in\mathbb{N}}$, where for each n, $\tilde{A}_n = A_n$ or $\overline{A}_n$ (the choice may vary with the index).*

Proof. Exercise 1.4.4. □

Example 1.3.4: The communications network. Two nodes A and B in a communications network are connected by three different routes, each containing a number of links that can fail, represented symbolically in the figure by switches that are lifted if the link is not operational. In the figure, the number associated with a switch is the probability of failure of the corresponding link. Link failures



occur independently. What is the probability that nodes A and B are connected, that is, that there exists at least one operational route between A and B?

Let U_1 be the event "no link failure in the upper route". Defining similarly U_2 and U_3, the probability to be computed is that of $U_1 \cup U_2 \cup U_3$, or by De Morgan's identity, that of the complement of $\overline{U}_1 \cap \overline{U}_2 \cap \overline{U}_3$:

$$1 - P(\overline{U}_1 \cap \overline{U}_2 \cap \overline{U}_3) = 1 - P(\overline{U}_1)P(\overline{U}_2)(P\overline{U}_3),$$

where the last equality follows from the independence assumption concerning the links. Letting now U_1^1 = "switch 1 (first from left) in the upper route is not lifted" and U_1^2 = "switch 2 in the upper route is not lifted", we have $U_1 = U_1^1 \cap U_1^2$, therefore, in view of the independence assumption,

$$P(\overline{U}_1) = 1 - P(U_1) = 1 - P(U_1^1)P(U_1^2).$$

With the given data, $P(\overline{U}_1) = 1-(0.75)^2$. Similarly, $P(\overline{U}_2) = 1-0.6$ and $P(\overline{U}_3) = 1-(0.9)^3$. The final result is $1-(0.4375)(0.4)(0.271) = 0.952575$.

The formula (1.9) extends to a countable number of events:

Theorem 1.3.5 *Let $\{C_n\}_{n\geq 1}$ be a sequence of independent events. Then*

$$P(\cap_{n=1}^{\infty} C_n) = \Pi_{n=1}^{\infty} P(C_n)\,. \tag{1.10}$$

Proof. Let $B_n = \cap_{k=1}^{n} C_k$. By independence $P(B_n) = P(\cap_{k=1}^{n} C_k) = \Pi_{k=1}^{n} P(C_k)$, a quantity which tends to $\Pi_{k=1}^{\infty} P(C_k)$ as $n \uparrow \infty$. Apply (1.10) to the decreasing events B_n to obtain the announced result. □

Next, we give the so-called converse Borel–Cantelli lemma. It is not strictly speaking a converse of the Borel–Cantelli lemma because it requires the additional assumption of independence.

Theorem 1.3.6 *Let $\{A_n\}_{n\geq 1}$ be a sequence of independent events. Then,*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \implies P(A_n \text{ } i.o.) = 1.$$

Proof. We may assume without loss of generality that $P(A_n) > 0$ for all $n \geq 1$ (why?). The divergence hypothesis implies that for all $n \geq 1$ (see Section A.1),

$$\prod_{k=n}^{\infty}(1 - P(A_k)) = 0\,.$$

This infinite product equals, in view of the independence assumption (see Theorem 1.3.5),

$$\prod_{k=n}^{\infty} P\left(\overline{A_k}\right) = P\left(\bigcap_{k=n}^{\infty} \overline{A_k}\right).$$

Therefore,

$$P\left(\bigcap_{k=n}^{\infty} \overline{A_k}\right) = 0\,.$$

Passing to the complement and using De Morgan's identity,

$$P\left(\bigcup_{k=n}^{\infty} A_k\right) = 1\,.$$

Therefore, by (1.6),

$$P(A_n\, i.o.) = \lim_{n\uparrow\infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) = 1\,.$$

□

1.3.2 Conditional Probability

We continue our heuristic discussion of probability in terms of empirical frequencies. Dependence between A and B occurs when $P(A \cap B) \neq P(A)P(B)$. In this case the relative frequency $n_{A\cap B}/n_B \approx P(A \cap B)/P(B)$, which represents what we expect concerning event A given that we already know that event B occured, is different from the frequency n_A/n. This suggests the following definition:

Definition 1.3.7 *The conditional probability of A given B is the number*

$$P(A \,|\, B) := \tfrac{P(A\cap B)}{P(B)}\,, \tag{1.11}$$

defined when $P(B) > 0$. If $P(B) = 0$, one defines $P(A \,|\, B)$ arbitrarily between 0 and 1.

The quantity $P(A \,|\, B)$ represents our expectation of A being realized when the only available information is that B is realized. Indeed, this expectation is based on the relative frequency $n_{A\cap B}/n_B$ alone. Of course, if A and B are independent, then $P(A \,|\, B) = P(A)$.

1.3.3 The Bayes Calculus

Probability theory is primarily concerned with the computation of probabilities of complex events. The following formulas, called the Bayes formulas, will be recurrently used.

Theorem 1.3.8 *For any events A and B of positive probability, we have the Bayes formula of retrodiction:*

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}\,. \tag{1.12}$$

Proof. Rewrite Definition 1.11 symmetrically in A and B:

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)\,.$$

□

We now give two more basic formulas useful in computing conditional probability.

Theorem 1.3.9 *Let the events $B_1, B_2, \ldots$ form a partition of Ω, that is, $\sum_{i=1}^{\infty} B_i = \Omega$. Then for any event A, we have the Bayes formula of exclusive and exhaustive causes, also called the Bayes formula of total causes:*

$$P(A) = \sum_{i=1}^{\infty} P(A \mid B_i)P(B_i)\,. \tag{1.13}$$

Proof. Decompose A as follows:

$$A = A \cap \Omega = A \cap \left(\sum_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty}(A \cap B_i).$$

Therefore (sigma-additivity and definition of conditional probability):

$$P(A) = P\left(\sum_{i=1}^{\infty}(A \cap B_i)\right) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A \mid B_i)P(B_i).$$

□

EXAMPLE 1.3.10: CAN WE ALWAYS BELIEVE DOCTORS? Doctors apply a test that gives a positive result in 99% of the cases where the patient is affected by the disease. However it happens in 2% of the cases that a healthy patient is "positive". Statistical data show that one individual out of 1000 has the disease. What is the probability that a patient with a positive test is affected by the disease?

Solution: Let M be the event "the patient is ill," and let $+$ and $-$ be the events "the test is positive" and "the test is negative" respectively. We have the data

$$P(M) = 0.001,\ P(+ \mid M) = 0.99,\ P(+ \mid \overline{M}) = 0.02,$$

and we must compute $P(M \mid +)$. By the retrodiction formula,

$$P(M \mid +) = \frac{P(+ \mid M)P(M)}{P(+)}.$$

By the formula of exclusive and exhaustive causes,

$$P(+) = P(+ \mid M)P(M) + P(+ \mid \overline{M})P(\overline{M}).$$

Therefore,

$$P(M \mid +) = \frac{(0.99)(0.001)}{(0.99)(0.001) + (0.02)(0.999)},$$

that is, approximately 0.05. What do you think?

EXAMPLE 1.3.11: HAFMORON UNIVERSITY ALUMNI. A student from the famous Veryhardvard University has with probability 0.25 a bright intelligence. Students from the Hafmoron State University have a probability 0.10 of being bright. You find yourself in an assembly with 10 Veryhardvard students and 20 Hafmoron State University students. You meet a handsome girl (*resp.*, boy) whose intelligence is obviously superior. What is the probability that she (*resp.*, he) registered at Hafmoron State University?

With obvious notation:

$$\begin{aligned} P(HM \mid BI) &= \frac{P(HM \cap BI)}{P(BI)} = \frac{P(BI \mid HM)P(HM)}{P(BI)} \\ &= \frac{P(BI \mid HM)P(HM)}{P(BI \mid HM)P(HM) + P(BI \mid VH)P(VH)}, \end{aligned}$$

that is, numerically,

$$P(HM \mid BI) = \frac{0.1 \times \frac{2}{3}}{0.1 \times \frac{2}{3} + 0.25 \times \frac{1}{3}} = \frac{20}{45}.$$

EXAMPLE 1.3.12: THE BALLOT PROBLEM. In an election, candidates I and II have obtained a and b votes respectively. Candidate I won, that is, $a > b$. What is the probability that in the course of the vote counting procedure, candidate I has always had the lead?

Solution: Let $p_{a,b}$ be the probability that A is always ahead. We have by the formula of exclusive and exhaustive causes, conditioning on the last vote:

$$\begin{aligned} p_{a,b} &= P(A \text{ always ahead } \mid A \text{ gets the last vote })P(A \text{ gets the last vote }) \\ &\quad + P(A \text{ always ahead } \mid B \text{ gets the last vote })P(B \text{ gets the last vote }) \\ &= p_{a-1,b}\frac{a}{a+b} + p_{a,b-1}\frac{b}{a+b}, \end{aligned}$$

with the convention that for $a = b + 1$, $p_{a-1,b} = p_{b,b} = 0$. The result follows by induction on the total number of votes $a + b$:

$$p_{a,b} = \frac{a-b}{a+b}\,.$$

Theorem 1.3.13 *For any sequence of events $A_1, \ldots, A_n$, we have the Bayes sequential formula:*

$$P\left(\cap_{i=1}^{k} A_i\right) = P(A_1)P(A_2 \,|\, A_1)P(A_3 \,|\, A_1 \cap A_2) \cdots P\left(A_k \,|\, \cap_{i=1}^{k-1} A_i\right)\,. \tag{1.14}$$

Proof. By induction. First observe that (1.14) is true for $k = 2$ by definition of conditional probability. Suppose that (1.14) is true for k. Write

$$\begin{aligned} P\left(\cap_{i=1}^{k+1} A_i\right) &= P\left(\left(\cap_{i=1}^{k} A_i\right) \cap A_{k+1}\right) \\ &= P\left(A_{k+1} \,|\, \cap_{i=1}^{k} A_i\right) P\left(\cap_{i=1}^{k} A_i\right), \end{aligned}$$

and replace $P\left(\cap_{i=1}^{k} A_i\right)$ by the assumed equality (1.14) to obtain the same equality with $k + 1$ replacing k. □

1.3.4 Conditional Independence

Definition 1.3.14 *Let A, B, and C be events, with $P(C) > 0$. One says that A and B are conditionally independent given C if*

$$P(A \cap B \,|\, C) = P(A \,|\, C)P(B \,|\, C)\,. \tag{1.15}$$

In other words, A and B are independent with respect to the probability P_C defined by $P_C(A) = P(A \,|\, C)$ (see Exercise 1.4.5).

EXAMPLE 1.3.15: CHEAP WATCHES. Two factories A and B manufacture watches. Factory A produces approximately one defective item out of 100, and B one out of 200. A retailer receives a container from one of the factories, but he does not know which. (It is however assumed that the two possible origins of the container are equiprobable.) The retailer checks the first watch. It works!

(a) What is the probability that the second watch he will check is good?

(b) Are the states of the first 2 watches independent?

Solution: (a) Let X_n be the state of the n-th watch in the container, with $X_n = 1$ if it works and $X_n = 0$ if it does not. Let Y be the factory of origin. We express our a priori ignorance of where the case comes from by

$$P(Y = A) = P(Y = B) = \frac{1}{2}.$$

Also, we assume that given $Y = A$ (resp., $Y = B$), the states of the successive watches are independent. For instance,

$$P(X_1 = 1, X_2 = 0 \,|\, Y = A) = P(X_1 = 1 \,|\, Y = A)P(X_2 = 0 \,|\, Y = A).$$

We have the data

$$P(X_n = 0 \,|\, Y = A) = 0.01\,, \qquad P(X_n = 0 \,|\, Y = B) = 0.005\,.$$

We are required to compute

$$P(X_2 = 1 \,|\, X_1 = 1) = \frac{P(X_1 = 1, X_2 = 1)}{P(X_1 = 1)}.$$

By the formula of exclusive and exhaustive causes, the numerator of this fraction equals

$$P(X_1 = 1, X_2 = 1 \,|\, Y = A)P(Y = A) + P(X_1 = 1, X_2 = 1 \,|\, Y = B)P(Y = B),$$

that is, $(0.99)^2(0.5) + (0.995)^2(0.5)$, and the denominator is

$$P(X_1 = 1 \,|\, Y = A)P(Y = A) + P(X_1 = 1 \,|\, Y = B)P(Y = B),$$

that is, $(0.99)(0.5) + (0.995)(0.5)$. Therefore,

$$P(X_2 = 1 \,|\, X_1 = 1) = \frac{(0.99)^2 + (0.995)^2}{0.99 + 0.995}.$$

(b) The states of the two watches are not independent. Indeed, if they were, then

$$P(X_2 = 1 \,|\, X_1 = 1) = P(X_2 = 1) = (0.5)\,(0.99 + 0.995)\,,$$

a result different from what we obtained. This shows that for some event C, two events A and B can very well be conditionally independent given C and conditionally independent given $\overline{C}$, and yet *not* be mutually independent.

1.4 Exercises

Exercise 1.4.1. Composed events
Let $\mathcal{F}$ be a sigma-field on some set Ω.
(1) Show that if $A_1, A_2, \ldots$ are in $\mathcal{F}$, then so is $\cap_{k=1}^{\infty} A_k$.
(2) Show that if A_1, A_2 are in $\mathcal{F}$, their *symmetric difference* $A_1 \triangle A_2 := A_1 \cup A_2 - A_1 \cap A_2$ is also in $\mathcal{F}$.

Exercise 1.4.2. Identities
Prove the set identities

$$P(A \cup B) = 1 - P(\overline{A} \cap \overline{B})$$

and

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Exercise 1.4.3. URNS

1. An urn contains 17 red balls and 19 white balls. Balls are drawn in succession at random and without replacement. What is the probability that the first 2 balls are red?

2. An urn contains N balls numbered from 1 to N. Someone draws n balls ($1 \leq n \leq N$) simultaneously from the urn. What is the probability that the lowest number drawn is k?

Exercise 1.4.4. ABOUT INDEPENDENCE

1. Give a simple example of a probability space $(\Omega, \mathcal{F}, P)$ with three events A_1, A_2, A_3 that are pairwise independent, but *not* globally independent (that is, the family $\{A_1, A_2, A_3\}$ is not independent).

2. If $\{A_i\}_{i \in I}$ is an independent family of events, is it true that $\{\tilde{A}_i\}_{i \in I}$ is also an independent family of events, where for each $i \in I$, $\tilde{A}_i = A_i$ or $\overline{A}_i$ (your choice; for instance, with $I = \mathbb{N}$, $\tilde{A}_0 = A_0, \tilde{A}_1 = \overline{A}_1, \tilde{A}_3 = A_3, \ldots$)?

Exercise 1.4.5. CONDITIONAL INDEPENDENCE AND THE MARKOV PROPERTY

1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Define for a fixed event C of positive probability, $P_C(A) := P(A \,|\, C)$. Show that P_C is a probability on $(\Omega, \mathcal{F})$ and that A and B are independent with respect to this probability if and only if they are conditionally independent given C.

2. Let A_1, A_2, A_3 be three events of positive probability. Show that events A_1 and A_3 are conditionally independent given A_2 if and only if the "Markov property" holds, that is, $P(A_3 \,|\, A_1 \cap A_2) = P(A_3 \,|\, A_2)$.

Exercise 1.4.6. HEADS AND TAILS AS USUAL

A person, A, tossing an *unbiased* coin N times obtains T_A tails. Another person, B, tossing her own unbiased coin $N+1$ times has T_B tails. What is the probability that $T_A \geq T_B$? *Hint:* Introduce H_A and H_B, the number of heads obtained by A and B respectively, and use a symmetry argument.

Exercise 1.4.7. APARTHEID UNIVERSITY

In the renowned Social Apartheid University, students have been separated into three social groups for "pedagogical" purposes. In group A, one finds students who individually have a probability of passing equal to 0.95. In group B this probability is 0.75, and in group C only 0.65. The three groups are of equal size. What is the probability that a student passing the course comes from group A? B? C?

Exercise 1.4.8. WISE BET

There are 3 cards. The first one has both faces red, the second one has both faces white, and the third one is white on one face, red on the other. A card is drawn

at random, and the colour of a randomly selected face of this card is shown to you (the other remains hidden). What is the winning strategy if you must bet on the colour of the hidden face?

Exercise 1.4.9. A sequence of liars
Consider a sequence $L_1, \ldots, L_n$ of liars. The first liar L_1 receives information about the occurrence of some event in the form "yes or no", and transmits it to L_2, who transmits it to L_3, etc... Each liar transmits what he hears with probability $p \in (0, 1)$ and the contrary with probability $q = 1 - p$. The decision of lying or not is made independently by each liar. What is the probability x_n of obtaining the correct information from L_n? What is the limit of x_n as n increases to infinity?

Exercise 1.4.10. The campus library complaint
You are looking for a book in the campus libraries. Each library has it with probability 0.60 but the book of each given library may have been stolen with probability 0.25. If there are 3 libraries, what are your chances of obtaining the book?

Exercise 1.4.11. Professor Nebulous
Professor Nebulous travels from Los Angeles to Paris with stopovers in New York and London. At each stop his luggage is transferred from one plane to another. In each airport, including Los Angeles, the probability that his luggage is not assigned to the right plane is p. Professor Nebulous finds that his suitcase has not reached Paris. What are the chances that the mishap took place in Los Angeles, New York, and London, respectively ?

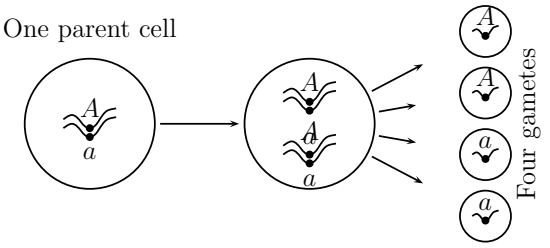
Exercise 1.4.12. Safari butchers
Three tourists participate in a safari in Africa. An elephant shows up, unaware of the rules of the game. The innocent beast is killed, having received two out of the three bullets simultaneously shot by the tourists. The respective probabilities of hit are: Tourist A: $\frac{1}{4}$, Tourist B: $\frac{1}{2}$, Tourist C: $\frac{3}{4}$. Give for each of the tourists the probability that he was the one who missed.

Exercise 1.4.13. The Hardy–Weinberg law
In diploid organisms, each heriditary character is carried by a pair of genes. Consider the situation in which each given *gene* can take two forms called *alleles*, denoted a and A. Such was the case in the historical experiments performed in 1865 by the Czech monk Gregory Mendel who studied the hereditary transmission of the nature of the skin in a species of green peas. The two alleles corresponding to the gene or character "nature of the skin" are a for "wrinkled" and A for "smooth". The genes are grouped into pairs and there are two alleles. Therefore, three *genotypes* are possible for the character under study: aa, Aa (same as aA), and AA. During the reproduction process, each of the two parents contributes to the genetic heritage of their descendant by providing *one* allele of their pair. This is done by the intermediary of the reproductive cells called *gametes*[1] which carry only one gene of the pair of genes characteristic of each parent. The gene carried

[1]In the human species, the spermatozoid and the ovula.

by the gamete is chosen at random among the pair of genes of the parent. The actual process occuring in the reproduction of diploid cells is called *meiosis*.

A given cell possesses two chromosomes. A chromosome can be viewed as a string of genes, each gene being at a specific location in the chain. The chromosomes double and four new cells are formed for every chromosome (see the figure below). One of the four gametes of a "mate" (say, the ovula) chosen at random selects randomly one of the four gametes of the other "partner" (here, the spermatozoid) and this gives "birth" to a pair of alleles.



Let us start from an idealistically infinite population where the genotypes are found in the following proportions:

$$AA \,:\, Aa \,:\, aa$$

$$x \,:\, 2z \,:\, y.$$

Here, x, y, and z are numbers between 0 and 1, and $x+2z+y = 1$. The two parents are chosen independently (random mating), and their gamete chooses an allele at random in the pair carried by the corresponding parent. What is the genotype distribution of the second generation?

Chapter 2

Random Variables

2.1 Probability Distribution and Expectation

2.1.1 Random Variables and their Distributions

The number of heads in a sequence of 10000 coin tosses, the number of days it takes until the next rain and the size of a genealogical tree are random numbers. All are functions of the outcome of a random experiment (performed either by man or by nature) and taking discrete values, that is, values in a countable set. These values are integers in the above examples, but they can be more complex mathematical objects, such as graphs for instance. This chapter gives the elementary rules for computing expectations, a list of famous discrete random variables or vectors (binomial, geometric, Poisson and multinomial), and the elementary theory of conditional expectation.

Definition 2.1.1 *Let E be a countable set. A function $X : \Omega \to E$ such that for all $x \in E$*

$$\{\omega; X(\omega) = x\} \in \mathcal{F}$$

is called a discrete random variable.

Since E is a countable set, it can always be identified with $\mathbb{N}$ or $\overline{\mathbb{N}}$, and therefore we shall often assume that either $E = \mathbb{N}$ or $\overline{\mathbb{N}}$.

Being in $\mathcal{F}$, the event $\{X = x\}$ can be assigned a probability.

Remark 2.1.2 Calling a random variable a *random number* is an innocuous habit as long as one is aware that it is *not the function* X that is random, but the outcome ω. This in turn makes the *number* $X(\omega)$ random.

EXAMPLE 2.1.3: TOSSING A DIE, TAKE 3. The sample space is the set $\Omega = \{1, 2, 3, 4, 5, 6\}$. Take for X the identity: $X(\omega) = \omega$. In that sense X is the random number obtained by tossing a die.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_2

EXAMPLE 2.1.4: HEADS AND TAILS, TAKE 4. The sample space Ω is the collection of all sequences $\omega = \{x_n\}_{n\geq 1}$, where $x_n = 1$ or 0. Define a random variable X_n by $X_n(\omega) = x_n$. It is the random number obtained at the n-th toss. It is indeed a random variable since for all $a_n \in \{0, 1\}$, $\{\omega\,;\, X_n(\omega) = a_n\} = \{\omega\,;\, x_n = a_n\} \in \mathcal{F}$, by definition of $\mathcal{F}$.

The following are elementary remarks.

Theorem 2.1.5 *Let E and F be countable sets. Let X be a random variable with values in E, and let $f : E \to F$ be an arbitrary function. Then $Y := f(X)$ is a random variable.*

Proof. Let $y \in F$. The set $\{\omega;\, Y(\omega) = y\}$ is in $\mathcal{F}$ since it is a countable union of sets in $\mathcal{F}$, namely:

$$\{Y = y\} = \sum_{x\in E;\, f(x)=y} \{X = x\}\,.$$

□

Theorem 2.1.6 *Let E_1 and E_2 be countable sets. Let X_1 and X_2 be random variable with values in E_1 and E_2 respectively. Then $Y := (X_1, X_2)$ is a random variable with values in $E = E_1 \times E_2$.*

Proof. Let $x = (x_1, x_2) \in E$. The set $\{\omega;\, X(\omega) = x\}$ is in $\mathcal{F}$ since it is the intersection of sets in $\mathcal{F}$: $\{X = x\} = \{X_1 = x_1\} \cap \{X_2 = x_2\}$. □

Definition 2.1.7 *From the probabilistic point of view, a discrete random variable X is described by its probability distribution function (or distribution, for short) $\{\pi(x)\}_{x\in E}$, where $\pi(x) := P(X = x)$.*

EXAMPLE 2.1.8: THE UNIFORM DISTRIBUTION. Let $\mathcal{X}$ be a finite set. The random variable with values in this set and having the distribution

$$P(X = x) = \frac{1}{|\mathcal{X}|} \text{ for all } x \in \mathcal{X}$$

is said to be uniformly distributed (or to have the uniform distribution) on $\mathcal{X}$.

EXAMPLE 2.1.9: IS THIS NUMBER THE LARGER ONE? Let a and b be two numbers in $\{1, 2, \ldots, 10,000\}$. Nothing is known about these numbers, except that they are not equal, say $a > b$. Only one of these numbers is shown to you, secretely chosen at random and equiprobably. Call X this random number. Is there a good strategy

for guessing if the number shown to you is the largest of the two? Of course, we would like to have a probability of success strictly larger than $\frac{1}{2}$.

Perhaps surprisingly, there is such a strategy, that we now describe. Select at random uniformly on $\{1, 2, \ldots, 10,000\}$ a number Y. If $X \geq Y$, say that X is the largest ($= a$), otherwise say that it is the smallest.

Let us compute the probability P_E of a wrong guess. An error occurs when either (i) $X \geq Y$ and $X = b$, or (ii) $X < Y$ and $X = a$. These events are exclusive of one another, and therefore

$$\begin{aligned} P_E &= P(X \geq Y, X = b) + P(X < Y, X = a) \\ &= P(b \geq Y, X = b) + P(a < Y, X = a) \\ &= P(b \geq Y)P(X = b) + P(a < Y)P(X = a) \\ &= P(b \geq Y)\frac{1}{2} + P(a < Y)\frac{1}{2} = \frac{1}{2}(P(b \geq Y) + P(a < Y)) \\ &= \frac{1}{2}(1 - P(Y \in [b+1, a]) = \frac{1}{2}\left(1 - \frac{a-b}{10,000}\right) < \frac{1}{2}\,. \end{aligned}$$

EXAMPLE 2.1.10: HEADS AND TAILS, TAKE 5. The number of occurrences of heads in n tosses is $S_n = X_1 + \cdots + X_n$. This random variable is the fortune at time n of a gambler systematically betting on heads. It takes the integer values from 0 to n. We have

$$P(S_n = k) = \tfrac{1}{2^n}\tbinom{n}{k}\,.$$

Proof. The event $\{S_n = k\}$ is "k among $X_1, \ldots, X_n$ are equal to 1." There are $\binom{n}{k}$ distinct ways of assigning k values 1 and $n-k$ values 0 to $X_1, \ldots, X_n$, and all have the same probability 2^{-n}. □

One sometimes needs to prove that a random variable X taking its values in $\overline{\mathbb{N}}$ (the value ∞ is *a priori* possible) is in fact almost surely finite, that is, one must prove that $P(X = \infty) = 0$ or, equivalently, $P(X < \infty) = 1$. Since $\{X < \infty\} = \sum_{n=0}^{\infty}\{X = n\}$, we have $P(X < \infty) = \sum_{n=0}^{\infty} P(X = n)$.

Remark 2.1.11 We seize this opportunity to recall that in an expression such as $\sum_{n=0}^{\infty}$, the sum is over $\mathbb{N}$ and does not include ∞ as the notation seems to suggest. A less ambiguous notation would be $\sum_{n\in\mathbb{N}}$. In case we want to sum over all integers plus ∞, we shall *always* use the notation $\sum_{n\in\overline{\mathbb{N}}}$.

The following result is highlighted as a theorem for the purpose of future reference:

Theorem 2.1.12 *Let X be an integer-valued random variable (in particular, the probability that $X = \infty$ is null). Then*

$$\lim_{n\uparrow\infty} P(X > n) = 0\,.$$

Proof. This follows by monotone sequential continuity since the sequence $\{X > n\}$, $n \geq 0$, is non-increasing and $\cap_{n\geq 0}\{X > n\} = \varnothing$ since X takes only finite values. □

Almost Surely, take 2

An expression like "$X = Y$ P-almost surely" means that $P(\{\omega \in \Omega;\, X(\omega) = Y(\omega)\}) = 1$. One interprets similarly expressions such as "$f(X) = 0$ P-almost surely" and so on.

2.1.2 Independent Random Variables

Definition 2.1.13 *Two discrete random variables X and Y are called independent if for all $i, j \in E$,*

$$P(X = i, Y = j) = P(X = i)P(Y = j)\,. \tag{2.1}$$

The extension of the definition to a finite number of random variables is natural:

Definition 2.1.14 *The discrete random variables $X_1, \ldots, X_k$ taking their values in $E_1, \ldots, E_k$ respectively are said to be independent if for all $i_1 \in E_1, \ldots, i_k \in E_k$,*

$$P(X_1 = i_1, \ldots, X_k = i_k) = P(X_1 = i_1) \cdots P(X_k = i_k)\,. \tag{2.2}$$

Theorem 2.1.15 *Let $X_1, \ldots, X_k$ be as in Definition 2.1.14. Then, for any $g_i : E_i \to \mathbb{R}$ $(1 \leq i \leq n)$, the random variables $g_i(X_i)$ $(1 \leq i \leq n)$ are independent.*

Proof. We do the proof in the case $n = 2$:

$$\begin{aligned} P(g_1(X_1) = j_1, g_2(X_2) = j_2) &= \sum_{i_1; g_1(i_1)=j_1} \sum_{i_2; g_1(i_2)=j_2} P(X_1 = i_1, X_2 = i_2) \\ &= \sum_{i_1; g_1(i_1)=j_1} \sum_{i_2; g_1(i_2)=j_2} P(X_1 = i_1)P(X_2 = i_2) \\ &= \left(\sum_{i_1; g_1(i_1)=j_1} P(X_1 = i_1) \right) \left(\sum_{i_2; g_1(i_2)=j_2} P(X_2 = i_2) \right) \\ &= P(g_1(X_1) = j_1)P(g_2(X_2) = j_2)\,. \end{aligned}$$

□

Definition 2.1.16 *A sequence $\{X_n\}_{n\geq 1}$ of discrete random variables indexed by the set of positive integers and taking their values in the sets $\{E_n\}_{n\geq 1}$ respectively is called independent if for all $n \geq 2$, the random variables $X_1, \ldots, X_n$ are independent. If in addition $E_n \equiv E$ for all $n \geq 1$ and the distribution of X_n does not depend on n, the sequence $\{X_n\}_{n\geq 1}$ is said to be* IID *(independent and identically distributed).*

EXAMPLE 2.1.17: HEADS AND TAILS, TAKE 6. We show that the sequence $\{X_n\}_{n\geq 1}$ is IID. Therefore, we have a model for *independent* tosses of an *unbiased* coin.

Proof. Event $\{X_k = a_k\}$ is the direct sum of events $\{X_1 = a_1, \ldots, X_{k-1} = a_{k-1}, X_k = a_k\}$ for all possible values of $(a_1, \ldots, a_{k-1})$. Since there are 2^{k-1} such values and each one has probability 2^{-k}, we have $P(X_k = a_k) = 2^{k-1}2^{-k}$, that is,

$$P(X_k = 1) = P(X_k = 0) = \frac{1}{2}.$$

Therefore,

$$P(X_1 = a_1, \ldots, X_k = a_k) = P(X_1 = a_1) \cdots P(X_k = a_k)$$

for all $a_1, \ldots, a_k \in \{0, 1\}$, from which it follows by definition that $X_1, \ldots, X_k$ are independent random variables, and more generally that $\{X_n\}_{n\geq 1}$ is a family of independent random variables. □

Definition 2.1.18 *Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be sequences of discrete random variables indexed by the positive integers and taking their values in the sets $\{E_n\}_{n\geq 1}$ and $\{F_n\}_{n\geq 1}$ respectively. They are said to be independent if for any finite collection of random variables $X_{i_1}, \ldots, X_{i_r}$ and $Y_{j_1}, \ldots, Y_{i_s}$ extracted from their respective sequences, the discrete random variables $(X_{i_1}, \ldots, X_{i_r})$ and $(Y_{j_1}, \ldots, Y_{i_s})$ are independent.*

(This means that

$$\begin{aligned} &P\left((\cap_{\ell=1}^{r}\{X_{i_\ell} = a_\ell\}) \cap (\cap_{m=1}^{s}\{Y_{j_m} = b_m\})\right) \\ &= P\left(\cap_{\ell=1}^{r}\{X_{i_\ell} = a_\ell\}\right) P\left(\cap_{m=1}^{s}\{Y_{j_m} = b_m\}\right) \end{aligned} \tag{2.3}$$

for all $a_1 \in E_1, \ldots, a_r \in E_r$, $b_1 \in F_1, \ldots, b_s \in F_s$.)

The notion of conditional independence for events (Definition 1.3.14) extends naturally to discrete random variables.

Definition 2.1.19 *Let X, Y, Z be random variables taking their values in the denumerable sets E, F, G, respectively. One says that X and Y are conditionally independent given Z if for all x, y, z in E, F, G, respectively, events $\{X = x\}$ and $\{Y = y\}$ are conditionally independent given $\{Z = z\}$.*

2.1.3 Expectation

Definition 2.1.20 *Let X be a discrete random variable taking its values in the countable set E and let $g : E \to \mathbb{R}$ be a function that is either non-negative or such that*

$$\sum_{x\in E} |g(x)|P(X = x) < \infty\,. \tag{2.4}$$

Then one defines $E[g(X)]$, the expectation of $g(X)$, by the formula

$$E[g(X)] = \sum_{x\in E} g(x)P(X = x)\,. \tag{2.5}$$

If the summability condition (2.4) is satisfied, we say that the random variable $g(X)$ is integrable, and in this case the expectation $E[g(X)]$ is a *finite* number. If it is only assumed that g is non-negative, the expectation may well be infinite.

EXAMPLE 2.1.21: HEADS AND TAILS, TAKE 7. Consider the random variable $S_n = X_1 + \cdots + X_n$ with values in $\{0, 1, \ldots, n\}$. Its expectation is $E[S_n] = n/2$. In fact,

$$\begin{aligned} E[S_n] &= \sum_{k=0}^{n} kP(S_n = k) = \frac{1}{2^n}\sum_{k=1}^{n} k\frac{n!}{k!(n-k)!} \\ &= \frac{n}{2^n}\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \\ &= \frac{n}{2^n}\sum_{j=0}^{n-1}\frac{(n-1)!}{j!(n-1-j)!} = \frac{n}{2^n}2^{n-1}. \end{aligned}$$

EXAMPLE 2.1.22: FINITE RANDOM VARIABLES WITH INFINITE EXPECTATIONS. It is important to realize that a discrete random variable taking *finite values* may have an *infinite expectation*. The canonical example is the random variable X with values in $E = \mathbb{N}_+$ and such that

$$P(X = n) = \frac{1}{cn^2} \qquad (n \in \mathbb{N}_+)$$

where the constant c is chosen such that X actually takes its values in $\mathbb{N}$:

$$P(X < \infty) = \sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty}\frac{1}{cn^2} = 1$$

(therefore $c = \sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$). In fact, the expectation of X is

$$E[X] = \sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} n\frac{1}{cn^2} = \sum_{n=1}^{\infty}\frac{1}{cn} = \infty.$$

Remark 2.1.23 Although the above example is artificial, there are many natural occurences of the phenomenon. Consider for instance Example 2.1.21, and let T be the first integer n such that $2S_n - n = 0$. Then, as it turns out, and as we shall prove in Subsection 8.1.1 that T is a *finite* random variable with *infinite expectation*. Note that the quantity $2S_n - n$ is the fortune at time n of a gambler systematically betting one *euro* on heads.

The telescope formula below gives an alternative way of computing the expectation of an integer-valued random variable.

Theorem 2.1.24 *For a random variable X taking its values in $\mathbb{N}$,*

$$E[X] = \sum_{n=1}^{\infty} P(X \geq n)\,.$$

Proof.

$$\begin{aligned} E[X] = P(X=1) &+ 2P(X=2) + 3P(X=3) + \dots \\ = P(X=1) \; &+ P(X=2) + P(X=3) + \dots \\ &+ P(X=2) + P(X=3) + \dots \\ &+ P(X=3) + \dots \end{aligned}$$

□

We now list a few elementary properties of expectation.

Theorem 2.1.25 *Let A be some event. The expectation of the indicator random variable $X = 1_A$ is*

$$E[1_A] = P(A)\,. \tag{2.6}$$

Proof. $X = 1_A$ takes the value 1 with probability $P(X = 1) = P(A)$ and the value 0 with probability $P(X = 0) = P(\overline{A}) = 1 - P(A)$. Therefore,

$$E[X] = 0 \times P(X = 0) + 1 \times P(X = 1) = P(X = 1) = P(A).$$

□

Theorem 2.1.26 *Let g_1 and g_2 be functions from E to $\overline{\mathbb{R}}$ such that $g_1(X)$ and $g_2(X)$ are integrable (resp., non-negative), and let $\lambda_1, \lambda_2 \in \mathbb{R}$ (resp., $\in \mathbb{R}_+$). Expectation is linear, that is,*

$$E[\lambda_1 g_1(X) + \lambda_2 g_2(X)] = \lambda_1 E[g_1(X)] + \lambda_2 E[g_2(X)]\,. \tag{2.7}$$

Also, expectation is monotone, in the sense that if $g_1(x) \leq g_2(x)$ for all x such that $P(X = x) > 0$ (in other words, $g_1(X) \leq g_2(X)$ almost surely)

$$E[g_1(X)] \leq E[g_2(X)]\,. \tag{2.8}$$

Also, we have the triangle inequality

$$|E[g(X)]| \leq E[|g(X)|]\,. \tag{2.9}$$

Proof. These properties follow from the corresponding properties of series. □

Theorem 2.1.27 *Let X be a random variable with values in E and let $g : E \to \overline{\mathbb{R}}_+$ be a non-negative function.*

(a) If $E[g(X)] = 0$, then $g(X) = 0$ P-almost surely.

(b) If $E[g(X)] < \infty$, then $g(X) < \infty$ P-almost surely.

Proof. (a) Condition $E[g(X)] = 0$ reads $\sum_{x\in E} g(x)P(X = x) = 0$. In particular $P(X = x) = 0$ whenever $g(x) > 0$. Therefore

$$P(g(X) > 0) = \sum_{x\in E;\, g(x)>0} P(X = x) = 0$$

or, equivalently, $P(g(X) = 0) = 1$.

(b) Condition $E[g(X)] < \infty$ reads $\sum_{x\in E} g(x)P(X = x) < \infty$. In particular $P(X = x) = 0$ whenever $g(x) = \infty$. Therefore

$$P(g(X) = \infty) = \sum_{x\in E;\, g(x)=\infty} P(X = x) = 0$$

or, equivalently, $P(g(X) < \infty) = 1$. □

Product Formula for Expectations

Theorem 2.1.28 *Let Y and Z be two independent random variables with values in the (denumerable) sets F and G respectively, and let $v : F \to \overline{\mathbb{R}}$, $w : G \to \overline{\mathbb{R}}$ be functions that are either non-negative, or such that $v(Y)$ and $w(Z)$ are both integrable. Then*

$$E[v(Y)w(Z)] = E[v(Y)]E[w(Z)]\,.$$

Proof. Consider the discrete random variable X with values in $E = F \times G$ defined by $X = (Y, Z)$, and consider the function $g : E \to \overline{\mathbb{R}}$ defined by $g(x) = v(y)w(z)$ where $x = (y, z)$. Under the above stated conditions, we have

$$\begin{aligned}
E[v(Y)w(Z)] = E[g(X)] &= \sum_{x\in E} g(x)P(X = x) \\
&= \sum_{y\in F}\sum_{z\in F} v(y)w(z)P(Y = y, Z = z) \\
&= \sum_{y\in F}\sum_{z\in F} v(y)w(z)P(Y = y)P(Z = z) \\
&= \left(\sum_{y\in F} v(y)P(Y = y)\right)\left(\sum_{z\in F} w(z)P(Z = z)\right) \\
&= E[v(Y)]E[w(Z)].
\end{aligned}$$

□

Mean, Variance and Covariance

Definition 2.1.29 *Let X be an integrable random variable. In this case, we define its mean as the (finite) number*

$$\mu = E[X]\,.$$

Let X be a square-integrable random variable. We then define its variance σ^2 by

$$\sigma^2 = E[(X-\mu)^2]\,.$$

(In the case of integer-valued random variables, the mean and variance, when they are well-defined, are therefore given by the following sums:

$$\mu = \sum_{n=0}^{+\infty} nP(X=n) \qquad \sigma^2 = \sum_{n=0}^{+\infty} (n-\mu)^2 P(X=n)\,.)$$

The variance is also denoted by $\mathrm{Var}\,(X)$. From the linearity of expectation, it follows that $E[(X-m)^2] = E[X^2] - 2mE[X] + m^2$, that is,

$$\mathrm{Var}\,(X) = E[X^2] - m^2\,.$$

The mean is the "center of inertia" of a random variable. More precisely,

Theorem 2.1.30 *Let X be a real integrable random variable with mean m and finite variance σ^2. Then, for all $a \in \mathbb{R}$, $a \neq \mu$,*

$$E[(X-a)^2] > E[(X-\mu)^2] = \sigma^2\,.$$

Proof.

$$\begin{aligned} E\left[(X-a)^2\right] &= E\left[((X-\mu)+(\mu-a))^2\right] \\ &= E\left[(X-\mu)^2\right] + (\mu-a)^2 + 2(\mu-a)E\left[(X-\mu)\right] \\ &= E\left[(X-\mu)^2\right] + (\mu-a)^2 > E\left[(X-\mu)^2\right] \end{aligned}$$

whenever $a \neq \mu$. □

The following consequence of the product rule is extremely important. It says that for *independent* random variables, variances add up.

Theorem 2.1.31 *Let $X_1, \ldots, X_n$ be independent square-integrable random variables. Then*

$$\sigma^2_{X_1+\cdots+X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n}\,.$$

Proof. Let $\mu_1, \ldots, \mu_n$ be the respective means of $X_1, \ldots, X_n$. The mean of the sum $X := X_1 + \cdots + X_n$ is $\mu := \mu_1 + \cdots + \mu_n$. If $i \neq k$, we have, by the product formula for expectations,

$$E\left[(X_i - \mu_i)(X_k - \mu_k)\right] = E\left[(X_i - \mu_i)\right] E\left[(X_k - \mu_k)\right] = 0.$$

Therefore

$$\begin{aligned}
\text{Var}\,(X) =& E\left[(X-\mu)^2\right] = E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right] \\
=& E\left[\sum_{i=1}^{n}\sum_{k=1}^{n}(X_i - \mu_i)(X_k - \mu_k)\right] \\
=& \sum_{i=1}^{n}\sum_{k=1}^{n} E\left[(X_i - \mu_i)(X_k - \mu_k)\right] \\
=& \sum_{i=1}^{n} E\left[(X_i - \mu_i)^2\right] = \sum_{i=1}^{n} \text{Var}\,(X_i).
\end{aligned}$$

□

Note that means always add up, even when the random variables are not independent.

Let X be an integrable random variable. Then, clearly, for any $a \in \mathbb{R}$, aX is integrable and its variance is given by the formula

$$\text{Var}\,(aX) = a^2\,\text{Var}\,(X)\,.$$

Example 2.1.32: Variance of the empirical mean. From this remark and Theorem 2.1.31, it immediately follows that if $X_1, \ldots, X_n$ are independent and identically distributed *integrable* random variables with values in $\mathbb{N}$ with common variance σ^2, then

$$\text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}\,.$$

2.1.4 Famous Distributions

A random variable X taking its values in $\{0, 1\}$ with distribution given by

$$P(X = 1) = p,$$

where $p \in (0, 1)$, is called a Bernoulli random variable with parameter p. This is denoted

$$X \sim Bern(p)\,.$$

Consider the following heads and tails framework which consists of an IID sequence $\{X_n\}_{n\geq 1}$ of Bernoulli variables with parameter p. It is called a Bernoulli sequence with parameter p.

Since $P(X_j = a_j) = p$ or $1-p$ depending on whether $a_i = 1$ or 0, and since there are exactly $h(a) := \sum_{j=1}^{k} a_j$ coordinates of $a = (a_1, \ldots, a_k)$ equal to 1,

$$P(X_1 = a_1, \ldots, X_k = a_k) = p^{h(a)} q^{k-h(a)} ,$$

where $q := 1 - p$. (The integer $h(a)$ is called the Hamming weight of the binary vector a.) Comparing with Examples 1.1.3 and 1.2.3, we see that we have a probabilistic model of a game of heads and tails, with a biased coin when $p \neq \frac{1}{2}$.

The heads and tails framework gives rise to two famous discrete random variables: the binomial random variable, and the geometric random variable.

The Binomial Distribution

Definition 2.1.33 *A random variable X taking its values in the set $E = \{0, 1, \ldots, n\}$ and with the distribution*

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

is called a binomial random variable of size n and parameter $p \in (0, 1)$.

This is denoted

$$X \sim \mathcal{B}(n, p) .$$

EXAMPLE 2.1.34: We place ourselves in the heads and tails framework. Define

$$S_n = X_1 + \cdots + X_n .$$

This random variable takes the values $0, 1, \ldots, n$. To obtain $S_n = i$ where $0 \leq i \leq n$, one must have $X_1 = a_1, \ldots, X_n = a_n$ with $\sum_{j=1}^{n} a_j = i$. There are $\binom{n}{i}$ distinct ways of having this, each one occuring with probability $p^i(1-p)^{n-i}$. Therefore, for $0 \leq i \leq n$,

$$P(S_n = i) = \binom{n}{i} p^i (1-p)^{n-i}.$$

Theorem 2.1.35 *The mean and the variance of a binomial random variable X of size n and parameter p are given by*

$$E[X] = np ,$$

$$\text{Var}\,(X) = np(1-p) .$$

Proof. This can be proven by a direct computation. Later on, in the Exercises section, you will prove this using generating functions. Another approach is to start from the random variable S_n of Example 2.1.34. This is a binomial random variable. We have

$$E[S_n] = \sum_{i=1}^{n} E[X_i] = nE[X_1]$$

and, since the X_i's are IID,

$$V(S_n) = \sum_{i=1}^{n} V(X_i) = nV(X_1).$$

Now,

$$E[X_1] = 0 \times P(X_1 = 0) + 1 \times P(X_1 = 1) = P(X_1 = 1) = p$$

and, since $X_1^2 = X_1$,

$$E\left[X_1^2\right] = E[X_1] = p.$$

Therefore

$$V(X_1) = E\left[X_1^2\right] - E[X_1]^2 = p - p^2 = p(1-p).$$

□

The following inequalities concerning the binomial coefficients are useful:

Theorem 2.1.36 *Let $p \in (0,1)$ and $H_2(p) := -p\log_2 p - q\log_2 q$, where $q := 1-p$. Then for $0 < p \leq \frac{1}{2}$,*

$$\binom{n}{\lfloor np \rfloor} \leq 2^{nH_2(p)}. \tag{2.10}$$

For $\frac{1}{2} \leq p < 1$,

$$\binom{n}{\lceil np \rceil} \leq 2^{nH_2(p)}. \tag{$\star$}$$

For $\frac{1}{2} \leq p < 1$,

$$\frac{2^{nH_2(p)}}{n+1} \leq \binom{n}{\lfloor np \rfloor}. \tag{2.11}$$

For $0 < p \leq \frac{1}{2}$,

$$\frac{2^{nH_2(p)}}{n+1} \leq \binom{n}{\lceil np \rceil}. \tag{$\dagger$}$$

The proof uses the following lemma.

Lemma 2.1.37 *Let n be an integer and let $p \in (0,1)$ be such that np is an integer. Then*

$$\frac{2^{nH_2(p)}}{n+1} \leq \binom{n}{np} \leq 2^{nH_2(p)}.$$

Proof. The inequality

$$\binom{n}{np} p^{np}(1-p)^{n(1-p)} \leq 1$$

follows from the fact that the left-hand side is a probability, namely $P(\mathcal{B}(n,p) = np)$. Therefore

$$\binom{n}{np} \leq p^{-np}(1-p)^{-n(1-p)} = 2^{nH_2(p)}\,.$$

The integer value $k = np$ will be shown to maximize $\binom{n}{k}$ among all integers k such that $0 \leq k \leq n$. Therefore

$$\begin{aligned} 1 = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} &\leq (n+1)\binom{n}{np} p^{np}(1-p)^{n(1-p)} \\ &= (n+1)\binom{n}{np} 2^{-nH_2(p)}\,. \end{aligned}$$

To prove that $k = np$ maximizes $\binom{n}{k}$, compare two adjacent terms. We have

$$\begin{aligned} &\binom{n}{k} p^k (1-p)^{n-k} - \binom{n}{k+1} p^{k+1}(1-p)^{n-k-1} \\ &\qquad = \binom{n}{k} p^k (1-p)^{n-k} \left(1 - \frac{p(n-k)}{(1-p)(k+1)}\right)\,. \end{aligned}$$

This difference is non-negative if and only if

$$1 - \frac{p(n-k)}{(1-p)(k+1)} \geq 0$$

or, equivalently, $k \geq pn - (1-p)$. This shows that the function $k \to \binom{n}{k}$ increases as k varies from 0 to pn and decreases afterwards. □

We now proceed to the proof of Theorem 2.1.36:

Proof. Proof of (2.10):

$$\begin{aligned} \binom{n}{\lfloor np \rfloor} p^{pn}(1-p)^{(1-p)n} &\leq \binom{n}{\lfloor np \rfloor} p^{\lfloor np \rfloor}(1-p)^{n-\lfloor np \rfloor} \\ &\leq \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1\,. \end{aligned}$$

Inequality $(\star)$ is proved in a similar way, or by an obvious symmetry argument. Inequality (2.11) follows from Lemma 2.1.37, since

$$\binom{n}{\lfloor np \rfloor} \geq \frac{2^{nH_2(\lfloor np \rfloor / n)}}{n+1} \geq \frac{2^{nH_2(p)}}{n+1}\,.$$

The proof of (†) is proved in a similar way, or by a symmetry argument. □

The Geometric Distribution

Definition 2.1.38 *A random variable X taking its values in $\mathbb{N}_+ := \{1, 2, \dots, \}$ and with the distribution*

$$P(T = k) = (1-p)^{k-1}p\,, \tag{2.12}$$

where $0 < p < 1$, is called a geometric random variable with parameter p.

This is denoted

$$X \sim \mathcal{G}\text{eo}(p)\,.$$

Of course, if $p = 1$, $P(T = 1) = 1$, and if $p = 0$, $P(T = \infty) = 1$. If $0 < p < 1$,

$$P(T < \infty) = \sum_{n=1}^{\infty}(1-p)^{k-1}p = \frac{1}{1-(1-p)} = \frac{p}{p} = 1\,,$$

and therefore $P(T = \infty) = 0$.

EXAMPLE 2.1.39: FIRST "HEADS" IN THE SEQUENCE. We are in the heads and tails framework. Define the random variable T to be the first time of occurrence of 1 in the sequence $X_1, X_2, \dots$, that is,

$$T = \inf\{n \geq 1; X_n = 1\},$$

with the convention that if $X_n = 0$ for all $n \geq 1$, then $T = \infty$. The event $\{T = k\}$ is exactly $\{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}$, and therefore,

$$P(T = k) = P(X_1 = 0) \cdots P(X_{k-1} = 0)P(X_k = 1),$$

that is, for $k \geq 1$,

$$P(T = k) = (1-p)^{k-1}p\,.$$

Theorem 2.1.40 *The mean of a geometric random variable X with parameter $p > 0$ is*

$$E[X] = \tfrac{1}{p}\,.$$

Proof.

$$E\,[X] = \sum_{k=1}^{\infty} k\,(1-p)^{k-1}\,p = \frac{1}{p^2} \times p = \frac{1}{p}\,.$$

□

Theorem 2.1.41 *A geometric random variable T with parameter $p \in (0,1)$ is memoryless in the sense that for any integers $k, k_0 \geq 1$, we have $P(T = k+k_0 \,|\, T > k_0) = P(T = k)$.*

Proof.

$$P(T > k_0) = \sum_{k=k_0+1}^{\infty} (1-p)^{k-1} p = (1-p)^{k_0}$$

and therefore

$$P(T = k_0 + k | T > k_0) = \frac{P(T = k_0 + k, T > k_0)}{P(T > k_0)} = \frac{P(T = k_0 + k)}{P(T > k_0)}$$
$$= \frac{p(1-p)^{k+k_0-1}}{(1-p)^{k_0}} = p(1-p)^k = P(T = k).$$

□

EXAMPLE 2.1.42: THE COUPON COLLECTOR, TAKE 1. In a certain brand of chocolate tablets one can find coupons, one in each tablet, randomly and independently chosen among n types. A prize may be claimed once the chocolate amateur has gathered a collection containing all the types of coupons. We seek to compute the average value of the number X of chocolate tablets bought when this happens for the first time.

For $0 \leq i \leq n-1$, let X_i be the number of tablets it takes after ($>$) i different types of coupons have been collected to find a new type of coupon (in particular, $X_0 = 1$), so that

$$X = \sum_{i=0}^{n-1} X_i,$$

where each X_i ($1 \leq i \leq n-1$) is a geometric random variable with parameter $p_i = 1 - \frac{i}{n}$. In particular,

$$E[X_i] = \frac{1}{p_i} = \frac{n}{n-i},$$

(still true for $i = 0$) and therefore

$$E[X] = \sum_{i=0}^{n-1} E[X_i] = n \sum_{i=0}^{n-1} \frac{1}{n-i} = n \sum_{i=1}^{n} \frac{1}{i}.$$

The sum $H(n) := \sum_{i=1}^{n} \frac{1}{i}$ (called the n-th harmonic number) satisfies the inequality

$$\log n \leq H(n) \leq \log n + 1, \tag{2.13}$$

as can be seen by expressing $\log n$ as the integral $\int_1^n \frac{1}{x}\,dx$, partitioning the domain of integration with segments of unit length, and using the fact that the integrand is a decreasing function, which gives the inequalities

$$\sum_{i=2}^{n} \frac{1}{i} \leq \int_1^n \frac{dx}{x} \leq \sum_{i=1}^{n-1} \frac{1}{i}.$$

Therefore,

$$E[X] = (1 + o(1)) n \log n,$$

where $o(1)$ is a symbolic representation of a function of the positive integers that tend to 0 as $n \uparrow \infty$ (Landau's notation; see Section A.5).

The Hypergeometric Distribution

Recall Example 1.2.17. There is an urn containing N_1 black balls and N_2 red balls. You draw successively without replacement and at random n balls from the urn ($n \leq N_1 + N_2$). The probability of having drawn k black balls ($0 \leq k \leq \inf(N_1, n)$) is:

$$p_k = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N_1+N_2}{n}}.$$

This probability distribution is called the hypergeometric distribution of parameters N_1 and N_2.

The Poisson Distribution

Definition 2.1.43 *A random variable X taking its values in $\mathbb{N}$ and such that for all $k \geq 0$,*

$$P(X = k) = e^{-\theta}\frac{\theta^k}{k!},$$

is called a Poisson random variable with parameter $\theta \geq 0$.

This is denoted by

$$X \sim \mathcal{P}\mathrm{oi}(\theta).$$

If $\theta = 0$, $X \equiv 0$ (the general formula applies if one uses the convention $0! = 1$).

EXAMPLE 2.1.44: THE POISSON LAW OF RARE EVENTS, TAKE 1. A veterinary surgeon in the Prussian cavalry once gathered data concerning the accidents due to horse kickbacks among soldiers. He deduced that the (random) number of accidents of the kind had a Poisson distribution. Here is an explanation.

Suppose that you play "heads and tails" for a large number n of (independent) tosses using a coin such that

$$P(X_i = 1) = \frac{\alpha}{n}.$$

In the Prussian army example, n is the (large) number of soldiers, and $X_i = 1$ if the i-th soldier has been hurt by a horse. Let S_n be the total number of heads (of wounded soldiers). We show that

$$\lim_{n\uparrow\infty} P(S_n = k) = e^{-\alpha}\frac{\alpha^k}{k!}, \qquad (\star)$$

and this explains the findings of the veterinary surgeon. (The average number of casualties is α and the choice $P(X_i = 1) = \frac{\alpha}{n}$ guarantees this. Letting $n \uparrow \infty$ accounts for n being large but unknown.) Here is the proof of the mathematical statement.

The random variable S_n follows a binomial law with mean $n \times \frac{\alpha}{n} = \alpha$:

$$P(S_n = k) = \binom{n}{k}\left(\frac{\alpha}{n}\right)^k\left(1 - \frac{\alpha}{n}\right)^{n-k}.$$

In particular $P(S_n = 0) = \left(1 - \frac{\alpha}{n}\right)^n \to e^{-\alpha}$ as $n \uparrow \infty$. Also,

$$\frac{P(S_n = k+1)}{P(S_n = k)} = \frac{\frac{n-k}{k+1}\frac{\alpha}{n}}{1 - \frac{\alpha}{n}}$$

tends to $\frac{\alpha}{k+1}$ as $n \uparrow \infty$, from which $(\star)$ follows.

Theorem 2.1.45 *The mean of a Poisson random variable with parameter θ is given by*

$$E[X] = \theta\,,$$

and its variance is

$$\text{Var}\,(X) = \theta\,.$$

Proof.

$$E\,[X] = e^{-\theta}\sum_{k=1}^{\infty}\frac{\theta^k}{k!}k = e^{-\theta}\theta\sum_{j=0}^{\infty}\frac{\theta^j}{j!} = e^{-\theta}\theta e^{\theta} = \theta$$

and

$$\begin{aligned} E\left[X^2 - X\right] &= e^{-\theta}\sum_{k=0}^{\infty}\left(k^2 - k\right)\frac{\theta^k}{k!} = e^{-\theta}\sum_{k=2}^{\infty}k\,(k-1)\,\frac{\theta^k}{k!} \\ &= e^{-\theta}\theta^2\sum_{k=2}^{\infty}\frac{\theta^{k-2}}{(k-2)!} = e^{-\theta}\theta^2\sum_{j=0}^{\infty}\frac{\theta^j}{j!} = e^{-\theta}\theta^2 e^{\theta} = \theta^2\,. \end{aligned}$$

Therefore

$$\begin{aligned} \text{Var}\ (X) &= E\left[X^2\right] - E\,[X]^2 \\ &= E\left[X^2 - X\right] + E\,[X] - E\,[X]^2 = \theta^2 + \theta - \theta^2 = \theta. \end{aligned}$$

□

Theorem 2.1.46 *Let X_1 and X_2 be two independent Poisson random variables with means $\theta_1 > 0$ and $\theta_2 > 0$, respectively. Then $X = X_1 + X_2$ is a Poisson random variable with mean $\theta = \theta_1 + \theta_2$.*

Proof. For $k \geq 0$,

$$\begin{aligned} P(X = k) = P(X_1 + X_2 = k) &= P\left(\sum_{i=0}^{k}\{X_1 = i, X_2 = k - i\}\right) \\ &= \sum_{i=0}^{k}P(X_1 = i, X_2 = k - i) = \sum_{i=0}^{k}P(X_1 = i)P(X_2 = k - i) \\ &= \sum_{i=0}^{k}e^{-\theta_1}\frac{\theta_1^i}{i!}e^{-\theta_2}\frac{\theta_2^{k-i}}{(k-i)!} = e^{-(\theta_1+\theta_2)}\frac{(\theta_1+\theta_2)^k}{k!}, \end{aligned}$$

where we used the binomial formula. □

The Multinomial Distribution

Consider the random vector $X = (X_1, \ldots, X_N)$ where all the random variables X_i take their values in the *same* (this restriction is not essential, but it simplifies the notation) denumerable space E. Let $p : E^N \to \mathbb{R}_+$ be a function such that

$$\sum_{x \in E^N} p(x) = 1\,.$$

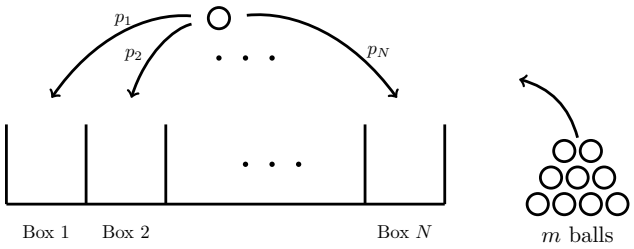
Definition 2.1.47 *The discrete random vector X above is said to admit the probability distribution p if for all sets $C \subseteq E^N$,*

$$P(X \in C) = \sum_{x \in C} p(x)\,.$$

In fact, there is nothing new here since X is a discrete random variable taking its values in the denumerable set $\mathcal{X} := E^N$.

Consider m balls to be placed in N boxes $B_1, \ldots, B_N$ independently of one another, with the probability p_i for a given ball to be assigned to box B_i. Of course,

$$\textstyle\sum_{i=1}^N p_i = 1\,.$$



After placing all the balls in the boxes, there are X_i balls in box B_i, where

$$\textstyle\sum_{i=1}^N X_i = m\,.$$

The random vector $X = (X_1, \ldots, X_N)$ is a multinomial vector of size (N, m) and parameters $p_1, \ldots, p_N$, that is, its probability distribution is

$$P(X_1 = m_1, \ldots, X_N = m_N) = \tfrac{k!}{\prod_{i=1}^N (m_i)!} \prod_{i=1}^N p_i^{m_i}\,,$$

where $m_1 + \cdots + m_N = m$.

Proof. Observe that (α): there are $m! / \prod_{i=1}^N (m_i)!$ distinct ways of placing m balls in N boxes in such a manner that m_1 balls are in box B_1, m_2 are in B_2, etc., and (β): each of these distinct ways occurs with the same probabilty $\prod_{i=1}^N p_i^{m_i}$. □

The Uniform Distribution on $[0,1]$

This subsection introduces non-discrete random variables. In fact, it gives just what is strictly necessary in this book, in particular, the notion of independent random numbers.

Definition 2.1.48 *A function* $X : \Omega \to \mathbb{R}$ *such that for all* $x \in \mathbb{R}$

$$\{\omega; X(\omega) \leq x\} \in \mathcal{F}$$

is called a real random variable.

Its cumulative distribution function is the function $F(x) := P(X \leq x)$. If

$$F(x) = \int_{-\infty}^{x} f(y)\, dy\,,$$

for all $x \in \mathbb{R}$ for some non-negative function f such that $\int_{-\infty}^{+\infty} f(y)\, dy = 1$, the latter is called the probability density function, or PDF, of X.

The following example is all we need in this book.

EXAMPLE 2.1.49: THE UNIFORM DISTRIBUTION. Let $[a,b] \in \mathbb{R}$. A real random variable X with the PDF

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x)$$

is called a uniform random variable on $[a,b]$. This is denoted by

$$X \sim \mathcal{U}([a,b])\,.$$

Uniform random variables are used in simulation, more precisely, to generate a discrete random variable Z with a prescribed distribution $P(Z = a_i) = p_i \quad (0 \leq i \leq K)$. The basic principle of the sampling algorithm is the following

Draw $U \sim \mathcal{U}([0,1])$.

Set $Z = a_\ell$ if $U \in I_\ell := (p_0 + p_1 + \ldots + p_{\ell-1}, p_0 + p_1 + \ldots + p_\ell]$.

Indeed, since the interval I_ℓ has length p_l, $P(Z = a_\ell) = P(U \in I_\ell) = p_\ell$.

This method is called the method of the inverse.

Definition 2.1.50 *A real random vector of dimension* d *is a mapping* $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}$ *such that each coordinate* X_i *is a real random variable.*

A non-negative function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}^d} f(x)\, dx = 1$ and

$$P(X_1 \leq x_1, \ldots, X_d \leq x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(x_1, \ldots, x_d)\, dx_1 \cdots dx_d$$

is called the probability distribution function of the random vector X.

Definition 2.1.51 *The real random variables $X_1, \dots, X_d$ admitting the respective* PDF*'s $f_1, \dots, f_d$ are said to be independent if the* PDF *of the random vector $X = (X_1, \dots, X_d)$ is of the form*

$$f(x_1, \dots, x_d) = f_1(x_1) \times \cdots \times f(x_d)$$

where the f_i's are non-negative functions such that $\int_{-\infty}^{+\infty} f_i(y)\, dy = 1$.

The f_i's are then the PDF's of the X_i's. For instance with $i = 1$,

$$\begin{aligned} P(X_1 \le x_1) &= P(X_1 \le x_1, X_2 < \infty \dots, X_d < \infty) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_1(x_1) f_2(x_2) \cdots f_d(x_d)\, dx_1 \cdots dx_d \\ &= \int_{-\infty}^{x_1} f_1(x_1)\, dx_1 \int_{-\infty}^{+\infty} f_2(x_2)\, dx_2 \cdots \int_{-\infty}^{+\infty} f_d(x_d) dx_d = \int_{-\infty}^{x_1} f_1(x_1)\, dx_1 \,. \end{aligned}$$

Definition 2.1.52 *The real random variables $X_1, X_2 \dots$ admitting the respective* PDF*'s $f_1, f_2, \dots$ are said to be independent if for all integers $k \ge 2$, the random variables $X_1, \dots, X_d$ are independent.*

EXAMPLE 2.1.53: SEQUENCE OF INDEPENDENT RANDOM NUMBERS. The sequence $\{U_n\}_{n \ge 1}$ is called a sequence of independent random numbers if for all $k \ge 1$, $U_1, \dots, U_k$ are independent random variables uniformly distributed on the interval $[0, 1]$.

The Gilbert–Erdös–Rényi Random Graphs

A graph is a discrete object and therefore random graphs are, from the purely formal point of view, discrete random variables. The random graphs considered in this book are in fact described by a finite collection of IID $\{0, 1\}$-valued random variables. They will be studied in more detail in Chapter 10. The basic definitions of graph theory below will be complemented as the need arises.

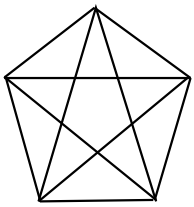
A (finite) graph $(V, \mathcal{E})$ consists of a finite collection V of vertices v and of a collection $\mathcal{E}$ of unordered pairs of distinct vertices, $\langle u, v \rangle$, called the edges. If $\langle u, v \rangle \in \mathcal{E}$, then u and v are called neighbours, and this is also denoted by $u \sim v$. The degree of vertex $v \in V$ is the number of edges stemming from it.

In a few occasions, some redundancy in the notation will be useful: V and $\mathcal{E}$ will be denoted by $V(G)$ and $\mathcal{E}(G)$.

A subgraph (or induced subgraph) of a graph $G = (V, \mathcal{E})$ is any graph $G' = (V', \mathcal{E}')$ with $V' \subseteq V$ and $\mathcal{E}' = \{\langle u, v \rangle \in \mathcal{E};\, u, v \in V'\}$. Such graph is also called the restriction of G to V' and is denoted by $G|_{V'}$.

A complete graph is one having all the possible $\binom{n}{2}$ edges. It will be denoted by K_n and its edge set by $\mathbf{E_n}$. Note that a subgraph of a complete graph is also complete.

A complete subgraph is called a clique of the graph. Note that a singleton of V is a clique.



The complete pythagorean graph

A graph is connected if for all pairs of distinct vertices v, w, there is a sequence $v_0 = v, v_1, \ldots, v_n = w$ (called a path from v to w) and such that $v_0 \sim v_1 \sim \ldots \sim v_n$.

A cycle of a graph is a sequence of distinct vertices $v_1, v_2, \ldots, v_n$ such that $v_1 \sim v_2 \sim \ldots \sim v_n \sim v_1$. A tree is a connected graph without cycles.

Let $G_1 = (V, \mathcal{E}_1)$ and $G_2 = (V, \mathcal{E}_2)$ be two graphs with the same set of vertices. The graph $G = G_1 \cup G_2$ is by definition the graph on the set of vertices V such that $e \in \mathcal{E}$ if and only if $e \in \mathcal{E}_1 \cup \mathcal{E}_2$. This graph is called the union of G_1 and G_2. One defines similarly the intersection of G_1 and G_2, $G = G_1 \cap G_2$, to be the graph on the set of vertices V such that $e \in \mathcal{E}$ if and only if $e \in \mathcal{E}_1 \cap \mathcal{E}_2$. One writes $G_2 \subseteq G_1$ if and only if $\mathcal{E}(G_1) \subseteq \mathcal{E}(G_2)$.

Some graph properties may be difficult to verify on a given graph. However, there exist results showing that they are satisfied (or not) for "large" and "typical" graphs. The question of course is: what is a typical graph? One possible choice is the Gilbert random graph (Definition 2.1.54 below).

Definition 2.1.54 *(Gilbert, 1959) Let n be a fixed positive integer and let $V = \{1, 2, \ldots, n\}$ be a finite set of vertices. To each unordered pair of distinct vertices $\langle u, v \rangle$, associate a random variable $X_{\langle u,v \rangle}$ taking its values in $\{0, 1\}$ and suppose that all such variables are* IID *with probability $p \in (0, 1)$ for the value 1. This defines a random graph denoted by $\mathcal{G}(n, p)$, a random element taking its values in the (finite) set of all graphs with vertices $\{1, 2, \ldots, n\}$ and admitting for edge the unordered pair of vertices $\langle u, v \rangle$ if and only if $X_{\langle u,v \rangle} = 1$.*

Note that $\mathcal{G}(n, p)$ is indeed a discrete random variable (taking its values in the finite set consisting of the collection of graphs with vertex set $V = \{1, 2, \ldots, n\}$). Similarly, the set $\mathcal{E}_{n,p}$ of edges of $\mathcal{G}(n, p)$ is also a discrete random variable. If we call any unordered pair of vertices $\langle u, v \rangle$ a potential edge (there are $\binom{n}{2}$ such edges forming the set $\mathbf{E_n}$), $\mathcal{G}(n, p)$ is constructed by accepting a potential edge as one of its edges with probability p independently of all other potential edges. The probability of occurence of a graph G with exactly m edges is then

$$P(\mathcal{G}(n, p) = G) = P(|\mathcal{E}_{n,p}| = m) = p^m (1 - p)^{\binom{n}{2} - m}.$$

Note that the degree of a given vertex, that is the number of edges stemming from it, is a binomial random variable $\mathcal{B}(n-1,p)$. In particular, the average degree is $d=(n-1)p$.

Another type of random graph is the Erdös–Rényi random graph (Definition 2.1.55 below). It is closely related to the Gilbert graph as we shall see below, in Theorem 2.1.56.

Definition 2.1.55 *(Erdös and Rényi, 1959) Consider the collection* $\mathbf{G_m}$ *of graphs* $G=(V,\mathcal{E})$ *where* $V=\{1,2,\ldots,n\}$ *with exactly* m *edges* ($|\mathcal{E}|=m$). *There are* $\binom{\binom{n}{2}}{m}$ *such graphs. The Erdös–Rényi random graph* $\mathcal{G}_{n,m}$ *is a random graph uniformly distributed on* $\mathbf{G_m}$.

(The notation is chosen for a quick differentiation between Gilbert graphs $\mathcal{G}_{n,m}$ and Erdös–Rényi graphs $\mathcal{G}(n,p)$.)

Denoting by $\mathcal{E}_{n,m}$ the (random) collection of edges of $\mathcal{G}_{n,m}$, the probability of obtaining a given graph $G\in\mathbf{G_m}$ is

$$P(G)=\binom{\binom{n}{2}}{m}^{-1}.$$

The random graph $\mathcal{G}_{n,m}$ can be constructed by including m edges successively at random. More precisely, denoting by G_k $(0\leq k\leq m)$ the successive graphs, and by $\mathcal{E}_k$ the collection of edges of G_k, $G_0=(V,\varnothing)$ and for $1\leq k\leq m$, $\mathcal{E}_k=\mathcal{E}_{k-1}\cup e_k$, where

$$P(e_k=e\mid G_0,\ldots,G_{k-1})=|\mathbf{E_n}\backslash\mathcal{E}_{k-1}|^{-1}$$

for all edges $e\in\mathbf{E_n}\backslash\mathcal{E}_{k-1}$.

Theorem 2.1.56 *The conditional distribution of* $\mathcal{G}(n,p)$ *given that the numner of edges is* $m\leq\binom{n}{2}$ *is uniform on the set* $\mathbf{G_m}$ *of graphs* $G=(V,\mathcal{E})$ *where* $V=\{1,2,\ldots,n\}$ *with exactly* m *edges.*

Proof. Let G be a graph with vertex set V have exactly m edges. Observing that $\{\mathcal{G}(n,p)=G\}\subseteq\{|\mathcal{E}_{n,p}|=m\}$, we have that

$$\begin{aligned}P(\mathcal{G}(n,p)=G\mid|\mathcal{E}_{n,p}|=m)&=\frac{P(\mathcal{G}(n,p)=G,|\mathcal{E}_{n,p}|=m)}{P(|\mathcal{E}_{n,p}|=m)}\\&=\frac{P(\mathcal{G}(n,p)=G)}{P(|\mathcal{E}_{n,p}|=m)}\\&=\frac{p^m(1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m}p^m(1-p)^{\binom{n}{2}-m}}=\binom{\binom{n}{2}}{m}^{-1}.\end{aligned}$$

□

Remark 2.1.57 In the sequel, we will follow the tradition of refering to Gilbert graphs as Erdös–Rényi graphs.

2.2 Generating functions

2.2.1 Definition and Properties

The computation of probabilities in discrete probability models often require an enumeration of all the possible outcomes realizing this particular event. Generating functions are very useful for this task, and more generally, for obtaining the probability distributions of integer-valued random variables. We first define the expectation of a complex-valued function of a random variable.

Let X be a discrete random variable with values in $\mathbb{N}$, and let $\varphi : \mathbb{N} \to \mathbb{C}$ be a complex function with real and imaginary parts φ_R and φ_I respectively. The expectation $E[\varphi(X)]$ is naturally defined by

$$E[\varphi(X)] := E[\varphi_R(X)] + iE[\varphi_I(X)]\,,$$

provided the expectations on the right-hand side are well-defined and finite. This is the case if $E\left[|\varphi(X)|\right] < \infty$.

Definition 2.2.1 *Let X be an integer-valued random variable. Its generating function (GF) is the function $g : \mathcal{D} \to \mathbb{C}$ defined by*

$$g(z) := E[z^X] = \sum_{k=0}^{\infty} P(X = k) z^k\,, \tag{2.14}$$

and where $\mathcal{D} := \overline{D}(0; R) := \{z \in \mathbb{C}; |z| \leq R\}$ is the closed disk of absolute convergence of the above series.

Since $\sum_{n=0}^{\infty} P(X = n) = 1 < \infty$, $R \geq 1$. In the next two examples, $R = \infty$.

EXAMPLE 2.2.2: GF OF THE BINOMIAL VARIABLE. For the binomial random variable of size n and parameter p,

$$g(z) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = \sum_{k=0}^{n} \binom{n}{k} (zp)^k (1-p)^{n-k}\,,$$

and therefore

$$g(z) = (1 - p + pz)^n\,.$$

EXAMPLE 2.2.3: GF OF THE POISSON VARIABLE. For the Poisson random variable of mean θ,

$$g(z) = e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta)^k}{k!} z^k = e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta z)^k}{k!},$$

and therefore

$$g(z) = e^{\theta(z-1)}\,.$$

Next is an example where the radius of convergence is finite.

EXAMPLE 2.2.4: GF OF THE GEOMETRIC VARIABLE. For the geometric random variable of (2.12),

$$g(z) = \sum_{k=0}^{\infty} p(1-p)^{k-1} z^k .$$

The radius of convergence of this generating function power series is $\frac{1}{1-p}$ and its sum is

$$g(z) = \sum_{k=0}^{\infty} pz((1-p)z)^{k-1} = \frac{pz}{1-qz} .$$

Theorem 2.2.5 *The generating function characterizes the distribution of a random variable.*

This means the following. Suppose that, without knowing the distribution of X, you have been able to compute its generating function g, and that, moreover, you are able to give its power series expansion in a neighborhood of the origin[1], say,

$$g(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Since g is the generating function of X,

$$g(z) = \sum_{n=0}^{\infty} P(X = n) z^n$$

and since the power series expansion around the origin is unique, $P(X = n) = a_n$ for all $n \geq 0$. Similarly, if two integer-valued random variables X and Y have the same generating function, they have the same distribution. Indeed, the identity in a neighborhood of the origin of two power series implies the identity of their coefficients.

Theorem 2.2.6 *Let X and Y be two independent integer-valued random variables with respective generating functions g_X and g_Y. Then the sum $X + Y$ has the* GF

$$g_{X+Y}(z) = g_X(z) \times g_Y(z).$$

Proof. Use the product formula for expectations:

$$g_{X+Y}(z) = E\left[z^{X+Y}\right] = E\left[z^X z^Y\right] = E\left[z^X\right] E\left[z^Y\right] .$$

□

EXAMPLE 2.2.7: SUM OF INDEPENDANT POISSON VARIABLES. Let X and Y be two *independent* Poisson random variables with means α and β respectively. The

[1] This is a common situation; see Theorem 2.2.10 for instance.

sum $X + Y$ is a Poisson random variable with mean $\alpha + \beta$. Indeed, by Theorem 2.2.6,

$$g_{X+Y}(z) = g_X(z) \times g_Y(z) = e^{\alpha(z-1)}\, e^{\beta(z-1)} = e^{(\alpha+\beta)(z-1)},$$

and the assertion follows directly from Theorem 2.2.5 since g_{X+Y} is the GF of a Poisson random variable with mean $\alpha + \beta$.

The next result gives concerns the shape of the generating function restricted to the interval $[0, 1]$.

Theorem 2.2.8 *(α) Let $g : [0, 1] \to \mathbb{R}$ be defined by $g(x) = E[x^X]$, where X is a non-negative integer-valued random variable. Then g is nondecreasing and convex. Moreover, if $P(X = 0) < 1$, it is strictly increasing, and if $P(X \leq 1) < 1$, it is strictly convex.*

(β) Suppose $P(X \leq 1) < 1$. If $E[X] \leq 1$, the equation $x = g(x)$ has a unique solution $x \in [0, 1]$, namely $x = 1$. If $E[X] > 1$, it has two solutions in $[0, 1]$, $x = 1$ and $x = x_0 \in (0, 1)$.

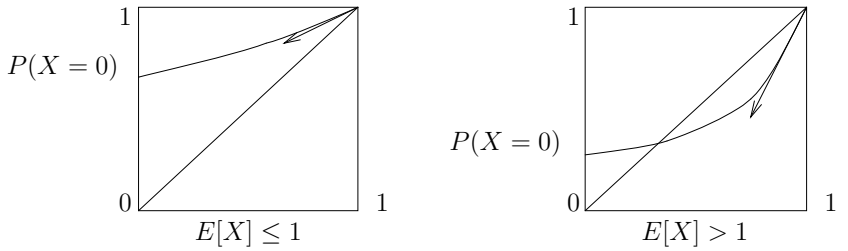
Proof. Just observe that for $x \in [0, 1]$,

$$g'(x) = \sum_{n=1}^{\infty} nP(X = n)x^{n-1} \geq 0,$$

and therefore g is nondecreasing, and

$$g''(x) = \sum_{n=2}^{\infty} n(n-1)P(X - n)x^{n-2} \geq 0,$$

and therefore g is convex. For $g'(x)$ to be null for some $x \in (0, 1)$, it is necessary to have $P(X = n) = 0$ for all $n \geq 1$, and therefore $P(X = 0) = 1$. For $g''(x)$ to be null for some $x \in (0, 1)$, one must have $P(X = n) = 0$ for all $n \geq 2$, and therefore $P(X = 0) + P(X = 1) = 1$.



Two aspects of the generating function

The graph of $g : [0, 1] \to \mathbb{R}$ has, in the strictly increasing strictly convex case $P(X = 0) + P(X = 1) < 1$, the general shape shown in the figure, where we distinguish two cases: $E[X] = g'(1) \leq 1$, and $E[X] = g'(1) > 1$. The rest of the proof is then easy. □

Moments from the Generating Function

Generating functions are powerful computational tools. First of all, they can be used to obtain moments of a discrete random variable.

Theorem 2.2.9 *We have*

$$g'(1) = E[X] \tag{2.15}$$

and

$$g''(1) = E[X(X-1)]. \tag{2.16}$$

Proof. Inside the open disk $D(0;R)$ centered at the origin and of radius R, the power series defining the generating function g is continuous, and differentiable at any order term by term. In particular, differentiating twice both sides of (2.14) inside the open disk $D(0;R)$ gives

$$g'(z) = \sum_{n=1}^{\infty} nP(X=n)z^{n-1}, \tag{2.17}$$

and

$$g''(z) = \sum_{n=2}^{\infty} n(n-1)P(X=n)z^{n-2}. \tag{2.18}$$

When the radius of convergence R is *strictly larger* than 1, we obtain the announced results by letting $z=1$ in the previous identities.

If $R=1$, the same is basically true but the mathematical argument is more subtle. The difficulty is not with the right-hand side of (2.17), which is always well-defined at $z=1$, being equal to $\sum_{n=1}^{\infty} nP(X=n)$, a non-negative and possibly infinite quantity. The difficulty is that g may be not differentiable at $z=1$, a boundary point of the disk (here of radius 1) on which it is defined. However, by *Abel's theorem* (Theorem A.1.3), the limit as the *real* variable x increases to 1 of $\sum_{n=1}^{\infty} nP(X=n)x^{n-1}$ is $\sum_{n=1}^{\infty} nP(X=n)$. Therefore g', as a function on the real interval $[0,1)$, can be extended to $[0,1]$ by (2.15), and this extension preserves continuity. With this *definition* of $g'(1)$, Formula (2.15) holds true. Similarly, when $R=1$, the function g'' defined on $[0,1)$ by (2.18) is extended to a continuous function on $[0,1]$ by *defining* $g''(1)$ by (2.16). □

2.2.2 Random Sums

How to compute the distribution of random sums? Here again, generating functions help.

Theorem 2.2.10 *Let $\{Y_n\}_{n\geq 1}$ be an* IID *sequence of integer-valued random variables with the common generating function g_Y. Let T be another random variable, integer-valued, independent of the sequence $\{Y_n\}_{n\geq 1}$, and let g_T be its generating function. The generating function of*

$$X = \sum_{n=1}^{T} Y_n \,,$$

where by convention $\sum_{n=1}^{0} = 0$, *is*

$$g_X(z) = g_T\left(g_Y(z)\right) \,. \tag{2.19}$$

Proof.

$$\begin{aligned}
\sum_{n\geq 0} z^n P(X=n) &= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(X=n, T=k\right) \right) \\
&= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(\sum_{j=1}^{k} Y_j = n, T=k \right) \right) \\
&= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(\sum_{j=1}^{k} Y_j = n, T=k \right) \right) \\
&= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(\sum_{j=1}^{k} Y_j = n \right) P\left(T=k\right) \right) \\
&= \sum_{k\geq 0} P\left(T=k\right) \left(\sum_{n\geq 0} z^n P\left(\sum_{j=1}^{k} Y_j = n \right) \right) \,.
\end{aligned}$$

But

$$\left(\sum_{n\geq 0} z^n P\left(\sum_{j=1}^{k} Y_j = n \right) \right) = g_{\sum_{j=1}^{k} Y_j}(z) = \left(g_Y(z)\right)^k \,.$$

Therefore,

$$\sum_{n\geq 0} z^n P(X=n) = \sum_{k\geq 0} P\left(T=k\right) \left(g_Y(z)\right)^k = g_T\left(g_Y(z)\right) \,.$$

□

By taking derivatives in (2.19),

$$E\left[X\right] = g_X^{'}(1) = g_Y^{'}(1) g_T^{'}\left(g_Y(1)\right) = E[Y_1]E[T].$$

This is Wald's formula. Exercise 2.4.16 gives more general conditions for its validity.

2.2.3 Counting with Generating Functions

The following example is typical of the use of generating functions in combinatorics (the art of counting).

EXAMPLE 2.2.11: LOTTERY. Let X_1, X_2, X_3, X_4, X_5, and X_6 be independent random variables uniformly distributed over $\{0, 1, \ldots, 9\}$. We shall compute the generating function of $Y = 27 + X_1 + X_2 + X_3 - X_4 - X_5 - X_6$ and use the result

to obtain the probability that in a 6-digit lottery the sum of the first three digits equals the sum of the last three digits. We have

$$E[z^{X_i}] = \frac{1}{10}(1 + z + \cdots + z^9) = \frac{1}{10}\frac{1 - z^{10}}{1 - z},$$

and therefore

$$E[z^{-X_i}] = \frac{1}{10}\frac{1}{z^9}\frac{1 - z^{10}}{1 - z},$$

and

$$\begin{aligned} E[z^Y] &= E\Big[\Big[z^{27+\sum_{i=1}^{3} X_i - \sum_{i=4}^{6} X_i}\Big] \\ &= E\left[z^{27}\prod_{i=1}^{3} z^{X_i}\prod_{i=4}^{6} z^{-X_i}\right] = z^{27}\prod_{i=1}^{3} E[z^{X_i}]\prod_{i=4}^{6} E[z^{-X_i}]\,. \end{aligned}$$

Therefore,

$$g_Y(z) = \frac{1}{10^6}\frac{(1 - z^{10})^6}{(1 - z)^6}.$$

But $P(X_1 + X_2 + X_3 = X_4 + X_5 + X_6) = P(Y = 27)$ is the factor of z^{27} in the power series expansion of $g_Y(z)$. Since

$$(1 - z^{10})^6 = 1 - \binom{6}{1}z^{10} + \binom{6}{2}z^{20} + \cdots$$

and

$$(1 - z)^{-6} = 1 + \binom{6}{5}z + \binom{7}{5}z^2 + \binom{8}{5}z^3 + \cdots$$

(recall the negative binomial formula:

$$(1 - z)^{-p} = 1 + \binom{p}{p-1}z + \binom{p+1}{p-1}z^2 + \binom{p+2}{p-1}z^3 + \cdots),$$

we find that

$$P(Y = 27) = \frac{1}{10^6}\left(\binom{32}{5} - \binom{6}{1}\binom{22}{5} + \binom{6}{2}\binom{12}{5}\right).$$

2.3 Conditional Expectation

2.3.1 Conditioning with Respect to an Event

Chapter 1 introduced the notion of conditional probability and the Bayes calculus associated with it. We now introduce the notion of conditional expectation and the set of rules accompanying it.

Let Z be a discrete random variable with values in E, and let $f : E \to \mathbb{R}$ be a non-negative function. Let A be some event of positive probability. The conditional expectation of $f(Z)$ given A, denoted by $E[f(Z) \mid A]$, is by definition the expectation when the distribution of Z is replaced by its conditional distribution given A, $P(Z = z \mid A)$. Therefore

$$E[f(Z) \mid A] := \sum_z f(z) P(Z = z \mid A) .$$

Let $\{A_i\}_{i \in \mathbb{N}}$ be a partition of the sample space. Then

$$E[f(Z)] = \sum_{i \in \mathbb{N}} E[f(Z) \mid A_i] P(A_i) .$$

Proof. This is a direct consequence of the Bayes formula of total causes:

$$\begin{aligned} E[f(Z)] &= \sum_z f(z) P(Z = z) = \sum_z \left(\sum_i f(z) P(Z = z \mid A_i) P(A_i) \right) \\ &= \sum_i \left(\sum_z f(z) P(Z = z \mid A_i) \right) P(A_i) = \sum_i E[f(Z) \mid A_i] P(A_i) . \end{aligned}$$

□

The following elementary result will often be used, and therefore, we shall promote it to the rank of theorem:

Theorem 2.3.1 *Let Z be a random variable with values in E, and let $f : E \mapsto \mathbb{R}$ be a non-negative function. Let A be some event of positive probability. Then*

$$E[f(Z) 1_A] = E[f(Z) \mid A] P(A) .$$

Proof.

$$E[f(Z) \mid A] P(A) = \left(\sum_{z \in E} f(z) P(Z = z \mid A) \right) P(A) = \sum_{z \in E} f(z) P(Z = z , A) .$$

Now, the random variable $f(Z) 1_A$ takes a non-null value if and only if this value is of the form $f(z) > 0$, and this happens with probability $P(Z = z , A)$. Therefore

$$E[f(Z) 1_A] = \sum_{z ; f(z) > 0} f(z) P(Z = z , A) = \sum_{z \in E} f(z) P(Z = z , A) .$$

□

EXAMPLE 2.3.2: POISSON BOUNDING OF MULTINOMIAL EVENTS. (Mitzenmacher and Upfal, 2005.) The computation of expectations concerning multinomial vectors

often turns out to be difficult, whereas it might be considerably simpler in the Poisson case. The result of this subsection gives, under certain conditions, a bound for the expectation of interest in terms of the expectation computed for the Poisson case. Before the precise statement of this result, some preliminary remarks are in order.

Balls are placed in N bins in the following manner. The number of balls in any given bin is a Poisson variable of mean $\frac{m}{N}$, and is independent of the numbers in the other bins. In particular, the total number of balls $Y_1 + \cdots + Y_N$ is, as the sum of independent Poisson random variables, a Poisson random variable whose mean is the sum of the means of the coordinates, that is m.

Let $f \geq 0$ be a function of N integer-valued arguments, and let $(X_1, \ldots, X_N)$ be a multinomial random vector of size (m, N) and with parameters $p_i = \frac{1}{N}$ (obtained by placing m balls independently and at random in N bins). Then, with the Y_i's as above,

$$E\left[f(X_1, \ldots, X_N)\right] \leq e\sqrt{m} E\left[f(Y_1, \ldots, Y_N)\right] . \tag{2.20}$$

In particular, with f the indicator of some subset E of $\mathbb{N}^N$, the probability that $(X_1, \ldots, X_N) \in E$ is less than $e\sqrt{m}$ times the probability that $(Y_1, \ldots, Y_N) \in E$. This can be rephrased in imprecise but suggestive terms as follows: An event that has probability P in the Poisson case happens with probability at most $e\sqrt{m}P$ in the multinomial case.

Proof. For a given arbitrary integer k, the conditional probability that there are k_1 balls in bin 1, k_2 balls in bin 2, ..., given that the total number of balls is $k_1 + \cdots + k_N = k$ is

$$\begin{aligned} &P\left(Y_1 = k_1, \ldots, Y_N = k_N \,|\, Y_1 + \cdots + Y_N = k\right) \\ &\qquad = \frac{P\left(Y_1 = k_1, \ldots, Y_N = k_N, Y_1 + \cdots + Y_N = k\right)}{P\left(Y_1 + \cdots + Y_N = k\right)} \\ &\qquad = \frac{P\left(Y_1 = k_1, \ldots, Y_N = k_N\right)}{P\left(Y_1 + \cdots + Y_N = k\right)} . \end{aligned}$$

By independence of the Y_i's and since they are Poisson variables with mean $\frac{m}{N}$,

$$P\left(Y_1 = k_1, \ldots, Y_N = k_N\right) = \prod_{i=1}^{N} \left(e^{-\frac{m}{N}} \frac{\left(\frac{m}{N}\right)^{k_i}}{k_i!} \right) .$$

Also, $P\left(Y_1 + \cdots + Y_N = k\right) = e^{-m} \frac{m^k}{k!}$. Therefore

$$P\left(Y_1 = k_1, \ldots, Y_N = k_N \,|\, Y_1 + \cdots + Y_N = k\right) = \frac{k!}{k_1! \cdots k_N!} \left(\frac{1}{N}\right)^N .$$

But this is equal to $P(Z_1 = k_1, \ldots, Z_N = k_N)$, where Z_i is the number of balls in bin i when $k = k_1 + \cdots + k_N$ balls are placed independently and at random in the N bins. Note that the above equality is independent of m.

Now:

$$\begin{aligned} E\left[f(Y_1,\dots,Y_N)\right] &= \sum_{k=0}^{\infty} E\left[f(Y_1,\dots,Y_N) \mid \sum_{i=1}^{N} Y_i = k\right] P\left(\sum_{i=1}^{N} Y_i = k\right) \\ &\geq E\left[f(Y_1,\dots,Y_N) \mid \sum_{i=1}^{N} Y_i = m\right] P\left(\sum_{i=1}^{N} Y_i = m\right) \\ &= E\left[f(X_1,\dots,X_N)\right] P\left(\sum_{i=1}^{N} Y_i = m\right) \\ &= E\left[f(X_1,\dots,X_N)\right] \frac{m^m e^{-m}}{m!}. \end{aligned}$$

The announced result will follow from the bound

$$m! \leq e\sqrt{m}\left(\frac{m}{e}\right)^m . \qquad (\star)$$

For this, use the fact that, by concavity of the function $x \to \log x$,

$$\int_{i-1}^{i} \log x\, dx \geq \frac{\log(i-1) + \log i}{2},$$

and therefore

$$\int_1^m \log x\, dx \geq \sum_{i=1}^{m} \log i - \frac{\log m}{2} = \log(m!) - \frac{\log m}{2}\,.$$

Integration by parts gives $m \log m - m + 1 = \int_1^m \log x\, dx$. Therefore $m \log m - m + 1 \geq \log(m!) - \frac{\log m}{2}$, from which the announced inequality follows by taking exponentials. □

There exists a stronger version of (2.20):

$$E\left[f(X_1,\dots,X_N)\right] \leq 4E\left[f(Y_1,\dots,Y_N)\right] ,$$

but this time it is required in addition that $E\left[f(X_1,\dots,X_N)\right]$ should be a quantity increasing with the number m of balls.

Proof.

$$\begin{aligned} E\left[f(Y)\right] &= \sum_{k=0}^{\infty} E\Big[f(Y) \mid \sum Y_i = k\Big] P\Big(\sum Y_i = k\Big) \\ &\geq \sum_{k=m}^{\infty} E\Big[f(Y) \mid \sum Y_i = k\Big] P\Big(\sum Y_i = k\Big) \\ &\geq E\Big[f(Y) \mid \sum Y_i = m\Big] P\Big(\sum Y_i = k\Big) \\ &\geq E\left[f(X)\right]\ P\Big(\sum Y_i = k\Big)\ \geq E\left[f(X)\right] \times \frac{1}{4}\,, \end{aligned}$$

since for any Poisson variable Z with a mean θ that is a positive integer, $P(Z \geq \theta) \geq \frac{1}{4}$. □

2.3.2 Conditioning with Respect to a Random Variable

Let X and Y be two discrete random variables taking their values in the denumerable sets F and G respectively. Let the function $g : F \times G \to \mathbb{R}$ be either non-negative, or such that $E[|g(X,Y)|] < \infty$. For each $y \in G$ such that $P(Y = y) > 0$, define

$$\psi(y) := \sum_{x \in F} g(x,y) P(X = x \,|\, Y = y)\,, \tag{2.21}$$

and let $\psi(y) := 0$ otherwise. The sum in (2.21) is well-defined (possibly infinite however) when g is non-negative. Note that in the non-negative case, we have that

$$\begin{aligned}\sum_{y \in G} \psi(y) P(Y = y) &= \sum_{y \in G} \sum_{x \in F} g(x,y) P(X = x \,|\, Y = y) P(Y = y) \\ &= \sum_{x} \sum_{y} g(x,y) P(X = x, Y = y) = E[g(X,Y)].\end{aligned}$$

In particular, if $E[g(X,Y)] < \infty$, $\sum_{y \in G} \psi(y) P(Y = y) < \infty$, which implies that (Theorem 2.1.27) $P(\psi(Y) < \infty) = 1$. Therefore, $E\left[E^Y\left[g(X,Y)\right]\right] < \infty$. Let now $g : F \times G \to \mathbb{R}$ be a function of arbitrary sign such that $E[|g(X,Y)|] < \infty$, and in particular $E[g^{\pm}(X,Y)] < \infty$. Denote by $\psi^{\pm}$ the functions associated with $g^{\pm}$ as in (2.21). As we just saw, for all $y \in G$, $\psi^{\pm}(y) < \infty$, and therefore $\psi(y) = \psi^{+}(y) - \psi^{-}(y)$ is well-defined (not an indeterminate $\infty - \infty$ form). Thus, the conditional expectation is well-defined also in the integrable case. From the observation made a few lines above, in this case,

$$|E^Y[g(X,Y)]| = |E^Y[g^+(X,Y)]| + |E^Y[g^-(X,Y)]| < \infty,\ P\text{-a.s.}$$

Definition 2.3.3 *The number $\psi(y)$ defined by (2.21) is called the conditional expectation of $g(X,Y)$ given $Y = y$, and is denoted by $E^{Y=y}[g(X,Y)]$ or, alternatively, by $E[g(X,Y) \,|\, Y = y]$. The random variable $\psi(Y)$ is called the conditional expectation of $g(X,Y)$ given Y, and is denoted by $E^Y[g(X,Y)]$ or $E[g(X,Y) \,|\, Y]$.*

Example 2.3.4: The hypergeometric distribution. Let X_1 and X_2 be independent binomial random variables of same size N and same parameter p. We are going to show that

$$E^{X_1+X_2}[X_1] = \psi(X_1 + X_2) = \frac{X_1 + X_2}{2}\,.$$

We have

$$\begin{aligned}P(X_1 = k | X_1 + X_2 = n) = \frac{P(X_1 = k, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \\ \frac{P(X_1 = k, X_2 = n - k)}{P(X_1 + X_2 = n)} \\ \frac{P(X_1 = k) P(X_2 = n - k)}{P(X_1 + X_2 = n)}\,.\end{aligned}$$

Inserting the values of the probabilities thereof, and using the fact that the sum of two independent binomial random variables with size N and parameter p is a binomial random variable with size $2N$ and parameter p, a straightforward computation gives

$$P(X_1 = k | X_1 + X_2 = n) = \frac{\binom{N}{k}\binom{N}{n-k}}{\binom{2N}{n}}.$$

This is the hypergeometric distribution. The right-hand side of the last display is the probability of obtaining k black balls when a sample of n balls is randomly selected from an urn containing N black balls and N red balls. The mean of such a distribution is (by symmetry) $\frac{n}{2}$, therefore

$$E^{X_1+X_2=n}[X_1] = \frac{n}{2} = \psi(n)$$

and this gives the announced result. A more elegant solution is given in Exercise 2.4.22 where the reader will also discover that the result is more general.

EXAMPLE 2.3.5: TWO POISSON VARIABLES. Let X_1 and X_2 be two independent Poisson random variables with respective means $\theta_1 > 0$ and $\theta_2 > 0$. We seek to compute $E^{X_1+X_2}[X_1]$, that is $E^Y[X]$, where $X = X_1$, $Y = X_1 + X_2$. For $y \geq x$, the same computations as in Example 2.3.4 give

$$P(X = x \,|\, Y = y) = \frac{P(X_1 = x)P(X_2 = y - x)}{P(X_1 + X_2 = y)}.$$

Inserting the values of the the probabilities thereof, and using the fact that the sum of two independent Poisson random variables with parameter θ_1 and θ_2 is a Poisson random variable with parameter $\theta_1 + \theta_2$, a simple computation yields

$$P(X = x \,|\, Y = y) = \binom{y}{x}\left(\frac{\theta_1}{\theta_1 + \theta_2}\right)^x \left(\frac{\theta_2}{\theta_1 + \theta_2}\right)^{y-x}.$$

Therefore, with $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$,

$$\psi(y) = E^{Y=y}[X] = \sum_{x=0}^{y} x\binom{y}{x}\alpha^x(1-\alpha)^{y-x} = \alpha y\,.$$

Finally, $E^Y[X] = \psi(Y) = \alpha Y$, that is,

$$E^{X_1+X_2}[X_1] = \frac{\theta_1}{\theta_1 + \theta_2}(X_1 + X_2)\,.$$

2.3.3 Basic Properties of Conditional Expectation

The first property of conditional expectation, *linearity*, is obvious from the definitions: For all $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$E^Y[\lambda_1 g_1(X,Y) + \lambda_2 g_2(X,Y)] = \lambda_1 E^Y[g_1(X,Y)] + \lambda_2 E^Y[g_2(X,Y)]$$

whenever the conditional expectations thereof are well-defined and do not produce $\infty - \infty$ forms. *Monotonicity* is equally obvious: if $g_1 \leq g_2$, then

$$E^Y[g_1(X,Y)] \leq E^Y[g_2(X,Y)].$$

Theorem 2.3.6 *If g is non-negative or such that $E[|g(X,Y)|] < \infty$, we have*

$$E[E^Y[g(X,Y)]] = E[g(X,Y)].$$

Proof.

$$\begin{aligned} E[E^Y[g(X,Y)]] = E[\psi(Y)]] &= \sum_{y \in G} \psi(y) P(Y = y) \\ &= \sum_{y \in G} \sum_{x \in F} g(x,y) P(X = x \,|\, Y = y) P(Y = y) \\ &= \sum_x \sum_y g(x,y) P(X = x, Y = y) = E[g(X,Y)]. \end{aligned}$$

□

Theorem 2.3.7 *If w is non-negative or such that $E[|w(Y)|] < \infty$,*

$$E^Y[w(Y)] = w(Y),$$

and more generally,

$$E^Y[w(Y)h(X,Y)] = w(Y) E^Y[h(X,Y)]\,,$$

assuming that the left-hand side is well-defined.

Proof. We prove the second (more general) identity. We do this for non-negative w and h, the general case following easily from this special case:

$$\begin{aligned} E^{Y=y}[w(Y)h(X,Y)] &= \sum_{x \in F} w(y) h(x,y) P(X = x \,|\, Y = y) \\ &= w(y) \sum_{x \in F} h(x,y) P(X = x \,|\, Y = y) \\ &= w(y) E^{Y=y}[h(X,Y)]. \end{aligned}$$

□

Theorem 2.3.8 *If X and Y are independent and if v is non-negative or such that $E[|v(X)|] < \infty$, then*

$$E^Y[v(X)] = E[v(X)].$$

Proof. We have

$$\begin{aligned} E^{Y=y}[v(X)] &= \sum_{x \in F} v(x) P(X = x \,|\, Y = y) \\ &= \sum_{x \in F} v(x) P(X = x) = E[v(X)]. \end{aligned}$$

□

Theorem 2.3.9 *If X and Y are independent and if $g : F \times G \to \mathbb{R}$ is non-negative or such that $E[|g(X,Y)|] < \infty$, then, for all $y \in G$,*

$$E[g(X, Y \,|\, Y = y] = E[g(X, y)].$$

Proof. Applying formula (2.21) with $P(X = x \,|\, Y = y) = P(X = x)$ (by independence), we obtain

$$\psi(y) = \sum_{x \in F} g(x, y) P(X = x) = E\left[g(X, y)\right] .$$

□

Successive Conditioning

Suppose that $Y = (Y_1, Y_2)$, where Y_1 and Y_2 are discrete random variables. In this situation, we use the more developed notation

$$E^Y[g(X, Y)] = E^{Y_1, Y_2}[g(X, Y_1, Y_2]\,.$$

Theorem 2.3.10 *Let $Y = (Y_1, Y_2)$ be as above, and let $g : F \times G \to \mathbb{R}$ be either non-negative or such that $E[|g(X,Y)|] < \infty$. Then*

$$E^{Y_2}[E^{Y_1, Y_2}[g(X, Y_1, Y_2)]] = E^{Y_2}[g(X, Y_1, Y_2)].$$

Proof. Let

$$\psi(Y_1, Y_2) = E^{Y_1, Y_2}[g(X, Y_1, Y_2)].$$

We must show that

$$E^{Y_2}[\psi(Y_1, Y_2)] = E^{Y_2}[g(X, Y_1, Y_2)].$$

But

$$\psi(y_1, y_2) = \sum_x g(x, y_1, y_2) P(X = x \,|\, Y_1 = y_1, Y_2 = y_2)$$

and

$$E^{Y_2=y_2}[\psi(Y_1, Y_2)] = \sum_{y_1} \psi(y_1, y_2) P(Y_1 = y_1 \,|\, Y_2 = y_2),$$

that is,

$$\sum_{y_1} \sum_x g(x, y_1, y_2) P(X = x \,|\, Y_1 = y_1, Y_2 = y_2) P(Y_1 = y_1 \,|\, Y_2 = y_2).$$

But

$$\begin{aligned} P(X = x \,|\, Y_1 = y_1, Y_2 = y_2) & P(Y_1 = y_1 \,|\, Y_2 = y_2) \\ &= \frac{P(X = x,\, Y_1 = y_1, Y_2 = y_2)}{P(Y_1 = y_1, Y_2 = y_2)} \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} \\ &= P(X = x, Y_1 = y_1 \,|\, Y_2 = y_2)\,. \end{aligned}$$

Therefore

$$\begin{aligned} E^{Y_2=y_2}[\psi(Y_1, Y_2)] &= \sum_{y_1} \sum_x g(x, y_1, y_2) P(X = x, Y_1 = y_1 \,|\, Y_2 = y_2) \\ &= E^{Y_2=y_2}[g(X, Y_1, Y_2)]. \end{aligned}$$

□

Conditional Jensen's Inequality

Theorem 2.3.11 *Let I, φ and X be as in Theorem 3.1.5. Let Y be another random variable. Then*

$$E\left[\varphi(X) \,|\, Y\right] \geq \varphi(E\left[X \,|\, Y\right])\,.$$

Proof. The proof follows exactly the same lines as that of Theorem 3.1.5. □

The FKG Inequality

Theorem 2.3.12 *Let $E \subseteq \mathbb{R}$ and let $f, g : E^n \to \mathbb{R}$ be two bounded functions that are non-decreasing in each of their arguments. Let $X_1^n := (X_1, \ldots, X_n)$ be a vector of independent variables with values in E. Then,*

$$E\left[f(X_0^n) g(X_0^n)\right] \geq E\left[f(X_0^n)\right] E\left[g(X_0^n)\right]\,. \tag{2.22}$$

In other words, $f(X_0^n)$ and $g(X_0^n)$ are positively correlated.

Proof. By induction. For $n = 1$: Let X_1 and Y_1 be two independent and identically distributed E-valued random variables, and let $f, g : E \to \mathbb{R}_+$ be two non-decreasing bounded functions. Since $f(X_1) - f(Y_1)$ and $g(X_1) - g(Y_1)$ have the same sign, their product is non-negative, and therefore

$$E\left[(f(X_1) - f(Y_1))(g(X_1) - g(Y_1))\right] \geq 0\,.$$

Developing the left-hand side

$$E\left[f(X_1)g(X_1)\right] + E\left[f(Y_1)g(Y_1)\right] \geq E\left[f(X_1)\right]E\left[g(Y_1)\right] + E\left[f(Y_1)\right]E\left[g(X_1)\right]\,.$$

As X_1 and Y_1 have the same distribution, the left-hand side equals $2E\left[f(X_1)g(X_1)\right]$. Since X_1 and Y_1 have the same distribution and are independent, the right-hand side equals $2E\left[f(X_1)\right]E\left[g(X_1)\right]$. Therefore

$$E\left[f(X_1)g(X_1)\right] \geq E\left[f(X_1)\right]E\left[g(X_1)\right].$$

We now suppose that the result is true for $n-1$ and show that it is then true for n. From the independence of X_0^{n-1} and X_n and Theorem 2.3.9,

$$E\left[f(X_0^n)g(X_0^n)\,|\,X_n = x_n\right] = E\left[f(X_0^{n-1}, x_n)g(X_0^{n-1}, x_n)\right]$$

and since, by the result assumed for $n-1$,

$$\begin{aligned} E\left[f(X_0^{n-1}, x_n)g(X_0^{n-1}, x_n)\right] &\geq E\left[f(X_0^{n-1}, x_n)\right]E\left[g(X_0^{n-1}, x_n)\right] \\ &= E\left[f(X_0^n)\,|\,X_n = x_n\right]E\left[g(X_0^n)\,|\,X_n = x_n\right]\,, \end{aligned}$$

we have that

$$E\left[f(X_0^n)g(X_0^n)\,|\,X_n = x_n\right] \geq E\left[f(X_0^n)\,|\,X_n = x_n\right]E\left[g(X_0^n)\,|\,X_n = x_n\right]\,,$$

or

$$E\left[f(X_0^n)g(X_0^n)\,|\,X_n\right] \geq E\left[f(X_0^n)\,|\,X_n\right]E\left[g(X_0^n)\,|\,X_n\right]\,.$$

Taking expectations

$$\begin{aligned} E\left[f(X_0^n)g(X_0^n)\right] &\geq E\left[E\left[f(X_0^n)\,|\,X_n\right]E\left[g(X_0^n)\,|\,X_n\right]\right] \\ &\geq E\left[E\left[f(X_0^n)\,|\,X_n\right]\right]E\left[E\left[g(X_0^n)\,|\,X_n\right]\right] \\ &= E\left[f(X_0^n\right]E\left[g(X_0^n\right]\,, \end{aligned}$$

where the last inequality follows from the case $n = 1$ applied to the functions $x_n \to E\left[f(X_0^n)\,|\,X_n = x_n\right] = E\left[f(X_0^{n-1}, x_n)\right]$ and $x_n \to E\left[g(X_0^n)\,|\,X_n = x_n\right] = E\left[f(X_0^{n-1}, x_n)\right]$ which are non-decreasing. □

Remark 2.3.13 A stronger version of the above FKG inequality will be given in Section 9.3.

An Alternative Point of View

This subsection presents another definition of conditional expectation. It is the starting point for a generalization to the case of random elements that are not discrete. Even in the discrete case, this new perspective is indispensable (see Exercise 2.4.24).

Let X and Y be two discrete random variables with values in E and F respectively. Let $g : E \times F \to \mathbb{R}_+$ be a function that is either non-negative or such that $g(X, Y)$ is integrable. For any non-negative bounded function $\varphi : F \to \mathbb{R}$, we have

$$E\left[E^Y\left[g(X,Y)\right]\varphi(Y)\right] = E\left[g(X,Y)\varphi(Y)\right]. \qquad (\star)$$

In fact,

$$\begin{aligned} E\left[E^Y\left[g(X,Y)\right]\varphi(Y)\right] = E\left[\psi(Y)\varphi(Y)\right] &= \sum_{y\in F}\psi(y)\varphi(y)P(Y=y) \\ &= \sum_{y\in F}\left(\sum_{x\in E} g(x,y)\frac{P(X=x,Y=y)}{P(Y=y)}\,dx\right)\varphi(y)P(Y=y) \\ &= \sum_{y\in F}\sum_{x\in E} g(x,y)\varphi(y)P(X=x,Y=y) = E\left[g(X,Y)\varphi(Y)\right]. \end{aligned}$$

This suggests to take $(\star)$ as a basis for an extension of the definition of conditional expectation. The conditioned variable is now any random element Z taking its values in E, a denumerable subset of $\mathbb{R}$.

Definition 2.3.14 *Let Z and Y be as above, and suppose that Z is either non-negative or integrable. A conditional expectation $E^Y[Z]$ is by definition a random variable of the form $\psi(Y)$ such that equality*

$$E\left[\psi(Y)\varphi(Y)\right] = E\left[Z\varphi(Y)\right] \qquad (2.23)$$

holds for any non-negative bounded function $\varphi : E \to \mathbb{R}$.

Theorem 2.3.15 *In the situation described in the above definition, the conditional expectation exists and is essentially unique.*

By "essentially unique" the following is meant. If there are two functions ψ_1 and ψ_2 that meet the requirement, then $\psi_1(Y) = \psi_2(Y)$ almost surely.

Proof. The proof of existence is by the construction at the begining of the section, replacing $g(X, Y)$ by Z (more explicitly, $h : E \to \mathbb{R}$, $X = Z$, $g(x, y) = h(z)$). For uniqueness, suppose that ψ_1 and ψ_2 meet the requirement. In particular $E\left[\psi_1(Y)\varphi(Y)\right] = E\left[\psi_2(Y)\varphi(Y)\right]$ $(= E\left[Z\varphi(Y)\right])$, or $E\left[(\psi_1(Y) - \psi_2(Y))\varphi(Y)\right] = 0$, for any non-negative bounded function $\varphi : \mathbb{R}^n \to \mathbb{R}$. Choose $\varphi(Y) = 1_{\{\psi_1(Y)-\psi_2(Y)>0\}}$ to obtain

$$E\left[(\psi_1(Y)-\psi_2(Y))1_{\{\psi_1(Y)-\psi_2(Y)>0\}}\right]=0\,.$$

Since the random variable $(\psi_1(Y)-\psi_2(Y))1_{\{\psi_1(Y)-\psi_2(Y)>0\}}$ is non-negative and has a null expectation, it must be almost surely null. In other terms $\psi_1(Y)-\psi_2(Y)\leq 0$ almost surely. Exchanging the roles of ψ_1 and ψ_2, we have that $\psi_1(Y)-\psi_2(Y)\geq 0$ almost surely. Therefore $\psi_1(Y)-\psi_2(Y)=0$ almost surely. □

EXAMPLE 2.3.16: Let Y be a positive integer-valued random variable.

$$E^Y[Z]=\sum_{n=1}^{\infty}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}1_{\{Y=n\}},$$

where, by convention, $\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}=0$ when $P(Y=n)=0$ (in other terms, the sum in the above display is over all n such that $P(Y=n)>0$).

Proof. We must verify (2.23) for all bounded measurable $\varphi:\mathbb{R}\to\mathbb{R}$. The right-hand side is equal to

$$\begin{aligned}
&E\left[\left(\sum_{n\geq 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}1_{\{Y=n\}}\right)\left(\sum_{k\geq 1}\varphi(k)1_{\{Y=k\}}\right)\right]\\
=&E\left[\sum_{n\geq 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}\varphi(n)1_{\{Y=n\}}\right]=\sum_{n\geq 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}\varphi(n)E[1_{\{Y=n\}}]\\
=&\sum_{n\geq 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}\varphi(n)P(Y=n)=\sum_{n\geq 1}E[Z1_{\{Y=n\}}]\varphi(n)\\
=&\sum_{n\geq 1}E[Z1_{\{Y=n\}}\varphi(n)]=E[Z(\sum_{n\geq 1}\varphi(n)1_{\{Y=n\}})]=E[Z\varphi(Y)]\,.
\end{aligned}$$

□

2.4 Exercises

Exercise 2.4.1. GEOMETRIC
Let T_1 and T_2 be two independent geometric random variables with the same parameter $p\in(0,1)$. Give the probability distribution of their sum $X=T_1+T_2$.

Exercise 2.4.2. VARIANCE OF THE COUPON'S COLLECTOR VARIABLE
In the coupon's collector problem of Example 2.1.42, compute the variance σ_X^2 of X (the number of chocolate tablets needed to complete the collection of the n different coupons) and show that $\frac{\sigma_X^2}{n^2}$ has a limit (to be identified) as $n\uparrow\infty$.

Exercise 2.4.3. POISSON
1. Let X be a Poisson random variable with mean $\theta>0$. Compute the mean of the random variable $X!$ (factorial, not exclamation mark).

2. Compute $E\left[\theta^X\right]$.

3. What is the probability that X is odd?

Exercise 2.4.4. RANDOM SUM
Let $\{X_n\}_{n\geq 1}$ be independent random variables taking the values 0 and 1 with probability $q = 1-p$ and p, respectively, where $p \in (0,1)$. Let T be a Poisson random variable with mean $\theta > 0$, independent of $\{X_n\}_{n\geq 1}$. Compute the probability distribution of $S := X_1 + \cdots + X_T$.

Exercise 2.4.5. THE BINOMIAL RANDOM VARIABLE
(a) Let $X \sim \mathcal{B}(n,p)$. Show that $Y := n - X \sim \mathcal{B}(n, 1-p)$.
(b) Let $X_1, \ldots, X_{2n}$ be independent random variables taking the values 0 or 1, and such that for all i, $P(X_i = 1) = p \in (0,1)$. Give the probability distribution of the random variable $Z := \sum_{i=1}^{n} X_i\, X_{n+i}$.

Exercise 2.4.6. NULL VARIANCE
Let X be a discrete random variable taking its values in E, with probability distribution $p(x)$, $x \in E$.
(i) Let $A := \{\omega;\, p(X(\omega)) = 0\}$. Show that $P(A) = 0$.
(ii) Prove that a real-valued random variable with null variance is almost surely constant.

Exercise 2.4.7. THE BLUE PINKO
The blue pinko is a bird owing its name to the fact that it lays eggs that are either blue or pink. Suppose that it lays T eggs, with probability p that a given egg is blue, and that the colours of the successive eggs are independent and independent of the total number of eggs. The conclusion of Exercise 2.4.4 was that if the number of eggs is Poisson with mean θ, then the number of blue eggs is a Poisson random variable with mean θp and the number of pink eggs is a Poisson random variable with mean $\theta(1-p)$. Prove that the number of blue eggs and the number of pink eggs are independent random variables.

Exercise 2.4.8. THE ENTOMOLOGIST
Each individual of a specific breed of insects has, independently of the others, the probability θ of being a male.

(A) An entomologist seeks to collect exactly $M > 1$ males, and therefore stops hunting as soon as she captures M males. What is the distribution of X, the number of insects she must catch to collect *exactly* M males?

(B) What is the distribution of X, the smallest number of insects that the entomologist must catch to collect *at least* M males and N females?

Exercise 2.4.9. MAXIMAL BIN LOAD
N balls are thrown independently and at random in N bins. This results in X_i balls in bin i ($1 \leq i \leq N$). Let $X_{max} = \max\{X_1, \ldots, X_N\}$ be the maximal bin load. Prove the following: For sufficiently large N,

$$P\left(X_{max} > \frac{\log N}{\log^{(2)} N}\right) \geq 1 - \frac{1}{N},$$

where $\log^{(2)} N := \log(\log N)$.

Exercise 2.4.10. The matchbox
A smoker has one matchbox with n matches in each pocket. He reaches at random for one box or the other. What is the probability that, having eventually found an empty matchbox, there will be k matches left in the other box?

Exercise 2.4.11. Biased dice and uniformity
Is it possible to have two biased dice such that tossing them independently results in a total number of points uniformly distributed on $\{2, 3, \ldots, 12\}$?

Exercise 2.4.12. Residual time
Let X be a random variable with values in $\mathbb{N}$ and with finite mean m. Show that $p_n = \frac{1}{m}P(X > n)$ $(n \geq 0)$ defines a probability distribution on $\mathbb{N}$ and compute its generating function in terms of the generating function of X.

Exercise 2.4.13. Mean and variance via generating functions
(a) Compute the mean and variance of the binomial random variable B of size n and parameter p from its generating function. Do the same for the Poisson random variable P of mean θ.

(b) What is the generating function g_T of the geometric random variable T with parameter $p \in (0, 1)$? Compute its first two derivatives and deduce from the result the variance of T.

(c) What is the n-th factorial moment $(E[X(X-1)\cdots(X-n+1)])$ of a Poisson random variable X of mean $\theta > 0$?

Exercise 2.4.14. From generating function to distribution
What is the probability distribution of the integer-valued random variable with generating function $g(z) = \frac{1}{(2-z)^2}$? Compute the fifth moment $(E[X^5])$ of this random variable.

Exercise 2.4.15. Throw a die
You perform three independent tosses of an unbiased die. What is the probability that one of these tosses results in a number that is the sum of the two other numbers? (You are required to find a solution using generating functions.)

Exercise 2.4.16. Generalized Wald's formula
Let $\{Y_n\}_{n\geq 1}$ be a sequence of integer-valued integrable random variables such that $E[Y_n] = E[Y_1]$ for all $n \geq 1$. Let T be an integer-valued random variable such that for all $n \geq 1$, the event $\{T \geq n\}$ is independent of Y_n. Let $X := \sum_{n=1}^{T} Y_n$. Prove that $E[X] = E[Y_1]E[T]$.

Exercise 2.4.17. When Wald's formula does not apply
Let $\{Y_n\}_{n\geq 1}$ be a sequence of integer-valued integrable random variables such that $E[Y_n] = E[Y_1]$ for all $n \geq 1$. Let T be an integer-valued random variable. Let $X := \sum_{n=1}^{T} Y_n$. It is not true in general that $E[X] = E[Y_1]E[T]$. Give a simple counterexample.

Exercise 2.4.18. The return of the entomologist
Recall the setup of Exercise 2.4.8. What is the expectation of X, the number of insects the entomologist must capture to collect *exactly* M males? (In Exercise 2.4.8, you computed the distribution of X, from which you can of course compute the mean. However, you can give the solution directly, and this is what is required in the present exercise.)

Exercise 2.4.19. Conditioning by sampling
Let Z be a discrete random variable with values in E and let $f : E \to \mathbb{R}$ be a non-negative function. Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of random variables with values in E and the same distribution as Z. Let A be some subset of E such that $P(Z \in A) > 0$.

(1) Define the random variable τ to be the first time $n \geq 1$ such that $Z_n \in A$. Prove that $P(\tau < \infty) = 1$.

(2) Let Z_τ be the random variable equal to Z_n when $\tau = n$. Prove that

$$E[f(Z_\tau)] = E[f(Z) \mid Z \in A] .$$

Exercise 2.4.20. Multinomial distribution and conditioning
Let $(X_1, \ldots, X_k)$ be a multinomial random vector with size n and parameters $p_1, \ldots, p_k$. Compute $E^{X_1}[X_2 + \cdots + X_{k-1}]$ and $E^{X_1}[X_2]$.

Exercise 2.4.21. XYZ
Let X, Y, and Z be three discrete random variables with values in E, F, and G, respectively. Prove the following: If for some function $g : E \times F \to [0, 1]$, $P(X = x \,|\, Y = y, Z = z) = g(x, y)$ for all x, y, z, then $P(X = x \,|\, Y = y) = g(x, y)$ for all x, y, and X and Z are conditionally independent given Y.

Exercise 2.4.22. A natural result
Let X_1 and X_2 be two integrable independent identically distributed discrete real-valued random variables. Prove that

$$E^{X_1+X_2}[X_1] = \frac{X_1 + X_2}{2}.$$

Exercise 2.4.23. Pólya's urn
There is an urn containing black balls and white balls, the number of which varies in time as follows. At time $n = 0$ there is one black ball and one white ball. At a given time one of the balls is selected at random, its colour is observed, and the ball is replaced in the urn together with a new ball of the same colour. In particular

the number of balls increases by one unit at each draw. Let B_k be the number of black balls after exactly k balls have been added. Prove that B_k is uniformly distributed on $\{1, 2, \ldots, k+1\}$.

Exercise 2.4.24. CONDITIONING BY THE SQUARE
Let X be a random variable with values in $\mathbb{Z}$ and probability distribution $(p(n)\,,\, n \in \mathbb{Z})$. Let $h : \mathbb{Z} \to \mathbb{R}$ be a function such that $E\left[|h(Z)|\right] < \infty$. Prove formally that

$$E\left[h(X)\,|\,X^2\right] = h(|X|)\frac{p(|X|)}{p(|X|)+p(-|X|)} + h(-|X|)\frac{p(-|X|)}{p(|X|)+p(-|X|)}\,.$$

Exercise 2.4.25. BAYESIAN TESTS OF HYPOTHESES
Let Θ be a discrete random variable with values in $\{1, 2, ..., K\}$ and let X be a discrete random variable with values in E. The joint distribution of Θ and X is specified in the following manner. For all $1 \leq i \leq K$,

$$P\left(\Theta = i\right) = \pi(i), \qquad P\left(X = x|\Theta = i\right) = p_i(x),$$

where π is a probability distribution on $\{1, 2, ..., K\}$ and the p_i's are probability disributions on E.

These random variables may be interpreted in terms of tests of hypotheses. The variable Θ represents the state of Nature, and X — called the observation — is the (random) result of an experiment that depends on the actual state of Nature. If Nature happens to be in state i, then X admits the distribution p_i.

In view of the observation X, we wish to infer the actual value of Θ. For this, we design a guess strategy, that is a function $g : E \to \{1, 2, ..., K\}$ with the interpretation that $\widehat{\Theta} := g(X)$ is our guess (based only on the observation X) of the (not directly observed) state Θ of Nature. An equivalent description of the strategy g is the partition $\mathcal{A} = \{A_1, \ldots, A_K\}$ of $\mathbb{R}^m$ given by $A_i := \{x \in E;\, g(x) = i\}$. The decision rule is then

$$X \in A_i \Rightarrow \widehat{\Theta} = i\,.$$

Prove the following: Any partition $\mathcal{A}^*$ such that

$$x \in A_i^* \Rightarrow \pi(i)p_i(x) = \max_k \left(\pi(k)p_k(x)\right)$$

minimizes the probability of error P_E.

Chapter 3

Bounds and Inequalities

3.1 The Three Basic Inequalities

3.1.1 Markov's Inequality

Bounding is the core of analysis and of probability. This chapter features the elementary inequalities and bounds, such as Markov's inequality and Jensen's inequality, the union bound and Chernoff's bounds. Other important bounds will be given as the need arises. For instance, Holley's inequality and its corollaries Harris' inequality and the FKG inequality will be presented in the context of random fields (Chapter 9). The coupling inequalities and Chen's Poisson approximation method are the objects of Chapter 16.

We begin with Markov's inequality, a simple consequence of the monotonocity and linearity properties of expectation.

Theorem 3.1.1 *Let Z be a non-negative random variable and let $a > 0$. Then*

$$P(Z \geq a) \leq \frac{E[Z]}{a}\,.$$

Proof. Take expectations in the inequality $a \times 1_{\{Z \geq a\}} \leq Z$. □

Corollary 3.1.2 *Let X be an integrable real random variable with mean m and finite variance σ^2. Then, for all $\varepsilon > 0$,*

$$P(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}\,.$$

Proof. This is a direct application of Theorem 3.1.1 with $Z = (X - m)^2$ and $a = \varepsilon^2$. It is called Chebyshev's inequality. □

EXAMPLE 3.1.3: WEAK LAW OF LARGE NUMBERS. Let $\{X_n\}_{n \geq 1}$ be an IID sequence of integrable real-valued discrete random variables with common mean m

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_3

and common variance $\sigma^2 < \infty$. The variance of the n-th order empirical mean $\overline{X}_n := \frac{X_1+\cdots+X_n}{n}$ equals $\frac{\sigma^2}{n}$, and therefore by Chebyshev's inequality, for all $\varepsilon > 0$,

$$P(\left|\overline{X}_n - m\right| \geq \varepsilon) = P\left(\left|\frac{\sum_{i=1}^n (X_i - m)}{n} \geq \varepsilon\right|\right) \leq \frac{\sigma^2}{n^2\varepsilon}\,. \qquad (\star)$$

In other words, the empirical mean $\overline{X}_n$ converges in probability to the mean m according to the following definition: A sequence of random variable $\{Z_n\}_{n\geq 1}$ is said to converge in probability to the random variable Z if, for all $\varepsilon > 0$,

$$\lim_{n\uparrow\infty} P(|Z_n - Z| \geq \varepsilon) = 0.$$

The specific result $(\star)$ is called the weak law of large numbers.

EXAMPLE 3.1.4: BERNSTEIN'S POLYNOMIAL APPROXIMATION. A continuous function f from $[0,1]$ into $\mathbb{R}$ can be approximated by a polynomial. More precisely, for all $x \in [0,1]$,

$$f(x) = \lim_{n\uparrow\infty} P_n(x)\,, \qquad (\star)$$

where

$$P_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k},$$

and the convergence of the series in the right-hand side is *uniform* in $[0,1]$. A proof of this classical theorem of analysis using probabilistic arguments is as follows.

Let $S_n := X_1 + \cdots + X_n$, where $\{X_n\}_{n\geq 1}$ is IID, with values in $\{0,1\}$, and such that $P(X_n = 1) = x$ $(n \geq 1)$. Since $S_n \sim \mathcal{B}(n,x)$,

$$E\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) P(S_n = k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k}.$$

The function f is continuous on the *bounded* interval $[0,1]$ and therefore *uniformly* continuous on this interval. Therefore to any $\varepsilon > 0$, one can associate a number $\delta(\varepsilon)$ such that if $|y-x| \leq \delta(\varepsilon)$, then $|f(x)-f(y)| \leq \varepsilon$. Being continuous on $[0,1]$, f is bounded on $[0,1]$ by some finite number, say M. Now

$$\begin{aligned} |P_n(x) - f(x)| &= \left|E\left[f\left(\frac{S_n}{n}\right) - f(x)\right]\right| \leq E\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\ &= E\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right| 1_A\right] + E\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right| 1_{\overline{A}}\right]\,, \end{aligned}$$

where $A := \{\omega\,;\, |S_n(\omega)/n) - x| \leq \delta(\varepsilon)\}$. Since $|f(S_n/n) - f(x)|1_{\overline{A}} \leq 2M1_{\overline{A}}$, we have

$$E\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right| 1_{\overline{A}}\right] \leq 2MP(\overline{A}) = 2MP\left(\left|\frac{S_n}{n} - x\right| \geq \delta(\varepsilon)\right)\,.$$

Also, by definition of A and $\delta(\varepsilon)$,

$$E\left[\left|f\left(\frac{S_n}{n}\right)-f(x)\right|1_A\right]\le\varepsilon\,.$$

Therefore

$$|P_n(x)-f(x)|\le\varepsilon+2MP\left(\left|\frac{S_n}{n}-x\right|\ge\delta(\varepsilon)\right).$$

But x is the mean of S_n/n, and the variance of S_n/n is $nx(1-x)\le n/4$. Therefore, by Chebyshev's inequality,

$$P\left(\left|\frac{S_n}{n}-x\right|\ge\delta(\varepsilon)\right)\le\frac{4}{n[\delta(\varepsilon)]^2}.$$

Finally

$$|f(x)-P_n(x)|\le\varepsilon+\frac{4}{n[\delta(\varepsilon)]^2}.$$

Since $\varepsilon>0$ is otherwise arbitrary, this suffices to prove the convergence in $(\star)$. The convergence is *uniform* since the right-hand side of the latter inequality does not depend on $x\in[0,1]$.

3.1.2 Jensen's Inequality

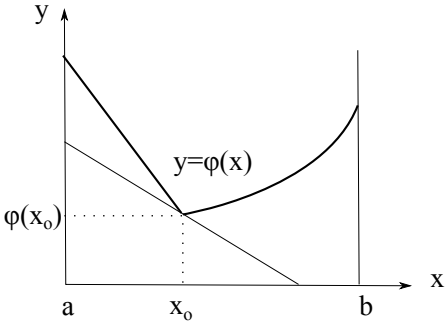
Jensen's inequality concerns the expectation of convex functions of a random variable. We therefore start by recalling the definition of a convex function. Let I be an interval of $\mathbb{R}$ (closed, open, semi-closed, infinite, etc.) with non-empty interior (a,b). The function $\varphi: I\to\mathbb{R}$ is called a convex function if for all $x,y\in I$, all $0<\theta<1$,

$$\varphi(\theta x+(1-\theta)y)\le\theta\varphi(x)+(1-\theta)\varphi(y)\,.$$

If the inequality is strict for all $x\neq y$ and all $0<\theta<1$, the function φ is said to be strictly convex.

Theorem 3.1.5 *Let I be as above and let $\varphi: I\to\mathbb{R}$ be a convex function. Let X be an integrable real-valued random variable such that $P(X\in I)=1$. Assume moreover that either φ is non-negative, or that $\varphi(X)$ is integrable. Then (Jensen's inequality)*

$$E\left[\varphi(X)\right]\ge\varphi(E\left[X\right]).$$



Proof. A convex function φ has the property that for any $x_0 \in (a,b)$, there exists a straight line $y = \alpha x + \beta$, passing through $(x_0, \varphi(x_0))$, that is

$$\varphi(x_0) = \alpha x_0 + \beta\,, \tag{$\star$}$$

and such that for all $x \in (a,b)$,

$$\varphi(x) \geq \alpha x + \beta\,, \tag{$\star\star$}$$

where the inequality is strict if φ is strictly convex. (The parameters α and β may depend on x_0 and may not be unique.) Take $x_0 = E[X]$. In particular $\varphi(E[X]) = \alpha E[X] + \beta$. By $(\star\star)$, $\varphi(X) \geq \alpha X + \beta$, and taking expectations using $(\star)$,

$$E[\varphi(X)] \geq \alpha E[X] + \beta = \varphi(E[X])\,.$$

□

EXAMPLE 3.1.6: THE ARITHMETIC-GEOMETRIC INEQUALITY. Let x_i $(1 \leq i \leq n)$ be positive numbers, and let p_i $(1 \leq i \leq n)$ be non-negative numbers such that $\sum_{i=1}^n p_i = 1$. Then

$$p_1 x_1 + p_2 x_2 + \cdots + p_n x_n \geq x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n}\,.$$

Proof. Letting X be a random variable taking the values x_i with probability p_i $(1 \leq i \leq n)$, Jensen's inequality applied to the convex function $\varphi = -\log$ gives

$$\log E[X] \geq E[\log X]\,,$$

that is

$$\log(p_1x_1+p_2x_2+\cdots+p_nx_n) \geq p_1 \log x_1+p_2 \log x_2+\cdots+p_n \log x_n = \log\left(x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n}\right)\,,$$

hence the result since log is an increasing function. The special case $p_i = \frac{1}{n}$ is worth highlighting:

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n) \geq (x_1 x_2 \cdots x_n)^{\frac{1}{n}}\,.$$

□

3.1.3 Schwarz's Inequality

Definition 3.1.7 *A random variable X taking real values is called integrable if $E[|X|] < \infty$ and square-integrable if $E\left[|X|^2\right] < \infty$.*

From the inequality $|a| \leq 1 + a^2$, true for all $a \in \overline{\mathbb{R}}$, we have that $|X| \leq 1 + X^2$, and therefore, by the monotonicity and linearity properties of expectation, and the fact that $E[1] = 1$, $E[|X|] \leq 1 + E[X^2]$. Therefore a square-integrable random variable is integrable.

Theorem 3.1.8 *Let X and Y be square-integrable real-valued random variables. Then, the random variable XY is integrable and (Schwarz's inequality)*

$$E\left[|XY|\right] \leq \left(E\left[X^2\right]\right)^{\frac{1}{2}} \left(E\left[Y^2\right]\right)^{\frac{1}{2}}$$

with equality if and only there exists a, $b \in \mathbb{R}$ such that $aX+bY = 0$ almost surely.

Proof. Taking expectations in the inequality $2|XY| \leq X^2 + Y^2$, we obtain $2E[|XY|] \leq E[X^2] + E[Y^2] < \infty$. We may suppose that $E\left[X^2\right] > 0$, since in the other case $X = 0$ almost surely and Schwarz's inequality is trivially satisfied. Also, we may suppose that X and Y are non-negative as they intervene only through their absolute values. For any $\lambda \in \mathbb{R}$,

$$E\left[X^2\right] + 2\lambda E\left[XY\right] + \lambda^2 E\left[Y^2\right] = E\left[\left(X + \lambda Y\right)^2\right] \geq 0.$$

Therefore the discriminant of this binomial in λ must be negative or null:

$$E\left[XY\right]^2 - E\left[X^2\right] E\left[Y^2\right] \leq 0\,,$$

and this is Schwarz's inequality. It is null if and only there exists a real root, that is if and only for some $\lambda_0 \in \mathbb{R}$, $E\left[\left(X + \lambda_0 Y\right)^2\right] = 0$ or, by Theorem 2.1.27, $X + \lambda_0 Y = 0$ almost surely. □

The Correlation Coefficient

Let X and Y be two real-valued square-integrable random variables with respective means μ_X and μ_Y, and respective variances σ_X^2 and σ_Y^2. Their covariance is, by definition, the number

$$\sigma_{XY} := E\left[(X - \mu_X)(Y - \mu_Y)\right]\,.$$

If X and Y both have positive variances, their intercorrelation is by definition the number

$$\rho_{XY} := \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

By Schwarz's inequality, $|\rho_{XY}| \leq 1$, and $\rho_{XY} = 0$ if and only if there exists $a, b \in \mathbb{R}$ such that

$$a(X - \mu_X) + b(Y - \mu_Y) = 0\,.$$

If $\rho_{XY} > 0$ (*resp.*, < 0, $= 0$) the random variables are said to be negatively correlated (*resp.*, positively correlated, uncorrelated).

3.2 Frequently Used Bounds

3.2.1 The Union Bound

This quite elementary and nevertheless very useful bound is just another name for the sub-sigma-additivity property of probability (Theorem 1.2.7)

$$P(\cup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k).$$

EXAMPLE 3.2.1: THE COUPON COLLECTOR, TAKE 2. Recall that the average value of the number X of chocolate tablets needed to have the complete collection of n coupons is $E[X] = nH(n)$, where

$$H(n) := \sum_{i=1}^{n} \frac{1}{i}.$$

Recall (18.1.11):

$$|H(n) - \log n| \leq 1.$$

Therefore $E[X] = (1 + o(1))n \log n$. More precisely

$$\log n - 1 \leq \frac{E[X]}{n} \leq \log n + 1.$$

We now estimate the deviation of X from its mean. For all $c > 0$,

$$P\left(X > \lceil n \log n + cn \rceil\right) \leq e^{-c}.$$

To prove this, define A_α to be the event that no coupon of type α shows up in the first $\lceil n \log n + cn \rceil$ tablets. Then

$$\begin{aligned} P\left(X > \lceil n \log n + cn \rceil\right) = P\left(\cup_{\alpha=1}^{n} A_\alpha\right) &\leq \sum_{\alpha=1}^{n} P\left(A_\alpha\right) \\ &= \sum_{\alpha=1}^{n} \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} = n \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil}, \end{aligned}$$

and therefore, since $1 + x \leq e^x$ for all $x \in \mathbb{R}$,

$$\begin{aligned} P\left(X > \lceil n \log n + cn \rceil\right) &\leq n \left(e^{-\frac{1}{n}}\right)^{\lceil n \log n + cn \rceil} \\ &\leq n \left(e^{-\frac{1}{n}}\right)^{n \log n + cn} \\ &= n e^{-\log n - c} = n e^{-\log n} e^{-c} = e^{-c}. \end{aligned}$$

3.2.2 The Chernoff Bounds

These powerful bounds are obtained by a clever use of the elementary Markov inequality.

Theorem 3.2.2 *Let X be a discrete real-valued random variable and let $a \in \mathbb{R}$. Then*

$$P\left(X \geq a\right) \leq \min_{t>0} \frac{E\left[e^{tX}\right]}{e^{ta}}\,, \tag{3.1}$$

and

$$P\left(X \leq a\right) \leq \min_{t<0} \frac{E\left[e^{tX}\right]}{e^{ta}}\,. \tag{3.2}$$

Proof. By the increasing monotonicity of the function $x \to e^x$ and Markov's inequality, for all $t > 0$,

$$P\left(X \geq a\right) = P(e^{tX} \geq e^{ta}) \leq \frac{E\left[e^{tX}\right]}{e^{ta}}\,,$$

and for all $t < 0$,

$$P\left(X \leq a\right) = P(e^{tX} \geq e^{ta}) \leq \frac{E\left[e^{tX}\right]}{e^{ta}}\,.$$

The announced result follows by minimizing the right-hand sides of the above inequalities with respect to $t > 0$ and $t < 0$ respectively. □

Theorem 3.2.3 *Let $X_1, \ldots, X_n$ be* IID *discrete real-valued random variables and let $a \in \mathbb{R}$. Then*

$$P\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-nh^+(a)}\,,$$

where

$$h^+(a) := \sup_{t>0}\{at - \log E\left[e^{tX_1}\right]\}\,. \tag{3.3}$$

Proof. First observe that since the X_i's are independent and identically distributed,

$$E\left[\exp\left\{t\sum_{i=1}^{n} X_i\right\}\right] = E\left[\exp\{tX_1\}\right]^n\,.$$

For all $t > 0$, Markov's inequality gives

$$\begin{aligned} P\left(\sum_{i=1}^{n} X_i \geq na\right) &= P\left(e^{t\sum_{i=1}^{n} X_i} \geq e^{nta}\right) \\ &\leq E\left[e^{t\sum_{i=1}^{n} X_i}\right] \times e^{-nta} \\ &= E\left[e^{tX_1}\right]^n \times e^{-nta} \\ &= \exp\{-n\left(at - \log E\left[e^{tX_1}\right]\right)\}\,, \end{aligned}$$

from which the result follows by optimizing this bound with respect to $t > 0$. □

Remark 3.2.4 Of course this bound is useful only if $h^+(a)$ is positive. Suppose for instance that the X_i's are bounded. Let x_i $(i \geq 1)$ be an enumeration of the values taken by X_1, and define $p_i = P(X = x_i)$, so that

$$at - \log E\left[e^{tX_1}\right] = at - \log\left(\sum_{i\geq 1} p_i e^{tx_i}\right).$$

The derivative with respect to t of this quantity is

$$a - \frac{\sum_{i\geq 1} p_i x_i e^{tx_i}}{\sum_{i\geq 1} p_i e^{tx_i}}$$

and therefore the function $t \to at - \log E\left[e^{tX_1}\right]$ is finite and differentiable on $\mathbb{R}$, with derivative at 0 equal to $a - E[X_1]$, which implies that when $a > E[X_1]$, $h^+(a)$ is positive[1].

Similarly to (3.3), we obtain that

$$P\left(\sum_{i=1}^{n} X_i \leq na\right) \leq e^{-nh^-(a)}, \tag{3.4}$$

where $h^-(a) := \sup_{t<0}\{at - \log E\left[e^{tX_1}\right]\}$, and moreover, $h^-(a)$ is positive if $a < E[X_1]$.

EXAMPLE 3.2.5: SIMPLIFIED CHERNOFF BOUND. The computation of the supremum in (3.3) may be fastidious, and shortcuts leading to practical bounds not as tight but nevertheless satisfactory are welcome. Suppose for instance that the X_i's take the values -1 and $+1$ equiprobably, and therefore, for all $t > 0$, $E\left[e^{tX}\right] = \frac{1}{2}e^{+t} + \frac{1}{2}e^{-t}$. We do not keep this expression as such but instead replace it by its upper bound $e^{\frac{t^2}{2}}$. (This bound is obtained from the following calculations:

$$\frac{1}{2}e^{-a} + \frac{1}{2}e^{+a} = \sum_{i\geq 0} \frac{a^{2i}}{(2i)!} \leq \sum_{i\geq 0} \frac{a^{2i}}{i!2^i} = e^{\frac{1}{2}a^2}.)$$

Therefore, for $a > 0$,

$$P\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-n\left(at - \log E\left[e^{tX_1}\right]\right)} \leq e^{-n\left(at - \frac{1}{2}t^2\right)},$$

so that, with $t = a$,

$$P\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-n\frac{1}{2}a^2}.$$

By symmetry of the distribution of $\sum_{i=1}^{n} X_i$, we have for $a > 0$,

[1] In fact, the boundedness assumption can be relaxed and replaced by $E\left[e^{tX_1}\right] < \infty$ for all $t \geq 0$, and even by any assumption guaranteeing that $t \to \sum_{i\geq 1} p_i e^{tx_i}$ is differentiable in a neighborhood of zero with a derivative equal to $\sum_{i\geq 1} p_i x_i e^{tx_i}$. See Exercise 4.3.9.

$$P\left(\sum_{i=1}^{n} X_i \leq -na\right) = P\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-n\frac{1}{2}a^2},$$

and therefore, combining the two inequalities above,

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq na\right) \leq 2e^{-n\frac{1}{2}a^2}.$$

EXAMPLE 3.2.6: A NON-EQUIPROBABLE CASE. Let $X = \sum_{i=1}^{n} X_i$ where the X_i's are independent (but not necessarily equiprobable) random variables taking their values in $\{0, 1\}$. Denote by μ the mean of X.

A. For all $\delta > 0$,

$$P(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}. \tag{3.5}$$

B. For all $\delta \in (0, 1]$,

$$P(X \geq (1+\delta)\mu) \leq e^{-\mu\frac{\delta^2}{3}}.$$

C. For all $\delta \in (0, 1]$,

$$P(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}.$$

D. For all $\delta \in (0, 1]$

$$P(X \leq (1-\delta)\mu) \leq e^{-\mu\frac{\delta^2}{3}}.$$

Combining bounds B and D yields for all $\delta \in (0, 1]$

$$P(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\frac{\delta^2}{3}}. \tag{3.6}$$

Proof. With $p_i := P(X_i = 1)$, we have that $\mu = \sum_{i=1}^{n} p_i$ and

$$E\left[e^{tX_i}\right] = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)},$$

where we have used the standard inequality $1 + x \leq e^x$. By the independence assumption,

$$E\left[e^{tX}\right] = E\left[e^{t(\sum_{i=1}^{n} X_i)}\right] = \prod_{i=1}^{n} E\left[e^{tX_i}\right]$$

and therefore

$$E\left[e^{tX}\right] \leq \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{\mu(e^t-1)}.$$

We now proceed to the proofs of A and B.

A. By Markov's inequality, for all $t > 0$

$$P(X \geq (1+\delta)\mu) = P(e^{tX} \geq e^{t(1+\delta)\mu}) \leq \frac{E\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} \leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}},$$

and the choice $t = \log(1+\delta)$ (> 0 when $\delta > 0$) gives the announced inequality.

B. It suffices to prove that, if $\delta \in (0, 1]$,

$$\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \leq e^{-\frac{\delta^2}{3}}$$

or, equivalently, passing to the log,

$$f(\delta) := \delta - (1+\delta)\log(1+\delta) + \frac{1}{3}\delta^2 \leq 0\,.$$

Computing the first derivative $f'(\delta) = -\log(1+\delta) + \frac{2}{3}\delta$ and the second derivative $f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$, we see that the latter is negative for $\delta \in [0, \frac{1}{2})$ and positive for $\delta > \frac{1}{2}$. Therefore, f' starts by increasing and then decreases. Since $f'(0) = 0$ and $f'(1) < 0$, it is non-negative for all $\delta \in [0, 1]$. Therefore f decreases on $[0, 1]$. As $f(0) = 0$, $f(\delta) \leq 0$ for all $\delta \in [0, 1]$.

C and D are proved with similar arguments. □

Remark 3.2.7 Concerning Chernoff bounds, the situation of interest is in general when n is large. However, there are situations where n is not necessarily large (tending to infinity). In this case, having the n in the numerator of the exponent is not what is important. Consider for example the bound obtained in Example 3.2.5. If we replace therein a by $\frac{a}{n}$, we obtain

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq a\right) \leq 2e^{-\frac{a^2}{2n}}\,. \qquad (\star)$$

The following example features an application of this bound.

EXAMPLE 3.2.8: SET BALANCING. (Mitzenmacher and Upfal, 2005) Let $A = \{a_{i,j}\}_{1\leq i\leq n, 1\leq j\leq m}$ be a given $n \times m$ matrix with entries in $\{0, 1\}$. Let $\{b_j\}_{1\leq j\leq m}$ be a column vector with entries in $\{-1, +1\}$ and let $\{c_i\}_{1\leq i\leq n}$ be the column vector defined by $c = Ab$.

An interpretation of the above mathematical objects is as follows. The j-th column of A, $\{a_{i,j}\}_{1\leq i\leq m}$ describes subject j in some statistical experiment, who has property $\mathcal{P}_i$ if and only $a_{i,j} = 1$. For instance, in a medical experiment, the properties, or features, of a subject could be "smoker", "drinker", or any aggravating factor in cardiovascular problems. Vector b defines two categories of subjects. If $b_i = -1$,

subject i is placed in the *placebo* group, whereas if $b_i = +1$, he is administered an experimental pill. Each feature should be, as much as possible, roughly equally represented in the two subgroups. A small value of the modulus $|c_i|$ of i-th entry therefore indicates a fair balance of feature i. One is therefore looking for a choice of vector b that minimizes $||c||_\infty := \max_{1\le i\le n} |c_i|$.

The following method for finding an approximate solution is rather simple: draw the values b_i's independently and equiprobably in $\{-1, +1\}$. How good is this solution? We prove that

$$P\left(||c||_\infty \ge \sqrt{4m\log n}\right) \le \frac{2}{n}\,.$$

Let k be the number of elements equal to 1 in the i-th row of A. If $k \le \sqrt{4m\log n}$, then $|c_i| \le \sqrt{4m\log n}$. If $k > \sqrt{4m\log n}$, c_i is the sum of k IID random variables taking equiprobably the values -1 and $+1$. Therefore, from the bound $(\star)$, since $k \le m$,

$$P\left(|c_i| \ge \sqrt{4m\log n}\right) \le 2e^{-\frac{4m\log n}{2k}} \le 2e^{-2\log n} = \frac{2}{n^2}\,.$$

By the union bound,

$$\begin{aligned} P\left(||c||_\infty \ge \sqrt{4m\log n}\right) &= P\left(\cup_{i=1}^n \{|c_i| \ge \sqrt{4m\log n}\}\right) \\ &\le \sum_{i=1}^n P\left(|c_i| \ge \sqrt{4m\log n}\right) \le \frac{2}{n}\,. \end{aligned}$$

3.2.3 The First- and Second-moment Bounds

These bounds will be particularly efficient in the asymptotic study of random graphs.

Lemma 3.2.9 *For any integer-valued random variable X,*

$$P(X \ne 0) \le E[X]\,.$$

Proof.

$$\begin{aligned} P(X \ne 0) &= P(X=1) + P(X=2) + P(X=3) + \cdots \\ &\le P(X=1) + 2P(X=2) + 3P(X=3) + \cdots = E[X]\,. \end{aligned}$$

□

Lemma 3.2.10 *For any square-integrable real-valued discrete random variable X,*

$$P(X=0) \le \frac{Var\,(X)}{E[X]^2}\,.$$

Proof. Since the event $X = 0$ implies the event $|X - E[X]| \geq E[X]$,

$$P(X = 0) \leq P(|X - E[X]| \geq E[X]) \leq \frac{\text{Var}\,(X)}{E[X]^2}\,,$$

where the last inequality is Chebyshev's inequality (Corollary 3.1.2). □

Lemma 3.2.11 *For a square-integrable integer-valued discrete random variable X,*

$$P(X = 0) \leq \frac{Var(X)}{E[X^2]}\,.$$

Proof. By Schwarz's inequality,

$$\begin{aligned} E\left[X\right]^2 &= E\left[\left(X 1_{\{X\geq 1\}}\right]^2 \leq E\left[X^2\right] E\left[(1_{\{X\geq 1\}})^2\right] \\ &= E\left[X^2\right] E\left[1_{\{X\geq 1\}}\right] = E\left[X^2\right] P(X \geq 1) = E\left[X^2\right](1 - P(X = 0))\,, \end{aligned}$$

and therefore $E\left[X^2\right] - E\left[X\right]^2 \geq P(X = 0) E\left[X^2\right]$. □

The bound of Lemma 3.2.11 is tighter than that of Lemma 3.2.10 since $E\left[X^2\right] \geq E\left[X\right]^2$ (by Jensen's inequality for instance, or Schwarz's inequality).

Remark 3.2.12 The above bounds will be used, mainly in the asymptotic analysis of random graphs, as follows. Suppose there is a sequence of integer-valued random variables $\{X_n\}_{n\geq 1}$, for instance, counting the number of cycles in a random graph $\mathcal{G}(n, p_n)$. If we can show that $\lim_{n\uparrow\infty} E\left[X_n\right] = 0$, then we can assert that $\lim_{n\uparrow\infty} P(X_n = 0) = 0$: "asymptotically" there exists no cycle in the random graph. This of course requires conditions on the asymptotic behaviour of the sequence $\{p_n\}_{n\geq 1}$. Under other circumstances, suppose that $\lim_{n\uparrow\infty} E\left[X_n\right] = \infty$. Does this imply that $\lim_{n\uparrow\infty} P(X_n > 0) = 1$? In fact, not (Exercise 3.3.1). Therefore we have to find another way, for instance via Lemma 3.2.10, of proving that $\lim_{n\uparrow\infty} \frac{\text{Var}(X_n)}{E[X_n]^2} = 0$.

Books for Furher Information

The classical bounds appear in virtually all textbooks of probability theory. [Mitzenmacher and Upfal, 2005], especially Chapter 4 therein on Chernoff bounds, contains examples of interest in the information and computing sciences.

3.3 Exercises

Exercise 3.3.1. When the first moment bound is not enough
Give a (very simple) example showing that for a sequence of integer-valued random variables $\{X_n\}_{n\geq 1}$, the fact that $E[X_n] \to +\infty$ is not sufficient to guarantee that $P(X_n \neq 0) \to 1$.

Exercise 3.3.2. Existence of triangles
A "triangle" of a graph is the obvious object: 3 vertices mutually linked by an edge. It is also called a 3-clique. Let $\mathcal{G}(n, p_n)$ be an Erdös–Rényi graph with $p_n = \frac{d}{n}$. Prove that "asymptotically almost surely", there is at least one triangle if $d \geq 6^{\frac{1}{3}}$. In other words, for such values of the average index d, $P(X_n = 0) < 1$ for sufficiently large n, where X_n is the number of triangles of the random graph. (Hint: Lemma 3.2.11.)

Exercise 3.3.3. Tail of the Poisson distribution
Let X be a Poisson variable with mean θ and therefore $E\left[e^{tX}\right] = e^{\theta(e^t-1)}$. Prove that for $c \geq 0$

$$P\left(X \geq \theta + c\right) \leq \exp\left\{-\frac{1}{e}\left(\frac{\theta + c}{e\theta}\right)^{\theta+c}\right\}.$$

Exercise 3.3.4. Large deviations without independence
Let $X = \sum_{i=1}^{n} X_i$ where the X_i's are independent but not necessarily identically distributed random variables taking their values in $\{0, 1\}$. Denote by μ the mean of X. Therefore, with $P(X_i = 1) = p_i$, $\mu = \sum_{i=1}^{n} p_i$. Then, for any $\varepsilon > 0$,

$$P(X - \mu \geq \varepsilon\mu) \leq e^{\mu h(\varepsilon)}$$

and

$$P(X - \mu \leq \varepsilon\mu) \leq e^{\mu h(-\varepsilon)}$$

where $h(\varepsilon) = (1 + \varepsilon)\log(1 + \varepsilon) - \varepsilon$.

Exercise 3.3.5. Degree of a random graph
Let $\mathcal{G}(n, p_n)$ be an Erdös–Rényi graph with set of vertices V (of size n). The degree $d(v)$ of any vertex v is a binomial random variable $\mathcal{B}(n, p_n)$ of mean $\overline{d} = (n-1)p_n$. Assume that $\lim_{n\uparrow\infty} \frac{\log n}{(n-1)p_n} = 0$. Prove that the probability that one cannot find a node that deviates from the mean $(n-1)p_n$ by a factor larger that $2\sqrt{\frac{\log n}{(n-1)p_n}}$ tends to 1 as $n \uparrow \infty$. (Hint: apply the bound of Exercise 3.3.4.)

Exercise 3.3.6. Large deviations of a fair coin
Prove that for a sequence of n independent coin flips of a fair coin,

$$P\left(|X - \tfrac{n}{2}| \geq \tfrac{1}{2}\sqrt{6n\log n}\right) \leq \tfrac{2}{n},$$

where X is the number of heads.

Chapter 4

Almost Sure Convergence

4.1 Conditions for Almost Sure Convergence

4.1.1 A Sufficient Condition

This chapter gives the basic theory of almost sure convergence and Kolmogorov's strong law of large numbers (1933) according to which the empirical mean of an IID sequence of integrable random variables converges almost surely to the probabilistic mean (the expectation).

The emblematic result of Probability theory features a game of *heads and tails* with a single, possibly biased, coin. Émile Borel proved in 1909 that the *empirical frequency* of occurences of heads in this sequence converges almost surely to the bias p of the coin. More precisely, let $\{X_n\}_{n\geq 1}$ be an IID sequence of random variables taking the values (1 for "heads") and 0 (for "tails") with $P(X_n = 1) = p$, where $p \in (0,1)$ is the bias of the coin. Then:

$$P\left(\lim_{n\uparrow\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = p\right) = 1\,. \tag{4.1}$$

Definition 4.1.1 *A sequence $\{Z_n\}_{n\geq 1}$ of random variables with values in $\mathbb{C}$ (resp., $\overline{\mathbb{R}}$), is said to converge P-almost surely (P-a.s.) to the random variable Z with values in $\mathbb{C}$ (resp., $\overline{\mathbb{R}}$) if*

$$P(\lim_{n\uparrow\infty} Z_n = Z) = 1. \tag{4.2}$$

Paraphrasing: For all ω outside a set N of null probability, $\lim_{n\uparrow\infty} Z_n(\omega) = Z(\omega)$.

The specification "discrete" for the random variables considered in this chapter and elsewhere in this book will be omitted only when the definitions, results or proofs apply to general random variables, although the proofs are given only for discrete random variables. This is the case for instance for the forthcoming Kolmogorov's strong law of large numbers and martingale convergence theorem.

A notion closely related to that of almost sure convergence is that of convergence in probability.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_4

Definition 4.1.2 *A sequence $\{Z_n\}_{n\geq 1}$ of variables is said to converge in probability to the random variable Z if for all $\varepsilon > 0$,*

$$\lim_{n\uparrow\infty} P(|Z_n - Z| \geq \varepsilon) = 0. \tag{4.3}$$

Theorem 4.1.7 below will show the links between convergence in probability and convergence almost sure.

The following is a useful sufficient condition of almost sure convergence.

Theorem 4.1.3 *Let $\{Z_n\}_{n\geq 1}$ and Z be random variables. If*

$$\sum_{n\geq 1} P(|Z_n - Z| \geq \varepsilon_n) < \infty \tag{4.4}$$

for some sequence of non-negative numbers $\{\varepsilon_n\}_{n\geq 1}$ converging to 0, then the sequence $\{Z_n\}_{n\geq 1}$ converges P-a.s. to Z.

Proof. If for a given ω, $|Z_n(\omega) - Z(\omega)| \geq \varepsilon_n$ finitely often (or *f.o.*; that is, for all but a finite number of indices n), then $\lim_{n\uparrow\infty} |Z_n(\omega) - Z(\omega)| \leq \lim_{n\uparrow\infty} \varepsilon_n = 0$. Therefore

$$P(\lim_{n\uparrow\infty} Z_n = Z) \geq P(|Z_n - Z| \geq \varepsilon_n \quad f.o.).$$

On the other hand,

$$\{|Z_n - Z| \geq \varepsilon_n \quad f.o.\} = \overline{\{|Z_n - Z| \geq \varepsilon_n \quad i.o.\}}.$$

Therefore

$$P(|Z_n - Z| \geq \varepsilon_n \quad f.o.) = 1 - P(|Z_n - Z| \geq \varepsilon_n \quad i.o.).$$

Hypothesis (4.4) implies (Borel–Cantelli lemma) that

$$P(|Z_n - Z| \geq \varepsilon_n \quad i.o.) = 0.$$

By linking the above facts, we obtain $P(\lim_{n\uparrow\infty} Z_n = Z) \geq 1$, and of course, the only possibility is $= 1$. □

EXAMPLE 4.1.4: BOREL'S STRONG LAW OF LARGE NUMBERS. Theorem 4.1.3 will now be applied to prove (4.1). For this, we bound the probability that $\left|\frac{S_n}{n} - p\right|$ exceeds some $\varepsilon > 0$. By Markov's inequality (Theorem 3.1.1)

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = P\left(\left(\frac{S_n}{n} - p\right)^4 \geq \varepsilon^4\right)$$
$$\leq \frac{E\left[\left(\frac{S_n}{n} - p\right)^4\right]}{\varepsilon^4} = \frac{E\left[\left(\sum_{i=1}^n (X_i - p)\right)^4\right]}{n^4\varepsilon^4}.$$

To simplify the notation, call Y_i the random variables $X_i - p$, and remember that

$$E[Y_i] = 0.$$

Also, in view of the independence hypothesis,

$$E[Y_1Y_2Y_3Y_4] = E[Y_1]E[Y_2]E[Y_3]E[Y_4] = 0\,,$$

$$E[Y_1Y_2^3] = E[Y_1]E[Y_2^3] = 0\,,$$

and the like. Finally, in the development

$$E\left[\left(\sum_{i=1}^{n} Y_i\right)^4\right] = \sum_{i,j,k,\ell=1}^{n} E[Y_iY_jY_kY_\ell]\,,$$

only the terms of the form $E[Y_i^4]$ and $E[Y_i^2Y_j^2]$ $(i \neq j)$ remain. There are n terms of the first type and $3n(n-1)$ terms of the second type. Therefore,

$$E\left[\left(\sum_{i=1}^{n} Y_i\right)^4\right] = nE[Y_1^4] + 3n(n-1)E[Y_1^2Y_2^2] \leq Kn^2\,,$$

for some finite K. Therefore

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{K}{n^2\varepsilon^4},$$

and in particular, with $\varepsilon = n^{-\frac{1}{8}}$,

$$P\left(\left|\frac{S_n}{n} - p\right| \geq n^{-\frac{1}{8}}\right) \leq \frac{K}{n^{\frac{3}{2}}}\,,$$

from which it follows that

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq n^{-\frac{1}{8}}\right) < \infty\,.$$

Therefore, by Theorem 4.1.3, $\left|\frac{S_n}{n} - p\right|$ converges almost surely to 0.

EXAMPLE 4.1.5: BOREL SEQUENCES. Let $(X_n,\, n \geq 1)$, be an IID sequence of 0's and 1's such that $P(X_1 = 1) = p \in (0,1)$ (a Bernoulli sequence). Let $(n_i,\, 1 \leq i \leq k)$ be a strictly increasing finite sequence of positive integers. Let $(\varepsilon_i,\, 1 \leq i \leq k)$ be a sequence of 0's and 1's. The pair $(n_i,\, \varepsilon_i,\, 1 \leq i \leq k)$ is called a pattern. Define for all $n \geq 1$ the random variable with values 0 or 1 such that

$$Y_n = 1 \text{ iff} \quad X_{n+n_i} = \varepsilon_i \text{ for all } i \quad (1 \leq i \leq k)\,.$$

Then it can be shown (Exercise 4.3.4) that

$$\frac{Y_1 + \ldots + Y_n}{n} \overset{\text{p.s.}}{\to} p^h q^{k-h} \qquad \text{where} \quad h := \sum_{i=1}^{k} \varepsilon_i\,. \tag{$\star$}$$

Since the Bernoulli sequence with $p = \frac{1}{2}$ — the random sequence "par excellence" — satisfies ($\star$) for all possible patterns, Borel had the idea of defining a deterministic sequence $(x_n,\, n \geq 1)$ of 0's and 1's as "random" if for all patterns

$$\lim_{n\uparrow\infty} \frac{y_1 + \ldots + y_n}{n} = \frac{1}{2^k}\ ,$$

where the y_n's are defined in the same way as the Y_n's above. Although this definition seems reasonable, it is not satisfactory. In fact, one can show that the so-called Champernowne sequence

$$0110111001011101111000\ \ldots\ ,$$

consisting of the integers written in natural order and in binary digits (starting with 0) is random in the Borel sense.

4.1.2 A Criterion

The sufficient condition of almost sure convergence of Theorem 4.1.3 is often all that one needs in practice. The following theorem is a *criterion* (necessary and sufficient condition) of convergence. Its interest is mainly theoretical. In particular it will be used in Theorem 4.1.7 below to prove that convergence in probability is a weaker notion than convergence almost sure convergence, but not much weaker.

Theorem 4.1.6 *The sequence $\{Z_n\}_{n\geq 1}$ of complex random variables converges P-a.s. to the complex random variable Z if and only if for all $\epsilon > 0$,*

$$P(|Z_n - Z| \geq \epsilon \ i.o.) = 0. \tag{4.5}$$

Proof. For the necessity, observe that

$$\{|Z_n - Z| \geq \epsilon \ i.o.\} \subseteq \overline{\{\omega; \lim_{n\uparrow\infty} Z_n(\omega) = Z(\omega)\}},$$

and therefore

$$P(|Z_n - Z| \geq \epsilon \ i.o.) \leq 1 - P(\lim_{n\uparrow\infty} Z_n = Z) = 0.$$

For the sufficiency, let N_k be the last index n such that $|Z_n - Z| \geq \frac{1}{k}$ (let $N_k = \infty$ if $|Z_n - Z| \geq \frac{1}{k}$ for an infinity of indices $n \geq 1$). By (4.5) with $\epsilon = \frac{1}{k}$, we have $P(N_k = \infty) = 0$. By sub-σ-additivity, $P(\cup_{k\geq 1}\{N_k = \infty\}) = 0$. Equivalently, $P(N_k < \infty, \ \text{for all } k \geq 1) = 1$, which implies $P(\lim_{n\uparrow\infty} Z_n = Z) = 1$. □

Theorem 4.1.7 *A. If the sequence $\{Z_n\}_{n\geq 1}$ of complex random variables converges almost surely to some complex random variable Z, it also converges in probability to the same random variable Z.*
B. If the sequence of complex random variables $\{Z_n\}_{n\geq 1}$ converges in probability to the complex random variable Z, one can find a sequence of integers $\{n_k\}_{k\geq 1}$, strictly increasing, such that $\{Z_{n_k}\}_{k\geq 1}$ converges almost surely to Z.

(B says, in other words: From a sequence converging in probability, one can extract a subsequence converging almost surely.)

Proof. A. Suppose almost sure convergence. By Theorem 4.1.6 , for all $\varepsilon > 0$,

$$P(|Z_n - Z| \geq \varepsilon \ i.o.) = 0,$$

that is

$$P(\cap_{n\geq 1} \cup_{k=n}^{\infty} (|Z_k - Z| \geq \varepsilon)) = 0,$$

or (sequential continuity of probability)

$$\lim_{n\uparrow\infty} P(\cup_{k=n}^{\infty} (|Z_k - Z| \geq \varepsilon)) = 0,$$

which in turn implies that

$$\lim_{n\uparrow\infty} P(|Z_n - Z| \geq \varepsilon) = 0\,.$$

B. By definition of convergence in probability, for all $\varepsilon > 0$,

$$\lim_{n\uparrow\infty} P\left(|Z_n - Z| \geq \varepsilon\right) = 0\,.$$

Therefore one can find n_1 such that

$$P\left(|Z_{n_1} - Z| \geq \frac{1}{1}\right) \leq \left(\frac{1}{2}\right)^1.$$

Then, one can find $n_2 > n_1$ such that

$$P\left(|Z_{n_2} - Z| \geq \frac{1}{2}\right) \leq \left(\frac{1}{2}\right)^2$$

and so on, until we have a strictly increasing sequence of integers n_k, $k \geq 1$ such that

$$P\left(|Z_{n_k} - Z| \geq \frac{1}{k}\right) \leq \left(\frac{1}{2}\right)^k.$$

It then follows from Theorem 4.1.3 that $\lim_{k\uparrow\infty} Z_{n_k} = Z$ a.s. □

Exercise 4.3.8 gives an example of a sequence converging in probability, but not almost surely. Thus, convergence in probability is a notion strictly weaker than almost sure convergence.

4.1.3 Convergence under the Expectation Sign

Lebesgue's Theorem for Series

Given a sequence $\{X_n\}_{n\geq 1}$ of random variables, one seeks conditions guaranteeing that, provided the limits thereafter exist,

$$\lim_{n\uparrow\infty} E\left[X_n\right] = E\left[\lim_{n\uparrow\infty} X_n\right]. \tag{4.6}$$

We start by giving a simple example where this is not true.

EXAMPLE 4.1.8: Let X be an integer-valued random variable with the probability distribution

$$P(X=k) = e^{-k\alpha}\left(1-e^{-\alpha}\right),$$

where $\alpha > 0$. Define for all $n \geq 1$,

$$X_n := e^{n\alpha} X 1_{\{X\geq n\}}.$$

Clearly, $\lim_{n\uparrow\infty} X_n := X = 0$. Also

$$E\left[X_n\right] = e^{n\alpha}\sum_{k=n}^{\infty} e^{-k\alpha}\left(1-e^{-\alpha}\right) = 1.$$

In particular,

$$\lim_{n\uparrow\infty} E\left[X_n\right] = 1 \neq 0 = E\left[\lim_{n\uparrow\infty} X_n\right].$$

In the case where the random variables involved and their limits are integer-valued, the anwers can be given as consequences of general results on series. We begin with the dominated convergence theorem for series.

Theorem 4.1.9 *Let $\{a_{nk}\}_{n\geq 1, k\geq 1}$ be an array of real numbers such that for some sequence $\{b_k\}_{k\geq 1}$ of non-negative numbers satisfying $\sum_{k=1}^{\infty} b_k < \infty$, it holds that for all $n \geq 1, k \geq 1$, $|a_{nk}| \leq b_k$. If moreover for all $k \geq 1$, $\lim_{n\uparrow\infty} a_{nk} = a_k$, then*

$$\lim_{n\uparrow\infty}\sum_{k=1}^{\infty} a_{nk} = \sum_{k=1}^{\infty} a_k.$$

Proof. See Section A.1. □

EXAMPLE 4.1.10: Let X be a discrete real-valued random variable that is integrable. Then

$$\lim_{n\uparrow\infty} E\left[|X| 1_{\{|X|\geq n\}}\right] = 0.$$

In fact, denoting by x_k, $k \in \mathbb{N}$, the values of X,

$$E\left[|X| 1_{\{|X|\geq n\}}\right] = \sum_{k\in\mathbb{N}} |x_k| 1_{\{x_k\geq n\}} P(X = x_k).$$

It suffices to apply Theorem 4.1.9 with $a_{nk} = x_k 1_{\{x_k\geq n\}} P(X = x_k)$ and $b_k = |x_k| 1_P(X = x_k)$, since $\sum_k b_k = E\left[|X|\right] < \infty$ and $\lim_{n\uparrow\infty} a_{nk} = 0$.

We now recall the monotone convergence theorem for series.

Theorem 4.1.11 *Let $\{a_{nk}\}_{n\geq 1,k\geq 1}$ be an array of non-negative real numbers such that for all $k \geq 1$, the sequence $\{a_{nk}\}_{n\geq 1}$ is non-decreasing with limit $a_k \leq \infty$. Then*

$$\lim_{n\uparrow\infty} \sum_{k=1}^{\infty} a_{nk} = \sum_{k=1}^{\infty} a_k \, .$$

Proof. See Section A.1. □

Finally, we have Fatou's lemma for series.

Theorem 4.1.12 *Let $\{a_{nk}\}_{n\geq 1,k\geq 1}$ be an array of non-negative real numbers. Then*

$$\sum_{k=1}^{\infty} \liminf_{n\uparrow\infty} a_{nk} \leq \liminf_{n\uparrow\infty} \sum_{k=1}^{\infty} a_{nk}.$$

Proof. See Section A.1. □

The Case of Random Variables

In probability theory, the analogue of the monotone convergence theorem for series is also called Beppo Levi's theorem:

Theorem 4.1.13 *Let $\{X_n\}_{n\geq 1}$ be a non-decreasing sequence of non-negative real-valued (including the infinite value) random variables converging to the real random variable X. Then (4.6) holds true.*

Proof. The proof will be given in the case of integer-valued random variables. From the telescope formula

$$E\left[X_n\right] = \sum_{k\geq 1} P(X_n \geq k).$$

By the assumption of non-decreasingness,

$$P(X \geq k) = \lim_{n\uparrow\infty} P(X_n \geq k)$$

and therefore, by the monotone convergence theorem for series,

$$E\left[X\right] = \sum_{k\geq 1} P(X \geq k) = \lim_{n\uparrow\infty} \sum_{k\geq 1} \uparrow P(X_n \geq k) = \lim_{n\uparrow\infty} E\left[X_n\right] \, .$$

□

From Beppo Levi's theorem, Fatou's lemma for expectations follows almost immediately:

Theorem 4.1.14 *Let $\{X_n\}_{n\geq 1}$ be a sequence of non-negative real-valued (including the infinite value) random variables. Then,*

$$E\left[\liminf_n X_n\right] \leq \liminf_n E\left[X_n\right] .$$

Proof. Let $Y = \liminf_n X_n := \lim_{n\uparrow\infty} \inf_{k\geq n} X_k$. By Beppo Levi's,

$$E\left[Y\right] = \lim_{n\uparrow\infty} E\left[\inf_{k\geq n} X_k\right] .$$

But for all $i \geq n$, by monotonicity of expectation,

$$E\left[\inf_{k\geq n} X_k\right] \leq E\left[X_i\right] ,$$

and therefore

$$E\left[\inf_{k\geq n} X_k\right] \leq \inf_{i\geq n} E\left[X_i\right] .$$

Therefore,

$$E\left[Y\right] = \lim_{n\uparrow\infty} E\left[\inf_{k\geq n} X_k\right] \leq \lim_{n\uparrow\infty} \inf_{i\geq n} E\left[X_i\right] := \liminf_n E\left[X_n\right] .$$

□

Finally, we have the dominated convergence theorem for expectations, also called Lebesgue's theorem:

Theorem 4.1.15 *Let $\{X_n\}_{n\geq 1}$ and X be real random variables such that*

(i) $\lim_{n\uparrow\infty} X_n = X$, and

(ii) there exists a non-negative real random variable Z with finite expectation such that $|X_n| \leq Z$ for all $n \geq 1$. Then (4.6) holds true.

Proof. Apply Fatou's lemma to the (non-negative) sequence $\{Z+X_n\}_{n\geq 1}$ to obtain

$$E\left[Z+X\right] = E\left[\liminf_n (Z+X_n)\right] \leq \liminf_n E\left[(Z+X_n)\right] \leq E\left[Z\right] + \liminf_n E\left[X_n\right] ,$$

that is $E\left[X\right] \leq \liminf_n E\left[X_n\right]$. Replacing Z and X_n by their opposites and using the same argument, we have that $E\left[X\right] \geq \limsup_n E\left[X_n\right]$. □

Remark 4.1.16 The last two results have been proved for random variables taking their values in the set of relative integers including the infinite values, since they depend on Beppo Levi's theorem which was proved only in this case. Note however that once the general version of the monotone convergence theorem is taken for granted, the proofs just given for Fatou's lemma and the dominated convergence theorem remain valid also in the general case, as long as we have a general definition of a random variable and of its expectation. Again, we shall not need this generality in this book.

4.2 Kolmogorov's Strong Law of Large Numbers

4.2.1 The Square-integrable Case

The proof of Borel's strong law of large numbers applies when the X_n's are just supposed uniformly bounded by a deterministic constant (this implying that the moments at any order are finite, and that is all we really need in the proof). In fact, there is a much stronger result, Kolmogorov's strong law of large numbers:

Theorem 4.2.1 *Let $\{X_n\}_{n\geq 1}$ be an* IID *sequence of random variables such that*

$$E[|X_1|] < \infty.$$

Then,

$$P\left(\lim_{n\uparrow\infty} \frac{S_n}{n} = E[X_1]\right) = 1.$$

We shall assume (without loss of generality) that $E[X_1] = 0$. The main ingredient of the proof is Kolmogorov's inequality.

Lemma 4.2.2 *Let $X_1, \ldots, X_n$ be independent random variables such that $E[|X_i|^2] < \infty$ and $E[X_i] = 0$ for all i, $1 \leq i \leq n$. Let $S_k = X_1 + \ldots + X_k$. Then for all $\lambda > 0$,*

$$P\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{E[S_n^2]}{\lambda^2}. \tag{4.7}$$

Proof. Let T be the first (random) index k, $1 \leq k \leq n$, such that $|S_k| \geq \lambda$, with $T = \infty$ if $\max_{1\leq k\leq n} |S_k| < \lambda$. For $k \leq n$,

$$\begin{aligned} E[1_{\{T=k\}} S_n^2] &= E\left[1_{\{T=k\}} \left\{(S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2\right\}\right] \\ &= E\left[1_{\{T=k\}} \left\{(S_n - S_k)^2 + S_k^2\right\}\right] \geq E[1_{\{T=k\}} S_k^2]. \end{aligned}$$

(We used the fact that $1_{\{T=k\}} S_k$ is a function of $X_1, \ldots, X_k$ and therefore independent of $S_n - S_k$, so that, by the product rule for expectations, $E[1_{\{T=k\}} S_k (S_n - S_k)] = E[1_{\{T=k\}} S_k] E[S_n - S_k] = 0$.) Therefore,

$$\begin{aligned} E[|S_n|^2] &\geq \sum_{k=1}^{n} E[1_{\{T=k\}} S_k^2] \\ &\geq \sum_{k=1}^{n} E[1_{\{T=k\}} \lambda^2] = \lambda^2 \sum_{k=1}^{n} P(T = k) \\ &= \lambda^2 P(T \leq n) = \lambda^2 P\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right). \end{aligned}$$

□

The next lemma is already the SLLN under the additional assumption $E[|X_1|^2] < \infty$.

Lemma 4.2.3 *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables such that $E[|X_n|^2] < \infty$ and $E[X_n] = 0$ for all $n \geq 1$. If*

$$\sum\nolimits_{n\geq 1} \frac{E[X_n^2]}{n^2} < \infty\,, \tag{4.8}$$

then $\frac{1}{n}\sum_{k=1}^n X_k \to 0$, P-a.s.

Proof. If $2^{k-1} \leq n \leq 2^k$, then $|S_n| \geq n\varepsilon$ implies $|S_n| \geq 2^{k-1}\varepsilon$. Therefore, for all $\varepsilon > 0$, and all $k \geq 1$,

$$\begin{aligned} P\left(\frac{|S_n|}{n} \geq \varepsilon \text{ for some } n \in [2^{k-1}, 2^k]\right) & \\ &\leq P\left(|S_n| \geq \varepsilon 2^{k-1} \text{ for some } n \in [2^{k-1}, 2^k]\right) \\ &\leq P\left(|S_n| \geq \varepsilon 2^{k-1} \text{ for some } n \in [1, 2^k]\right) \\ &= P\left(\max_{1\leq n\leq 2^k} |S_n| \geq \varepsilon 2^{k-1}\right) \leq \frac{4}{\varepsilon^2}\frac{1}{(2^k)^2}\sum_{n=1}^{2^k} E[X_n^2], \end{aligned}$$

where the last inequality follows from Kolmogorov's inequality. But

$$\begin{aligned} \sum_{k=1}^{\infty}\frac{1}{(2^k)^2}\sum_{n=1}^{2^k} E[X_n^2] &= \sum_{n=1}^{\infty} E[X_n^2]\sum_{k=1}^{\infty} 1_{\{2^k\geq n\}}\frac{1}{(2^k)^2} \\ &\leq \sum_{n=1}^{\infty} E[X_n^2]\frac{K}{n^2} \end{aligned}$$

for some finite K, since

$$\sum_{k=1}^{\infty} 1_{\{2^k\geq n\}}\frac{1}{(2^k)^2} = \sum_{k\geq \log_2 n}\frac{1}{4^k} \leq \frac{1}{n^2}\sum_{k\geq 0}\frac{1}{4^k}\,.$$

Therefore, by (4.8),

$$\sum_{k=1}^{\infty} P\left(\frac{|S_n|}{n} \geq \varepsilon \text{ for some } n \in [2^{k-1}, 2^k]\right) < \infty,$$

and by the Borel–Cantelli lemma,

$$P\left(\frac{|S_n|}{n} \geq \varepsilon \text{ i.o.}\right) = 0.$$

The result then follows from Theorem 4.1.6. □

4.2.2 The General Case

We are now ready for the proof of Theorem 4.2.1.

Proof. It remains to get rid of the assumption $E[|X_n|^2] < \infty$, and the natural technique for this is truncation. Define

$$\tilde{X}_n = \begin{cases} X_n \text{ if } |X_n| \le n, \\ 0 \text{ otherwise.} \end{cases}$$

A. We first show that

$$\lim_{n\uparrow\infty} \frac{1}{n} \sum_{k=1}^{n} (\tilde{X}_k - E[\tilde{X}_k]) = 0\,. \qquad (\star)$$

In view of the preceding corollary, it suffices to prove that

$$\sum_{n=1}^{\infty} \frac{E[(\tilde{X}_n - E[\tilde{X}_n])^2]}{n^2} < \infty.$$

But

$$E[(\tilde{X}_n - E[\tilde{X}_n])^2] = E[(\tilde{X}_n)^2] - E[\tilde{X}_n]^2 \le E[\tilde{X}_n^2] = E[X_1^2 1_{\{|X_1|\le n\}}].$$

It is therefore enough to show that

$$\sum_{n=1}^{\infty} \frac{E[X_1^2 1_{\{|X_1|\le n\}}]}{n^2} < \infty.$$

The left-hand side of the above inequality equals

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} E[X_1^2 1_{\{k-1<|X_1|\le k\}}] = \sum_{k=1}^{\infty} \left(\sum_{n=k}^{\infty} \frac{1}{n^2} \right) E[X_1^2 1_{\{k-1<|X_1|\le k\}}].$$

Using the fact that

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \le \frac{1}{k^2} + \int_k^{\infty} \frac{1}{x^2} dx = \frac{1}{k^2} + \frac{1}{k} \le \frac{2}{k}$$

(draw the graph of $x \to x^{-2}$), this is less than or equal to

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{2}{k} E[X_1^2 1_{\{k-1<|X_1|\le k\}}] &= 2 \sum_{k=1}^{\infty} E\left[\frac{X_1^2}{k} 1_{\{k-1<|X_1|\le k\}} \right] \\ &\le 2 \sum_{k=1}^{\infty} E[|X_1| 1_{\{k-1<|X_1|\le k\}}] = 2E[|X_1|] < \infty. \end{aligned}$$

B. Since X_1 is integrable, $\lim_{n\uparrow\infty} E[X_1 1_{\{|X_1|\le n\}}] = E[X_1]$ by dominated convergence (Example 4.1.10). Now X_n has the same distribution as X_1, and therefore

$$\lim_{n\uparrow\infty} E[\tilde{X}_n] = \lim_{n\uparrow\infty} E[X_n 1_{\{|X_n|\le n\}}] = \lim_{n\uparrow\infty} E[X_1 1_{\{|X_1|\le n\}}] = E[X_1].$$

In particular (Cesaro's lemma),

$$\lim_{n\uparrow\infty} \frac{1}{n} \sum_{k=1}^{n} E[\tilde{X}_k] = 0\,. \qquad (\star\star)$$

C. In view of $(\star)$ and $(\star\star)$, it remains to show that

$$\lim_{n\uparrow\infty} \frac{\tilde{S}_n}{n} = \lim_{n\uparrow\infty} \frac{S_n}{n}\,. \qquad (\star\star\star)$$

We have, by the telescope formula (Theorem 2.1.24),

$$\sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) \leq E[|X_1|] < \infty,$$

and therefore, by Borel–Cantelli's lemma,

$$P(\tilde{X}_n \neq X_n \text{ i.o.}) = P(X_n > n \text{ i.o.}) = 0\,,$$

which implies $(\star\star\star)$. □

Remark 4.2.4 The statement of the strong law of large numbers does not mention any restriction on the range of the random variables concerned, which in this book are discrete. The reason for this omission is that it remains true in the general case, with exactly the same proof. Note however that the above proof of Theorem 4.2.1 features only discrete random variables if the sequence $\{X_n\}_{n\geq 0}$ takes discrete values.

4.3 Exercises

Exercise 4.3.1. The geometric process
Let $\{X_n\}_{n\geq 1}$ be a sequence of IID variables with values in $\{0, 1\}$ such that $P(X_1 = 1) = p$ $(0 < p < 1)$. Let T_k be the k-th index $n \geq 1$ such that $X_n = 1$. Prove that $\lim_{k\uparrow\infty} \frac{T_k}{k} = \frac{1}{p}$.

Exercise 4.3.2. The repair shop
(The title of this exercise refers to a model that will be studied later in more detail.) Consider the recurrence equation

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1} \qquad (n \geq 0)$$

$(a^+ := \sup(a, 0))$ where $X_0 = 0$ and where $\{Z_n\}_{n\geq 1}$ is an IID sequence of random variables with values in $\mathbb{N}$. Denote by T_0 the first index $n \geq 1$ such that $X_n = 0$ ($T_0 = \infty$ if such index does not exist). Show that if $E[Z_1] < 1$, $P(T_0 < \infty) = 1$.

Exercise 4.3.3. Convergence in the quadratic mean
A sequence of square-integrable real random variables $\{Z_n\}_{n\geq 1}$ is said to converge in quadratic mean to the square-integrable real random variable Z if

$\lim_{n\uparrow\infty} E\left[|Z_n - Z|^2\right] = 0$. Show that such a sequence also converges to Z in probability.

Exercise 4.3.4. BOREL SEQUENCES
Prove the convergence result stated in Example 4.1.5.

Exercise 4.3.5. ASYMPTOTICS OF THE POISSON PROCESS
Let $\{S_n\}_{n\geq 1}$ be an IID sequence of real random variables such that $P(0 < S_1 < +\infty) = 1$ and $E[S_1] < \infty$, and let for each $t \geq 0$, $N(t) := \sum_{n\geq 1} 1_{(0,t]}(T_n)$, where $T_n := S_1 + \cdots + S_n$.

(a) Prove that P-almost surely $\lim_{n\uparrow\infty} T_n = \infty$ and $\lim_{t\uparrow\infty} N(t) = \infty$.

(b) Prove that P-almost surely $\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{E[S_1]}$.

Exercise 4.3.6. SLNN, THE INFINITE EXPECTATION CASE
Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of non-negative random variables such that $E\left[Z_1\right] = \infty$. Show that

$$\lim_{n\uparrow\infty} \frac{Z_1 + \ldots + Z_n}{n} = \infty \quad (= E\left[Z_1\right]).$$

Exercise 4.3.7. EXPECTATION OF SERIES
Prove the following.

(A) Let $\{Z_n\}_{n\geq 1}$ be a sequence of real-valued non-negative random variables. Then

$$E\left[\sum_{n\geq 1} Z_n\right] = \sum_{n\geq 1} E\left[Z_n\right]. \tag{4.9}$$

(B) Let $\{Z_n\}_{n\geq 1}$ be a sequence of real-valued random variables such that

$$\sum_{n\geq 1} E\left[|Z_n|\right] < \infty.$$

Prove that (4.9) holds true.

Exercise 4.3.8. CONVERGENCE A.S. *versus* CONVERGENCE IN PROBABILITY
Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with values in $\{0, 1\}$.

(A) Prove that a necessary and sufficient condition of almost sure convergence to 0 is that

$$\sum_{n\geq 1} P(X_n = 1) < \infty.$$

(B) Prove that a necessary and sufficient condition of convergence in probability to 0 is that

$$\lim_{n\uparrow\infty} P(X_n = 1) = 0.$$

(C) Deduce from the above that convergence in probability does not imply almost sure convergence.

Exercise 4.3.9. Derivative of the Laplace transform
Let X be a discrete random variable with values $x_i \in \mathbb{R}_+$ $(i \geq 1)$ and of distribution $p_i = P(X = x_i)$ $(i \geq 1)$. Suppose that for some $t_0 > 0$, $\sum_{i\geq 1} p_i e^{t_0 x_i} < \infty$. Prove that the function $g : t \to \sum_{i\geq 1} p_i e^{tx_i}$ is differentiable in a neighborhood of 0, with derivative $\sum_{i\geq 1} p_i x_i e^{tx_i}$.

Chapter 5

The probabilistic Method

5.1 Proving Existence

5.1.1 The Counting Argument

The techniques presented in this chapter for asserting the existence of mathematical objects with a certain property will at first sight appear as a collection of tricks, but as one gets used to it, some unity emerges from the recurrence of these tricks that henceforth deserve to be called "tools". One can of course strive to give a unified theoretical treatment to the probabilistic method. Even though this may be fruitful at a more advanced level, we have chosen to proceed by means of examples that will introduce the reader to what is an art as well as a science.

The original ideas involved in the probabilistic method are as ingenious as they are simple. Suppose for instance that you are dealing with a countable collection a_i $(i \in I)$ of "objects" and that you want to know if at least one of them has a certain property $\mathcal{P}$. It is sometimes convenient and efficient to imagine a random element X taking the values a_i $(i \in I)$ and see if $P(X \text{ satisfies property } \mathcal{P}) > 0$, in which case the answer is yes. In fact,

$$P(X \text{ satisfies property } \mathcal{P}) = \sum_{i \in I} P(X = a_i) 1_{\{a_i \text{ satisfies property } \mathcal{P}\}}$$

a quantity which cannot be positive if $1_{\{a_i \text{ satisfies property } \mathcal{P}\}} = 0$ for all $i \in I$. This argument is called the counting argument.

The following examples of application of the probabilistic method concern graphs.

Example 5.1.1: Graph colouring. Let K_n be a complete graph with n vertices. Coloring this graph consists in assigning a colour to each of its edges. Here we consider the case with 2 colours, say, red and blue. Let $k < n$. We ask the following question: does there exist a 2-colouring such that one cannot find a subgraph of size k with all its edges of the same colour?

We are going to prove that if

$$\binom{n}{k} 2^{-\binom{k}{2}+1} < 1\,, \tag{5.1}$$

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_5

there is no such colouring. For this, we use the probabilistic method, considering a random colouring of the graph obtained by choosing independently the colours of each edge, blue with probability $\frac{1}{2}$, red with probability $\frac{1}{2}$. There are exactly $\binom{n}{k}$ subgraphs of size k enumerated from 1 to $\binom{n}{k}$. Let A_i be the event that the i-th subgraph is monochromatic. This occurs if, one of its edge being of any given colour, the remaining $\binom{k}{2} - 1$ edges are of the same colour. Therefore

$$P(A_i) = 2^{-\binom{k}{2}+1}.$$

Now the probability that there is no monochromatic subgraph of size k is

$$P\left(\bigcap_{i=1}^{\binom{n}{k}} \overline{A_i}\right) = 1 - P\left(\bigcup_{i=1}^{\binom{n}{k}} A_i\right).$$

But, by sub-sigma-additivity,

$$P\left(\bigcup_{i=1}^{\binom{n}{k}} A_i\right) \leq \sum_{i=1}^{\binom{n}{k}} P(A_i) = \binom{n}{k} 2^{-\binom{k}{2}+1},$$

a quantity which is, under assumption (5.1), strictly less than 1. Therefore the probability that there is no monochromatic subgraph of size k is strictly positive. In particular, there must exist at least one 2-colouring without monochromatic subgraph of size k.

The next example features hypergraphs. A hypergraph is a pair $H = (V, \mathcal{E})$ where V is a finite set of vertices and $\mathcal{E}$ is a collection of subsets of V called the hyperedges. If all the hyperedges have cardinality k, the hypergraph is called k-uniform. It is called d-regular if each vertex is present in exactly d hyperedges. Let $\{1, 2, \ldots, L\}$ be a set of "colours". A L-coloring of an hypergraph is an assignment of a colour to each vertex. The hypergraph is called L-colourable if no hyperedge is monochromatic.

EXAMPLE 5.1.2: EXISTENCE OF A 2-COLOURING. Any k-uniform hypergraph with less than 2^{k-1} hyperedges is 2-colourable. To see this, colour independently the vertices with one of two colours, say RED and BLUE, equiprobably. Let A_e be the event that hyperedge e is monochromatic. Since $P(A_e) = \frac{1}{2^{k-1}}$, the probability that there exists a monochromatic hyperedge is, by the union bound,

$$P(\cup_e A_e) \leq \sum_e P(\cup_e A_e) = |\mathcal{E}| \frac{1}{2^{k-1}} < 1.$$

In particular, the probability that there exists no monochromatic hyperedge is strictly positive. Consequently, by the counting argument, there exists a 2-colouring with no monochromatic edge.

The next example features tournaments. A tournament is a complete oriented graph. More precisely a tournament T_n of size n is a complete graph K_n with the additional feature that each edge $\langle u, v\rangle$ is oriented, either from u to v or from v to u.

EXAMPLE 5.1.3: TOURNAMENT WITH THE S_k PROPERTY. A tournament T_n is said to have the property S_k ($k \leq n$) if for any set of k vertices, there exists a vertex that has an oriented edge towards each of these k vertices. Erdös has shown that for a given k, there exists tournaments T_n with the S_k property provided $n > k^2 2^k$.

Proof. Let T_n be a random tournament, that is, a complete graph K_n where the directions of the edges are chosen independently and equiprobably, with probability $\frac{1}{2}$ for each direction. Let S be a set of k vertices and let u be a vertex not in S. The probability that u has an oriented edge to each vertex of S is $\frac{1}{2^k}$, or, equivalently, the probability that u fails to have an oriented edge to each vertex of S is $1 - \frac{1}{2^k}$. For different vertices, these events are independent and therefore the probability that for all $u \notin S$, $u \not\to S$ is equal to $(1 - \frac{1}{2^k})^{n-k}$. There are $\binom{n}{k}$ sets of k vertices, and therefore, by the union bound, the probability that there exists a set S of k vertices such that for all $u \notin S$, $u \not\to S$ (call it a "bad" set), is, using the bound $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ (see (1.8)) and the inequality $1 - x \leq e^{-x}$

$$\leq \binom{n}{k}\left(1 - \frac{1}{2^k}\right)^{n-k} \leq \left(\frac{en}{k}\right)^k e^{-\frac{n-k}{2^k}}.$$

If the last quantity is strictly less than 1, this means that there exists at least one tournament on which no bad set exists, that is, a tournament with the S_k property. Now

$$\left(\frac{en}{k}\right)^k e^{-\frac{n-k}{2^k}} < 1 \Leftrightarrow \left(\frac{en}{k}\right)^k < e^{\frac{n-k}{2^k}},$$

and this is in turn equivalent to

$$k(1 + \log(n/k))2^k + k < n.$$

If $n > 2^k$,

$$\begin{aligned} k(1 + \log(n/k))2^k + k &< k(1 + \log(2^k))2^k + k \\ &= 2^k k^2 \log 2\left(1 + \frac{1}{k \log 2} + \frac{1}{k 2^k \log 2}\right) = 2^k k^2 \log 2(1 + O(1)) \end{aligned}$$

Therefore, the S_k property is satisfied if $n > 2^k k^2$. □

5.1.2 The Expectation Argument

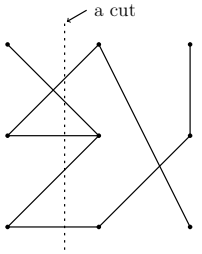
The following is another instance of the probabilistic method. Suppose again that you have a countable collection a_i ($i \in I$) of "objects". In addition, there is a

"performance index" for comparing them, that is, a function $f : \{a_i, i \in I\} \to \mathbb{R}$. You want to check if there is at least one object whose performance is larger or equal to some threshold α. For this again, you might imagine a random element X taking the values $a_i (i \in I)$ and such that $E[f(X)] \geq \alpha$, in which case the answer is yes. In fact,

$$E[f(X)] = \sum_{i \in I} P(X = a_i) f(a_i)\,,$$

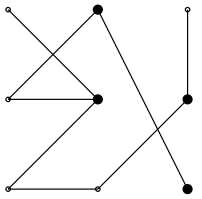
a quantity which cannot be $\geq \alpha$ if $f(a_i) < \alpha$ for all $i \in I$. This is the expectation argument.

EXAMPLE 5.1.4: LARGE CUTS, TAKE 1. Let $G = (V, \mathcal{E})$ be a graph with n vertices and m edges. A cut of the graph is a partition (A, B), where $B = \overline{A}$, of the set of vertices. An edge $e = \langle u, v \rangle$ is said to connect A and B if either $u \in A$ and $v \in B$ or $v \in A$ and $u \in B$. The number $N(A, B)$ of such edges is called the size of the cut.



The size of this cut is 5

We shall prove that there exists at least one cut of size larger than $\frac{1}{2}m$ (half the number of edges). For this, we colour all the vertices independently of one another, white or black with probability $\frac{1}{2}$. Call B the random set of black vertices.



Here, B is is the set of fat points and $N(A, B) = 6$

Since $N(A, B) = \sum_{e \in \mathcal{E}} 1_{\{e \text{ connects } A \text{ and } B\}}$, we have that $E[N(A, B)] = \frac{1}{2}m$ because the probability that a given edge $e \in E$ connects A and B (the probability that the end vertices of e have a different colour) is $\frac{1}{2}$. Therefore, by the expectation argument, there exists at least one cut of size at least $\frac{1}{2}m$.

EXAMPLE 5.1.5: CROSSING NUMBER. The crossing number $cr(G)$ of a graph $G = (V, \mathcal{E})$ with n vertices and m edges is the least number of edge crossings among

all the planar representations of this graph. Recall that a planar graph is, by definition, a graph with no edge crossing. We show that if $m \geq 4n$, then

$$cr(G) \geq \frac{1}{64}\frac{m^3}{n^2}\,.$$

We first recall the following inequality:

$$m - cr(G) \leq 3n\,. \qquad (\star)$$

To prove it, we use Euler's formula for planar graphs

$$n - m + f = 2\,,$$

where f is the number of faces of the graph, that is the number of subsets of the plane delimited by edges, including the exterior infinite one. To prove this formula, observe that the removal of an edge is accompanied by the diminution by 1 of the number of faces. Therefore the quantity $n - m + f$ remains constant through successive edge removals. Choose edge removals that each time break a cycle, so that at the end we have a tree with n vertices, $m = n - 1$ and $f = 1$ faces, and therefore $n - m + f = 2$. The same value of $n - m + f$ applies to the original planar graph, and this yields Euler's formula.

For the proof of $(\star)$, remove for each edge crossing one of the edges involved, to obtain a planar graph with $m - cr(G)$ edges, to which Euler's formula applies. This yields the announced result since by hypothesis $m \geq 4n$.

We can now proceed to the probabilistic argument. Construct a subgraph H of G whose vertex set $V(H)$ is obtained by random independent thinning of V: a vertex $e \in V$ is accepted as a vertex of H with probability $p = \frac{4n}{m}$ (< 1 since by assumption $m > 4n$) independently of the other vertices. Then, by $(\star)$ applied to H and with an obvious notation, $cr(H) \geq m_H - 3n_H$. Taking expectations,

$$E\left[cr(H)\right] \geq E\left[m_H\right] - 3E\left[n_H\right]\,.$$

Now, $E\left[n_H\right] = np$, $E\left[m_H\right] = p^2 m$ (each edge is retained with probability p^2) and $E\left[cr(H)\right] = p^4 E\left[cr(H)\right]$ (an edge crossing involves 4 vertices). Therefore

$$p^4 E\left[cr(G)\right] \geq p^2 m - 3pn\,.$$

With $p = \frac{4n}{m}$, this is the announced result.

EXAMPLE 5.1.6: INDEPENDENT SET OF VERTICES. A subset $B \subset V$ of vertices of a graph $G = (V, \mathcal{E})$ such that any pair of vertices in B has no edge linking them is, by definition, an independent set of vertices. Denote by $\alpha(G)$ the largest size of an independent set. Let n be the number of vertices and suppose that the number of edges is $m = n\frac{d}{2}$ (therefore, d is the average index of a vertex). We shall henceforth assume that $d \geq 1$ and show that $\alpha(G) \geq \frac{n}{2d}$.

Proof. Select a random set of vertices $S \subseteq V$ as follows. Let $\{Z_v\}_{v \in V}$ be a collection of IID $\{0,1\}$-valued random variables with common distribution $P(Z_v = 1) = p$.

Then decide to include v in S if and only if $Z_v = 1$. The random number $X = |S|$ of vertices in S therefore has the mean

$$E[X] = np\,.$$

The number of edges of $G|_S$, the restriction of G to S, is $Y = \sum_{e\in\mathcal{E}} Y_e$, where for $e = \langle u, v\rangle \in \mathcal{E}$, $Y_e = Z_u Z_v$. In particular $E[Y_e] = p^2$ and by linearity of expectation,

$$E[Y] = \frac{nd}{2}p^2\,.$$

Therefore

$$E[X - Y] = np - \frac{nd}{2}p^2\,,$$

a quantity that is maximal for $p = \frac{1}{d}$ (remember that d is assumed ≥ 1). With this value,

$$E[X - Y] = \frac{n}{2d}.$$

In particular, by the expectation argument, there exists a subset of vertices S for which $G|_S$ has a number of vertices exceeding the number of edges by at least $\frac{n}{2d}$. Construct now an independent set B from S by deleting for each edge of $G|_S$ one vertex so that B is an independent set of vertices (to help understanding what has just been said, maybe you can draw a graph with, say, 10 vertices and 7 edges, and suppress minimally vertices so as to obtain a set of at least 3 independent vertices). There are $\leq Y$ such edges, the number of vertices in B is therefore $X - Y \geq \frac{n}{2d}$. □

The trick of eliminating vertices so as to obtain a subset of vertices with a desired property is called the thinning argument.

EXAMPLE 5.1.7: DENSE GRAPHS WITH LARGE GIRTH. A cycle of a graph $G = (V, \mathcal{E})$ is an ordered sequence of distinct vertices $v_1, \ldots, v_\ell$ ($\ell \geq 3$) such that $v_i \sim v_{i+1}$ ($1 \leq i \leq \ell - 1$) and $v_\ell \sim v_1$. The length of the cycle is ℓ. The girth of the graph is the smallest length of a cycle of this graph. At first sight, one expects that the more dense the graph, the smaller its girth. This is roughly true, however there exists dense graphs with large girth. For instance:

Let $k \geq 3$ be a fixed integer. For n sufficiently large, there exists a graph with n nodes, at least $m := \frac{1}{4}n^{1+\frac{1}{k}}$ edges and girth $\geq k$. (The average index $d := \frac{1}{2}\frac{m}{n}$ is in this case $\frac{1}{2}n^{\frac{1}{k}}$ and grows to infinity with the number of edges. Therefore we are dealing with a "dense" graph.)

Proof. We apply the probabilistic method and the thinning argument of Example 5.1.6. Consider the Erdös–Rényi graph $\mathcal{G}(n, p_n)$ with $p_n = n^{\frac{1}{k}-1}$ (its average index is therefore of the order of $n^{\frac{1}{k}}$). Let X be the number of its edges:

$$E[X] = p_n\binom{n}{2} = \frac{1}{2}\left(1 - \frac{1}{n}\right)n^{1+\frac{1}{k}}.$$

Let Y be the number of cycles of length $\leq k-1$. Any specific cycle of length i occurs in $\mathcal{G}(n,p)$ with probability p^i. Also there are $\binom{n}{i}\frac{(i-1)!}{2}$ possible cycles of length i. In fact, one first chooses the i vertices: $\binom{n}{i}$ choices, then, their order: $(i-1)!$ choices (not $i!$ because a cycle such as $v_1, v_2, \ldots, v_i$ produces i identical cycles: $v_k, v_{k+1} \ldots, v_{k+i-1}$, $0 \leq k \leq i$). The $\frac{1}{2}$ factor comes from the fact that one does not distinguish the order from the reversed order in the definition of a cycle. Therefore

$$\begin{aligned} E[Y] &= \sum_{i=3}^{k-1} \binom{n}{i} \frac{(i-1)!}{2} p^i \\ &\leq \sum_{i=3}^{k-1} n^i p^i = \sum_{i=3}^{k-1} n^{\frac{i}{k}} < k n^{\frac{k-1}{k}} . \end{aligned}$$

Now, we eliminate from the graph one edge from each cycle of length $\leq k-1$, so that the resulting graph has a girth $\geq k$. The number of edges of the modified graph is $X - Y$ and

$$E[X-Y] \geq \frac{1}{2}\left(1 - \frac{1}{n}\right) n^{1+\frac{1}{k}} - k n^{\frac{k-1}{k}}.$$

When n is sufficiently large, this quantity is larger than $\frac{1}{4} n^{1+\frac{1}{k}}$. Therefore, by the expectation argument, there exists a graph with at least $\frac{1}{4} n^{1+\frac{1}{k}}$ edges and girth $\geq k$. □

EXAMPLE 5.1.8: DOMINATING SET. A graph $G = (V, \mathcal{E})$ being given, a subset D of vertices is called dominating if every vertex $v \notin D$ is adjacent to D. Let δ be the smallest vertex degree, assumed positive. Then, there exists a dominating set of size $\leq n \frac{\log(1+\delta)+1}{1+\delta}$.

Proof. Let S be a random set of vertices formed by the vertices selected independently with probability p. Let T be the collection of vertices outside S without neighbours in S. Note that $D := S \cup T$ is a dominating set. Then

$$E[D] = E[S] + E[T] = np + E[T]\,,$$

and

$$\begin{aligned} E[T] &= \sum_{v \in V} 1_{v \notin S\,;\,(u \sim v) \Rightarrow (u \notin S)} \\ &= \sum_{v \in V} E\left[(1-p)^{\mathrm{d}(v)+1}\right] \\ &\leq \sum_{v \in V} (1-p)^{\delta+1} = n(1-p)^{\delta+1}\,. \end{aligned}$$

Therefore,

$$E[D] \leq n\left((1-p)^{\delta+1}+p\right) \leq n\left(e^{-p(\delta+1)}+p\right) .$$

This function of δ has a minimum at $p = \frac{\log(1+\delta)}{1+\delta}$. For this value of p,

$$\left((1-p)^{\delta+1}+p\right) \leq n\left(e^{-p(\delta+1)}+p\right) = \frac{\log(1+\delta)+1}{1+\delta} .$$

□

Remark 5.1.9 Chapter 12 will give yet another spectacular application of the probabilistic method, namely to the fundamental result of information theory, Shannon's capacity theorem.

5.1.3 Lovasz's Local Lemma

In order to prove existence of objects with a given property, the probabilistic method requires to show that the probability of some event is positive. For instance, let $A_1, \ldots, A_n$ be events, each of them having probability < 1. If they are mutually independent, so are their complements, and therefore

$$P\left(\cap_{i=1}^{n} \overline{A}_i\right) > 0 . \tag{5.2}$$

One may think of the A_i's as "undesirable", or "bad", events (see the next example) and therefore, if the above condition is satisfied, there exists at least one event (namely $\cap_{i=1}^{n} \overline{A}_i$) not included in any of the bad events.

The statement is equivalent to $P\left(\cup_{i=1}^{n} A_i\right) < 1$. If the independence assumption does not hold, one may think of applying the sub-σ-additivity property of probability and see if $\sum_{i=1}^{n} P(A_i) < 1$. This rough bound may be too coarse, and one therefore has to resort to other methods. The following one applies when the events are not independent, but "weakly" dependent.

In order to state the corresponding conditions, we introduce the notion of dependency graph.

Definition 5.1.10 *(i) An event B is said to be mutually independent of the events $B_1, \ldots, B_\ell$ if*

$$P\left(B \middle| \cap_{j \subseteq I} B_j\right) = P(B)$$

for all subsets $I \subseteq \{1, 2, \ldots, \ell\}$.

(ii) A dependency graph for $A_1, \ldots, A_n$ is a graph $G = (V, \mathcal{E})$ where $V = \{1, 2, \ldots, n\}$ and for all $1 \leq i \leq n$, A_i is mutually independent of $\{A_j\,;\, \langle i, j\rangle \notin \mathcal{E}\}$.

Paraphrasing (ii): A_i is mutually independent of the events A_j such that j is not directly connected to i by an edge of the dependency graph.

Observe (Exercise 5.3.5) that if B is mutually independent of the events $B_1, \ldots, B_\ell$, it is also mutually independent of the events $\tilde{B}_1, \ldots, \tilde{B}_\ell$, where either $\tilde{B}_i = B_i$ or $\tilde{B}_i = \overline{B}_i$, the choice varying arbitrarily from one index to the other.

We may now state Lovasz's local lemma:

Theorem 5.1.11 *(Lovasz, 1993) Let $A_1, \ldots, A_n$ be events with dependency graph $G = (V, \mathcal{E})$. Suppose there exist numbers $x_i \in (0,1)$ such that for all $1 \leq i \leq n$,*

$$P(A_i) \leq x_i \prod_{\langle i,j \rangle \in \mathcal{E}} (1 - x_j)\,.$$

Then

$$P\left(\cap_{i=1}^n \overline{A}_i\right) \geq \prod_{i=1}^n (1 - x_i)\,.$$

Proof. Let $S \subset \{1, 2, \ldots, n\}$. If we can show that

$$P\left(A_k \mid \cap_{j \in S} \overline{A}_j\right) \leq x_k \qquad (\star)$$

for all $k \notin S$, then the proof follows from

$$\begin{aligned} P\left(\cap_{i=1}^n \overline{A}_i\right) &= \prod_{i=1}^n P\left(\overline{A}_i \mid \cap_{j=1}^{i-1} \overline{A}_j\right) \\ &= \prod_{i=1}^n \left(1 - P\left(A_i \mid \cap_{j=1}^{i-1} \overline{A}_j\right)\right) \\ &\geq \prod_{i=1}^n (1 - x_i)\,, \end{aligned}$$

a strictly positive quantity.

We now proceed to prove $(\star)$ by induction on $s := |S|$.

Step 1. For $s = 0$, $(\star)$ follows from the hypothesis since

$$P(A_k \,|\, \overline{\varnothing}) = P(A_k \,|\, \Omega) = P(A_k) \leq x_k \prod_{\langle k,j \rangle \in E} (1 - x_j) \leq x_k\,.$$

Step 2. For $s \geq 1$, we must first verify that $P\left(\cup_{j \in S} \overline{A}_j\right) > 0$ so that the left-hand side of $(\star)$ is meaningful. This is true for $s = 1$ because $P\left(\overline{A}_j\right) \geq 1 - x_j > 0$. For $s \geq 2$, rename the elements of $\{1, 2, \ldots, n\}$ in such a way that $S = \{1, 2, \ldots, s\}$. Then

$$\begin{aligned} P\left(\cap_{i=1}^s \overline{A}_i\right) &= \prod_{i=1}^s P\left(\overline{A}_i \mid \cap_{j=1}^{i-1} \overline{A}_j\right) \\ &= \prod_{i=1}^s \left(1 - P\left(A_i \mid \cap_{j=1}^{i-1} \overline{A}_j\right)\right) \\ &\geq \prod_{i=1}^s \left(1 - x_i(1 - x_1) \cdots (1 - x_{i-1})\right) > 0\,, \end{aligned}$$

according to the recurrence hypothesis.

Step 3. Let $S_1 = \{j \in S\,;\, k \sim j\}$ ($k \sim j$ means that $\langle i,j\rangle$ is an edge of the dependency graph of $A_1,\ldots,A_n$), and $S_2 = S\backslash S_1$.

First case: $S_2 = S$. This means that A_k is mutually independent of $\overline{A}_i, i \in S$, and therefore $P\left(A_k|\cap_{j\in S}\overline{A}_j\right) = P(A_k) \le x_k$.

Second case: $|S_2| < s$. With the notation $B_S = \cap_{j\in S}\overline{A}_j$, we have $B_S = B_{S_1} \cap B_{S_2}$. Then

$$\begin{aligned} P\left(A_k \mid \cap_{j\in S}\overline{A}_j\right) &:= P\left(A_k \mid B_S\right) \\ &= \frac{P(A_k \cap B_S)}{P(B_S)} = \frac{P(A_k \cap B_{S_1} \cap B_{S_2})}{P(B_{S_1} \cap B_{S_2})} \\ &= \frac{P(A_k \cap B_{S_1} \mid B_{S_2})P(B_{S_2})}{P(B_{S_1} \mid B_{S_2})P(B_{S_2})} \\ &= \frac{P(A_k \cap B_{S_1} \mid B_{S_2})}{P(B_{S_1} \mid B_{S_2})} := \frac{N}{D}\,. \end{aligned}$$

But, by the definition of a dependency graph and the hypothesis,

$$N \le P(A_k \mid B_{S_2}) = P(A_k) \le x_k \prod_{\langle k,j\rangle\in E}(1-x_j)\,.$$

Also, letting $S_1 := \{j_1,\ldots,j_r\}$,

$$\begin{aligned} D &= P\left(\cap_{i\in S_1}\overline{A}_i \mid \cap_{j\in S_2}\overline{A}_j\right) \\ &= \prod_{\ell=1}^{r}\left(1 - P\left(A_{j_\ell} \mid \left(\cap_{t=1}^{\ell-1}\overline{A}_{j_t}\right) \cap \left(\cap_{j\in S_2}\overline{A}_j\right)\right)\right) \\ &\ge \prod_{\ell=1}^{r}(1-x_{j_\ell}) \ge \prod_{\langle k,j\rangle\in E}(1-x_j)\,, \end{aligned}$$

where we used the induction hypothesis for the last inequality since in this second case $|S_2| < |S| = s$. Combining the bounds for N and D gives the announced result. □

Corollary 5.1.12 *Let $A_1,\ldots,A_n$ be events. A sufficient condition for inequality (5.2) is that the three conditions below be satisfied*

(a) $P(A_i) \le p$ $(1 \le i \le n)$ for some $p \in (0,1)$,

(b) the largest degree of a vertex of the dependency graph is d (assumed ≥ 1, otherwise we are in the known independent case), and

(c) $ep(d+1) \le 1$.

Proof. Take $x_i = \frac{1}{d+1}$ so that

$$x_i \prod_{\langle i,j\rangle\in\mathcal{E}}(1-x_j) \ge \frac{1}{d+1}\left(1-\frac{1}{d+1}\right)^d = \frac{1}{d+1}\left(\frac{d}{d+1}\right)^d$$

and observe that $\left(\frac{d}{d+1}\right)^d = \left(1+\frac{d}{d+1}\right)^d \leq e$. Therefore, for each i,

$$x_i \prod_{\langle i,j\rangle \in \mathcal{E}} (1-x_j) \geq \quad \geq p \geq P(A_i)\,.$$

The result follows by applying the general version of Lovasz's lemma. □

Since $1 \leq \frac{d}{2}$, condition (c) is satisfied if

(c') $6dp \leq 1$.

EXAMPLE 5.1.13: THE SATISFIABILITY PROBLEM. A logical formula is, roughly speaking, an expression involving literals $x_1, x_2, \ldots$ with values in $\{0,1\}$ (where 1 and 0 mean TRUE and FALSE, respectively), their negations $\overline{x}_1, \overline{x}_2, \ldots$, and the operations AND (conjunction, represented by the symbol $\wedge$) and OR (disjunction, represented by the symbol $\vee$). In this example we shall consider the so-called SAT formulas. Such a formula is by definition a conjunction of clauses, a clause being a disjunction of literals and their negations. For instance, $(x_1 \vee \overline{x}_2)$, $(\overline{x}_1 \vee \overline{x}_4 \vee x_3)$ and x_2 are clauses, and

$$(x_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee \overline{x}_4 \vee x_3) \wedge x_2$$

is a SAT formula. A solution of the SAT formula is any assignment of values to the literals resulting in the value 1 (TRUE). Equivalently, each clause must take the value 1. In the above example, there is a solution (in fact several), for instance $x_1 = x_2 = 1, x_3 =$ arbitrary, $x_4 = 0$. There are SAT formulas that do not have a solution, for instance,

$$(x_1 \vee x_3) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_3).$$

In general, determining if a SAT formula has a solution is NP-hard. Here we shall be interested in finding a sufficient condition ensuring the satisfiability of a SAT formula. It will be assumed that no clause contains a literal and its negation (in this case the clause is trivially satisfied).

A k-SAT formula is a SAT formula in which each clause features exactly k literals. We prove the following: In a k-SAT formula with m clauses, if no variable appears in more than $\frac{2^k}{6k}$ clauses, then it is satisfiable.

Proof. Let X_j $(1 \leq j \leq k)$ be IID random variables taking equiprobably the values 0 and 1. These are random values assigned to the literals x_j $(1 \leq j \leq k)$. Let A_i be the event that clause i is not satisfied. Since a clause has k elements, $p := P(A_i) = 2^{-k}$. The event A_i is mutually independent of all the A_ℓ relative to a clause ℓ that shares no literal with clause i. Since by hypothesis no literal appears in more that $\frac{2^k}{6k}$ clauses, the maximal degree d of the dependency graph of $(A_\ell\,;\, 1 \leq \ell \leq m)$ is such that $d < k\frac{2^k}{6k} = \frac{2^k}{6}$ and therefore $6pd \leq 1$. The conditions of Lovasz's lemma are therefore satisfied and we can conclude that the event $\cap_{i=1}^m \overline{A}_i$ (all clauses are satisfied) has a positive probability. □

EXAMPLE 5.1.14: NON-COLLIDING PATHS IN A COMMUNICATIONS NETWORK. We start with the following problem. There are n customers, and each consumer i ($1 \leq i \leq n$) can choose goods from a list L_i of m items. The lists are not disjoint, but each given list shares no more that k items with any other given list. We shall prove that under the condition $12nk \leq m$, there is at least one non-conflictual assignment of goods (two customers who want the same object).

For this, let X_i be a random variable uniformly distributed on L_i, and suppose that the X_i's are mutually independent. (The customers choose at random an item from their list independently from one another.) Let $E_{i,j} = \{X_i = X_j\}$ be the event (of probability $P(E_{i,j}) \leq \frac{k}{m} := p$) that there is a conflict between customers i and j. Observe that $E_{i,j}$ is mutually independent of the $E_{r,s}$ whenever $r, s \notin \{i, j\}$. Therefore, the events $E_{i,j}$, $1 \leq i, j \leq n$, have a dependency graph with maximal degree $d \leq 2(n-1) < 2n$ ($2(n-1)$ is the number of unordered pairs (r, s) such that one of them at least is not in $\{i, j\}$), so that $6pd < \frac{12nk}{m}$, which is < 1 by assumption. The conditions of Lovasz's lemma are therefore satisfied and one can therefore conclude that the event $\cap_{i=1}^{m} \overline{E}_i$ (no conflict between customers) has a positive probability.

We return to the title of the example, which refers to a communication network with nodes and links between some pairs of distinct nodes. Such a network can therefore be identified with a graph $G = (V, \mathcal{E})$, the nodes and links being respectively the vertices and edges. Let be given n distinct pairs of distinct nodes. Each of these pairs of nodes seeks to establish a communications path between them along the links. The possible paths available to the pair i ($1 \leq i \leq n$) form a list L_i of cardinal m. If a given path in L_i shares at least a link with a path in L_j, these paths are said to be colliding. Under the condition $12nk \leq m$, there is at least one non-colliding assignment of paths, that is each one of the n pairs of nodes can select a path of his list in such a way that all the paths selected are non-colliding. The proof is the same as before, with a slight reinterpretation of the mathematical objects involved (the event $E_{i,j}$ is now "path X_i shares no link with path X_j").

EXAMPLE 5.1.15: 2-COLOURABILITY OF UNIFORM REGULAR HYPERGRAPHS. (a) If

$$e\frac{1}{2^{k-1}}(d-1)k + 1 \leq 1\,, \qquad (\star)$$

a k-uniform, d-regular hypergraph is 2-colourable.

(b) If $k \geq 9$, a k-uniform, k-regular hypergraph is 2-colourable.

Proof. (a) Consider a random uniform 2-colouring of the set of vertices. Let A_e denote the event that hyperedge e is monochrome. We have that $P(A_e) = \frac{1}{2^{k-1}}$. We want to show that the event $\cap_{e \in \mathcal{E}} \overline{A_e}$ has a positive probability. Event A_e and A_f are dependent if $e \cap f \neq \varnothing$. Since e contains k vertices and each of them is contained in $d-1$ other hyperedges, an upper bound for the degree of the dependency graph is given by

$$|\{f \in \mathcal{E}\,;\, f \cap e \neq \varnothing\}| \leq (d-1)k\,.$$

By Corollary 5.1.12, if $(\star)$ is satisfied, then

$$P\left(\cap_{e\in\mathcal{E}}\overline{A_e}\right) > 0\,.$$

(b) With $d = k$, inequality $(\star)$ reads

$$e\frac{1}{2^{k-1}}(k-1)k + 1 \leq 1\,,$$

and one checks that it is satisfied for $k \geq 9$. □

5.2 Random Algorithms

5.2.1 Las Vegas Algorithms

Another aspect of the probabilistic method concerns random algorithms, that is algorithms involving random steps, which produce correct results with high probability (even probability 1) in situations where deterministic algorithms are either not available or too costly from a computational point of view.

Algorithms are devised to find a given mathematical object, for instance the greatest common divisor of two integers, or to check if some property is satisfied, for instance if a given integer is a prime number. Random algorithms are used when the computational burden of the classical deterministic algorithms is too heavy. This section consists of a collection of examples that will give some feeling as to what is meant by a random algorithm. However, we shall not perform the complete analysis of efficiency of these algorithms, for instance we shall not discuss in detail their computational cost, in terms of time or of memory requirement, but only deal with the purely probabilistic aspects, such as the performance analysis in terms of probability of error.

We shall distinguish two types of random algorithms. The output of a Monte Carlo algorithm is correct only with some probability (hopefully close to 1, but not necessarily), whereas a Las Vegas algorithm, even though it involves some kind of randomization, eventually gives the correct answer.

EXAMPLE 5.2.1: QUICKSORT. Suppose that we need to sort a sequence of numbers in increasing order. For example $7, 6, 4, 2, 9, 3, 1, 8, 5$. The quicksort algorithm proposes to choose one at random, say, 4, called the pivot. It then scans the list from left to right, comparing each number to the pivot, placing the ones that are smaller than the pivot to the left, the others to the right. This creates three sets:

$$\{2, 1, 3\}, 4, \{7, 6, 9, 8, 5\}$$

It operates likewise on the two unordered subsets of the last list. For instance, starting with subset $\{2, 1, 3\}$, and choosing at random the pivot for this sublist, say 1, and then continuing with the subset $\{7, 6, 9, 8, 5\}$ with the pivot 7, we obtain:

$$1, \{2,3\}, 4, \{6,5\}, 7, \{9,8\}.$$

We keep doing this until all the subsets have only one member. In this particular example just one more iteration is needed.

The number of comparisons needed is $8 + (2+4) + (1+1+1) = 17$. One would like to know how well this algorithm does in terms of the number of comparisons. The best case would be if at each splitting the median number is chosen, resulting in a number of comparisons approximately equal to

$$n + 2\frac{n}{2} + 4\frac{n}{4} + \cdots$$

where there are approximately $\log_2 n$ terms in the sum. Therefore, one should compare the average number of comparisons in the random quicksort to $n \log_2 n$.

Let C_n be the number of comparisons needed and let X be the rank of the initial value selected. We have, with $M_n = E[C_n]$,

$$\begin{aligned} M_n &= \sum_{j=1}^{n} E[C_n | X = j] \, P(X = j) \\ &= \sum_{j=1}^{n} (n - 1 + M_{j-1} + M_{n-j}) \times \frac{1}{n} = n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} M_k \,, \end{aligned}$$

and therefore

$$nM_n = n(n-1) + 2\sum_{k=1}^{n-1} M_k \,.$$

Subtracting the same expression with $n-1$ instead of n, we have

$$nM_n = (n+1)M_{n-1} + 2(n-1) \,,$$

or

$$\frac{M_n}{n+1} = \frac{M_{n-1}}{n} + \frac{2(n-1)}{n(n+1)}.$$

By iteration,

$$\frac{M_n}{n+1} = 2\sum_{k=1}^{n} \frac{k-1}{k(k+1)} = 2\sum_{k=1}^{n} \left(\frac{2}{k+1} - \frac{1}{k} \right)$$

and therefore, finally, using the bounds

$$\log n \le H(n) := \sum_{i=1}^{n} \frac{1}{i} \le \log n + 1$$

(see Example 2.1.42), $M_n \sim 2n \log n$. This compares quite well with the idyllic best case.

EXAMPLE 5.2.2: LARGE CUTS, TAKE 2. We know that in a graph $G = (V, \mathcal{E})$ with m edges, there exists at least one cut of size $\frac{m}{2}$. If the cut is obtained by random

selection in Example 5.1.4, the probability of success is $p = P\left(N(A,B) \geq \frac{m}{2}\right)$. In the following computations, we suppose that m is even. We have seen that

$$\frac{m}{2} = E\left[N(A,B)\right] = \sum_{i \leq \frac{m}{2}-1} iP(N(A,B)=i) + \sum_{i \geq \frac{m}{2}} iP(N(A,B)=i)\,.$$

But

$$\begin{aligned}\sum_{i \leq \frac{m}{2}} iP(N(A,B)=i) &\leq \left(\frac{m}{2}-1\right)\left(\sum_{i \leq \frac{m}{2}-1} P(N(A,B)=i)\right)\\ &= \left(\frac{m}{2}-1\right)P(N(A,B) \leq \frac{m}{2}-1) = \left(\frac{m}{2}-1\right)(1-p)\end{aligned}$$

and observing that $N(A,B) \leq m$,

$$\sum_{i \geq \frac{m}{2}} iP(N(A,B)=i) \leq m \sum_{i \geq \frac{m}{2}} P(N(A,B)=i) = mP(N(A,B) \geq \frac{m}{2}) = mp\,.$$

Therefore $\frac{m}{2} \leq \left(\frac{m}{2}-1\right)(1-p) + mp$, which gives $p \geq \frac{1}{\frac{m}{2}+1}$. Observe that the time needed to check if a given cut is at least of size $\frac{m}{2}$ is polynomial in m (counting the edges linking the partitioning sets A and B).

The above algorithm is, as such, a Monte Carlo algorithm. It only has a positive probability of success. However, it can be iterated with independent random cuts until the desired result is attained and therefore, its iterated version is a Las Vegas algorithm. The random number of iterations until success is geometric and therefore has a mean equal to $\frac{1}{p} \leq 1 + \frac{m}{2}$.

5.2.2 Monte Carlo Algorithms

The next example is that of a Monte Carlo algorithm that cannot be transformed into a Las Vegas algorithm, although iterations of it eventually give arbitrarily small probability as their number increases.

Example 5.2.3: Checking a polynomial identity, 1 variable. Let P_1 and P_2 be two polynomials of degree d. We wish to check if $P_1 \equiv P_2$. Of course the query is meaningful only if the two polynomials are not in canonical form, in which case a simple inspection of the coefficients is enough, and this requires $O(d)$ operations in the worst case. If one of the polynomials is presented in the form of a product of d monomials, the reduction to canonical form requires $O(d^2)$ multiplications (at each step i, $2 \leq i \leq n$, of the procedure, one has to multiply the i-th monomial with the product of the $i-1$ first monomials).

There is a random algorithm requiring $O(d)$ operations which gives an answer, but this answer may be the wrong one. We present this algorithm and estimate the probability of a wrong decision. Then we discuss what can be done to make the probability of error small.

The algorithm consists in choosing a number r at random in the set $\{1, 2, \ldots, 100d\}$ and then comparing the values of these polynomials for the value r of the argument (which can be done in $O(d)$ time).

If $P_1(r) \neq P_2(r)$ say that $P_1 \not\equiv P_2$. If $P_1(r) = P_2(r)$ say that $P_1 \equiv P_2$.

The answer of the algorithm is wrong if and only if $P_1 \not\equiv P_2$ and $P_1(r) = P_2(r)$. But given that $P_1 \not\equiv P_2$, the event $P_1(r) = P_2(r)$ is implied by the event that the polynomial $P := P_1 - P_2$ (whose degree is at most d) has r for a root. But among the $100d$ possible values of r, at most d are the roots of D, therefore, since r is chosen at random among $100d$ values, the probability of error is $\frac{d}{100d} = \frac{1}{100}$.

Accurateness can be improved by repeating the algorithm k times independently. If p is the probability of error of a single run of the algorithm, the probability that the k runs give the wrong answer is p^k. Already 3 runs give a probability of error of one millionth.

Remark 5.2.4 An argument against the above method is that there exist good deterministic algorithms for the same purpose. Indeed, if we take $d + 1$ samples $r_1, \ldots, r_{d+1}$ without replacements (that is, distinct), then if $P_1 \not\equiv P_2$, one of these values will give different evaluations of the polynomial (a polynomial of degree d has at most d roots), and therefore, a correct answer. Note however that the number of operations required is $O(d^2)$. The choice is between complexity and accurateness (measured by the probability of a correct answer).

EXAMPLE 5.2.5: CHECKING A POLYNOMIAL IDENTITY, n VARIABLES. (Schwartz, 1980, and Zippel, 1979.) This example is an extension of Example 5.2.3 to polynomials of n real variables. One wishes to know if the polynomial $Q(x_1, x_2, \ldots, x_n)$ of degree d is identically null or not. The algorithm that is proposed is the following. Let S be a set of real numbers. Choose n values $r_1, r_2, \ldots, r_n$ independently and uniformly in S.

(1) If $Q(r_1, r_2, \ldots, r_n) = 0$ claim that $Q \equiv 0$.

(2) If $Q(r_1, r_2, \ldots, r_n) \neq 0$ claim that $Q \not\equiv 0$.

An error may occur only in case (1): $Q \not\equiv 0$ and $Q(r_1, r_2, \ldots, r_n) = 0$. We show that if $Q \not\equiv 0$, then

$$P(Q(r_1, r_2, \ldots, r_n) = 0) \leq \frac{d}{|S|}.$$

This is proved by induction on the number of variables. Example 5.2.3 has shown that the claim is true for $n = 1$. Suppose now that the result has been proved for $n - 1 \geq 1$ variables. Let k be the largest degree of x_1 in Q. The polynomial Q can then be written as

$$Q(x_1, x_2, \ldots, x_n) = R(x_2, x_3, \ldots, x_n)x_1^k + T(x_1, x_2, \ldots, x_n)$$

where the maximum degree of R is $\leq d - k$ and the maximum degree of x_1 in T is $< k$.

Now choose $r_2, r_3, \ldots, r_n$ uniformly and independently in S. Let

$$A := \{R(r_2, r_3, \ldots, r_n) = 0\}\,.$$

The induction hypothesis tells us that $P(A) \leq \frac{d-k}{|S|}$.

Define $Q'(x_1) := Q(x_1, r_2, \ldots, r_n)$. Outside A, this is a non-null polynomial in one variable of degree k. It has at most k roots and therefore

$$P(Q'(r_1) = Q(r_1, r_2, \ldots, r_n) = 0 \,|\, \overline{A}) \leq \frac{k}{|S|}\,.$$

Now

$$\begin{aligned} &P(Q(r_1, r_2, \ldots, r_n) = 0) \\ &\quad = P(Q(r_1, r_2, \ldots, r_n) = 0 \,|\, A)P(A) + P(Q(r_1, r_2, \ldots, r_n) = 0 \,|\, \overline{A})P(\overline{A}) \\ &\quad \leq \frac{d-k}{|S|} + \frac{k}{|S|} = \frac{d}{|S|}\,. \end{aligned}$$

EXAMPLE 5.2.6: A PRIMALITY TEST. (Rabin, 1980) Let $x \leq n$ be a positive integer. Define $x_0 := x^p \,(\mathrm{mod}\, n)$, $x_j = x_{j-1}^2 \,(\mathrm{mod}\, n)$ $(1 \leq j \leq p)$. A test for checking the primality of n is based on the following number-theoretic lemma.: Let $n > 4$ be an odd number (necessarily of the form $n = 1 + 2^p m$ where m is odd).

(i) If n is composite, there are at least $\frac{3}{4}(n-1)$ integers $x \in \{1, 2, \ldots, n\}$ such that either: (a) $x_p \neq 1$, or: (b) for some j $(1 \leq j \leq p)$, $x_j = 1$ and $x_{j-1} \neq n-1$.

(ii) If n is prime, neither (a) nor (b) holds for any $x \in \{1, 2, \ldots, n\}$.

This leads to the following algorithm. Select uniformly at random an integer $x \in \{1, 2, \ldots, n\}$. If neither (a) nor (b) is satisfied, say that n is prime, otherwise say that it is composite. According to the lemma, an error occurs only if n is composite and neither (a) nor (b) is satisfied, which occurs with probability less than $\frac{1}{4}$. Here again, by repeating k times this test, we can attain a probability of error smaller than $\frac{1}{4^k}$.

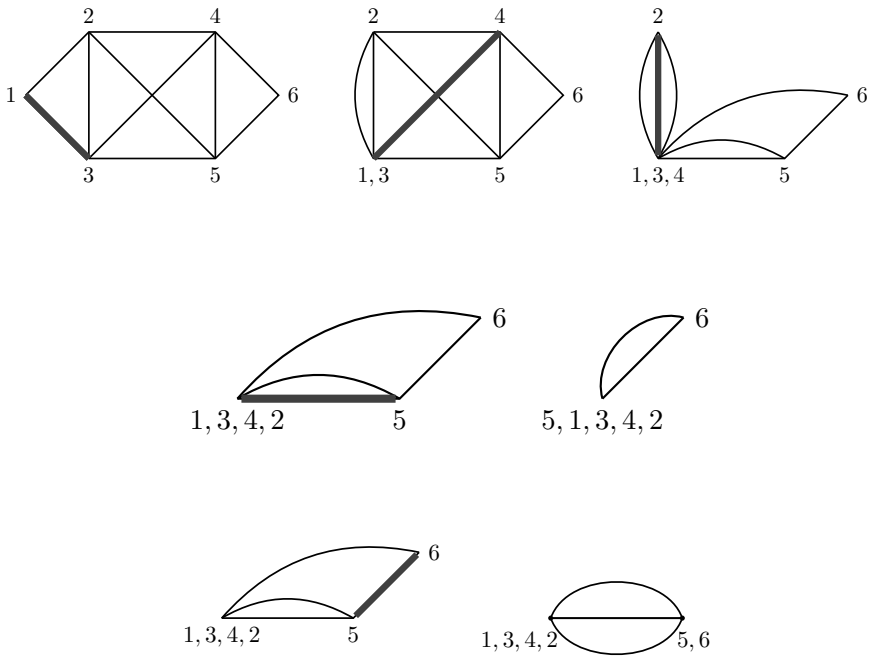
Properties (a) and (b) can be checked in $\log n$ time. This must be taken into account when comparing the randomized algorithm to a deterministic algorithm.

The last examples are of the same kind. One wishes to check if a given property $\mathcal{P}$ relative to the elements of a collection $\mathcal{C}$ of objects is satisfied by a given particular object $x \in \mathcal{C}$. One can subject this object to random experiments resulting in a random variable Z taking two values, say yes or no, with the following property (a) if no, property $\mathcal{P}$ is not satisfied for this x, and (b) if yes, property $\mathcal{P}$ is satisfied with a certain probability $p < \frac{1}{2}$. The experiments have the following features. If no, property $\mathcal{P}$ is not satisfied, and one has obtained the correct answer

in a single iteration of the random experiment. Whereas in the other case, the successive independent repetitions of the experiment will never end up with a clear unambiguous answer, but they decrease to any small fixed value the probability of making an error. Such random algorithms are, by essence, pure Monte Carlo algorithms.

EXAMPLE 5.2.7: AN ALGORITHM FOR THE MIN-CUT SIZE OF A MULTIGRAPH. In a multigraph $G = (V, \mathcal{E})$ (a graph with possibly multiple edges) — assumed connected and loop free — a cut is, by definition, a set of edges whose suppression disconnects the graph. The present goal is to find the min-cut size of the graph, that is, the minimal size of a cut. The random algorithm proposed below may or may not yield the correct answer, but nevertheless one can compute a lower bound for the probability of success that can be exploited, as we shall see later.

The algorithm consists in a succession of edge contractions: at each step, as long as there remains two vertices in the graph, one applies to the current graph the following operation. An edge e is selected at random amongst its edges, and the two corresponding vertices are merged while suppressing all the edges between these two vertices. In particular, there never appear loops in the process. Note that an edge contraction does not reduce the min-cut size because every cut in the graph at an intermediate size is a cut of the original graph. When there are only two vertices left, the answer (right or wrong) proposed by the algorithm is just the number of edges linking them. The following pictures describe the actual procedure in a particular case. The second and third level of pictures describe two possible paths, with different outcomes, showing that the algorithm might err.



Let n be the number of vertices of the original graph and k be its min-cut size. Let $\mathcal{C}$ be a particular cut of minimal size. Then G has at least $\frac{kn}{2}$ edges, otherwise there would be a vertex of degree strictly less than k and its incident edges would then form a cut of size strictly less than k.

One circumstance leading to a correct answer is when no edge of $\mathcal{C}$ is ever contracted, so that the surviving edges are exactly those of $\mathcal{C}$. The probability of this event that we proceed to estimate is then a lower bound for the probability of success.

Denote by A_i the probability that no edge of $\mathcal{C}$ is picked for contraction at stage i $(1 \leq i \leq n-2)$. The probability that the first edge chosen is in $\mathcal{C}$ is at most $\frac{k}{nk/2}$ so that $P(A_1) \geq 1 - \frac{2}{n}$. If A_1 occurs, there are at least $\frac{k(n-1)}{2}$ edges in the contracted graph, so that, as above $P(A_2 \,|\, A_1) \geq 1 - \frac{2}{n-1}$. More generally,

$$P(A_i \,|\, \cap_{\ell=1}^{i-1} A_\ell) \geq 1 - \frac{2}{n-i+1}$$

and therefore, by the Bayes sequential rule,

$$P\left(\cap_{i=1}^{n-1} A_i\right) \geq \prod_{i=1}^{n-1} \left(1 - \frac{2}{n-i+1}\right) = \frac{2}{n(n-1)}\,.$$

Therefore, the probability of a correct answer is $\geq \frac{2}{n^2}$

This is a low probability for large graphs. However if we repeat the algorithm $\frac{n^2}{2}$ times, and keep the minimal min-cut size proposals of the $\frac{2}{n^2}$ iterations, the probability of not having found the right answer is at most

$$\left(1 - \frac{2}{n^2}\right)^{\frac{n^2}{2}} < \frac{1}{e}.$$

EXAMPLE 5.2.8: FINGERPRINTING. (Rabin, 1981) Two numbers, repectively a and b, with binary representations

$$a = a_1 a_2 \cdots a_n \text{ and } b = b_1 b_2 \cdots b_n$$

have been recorded by two individuals, respectively A and B, in different locations. They both wish to know if these two sequences are identical. For this purpose, one of them (say, A) transmits his number to the other (in this case, B) so that the latter can compare the sequences. But, in view of saving communication costs A sends a shorter sequence, a fingerprint of his sequence, namely $F_p(a) := a(\mod p)$ where p is a prime number chosen uniformly at random among the prime numbers smaller than or equal to some number N. He also sends p, so that receiver B can compute the fingerprint $F_p(b)$ of his own number and check if $F_p(a) = F_p(b)$. If this is not the case, he concludes that the sequences are different, making the right decision since the communications channel is assumed error-free. If the fingerprints

coincide, he concludes that the original sequences are the same, thus potentially making an error, the probability of which we now estimate.

If $F_p(a) = F_p(b)$ and $a \neq b$, then p divides $|a - b|$, and therefore

$$P(Err.) \leq P(p \text{ divides } |a - b|)\,.$$

Let $\alpha(n)$ be an upper bound of the number of primes that divides a given n-bit number, and let $\pi(N)$ be the number of primes smaller than or equal to N. Since p is chosen uniformly at random among the prime numbers smaller than or equal to N,

$$P(p \text{ divides } |a - b|) \leq \frac{\alpha(n)}{\pi(N)}\,.$$

Recall (or else, believe) the following fact from prime number theory. For any number $x \geq 17$,

$$\frac{x}{\log x} \leq \pi(x) \leq 1.26\frac{x}{\log x}\,.$$

Therefore

$$P(Err.) \leq 1.26\frac{n \log N}{N \log n}.$$

Choosing $N = cn$, we therefore obtain the bound

$$P(Err.) \leq \frac{1.26}{c}\left(1 + \frac{\log c}{\log n}\right)\,.$$

This is a small number even for small c. For instance with $n = 2^{23}$ and $N = 2^{32}$,

$$P(Err.) \leq 1.26\frac{n \log N}{N \log n} = 1.26\frac{2^{23} \times 32}{2^{32} \times 23} < 0.0035\,.$$

The prime number p is an $O(n)$. Therefore the number of bits required to transmit p and a is an $O(\log_2 n)$, to be compared with the crude method (transmitting a instead of its fingerprint). The algorithm requires to find a prime number $\leq N$. For this, we may pick at random a number $\leq N$, which is a prime number with probability $\pi(N)$. It will therefore take on the average $\log N$ primality tests before we find a prime number. Efficient randomized primality tests can be used, such that the one in Example 5.2.6.

Books for Further Information

For this chapter, see [Alon and Spencer, 1991, 2010], [Mitzenmacher and Upfal, 2005] and [Motwani and Raghavan, 1995]. The first reference is the main one on the subject and contains a wealth of examples. It is mathematicaly oriented, whereas the second one contains applications, mostly in the information and communications sciences. The third item is devoted to random algorithms. [Sinclair, 2011] treats most subjects and emphasizes the important algorithmic complexity aspects.

5.3 Exercises

Exercise 5.3.1. Another simple inequality
Let X be a non-negative random variable with mean μ. Prove that $P(X \leq \mu) > 0$.

Exercise 5.3.2. Graph colouring
Refer to Example 5.1.1. Show that if $n \leq 2^{k/2}$ and $k \geq 3$, there exists a 2-coloring of the complete graph K_n with no monochromatic complete subgraph of size k.

Exercise 5.3.3. A variant of the expectation argument
(a) Let X be an integer-valued random variable and r be an integer. Show that

$$E[X] \leq r \Longrightarrow P(X \leq r) > 0\,.$$

(b) Prove that a k-uniform hypergraph with m hyperedges can be coloured in such a way that at most $\frac{m}{2^{k-1}}$ hyperedges are monochromatic.

Exercise 5.3.4. Satisfiability
Consider the satisfiability problem of Example 5.1.13 with k literals and m clauses. Suppose that clause i features k_i distinct literals ($1 \leq i \leq m$). In the example just mentioned, $k_i \equiv k$, whereas here, we define $k = \min k_i$. Show that there exists at least one assignment of the literals such that at least $m\left(1 - 2^{-k}\right)$ clauses are satisfied.

Exercise 5.3.5. The dependency graph
Show that if B is mutually independent of the events $B_1, \ldots, B_\ell$ (see Definition 5.1.10), it is also mutually independent of the events $\tilde{B}_1, \ldots, \tilde{B}_\ell$, where either $\tilde{B}_i = B_i$ or $\overline{B}_i$, the choice varying arbitrarily from an index to the other.

Exercise 5.3.6. 2-colourable k-uniform hypergraph, take 1
A 2-colouring of a hypergraph H is the attribution to each vertex a colour, blue or red. A hyperedge of this hypergraph is called monochromatic if all its vertices have the same colour. A 2-colouring is called is proper if none of its hyperedges is monochromatic. A hypergraph is 2-colourable if it admits a proper 2-colouring.

Prove the following: A k-uniform hypergraph with less than 2^{k-1} hyperedges is 2-colourable.

Exercise 5.3.7. 2-colourable k-uniform hypergraph, take 2
Prove that a k-uniform hypergraph with m hyperedges can be 2-coloured in such a way that at most $\frac{m}{2^{k-1}}$ hyperedges are monochromatic.

Exercise 5.3.8. Light switches
There is an $n \times n$ array of lights that are either "on" or "off". There is for each row i a $\{0,1\}$-valued switch variable Y_i, with the following effect: if $Y_i = 0$ all the lamps in row i change states, otherwise they stay as they are. Similarly for each column j there is a $\{0,1\}$-valued switch variable Z_j, with the following effect: if $Z_j = 0$ all the lamps in column j change states, otherwise they stay as they are.

The aim of this exercise is to show that for any initial "on/off" configuration of lights, the states of the switches can be chosen such that the number of on lights is asymptotically $\frac{n^2}{2} + \sqrt{\frac{1}{2\pi}} \times n^{\frac{3}{2}}$. For this, starting with an arbitrary "on/off" configuration of lights, set randomly and uniformly the column switches. In other words the Z_j's form an IID family of variables, each of them uniformly distributed ($P(Z_j = 1) = \frac{1}{2}$). If light (i, j) is "on", we set $X(i, j) = 1$, and 0 otherwise. Show that

$$E\left[\left|\sum_j X(i,j)\right|\right] \sim \sqrt{\frac{2}{\pi}} n^{\frac{1}{2}}.$$

Now set the switch of row i so as to obtain a majority of lights "on" in this row and compute the resulting expectation of the excess number of "on" lights with respect to "off" lights. Conclude. (You may admit the following fact concerning the sum X of IID $\{-1, +1\}$-valued variables $B_1, \ldots, B_n$ uniformly distributed: $E[|X|] \sim \sqrt{\frac{2}{\pi}} n^{\frac{1}{2}}$.)

Exercise 5.3.9. TOURNAMENT 1
Players of a given game (say, tennis) are ranked, and this ranking is supposed to be strict (no ex-aequo). A tournament between n players is represented by a complete oriented graph with vertex set $V = \{1, 2, \ldots, n\}$, that is, a complete graph K_n where each edge is oriented. Here, an edge $\langle u, v\rangle$ represents a game between players u and v, and the arrow on this edge points to the looser. Show that for every tournament, there exists a ranking that disagrees with less than half the edges.

Exercise 5.3.10. TOURNAMENT 2
Recall that a random tournament on the complete graph K_n is the tournament for which the directions of the edges are chosen independently with probability $\frac{1}{2}$ for each direction. Show that for all $n \geq 1$, there is a tournament on n vertices with at least $n!2^{-(n-1)}$ Hamiltonian cycles.

Exercise 5.3.11. COLOURING
Let $G = (V, \mathcal{E})$ be a graph and let $L(v)$ be for each vertex v a list of colours. This defines a list assignment L. This list assignment is said to be of size k if $L(v) \leq k$ for all $v \in V$. A L-colouring of G is a colouring that assigns to each vertex v a colour $c(v)$ in the list $L(v)$. The graph G is called L-colourable if there exists a L-colouring that is proper, that is such that there exists no adjacent peir of vertices with the same colour.

Prove the following: Let L be a list assignment of size k. If for every vertex v, every colour $q \in L(v)$ appears in at most $\frac{1}{e}$ neighbours of v, then G is L-colourable.

Exercise 5.3.12. PATTERN MATCHING
The following situation occurs in DNA analysis. There is a long chain of symbols $x = x_1 x_2 \cdots x_n$ in which we try to detect the presence of a shorter sequence $y = y_1 y_2 \cdots y_m$. In the DNA context, the symbols are G, A, T, C, representing

the four nucleobases. By an obvious binary encoding of these four symbols, one can reduce the problem to the case where the symbols forming the chains x and y are binary digits.

The problem is therefore the following. Is there some j $(1 \leq j \leq n - m + 1)$ such that $x(j) := x_j x_{j+1} \cdots x_{j+m-1} = y_1 y_2 \cdots y_m$? The following algorithm is proposed. Select a prime number p uniformly at random among the prime numbers smaller than or equal to some number N. Then check if there is some j $(1 \leq j \leq m-n+1)$ such that $F_p(x(j)) = F_p(y)$ or, equivalently, such that p divides $|F_p(x(j)) - F_p(y)|$ (the definition of the function F_p is given in Example 5.2.8). If not, the return of the algorithm is that there is no match, and the answer is correct. If yes, the return is that there is a match, and this may not be true. Prove that with the choice $N = cnm$, the probability of error is bounded as follows:

$$P(\mathcal{E}) \leq 1.26 \frac{mn \log N}{\log(mn)N}.$$

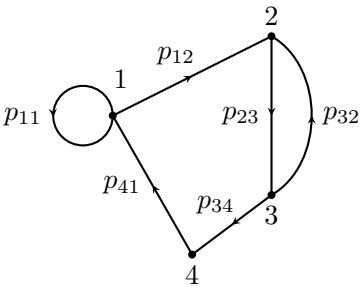
Chapter 6

Markov Chain Models

6.1 The Transition Matrix

6.1.1 Distribution of a Markov Chain

A particle on a denumerable set E. If at time n, the particle is in position $X_n = i$, it will be at time $n+1$ in a position $X_{n+1} = j$ chosen independently of the past trajectory X_{n-1}, X_{n-2} with probability p_{ij}. This can be represented by a labeled directed graph, called the transition graph, whose set of vertices is E, and for which there is a directed edge from $i \in E$ to $j \in E$ with label p_{ij} if and only the latter quantity is positive. Note that there may be "self-loops", corresponding to positions i such that $p_{ii} > 0$.



This graphical interpretation of as Markov chain in terms of a "random walk" on a set E is adapted to the study of random walks on graphs (see Chapter 8). Since the interpretation of a Markov chain in such terms is not always the natural one, we proceed to give a more formal definition.

Recall that a sequence $\{X_n\}_{n\geq 0}$ of random variables with values in a set E is called a discrete-time stochastic process with state space E. In this chapter, the state space is countable, and its elements will be denoted by i, j, k,... If $X_n = i$, the process is said to be in state i at time n, or to visit state i at time n.

Definition 6.1.1 *If for all integers $n \geq 0$ and all states $i_0, i_1, \ldots, i_{n-1}, i, j$,*

$$P(X_{n+1} = j \,|\, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j \,|\, X_n = i)\,,$$

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_6

this stochastic process is called a Markov chain, and a homogeneous Markov chain (HMC) *if, in addition, the right-hand side is independent of* n.

The matrix $\mathbf{P} = \{p_{ij}\}_{i,j\in E}$, where

$$p_{ij} = P(X_{n+1} = j \,|\, X_n = i),$$

is called the transition matrix of the HMC. Since the entries are probabilities, and since a transition from any state i must be to some state, it follows that

$$p_{ij} \geq 0, \text{ and } \sum_{k\in E} p_{ik} = 1$$

for all states i, j. A matrix $\mathbf{P}$ indexed by E and satisfying the above properties is called a stochastic matrix. The state space may be infinite, and therefore such a matrix is in general not of the kind studied in linear algebra. However, the basic operations of addition and multiplication will be defined by the same formal rules. The notation $x = \{x(i)\}_{i\in E}$ formally represents a column vector, and x^T is the corresponding row vector.

The Markov property easily extends (Exercise 6.4.2) to

$$P(A \,|\, X_n = i, B) = P(A \,|\, X_n = i)\,,$$

where

$$A = \{X_{n+1} = j_1, \ldots, X_{n+k} = j_k\}, B = \{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\}.$$

This is in turn equivalent to

$$P(A \cap B \,|\, X_n = i) = P(A \,|\, X_n = i)P(B \,|\, X_n = i).$$

That is, A and B are conditionaly independent given $X_n = i$. In other words, the future at time n and the past at time n are conditionally independent given the present state $X_n = i$. In particular, the Markov property is independent of the direction of time.

Notation. We shall from now on abbreviate $P(A \,|\, X_0 = i)$ as $P_i(A)$. Also, if μ is a probability distribution on E, then $P_\mu(A)$ is the probability of A given that the initial state X_0 is distributed according to μ.

The distribution at time n of the chain is the vector $\nu_n := \{\nu_n(i)\}_{i\in E}$, where

$$\nu_n(i) := P(X_n = i).$$

From the Bayes rule of exclusive and exhaustive causes, $\nu_{n+1}(j) = \sum_{i\in E} \nu_n(i)p_{ij}$, that is, in matrix form, $\nu_{n+1}^T = \nu_n^T\mathbf{P}$. Iteration of this equality yields

$$\nu_n^T = \nu_0^T\mathbf{P}^n. \tag{6.1}$$

The matrix $\mathbf{P}^m$ is called the m-step transition matrix because its general term is

$$p_{ij}(m) = P(X_{n+m} = j \,|\, X_n = i).$$

In fact, by the Bayes sequential rule and the Markov property, the right-hand side equals $\sum_{i_1,\ldots,i_{m-1}\in E} p_{ii_1}p_{i_1i_2}\cdots p_{i_{m-1}j}$, which is the general term of the m-th power of $\mathbf{P}$.

The probability distribution ν_0 of the initial state X_0 is called the initial distribution. From the Bayes sequential rule and in view of the homogeneous Markov property and the definition of the transition matrix,

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = \nu_0(i_0)p_{i_0i_1}\cdots p_{i_{k-1}i_k}\,.$$

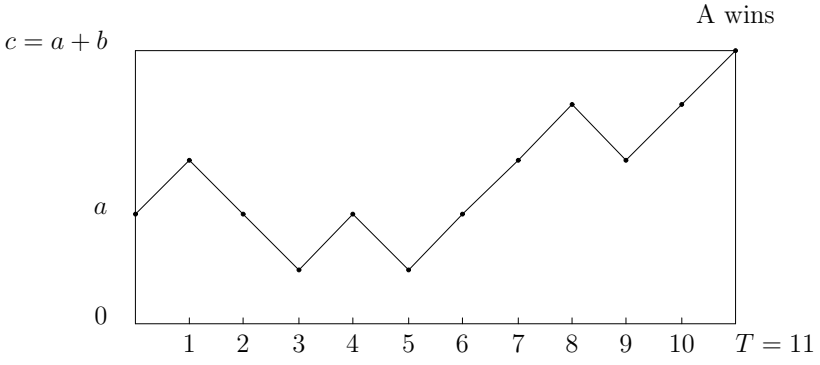
Therefore,

Theorem 6.1.2 *The distribution of a discrete-time* HMC *is uniquely determined by its initial distribution and its transition matrix.*

First-step Analysis

Some functionals of homogeneous Markov chains such as probabilities of absorption by a closed set (A is called closed if $\sum_{j\in A} p_{ij} = 1$ for all $i \in A$) and average times before absorption can be evaluated by a technique called first-step analysis.

EXAMPLE 6.1.3: THE GAMBLER'S RUIN, TAKE 1. Two players A and B play "heads or tails", where heads occur with probability $p \in (0,1)$, and the successive outcomes form an IID sequence. Calling X_n the fortune in dollars of player A at time n, then $X_{n+1} = X_n + Z_{n+1}$, where $Z_{n+1} = +1$ (*resp.*, -1) with probability p (*resp.*, $q := 1-p$), and $\{Z_n\}_{n\geq 1}$ is IID. In other words, A bets \$1 on heads at each toss, and B bets \$1 on tails. The respective initial fortunes of A and B are a and b (positive integers). The game ends when a player is ruined, and therefore the process $\{X_n\}_{n\geq 1}$ is a random walk as described in Example 6.1.5, except that it is restricted to $E = \{0,\ldots,a,a+1,\ldots,a+b=c\}$. The duration of the game is T, the first time n at which $X_n = 0$ or c, and the probability of winning for A is $u(a) = P(X_T = c \,|\, X_0 = a)$.



The gambler's ruin

Instead of computing $u(a)$ alone, first-step analysis computes

$$u(i) = P(X_T = c \,|\, X_0 = i)$$

for all states i ($0 \leq i \leq c$) and for this, it first generates a recurrence equation for $u(i)$ by breaking down event "A wins" according to what can happen after the first step (the first toss) and using the rule of exclusive and exhaustive causes. If $X_0 = i$, $1 \leq i \leq c-1$, then $X_1 = i+1$ (resp., $X_1 = i-1$) with probability p (resp., q), and the probability of winning for A with updated initial fortune $i+1$ (resp., $i-1$) is $u(i+1)$ (resp., $u(i-1)$). Therefore, for i ($1 \leq i \leq c-1$)

$$u(i) = pu(i+1) + qu(i-1),$$

with the boundary conditions $u(0) = 0$, $u(c) = 1$.

The characteristic equation associated with this linear recurrence equation is $pr^2 - r + q = 0$. It has two distinct roots, $r_1 = 1$ and $r_2 = \frac{q}{p}$, if $p \neq \frac{1}{2}$, and a double root, $r_1 = 1$, if $p = \frac{1}{2}$. Therefore, the general solution is $u(i) = \lambda r_1^i + \mu r_2^i = \lambda + \mu \left(\frac{q}{p}\right)^i$ when $p \neq q$, and $u(i) = \lambda r_1^i + \mu i r_1^i = \lambda + \mu i$ when $p = q = \frac{1}{2}$. Taking into account the boundary conditions, one can determine the values of λ and μ. The result is, for $p \neq q$,

$$u(i) = \frac{1 - (\frac{q}{p})^i}{1 - (\frac{q}{p})^c},$$

and for $p = q = \frac{1}{2}$,

$$u(i) = \frac{i}{c}.$$

In the case $p = q = \frac{1}{2}$, the probability $v(i)$ that B wins when the initial fortune of B is $c-i$ is obtained by replacing i by $c-i$ in expression for $u(i)$: $v(i) = \frac{c-i}{c} = 1 - \frac{i}{c}$. One checks that $u(i) + v(i) = 1$, which means in particular that the probability that the game lasts forever is null. The reader is invited to check that the same is true in the case $p \neq q$.

First-step analysis can also be used to compute average times before absorption (Exercise 6.4.7).

6.1.2 Sample Path Realization

Many HMC's receive a natural description in terms of a recurrence equation.

Theorem 6.1.4 *Let $\{Z_n\}_{n\geq 1}$ be an* IID *sequence of random variables with values in an arbitrary space F. Let E be a countable space, and $f : E \times F \to E$ be some function. Let X_0 be a random variable with values in E, independent of $\{Z_n\}_{n\geq 1}$. The recurrence equation*

$$X_{n+1} = f(X_n, Z_{n+1}) \tag{6.2}$$

then defines a HMC.

Proof. Iteration of recurrence (6.2) shows that for all $n \geq 1$, there is a function g_n such that $X_n = g_n(X_0, Z_1, \ldots, Z_n)$, and therefore $P(X_{n+1} = j \,|\, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(f(i, Z_{n+1}) = j \,|\, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(f(i, Z_{n+1}) = j)$, since the event $\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\}$ is expressible in terms of $X_0, Z_1, \ldots, Z_n$ and is therefore independent of Z_{n+1}. Similarly, $P(X_{n+1} = j \,|\, X_n = i) = P(f(i, Z_{n+1}) = j)$. We therefore have a Markov chain, and it is homogeneous since the right-hand side of the last equality does not depend on n. Explicitly:

$$p_{ij} = P(f(i, Z_1) = j)\,. \tag{6.3}$$

□

EXAMPLE 6.1.5: 1-D RANDOM WALK, TAKE 1. Let X_0 be a random variable with values in $\mathbb{Z}$. Let $\{Z_n\}_{n\geq 1}$ be a sequence of IID random variables, independent of X_0, taking the values $+1$ or -1, and with the probability distribution

$$P(Z_n = +1) = p,$$

where $p \in (0, 1)$. The process $\{X_n\}_{n\geq 1}$ defined by

$$X_{n+1} = X_n + Z_{n+1}$$

is, in view of Theorem 6.1.4, an HMC, called a random walk on $\mathbb{Z}$. It is called a "symmetric" random walk if $p = \frac{1}{2}$.

EXAMPLE 6.1.6: THE REPAIR SHOP, TAKE 1. During day n, Z_{n+1} machines break down, and they enter the repair shop on day $n+1$. Every day one machine among those waiting for service is repaired. Therefore, denoting by X_n the number of machines in the shop on day n,

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1}, \tag{6.4}$$

where $a^+ = \max(a, 0)$. The sequence $\{Z_n\}_{n\geq 1}$ is assumed to be an IID sequence, independent of the initial state X_0, with common probability distribution

$$P(Z_1 = k) = a_k,\ k \geq 0$$

of generating function g_Z.

This may also be interpreted in terms of communications. The model then describes a communications link in which time is divided into successive intervals (the "slots") of equal length, conventionally taken to be equal to 1. In slot n (extending from time n included to time $n+1$ excluded), there arrive Z_{n+1} messages requiring transmission. Since the link can transmit at most one message in a given slot, the messages may have to be buffered, and X_n represents the number of messages in the buffer (supposed of infinite capacity) at time n. The dynamics of the buffer content are therefore those of Eqn. (6.5).

The stochastic process $\{X_n\}_{n\geq 0}$ is a HMC of transition matrix

$$\mathbf{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

Indeed, by formula (6.3), $p_{ij} = P((i-1)^+ + Z_1 = j) = P(Z_1 = j - (i-1)^+)$.

Stochastic Automata

Many random sequences intervening in digital communications systems are described in terms of stochastic automata (see Exercise 6.4.14 for instance). Stochastic automata may also be useful in problems of pattern recognition (see Example 6.1.7 below).

A finite automaton $(E, \mathcal{A}, f)$ can read sequences of letters from a finite alphabet $\mathcal{A}$ written on some infinite tape. It can be in any state of a finite set E, and its evolution is governed by a function $f : E \times \mathcal{A} \to E$, as follows. When the automaton is in state $i \in E$ and reads letter $a \in \mathcal{A}$, it switches from state i to state $j = f(i, a)$ and then reads on the tape the next letter to the right.

An automaton can be represented by its transition graph G having for nodes the states of E. There is an oriented edge from the node (state) i to the node j if and only if there exists $a \in \mathcal{A}$ such that $j = f(i, a)$, and this edge then receives label a. If $j = f(i, a_1) = f(i, a_2)$ for $a_1 \neq a_2$, then there are two edges from i to j with labels a_1 and a_2, or, more economically, one such edge with label (a_1, a_2). More generally, a given oriented edge can have multiple labels of any order.

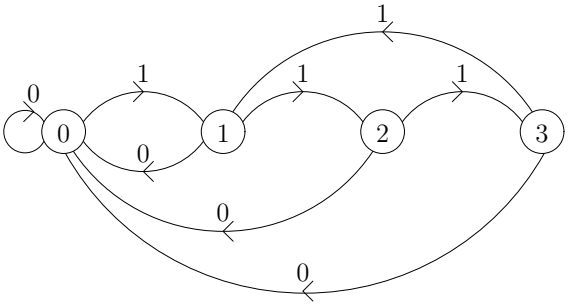
EXAMPLE 6.1.7: PATTERN DETECTION. Consider the automaton with alphabet $\mathcal{A} = \{0, 1\}$ corresponding to the transition graph of Figure (a). As the automaton, initialized in state 0, reads the sequence of Figure (b) from left to right, it passes successively through the states (including the initial state 0)

$$0\,1\,0\,0\,1\,2\,3\,1\,0\,0\,1\,2\,3\,1\,2\,3\,0\,1\,0\,.$$

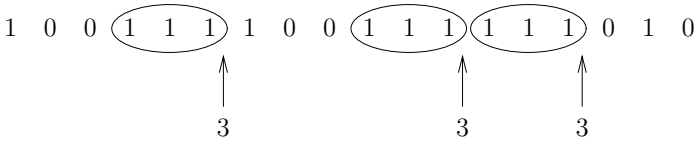
Rewriting the sequence of states below the sequence of letters, it appears that the automaton is in state 3 after it has seen three consecutive 1's. This automaton is therefore able to recognize and count such blocks of 1's. However, it does not take into account overlapping blocks (Figure (b)).

If the sequence of letters read by the automaton is $\{Z_n\}_{n\geq 1}$, the sequence of states $\{X_n\}_{n\geq 0}$ is then given by the recurrence equation $X_{n+1} = f(X_n, Z_{n+1})$ and therefore, if $\{Z_n\}_{n\geq 1}$ is i.i.d and independent of the initial state X_0, then $\{X_n\}_{n\geq 1}$ is, according to Theorem 6.1.4 an HMC.

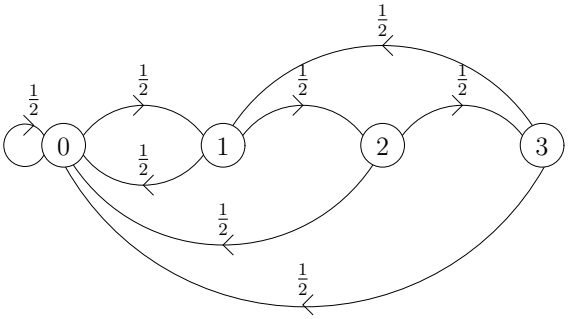
Not all homogeneous Markov chains receive a "natural" description of the type featured in Theorem 6.1.4. However, it is always possible to find a "theoretical" description of the kind. More exactly,



a



b



c

The automaton: the recognition process and the Markov chain

Theorem 6.1.8 *For any transition matrix* $\mathbf{P}$ *on* E*, there exists a homogeneous Markov chain with this transition matrix and with a representation such as in Theorem 6.1.4.*

Proof. Define

$$X_{n+1} := j \text{ if } \sum_{k=0}^{j-1} p_{X_n k} \leq Z_{n+1} < \sum_{k=0}^{j} p_{X_n k}\,,$$

where $\{Z_n\}_{n\geq 1}$ is IID, uniform on $[0,1]$. By application of Theorem 6.1.4 and of formula (6.3), we check that this HMC has the announced transition matrix. □

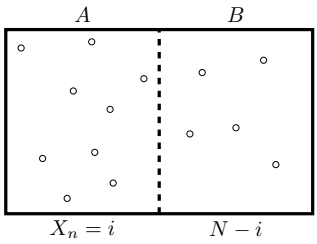
As we already mentioned, not all homogeneous Markov chains are naturally described by the model of Theorem 6.1.4. A slight modification of this result considerably enlarges its scope.

Theorem 6.1.9 *Let things be as in Theorem 6.1.4 except for the joint distribution of* $X_0, Z_1, Z_2, \ldots$*. Suppose instead that for all* $n \geq 0$*,* Z_{n+1} *is conditionally independent of* $Z_n, \ldots, Z_1, X_{n-1}, \ldots, X_0$ *given* X_n*, and that for all* $i, j \in E$*,* $P(Z_{n+1} = k \,|\, X_n = i)$ *is independent of* n*. Then* $\{X_n\}_{n\geq 0}$ *is a* HMC*, with transition probabilities*

$$p_{ij} = P(f(i, Z_1) = j \,|\, X_0 = i).$$

Proof. The proof is similar, *mutandis mutatis* to that of Theorem 6.1.4 and is left to the reader. □

EXAMPLE 6.1.10: THE EHRENFEST URN, TAKE 1. This idealized model of diffusion through a porous membrane, proposed in 1907 by the Austrian physicists Tatiana and Paul Ehrenfest to describe in terms of statistical mechanics the exchange of heat between two systems at different temperatures, considerably helped understanding the phenomenon of thermodynamic irreversibility (see Example 15.1.3). It features N particles that can be either in compartment A or in compartment B.



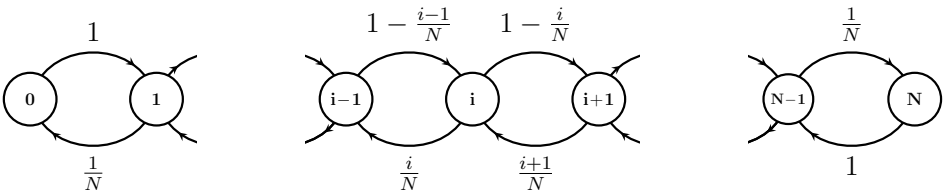
Suppose that at time $n \geq 0$, $X_n = i$ particles are in A. One then chooses a particle at random, and this particle is moved at time $n+1$ from where it is to the other compartment. Thus, the next state X_{n+1} is either $i-1$ (the displaced particle was

found in compartment A) with probability $\frac{i}{N}$, or $i+1$ (it was found in B) with probability $\frac{N-i}{N}$. This model pertains to Theorem 6.1.9. For all $n \geq 0$,

$$X_{n+1} = X_n + Z_{n+1},$$

where $Z_n \in \{-1, +1\}$ and $P(Z_{n+1} = -1 \,|\, X_n = i) = \frac{i}{N}$. The nonzero entries of the transition matrix are therefore

$$p_{i,i+1} = \frac{N-i}{N}, \quad p_{i,i-1} = \frac{i}{N}.$$



EXAMPLE 6.1.11: ALOHA, TAKE 1. A typical situation in a multiple-access satellite communications system is the following. Users—each one identified with a message—contend for access to a single-channel communications link. Two or more messages in the air at the same time jam each other, and are not successfully transmitted. The users are somehow able to detect a collision of this sort and will try to retransmit later the message involved in a collision. The difficulty in such communications systems resides mainly in the absence of cooperation among users, who are all unaware of the intention to transmit of competing users.

The *slotted* ALOHA protocol imposes on the users the following rules (see the figure below):

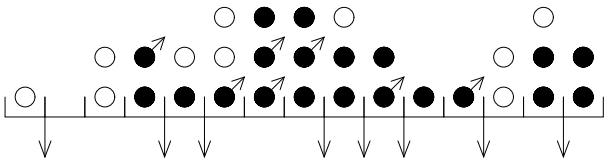
(i) Transmissions and retransmissions of messages can start only at equally spaced times; the interval between two consecutive (re-)transmission times is called a *slot*; the duration of a slot is always larger than that of any message.

(ii) All *backlogged* messages, that is, those messages having already tried unsuccessfully (maybe more than once) to get through the link, require retransmission independently of one another with probability $\nu \in (0, 1)$ at each slot. This is the so-called *Bernoulli retransmission policy.*

(iii) The *fresh messages*—those presenting themselves for the first time—immediately attempt to get through.

Let X_n be the number of backlogged messages at the beginning of slot n. The backlogged messages behave independently, and each one has probability ν of attempting retransmission in slot n. In particular, if there are $X_n = k$ backlogged messages, the probability that i among them attempt to retransmit in slot n is

$$b_i(k) = \binom{k}{i} \nu^i (1-\nu)^{k-i}.$$



○ fresh message

● backlogged message, not authorized to attempt retransmission

 backlogged message, authorized to attempt retransmission

 successful transmission (or retransmission)

The ALOHA protocol

Let A_n be the number of fresh requests for transmission in slot n. The sequence $\{A_n\}_{n\geq 0}$ is assumed IID with the distribution $P(A_n = j) = a_j$. The quantity $\lambda := E[A_n] = \sum_{i=1}^{\infty} i a_i$ is called the *traffic intensity*. We suppose that $a_0 + a_1 \in (0, 1)$, so that $\{X_n\}_{n\geq 0}$ is an irreducible HMC. Its transition matrix is

$$p_{ij} = \begin{cases} b_1(i)a_0 & \text{if } j = i-1, \\ [1-b_1(i)]a_0 + b_0(i)a_1 & \text{if } j = i, \\ [1-b_0(i)]a_1 & \text{if } j = i+1, \\ a_{j-i} & \text{if } j \geq i+2. \end{cases}$$

The proof is by accounting. For instance, the first line corresponds to one among the i backlogged messages having succeeded to retransmit, and for this there should be no fresh arrival (probability a_0) and only one of the i backlogged messages allowed to retransmit (probability $b_1(i)$). The second line corresponds to one of the two events "no fresh arrival and zero or strictly more than two retransmission requests from the backlog" and "zero retransmission request from the backlog and one fresh arrival."

Aggregation of States

Let $\{X_n\}_{n\geq 0}$ be a HMC with state space E and transition matrix $\mathbf{P}$, and let $(A_k, k \geq 1)$ be a countable partition of E. Define the process $\{\hat{X}_n\}_{n\geq 0}$ with state space $\hat{E} = \{\hat{1}, \hat{2}, \ldots\}$ by $\hat{X}_n = \hat{k}$ if and only if $X_n \in A_k$.

Theorem 6.1.12 *If $\sum_{j\in A_\ell} p_{ij}$ is independent of $i \in A_k$ for all k, ℓ, then $\{\hat{X}_n\}_{n\geq 0}$ is a* HMC *with transition probabilities $\hat{p}_{\hat{k}\hat{\ell}} = \sum_{j\in A_\ell} p_{ij}$ (any $i \in A_k$).*

Proof. a) Sufficiency. We have

$$P\left(\widehat{X}_{n+1}=\widehat{j}|\widehat{X}_n=\widehat{i},\widehat{X}_{n-1}=\widehat{i}_{n-1},\dots,\widehat{X}_0=\widehat{i}_0\right)$$
$$=\frac{P\left(\widehat{X}_{n+1}=\widehat{j},\widehat{X}_n=\widehat{i},\widehat{X}_{n-1}=\widehat{i}_{n-1},\dots,\widehat{X}_0=\widehat{i}_0\right)}{P\left(\widehat{X}_n=\widehat{i},\widehat{X}_{n-1}=\widehat{i}_{n-1},\dots,\widehat{X}_0=\widehat{i}_0\right)}=\frac{A}{B}$$

$$\begin{aligned}A=&P\left(X_{n+1}\in A_j,X_n\in A_i,X_{n-1}\in A_{i_{n-1}},\dots,X_0\in A_{i_0}\right)\\&=\sum_{k\in A_i}P\left(X_{n+1}\in A_j,X_n=k,X_{n-1}\in A_{i_{n-1}},\dots,X_0\in A_{i_0}\right)\\&=\sum_{k\in A_i}P\left(X_{n+1}\in A_j|X_n=k\right)P\left(X_n=k,X_{n-1}\in A_{i_{n-1}},\dots,X_0\in A_{i_0}\right)\end{aligned}$$

If we suppose that $\sum_{l\in A_j}p_{kl}$ is independent of $k\in A_i$, and if we denote this quantity by $\widehat{p}_{\widehat{i}\widehat{j}}$, we have

$$A=\widehat{p}_{\widehat{i}\widehat{j}}\sum_{k\in A_i}P\left(X_n=k,X_{n-1}\in A_{i_{n-1}},\dots,X_0\in A_{i_0}\right)=\widehat{p}_{\widehat{i}\widehat{j}}B$$

and therefore

$$P\left(\widehat{X}_{n+1}=\widehat{j}|\widehat{X}_n=\widehat{i},\widehat{X}_{n-1}=\widehat{i}_{n-1},\dots,\widehat{X}_0=\widehat{i}_0\right)=\widehat{p}_{\widehat{i}\widehat{j}}$$

By the result of Exercise 2.4.21, this suffices to show that the process $\left\{\widehat{X}_n\right\}_{n\geq 0}$ is a HMC with transition probabilities $\widehat{p}_{\widehat{i}\widehat{j}}$.

b) Necessity. By hypothesis, $\{\widehat{X}_n\}_{n\geq 0}$ is a HMC, for all initial distribution μ of $\{X_n\}_{n\geq 0}$. In particular, with an initial distribution putting all its mass on a fixed $\ell\in A_i$,

$$\widehat{p}_{\widehat{i}\widehat{j}}=P\left(\widehat{X}_1=\widehat{j}|\widehat{X}_0=\widehat{i}\right)=P\left(X_1\in A_j|X_0=l\right)=\sum_{k\in A_j}p_{lk}$$

Therefore for all j, the quantity $\sum_{k\in A_j}p_{lk}$ is independent of $l\in A_i$. □

EXAMPLE 6.1.13: THE EHRENFEST URN, TAKE 2. Let $\{\widetilde{X}_n\}_{n\geq 0}$ be the Markov chain with state space $\widetilde{E}:=\{0,1\}^N$ and denote by $x=(x_1,\dots,x_N)$ the generic state. The general term $\widetilde{p}_{x,y}$ of the transition matrix $\widetilde{\mathbf{P}}$ is non null only if x and y differ in exactly one position, and in this case

$$\widetilde{p}_{x,y}=\frac{1}{N}\,.$$

One immediately verifies that the uniform distribution is a stationary distribution and satisfies the detailed balance equation. Aggregating with respect to the partition

$$A_i:=\left\{x\in\{0,1\}^N\,;\,h(x)=i\right\}\ (1\leq i\leq N)$$

gives the Ehrenfest model.

6.1.3 Communication and Period

Communication and period are topological properties in the sense that they concern only the naked transition graph (with only the arrows, without the labels).

Communication Classes and Irreducibility

Definition 6.1.14 *State j is said to be accessible from state i if there exists $M \geq 0$ such that $p_{ij}(M) > 0$. States i and j are said to communicate if i is accessible from j and j is accessible from i, and this is denoted by $i \leftrightarrow j$.*

In particular, a state i is always accessible from itself, since $p_{ii}(0) = 1$ ($\mathbf{P}^0 = I$, the identity).

For $M \geq 1$, $p_{ij}(M) = \sum_{i_1,\ldots,i_{M-1}} p_{ii_1} \cdots p_{i_{M-1}j}$, and therefore $p_{ij}(M) > 0$ if and only if there exists at least one path $i, i_1, \ldots, i_{M-1}, j$ from i to j such that

$$p_{ii_1} p_{i_1 i_2} \cdots p_{i_{M-1}j} > 0,$$

or, equivalently, if there is a directed path from i to j in the transition graph G. Clearly,

$$\begin{aligned} i \leftrightarrow i \qquad & \text{(reflexivity)}, \\ i \leftrightarrow j \Rightarrow j \leftrightarrow i \qquad & \text{(symmetry)}, \\ i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k \qquad & \text{(transivity)}. \end{aligned}$$

Therefore, the communication relation ($\leftrightarrow$) is an equivalence relation, and it generates a partition of the state space E into disjoint equivalence classes called communication classes.

Definition 6.1.15 *A state i such that $p_{ii} = 1$ is called closed. More generally, a set C of states such that for all $i \in C$, $\sum_{j \in C} p_{ij} = 1$ is called closed.*

Definition 6.1.16 *If there exists only one communication class, then the chain, its transition matrix, and its transition graph are said to be irreducible.*

EXAMPLE 6.1.17: THE REPAIR SHOP, TAKE 3. Recall that this Markov chain satisfies the recurrence equation

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1}, \tag{6.5}$$

where $a^+ = \max(a, 0)$. The sequence $\{Z_n\}_{n \geq 1}$ is assumed to be IID, independent of the initial state X_0, and with common probability distribution

$$P(Z_1 = k) = a_k,\ k \geq 0$$

of generating function g_Z.

This chain is irreducible if and only if $P(Z_1 = 0) > 0$ and $P(Z_1 \geq 2) > 0$ as we now prove formally. Looking at (6.5), we make the following observations. If

$P(Z_{n+1} = 0) = 0$, then $X_{n+1} \geq X_n$ a.s. and there is no way of going from i to $i-1$. If $P(Z_{n+1} \leq 1) = 1$, then $X_{n+1} \leq X_n$, and there is no way of going from i to $i+1$. Therefore, the two conditions $P(Z_1 = 0) > 0$ and $P(Z_2 \geq 2) > 0$ are *necessary* for irreducibility. They are also sufficient. Indeed if there exists an integer $k \geq 2$ such that $P(Z_{n+1} = k) > 0$, then one can jump with positive probability from any $i > 0$ to $i + k - 1 > i$ or from $i = 0$ to $k > 0$. Also if $P(Z_{n+1} = 0) > 0$, one can step down from $i > 0$ to $i - 1$ with positive probability. In particular, one can go from i to $j < i$ with positive probability. Therefore, one way to travel from i to $j \geq i$, is by taking several successive steps of height at least $k-1$ in order to reach a state $l \geq i$, and then (in the case of $l > i$) stepping down one stair at a time from l to i. All this with positive probability.

EXAMPLE 6.1.18: HARMONIC FUNCTIONS OF AN IRREDUCIBLE HMC. Consider an irreducible HMC with finite state space E and let $h : E \to \mathbb{R}$ be a harmonic function. We show that h is necessarily a constant. (This result will be generalized in Theorem 17.3.8.)

Proof. Let $G := \{i \in E\,;\, h(i) = \max_{j \in E} h(j)\}$. Since E is finite, G is non-empty. For $i \in G$, let $N_i := \{j\,;\, p_{ij} > 0\}$. We show that if $j_0 \in N_i$, then $h(j_0) = h(i)$. Otherwise, $h(j_0) < h(i)$ and

$$h(i) = p_{i,j_0} h(j_0) + \sum_{j \in N_i \setminus j_0} p_{ij} h(j) < h(i)\,,$$

a contradiction. Therefore $h(j) = h(i)$ for all j such that $p_{ij} > 0$.

Since the chain is irreducible, from any $j \neq i$ there is a path $i_0 = i$, $j_1 \ldots$, $j_k := j$ such that $p_{i_{\ell-1} i_\ell} > 0$ $(1 \leq \ell \leq k)$. In particular, since $h(i)$ is a maximum, $h(i_1) = h(i)$ is also a maximum, and so is $h(i_2) = h(i_1)$, and so on, so that $h(j) = h(i)$. Therefore h is a constant. □

Period and Aperiodicity

Consider the random walk on $\mathbb{Z}$ (Example 6.1.5). Since $0 < p < 1$, it is irreducible. Observe that $E = C_0 + C_1$, where C_0 and C_1, the set of even and odd relative integers respectively, have the following property. If you start from $i \in C_0$ (resp., C_1), then in one step you can go only to a state $j \in C_1$ (resp., C_0). The chain $\{X_n\}$ passes alternately from one cyclic class to the other. In this sense, the chain has a periodic behavior, with a period equal to 2. More generally, for any irreducible Markov chain, one can find a unique partition of E into d classes $C_0, C_1, \ldots, C_{d-1}$ such that for all $k, i \in C_k$,

$$\sum_{j \in C_{k+1}} p_{ij} = 1,$$

where by convention $C_d = C_0$, and where d is maximal (that is, there is no other such partition $C'_0, C'_1, \ldots, C'_{d'-1}$ with $d' > d$). The proof follows directly from Theorem 6.1.21 below.

The number $d \geq 1$ is called the period of the chain (resp., of the transition matrix, of the transition graph). The classes $C_0, C_1, \ldots, C_{d-1}$ are called the cyclic classes. The chain moves cyclically from one class to the next.

We now give the formal definition of period. It is based on the notion of greatest common divisor of a set of positive integers.

Definition 6.1.19 *The period d_i of state $i \in E$ is, by definition,*

$$d_i = \text{GCD}\{n \geq 1\,;\, p_{ii}(n) > 0\},$$

with the convention $d_i = +\infty$ if there is no $n \geq 1$ with $p_{ii}(n) > 0$. If $d_i = 1$, the state i is called aperiodic .

Theorem 6.1.20 *If states i and j communicate they have the same period.*

Proof. As i and j communicate, there exist integers N and M such that $p_{ij}(M) > 0$ and $p_{ji}(N) > 0$. For any $k \geq 1$,

$$p_{ii}(M + nk + N) \geq p_{ij}(M)(p_{jj}(k))^n p_{ji}(N)$$

(indeed, the path $X_0 = i, X_M = j, X_{M+k} = j, \ldots, X_{M+nk} = j, X_{M+nk+N} = i$ is just one way of going from i to i in $M + nk + N$ steps). Therefore, for any $k \geq 1$ such that $p_{jj}(k) > 0$, we have $p_{ii}(M + nk + N) > 0$ for all $n \geq 1$. Therefore, d_i divides $M+nk+N$ for all $n \geq 1$, and in particular, d_i divides k. We have therefore shown that d_i divides all k such that $p_{jj}(k) > 0$, and in particular, d_i divides d_j. By symmetry, d_j divides d_i, so that finally, $d_i = d_j$. □

We can therefore speak of *the* period of a communication class or of an irreducible chain.

The important result concerning periodicity is the following.

Theorem 6.1.21 *Let $\mathbf{P}$ be an irreducible stochastic matrix with period d. Then for all states i, j there exist $m \geq 0$ and $n_0 \geq 0$ (m and n_0 possibly depending on i, j) such that*

$$p_{ij}(m + nd) > 0, \quad \textit{for all } n \geq n_0.$$

Proof. It suffices to prove the theorem for $i = j$. Indeed, there exists m such that $p_{ij}(m) > 0$, because j is accessible from i, the chain being irreducible, and therefore, if for some $n_0 \geq 0$ we have $p_{jj}(nd) > 0$ for all $n \geq n_0$, then $p_{ij}(m+nd) \geq p_{ij}(m)p_{jj}(nd) > 0$ for all $n \geq n_0$.

The rest of the proof is an immediate consequence of a classical result of number theory. Indeed, the GCD of the set $A = \{k \geq 1; p_{jj}(k) > 0\}$ is d, and A is closed under addition. The set A therefore contains all but a finite number of the positive multiples of d. In other words, there exists an integer n_0 such that $n > n_0$ implies $p_{jj}(nd) > 0$. □

6.2 Stationary Distribution and Reversibility

6.2.1 The Global Balance Equation

The central notion of the stability theory of discrete-time HMC's is that of a stationary distribution.

Definition 6.2.1 *A probability distribution π satisfying*

$$\pi^T = \pi^T \mathbf{P} \tag{6.6}$$

is called a stationary distribution of the transition matrix $\mathbf{P}$, or of the corresponding HMC.

The global balance equation (6.6) says that for all states i,

$$\pi(i) = \sum_{j \in E} \pi(j) p_{ji}.$$

Iteration of (6.6) gives $\pi^T = \pi^T \mathbf{P}^n$ for all $n \geq 0$, and therefore, in view of (6.1), if the initial distribution $\nu = \pi$, then $\nu_n = \pi$ for all $n \geq 0$. Thus, if a chain is started with a stationary distribution, it keeps the same distribution forever. But there is more, because then,

$$\begin{aligned} P(X_n = i_0, X_{n+1} = i_1, \ldots, X_{n+k} = i_k) &= P(X_n = i_0) p_{i_0 i_1} \cdots p_{i_{k-1} i_k} \\ &= \pi(i_0) p_{i_0 i_1} \cdots p_{i_{k-1} i_k} \end{aligned}$$

does not depend on n. In this sense the chain is stationary. One also says that the chain is in a stationary regime, or in equilibrium, or in steady state. In summary:

Theorem 6.2.2 *An* HMC *whose initial distribution is a stationary distribution is stationary.*

The balance equation $\pi^T \mathbf{P} = \pi^T$, together with the requirement that π be a probability vector, that is, $\pi^T \mathbf{1} = 1$ (where $\mathbf{1}$ is a column vector with all its entries equal to 1), constitute $|E| + 1$ equations for $|E|$ unknown variables. One of the $|E|$ equations in $\pi^T \mathbf{P} = \pi^T$ is superfluous given the constraint $\pi^T \mathbf{1} = 1$. Indeed, summing up all equalities of $\pi^T \mathbf{P} = \pi^T$ yields the equality $\pi^T \mathbf{P} \mathbf{1} = \pi^T \mathbf{1}$, that is, $\pi^T \mathbf{1} = 1$.

EXAMPLE 6.2.3: TWO-STATE MARKOV CHAIN. Take $E = \{1, 2\}$ and define the transition matrix

$$\mathbf{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix},$$

where $\alpha, \beta \in (0, 1)$. The global balance equations are

$$\pi(1) = \pi(1)(1 - \alpha) + \pi(2)\beta\,, \qquad \pi(2) = \pi(1)\alpha + \pi(2)(1 - \beta)\,.$$

These two equations are dependent and reduce to the single equation $\pi(1)\alpha = \pi(2)\beta$, to which must be added the constraint $\pi(1) + \pi(2) = 1$ expressing that π is a probability vector. We obtain

$$\pi(1) = \frac{\beta}{\alpha+\beta}, \quad \pi(2) = \frac{\alpha}{\alpha+\beta}.$$

EXAMPLE 6.2.4: THE EHRENFEST URN, TAKE 3. The global balance equations are, for $i \in [1, N-1]$,

$$\pi(i) = \pi(i-1)\left(1 - \frac{i-1}{N}\right) + \pi(i+1)\frac{i+1}{N}$$

and, for the boundary states, $\pi(0) = \pi(1)\frac{1}{N}$, $\pi(N) = \pi(N-1)\frac{1}{N}$. Leaving $\pi(0)$ undetermined, one can solve the balance equations for $i = 0, 1, \ldots, N$ successively, to obtain $\pi(i) = \pi(0)\binom{N}{i}$. The value of $\pi(0)$ is then determined by writing down that π is a probability vector: $1 = \sum_{i=0}^{N} \pi(i) = \pi(0)\sum_{i=0}^{N}\binom{N}{i} = \pi(0)2^N$. This gives for π the binomial distribution of size N and parameter $\frac{1}{2}$:

$$\pi(i) = \frac{1}{2^N}\binom{N}{i}.$$

This is the distribution one would obtain by assigning independently to each particle a compartment, with probability $\frac{1}{2}$ for each compartment.

There may exist several stationary distributions. Take the identity as transition matrix. Then any probability distribution on the state space is a stationary distribution. Also, it may occur that the chain has no stationary distribution. See Exercise 6.4.12.

EXAMPLE 6.2.5: THE LAZY MARKOV CHAIN, TAKE 1. Let $\mathbf{P}$ be the transition matrix of a HMC with state space E. The matrix

$$\mathbf{Q} := \frac{I + \mathbf{P}}{2}$$

is clearly a transition matrix, that of an HMC called the lazy version of the original one. In the lazy version, a move is decided after tossing a fair coin. If heads, the lazy traveler stays still, otherwise, he moves according to $\mathbf{P}$. Clearly, a stationary distribution of $\mathbf{P}$ is also a stationary distribution of $\mathbf{Q}$.

Both chains are simultaneously irreducible or not irreducible. However, in the irreducible case, the lazy chain is always aperiodic (since $q_{ii} > 0$) whereas the original chain may be periodic.

EXAMPLE 6.2.6: LAZY WALK ON THE HYPERCUBE, TAKE 1. The N-hypercube is a graph whose set of vertices is $V = \{0,1\}^N$ and its set of edges $\mathcal{E}$ consists of the pairs of vertices $\langle x, y\rangle$ that are adjacent in the sense that there exists an index i $(1 \le i \le N)$ such that $y = x^{(i)} := (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_N)$. The (pure) random walk on the hypercube is the HMC describing the motion of a particle along the edges at random. That is to say, if the position at a given time is x, the next position is $x^{(i)}$ where i is chosen uniformly at random among $\{1, 2, \ldots, N\}$ independently of all that happened before.

To avoid periodicity, we consider the *lazy* random walk, for which the decision to move depends on the result of a fair coin toss. More precisely, $p_{x,x} = \frac{1}{2}$ and if y is adjacent to x, $p_{xy} = \frac{1}{2N}$. This modification does not change the stationary distribution, which is the uniform distribution.

We may always describe, distributionwise, the HMC $\{X_n\}_{n\ge0}$ in the manner of Theorem 6.1.4, that is $X_{n+1} = f(X_n, Z_{n+1})$ where $\{Z_n\}_{n\ge1}$ is an IID sequence of random variables uniformly distributed on $\{1, \ldots, N\}$ independent of the initial state X_0: take $Z_n = (U_n, B_n)$ where the sequence $\{(U_n, B_n)\}_{n\ge1}$ is IID and uniformly distributed on $\{1, 2, \ldots, N\} \times \{0, 1\}$. The position at time $n+1$ is that of X_n except that the bit in position U_{n+1} is replaced by B_{n+1}.

6.2.2 Reversibility and Detailed Balance

The notions of time-reversal and time-reversibility are very productive, as we shall see in several occasions in the sequel.

Let $\{X_n\}_{n\ge0}$ be an HMC with transition matrix $\mathbf{P}$ and admitting a stationary distribution $\pi > 0$ (meaning $\pi(i) > 0$ for all states i). Define the matrix $\mathbf{Q}$, indexed by E, by

$$\pi(i)q_{ij} = \pi(j)p_{ji}. \tag{6.7}$$

This is a stochastic matrix since

$$\sum_{j\in E} q_{ij} = \sum_{j\in E} \frac{\pi(j)}{\pi(i)} p_{ji} = \frac{1}{\pi(i)} \sum_{j\in E} \pi(j)p_{ji} = \frac{\pi(i)}{\pi(i)} = 1\,,$$

where the third equality uses the global balance equations. Its interpretation is the following: Suppose that the initial distribution of the chain is π, in which case for all $n \ge 0$ and all $i \in E$, $P(X_n = i) = \pi(i)$. Then, from the Bayes retrodiction formula,

$$P(X_n = j \,|\, X_{n+1} = i) = \frac{P(X_{n+1} = i \,|\, X_n = j)P(X_n = j)}{P(X_{n+1} = i)},$$

that is, in view of (6.7),

$$P(X_n = j \,|\, X_{n+1} = i) = q_{ji}\,.$$

We see that $\mathbf{Q}$ is the transition matrix of the initial chain when time is reversed.

The following is a very simple observation that will be promoted to the rank of a theorem in view of its usefulness and also for the sake of easy reference.

Theorem 6.2.7 *Let* $\mathbf{P}$ *be a stochastic matrix indexed by a countable set* E*, and let* π *be a probability distribution on* E*. Define the matrix* $\mathbf{Q}$ *indexed by* E *by (6.7). If* $\mathbf{Q}$ *is a stochastic matrix, then* π *is a stationary distribution of* $\mathbf{P}$*.*

Proof. For fixed $i \in E$, sum equalities (6.7) with respect to $j \in E$ to obtain

$$\sum_{j \in E} \pi(i) q_{ij} = \sum_{j \in E} \pi(j) p_{ji}.$$

This is the global balance equation since the left-hand side is equal to $\pi(i) \sum_{j \in E} q_{ij} = \pi(i)$. □

Definition 6.2.8 *One calls reversible a stationary Markov chain such that the initial distribution* π *(a stationary distribution) satisfies the so-called detailed balance equations*

$$\pi(i) p_{ij} = \pi(j) p_{ji} \qquad (i, j \in E)\,. \tag{6.8}$$

One also says: the pair $(\mathbf{P}, \pi)$ *is reversible.*

In this case, $q_{ij} = p_{ij}$, and therefore the chain and the time-reversed chain are statistically the same, since the distribution of a homogeneous Markov chain is entirely determined by its initial distribution and its transition matrix.

The following is an immediate corollary of Theorem 6.2.7.

Theorem 6.2.9 *Let* $\mathbf{P}$ *be a transition matrix on the countable state space* E*, and let* π *be some probability distribution on* E*. If for all* $i, j \in E$*, the detailed balance equations (6.8) are satisfied, then* π *is a stationary distribution of* $\mathbf{P}$*.*

EXAMPLE 6.2.10: THE EHRENFEST URN, TAKE 4. The verification of the detailed balance equations $\pi(i) p_{i,i+1} = \pi(i+1) p_{i+1,i}$ is immediate.

Random Walk on a Group

Let G be a finite associative group with respect to the operation $*$ and let the inverse of $a \in G$ be denoted by a^{-1} and the identity by id. Let μ be a probability distribution on G. Let X_0 be an arbitrary random element of G, and let $\{Z_n\}_{n \geq 1}$ be a sequence of IID random elements of G, independent of X_0, with common distribution μ. The recurrence equation

$$X_{n+1} = Z_{n+1} * X_n \tag{6.9}$$

defines according to Theorem 6.1.4 an HMC whose transition probabilities are

$$P_{g,h*g} = \mu(h)$$

for all $g, h \in G$.

For $H \subset G$, denote by $\langle H \rangle$ the smallest subgroup of G containing H. Recall that $\langle H \rangle$ consists of all elements of the type $b_r * b_{r-1} * \cdots * b_1$ where the b_i's are elements of H or inverses of elements of H. Let $S = \{g \in G;\ \mu(g) > 0\}$.

Theorem 6.2.11 *(a) The random walk is irreducible if and only if S generates G, that is, $\langle S \rangle = G$.*

(b) The uniform distribution U on G is a stationary distribution of the chain.

Proof. (a) Assume irreducibility. Let $a \in G$. There exists $r > 0$ such that $p_{e,a}(r) > 0$, that is, there exists a sequence $s_1, \ldots, s_r$ of S such that $a = s_r * \cdots * s_1$. Therefore $a \in \langle S \rangle$. Conversely, suppose that S generates G. Let $a, b \in G$. The element $b * a^{-1}$ is therefore of the type $u_r * u_{r-1} * \cdots * u_1$ where the u_i's are elements of S or inverses of elements of S. Now, every element of G is of finite order, that is, can be written as a power of some element of G. Therefore $b * a^{-1}$ can be written as $b * a^{-1} = s_r * \cdots * s_1$ where the s_i's are in S. In particular, $p_{a,b}(r) > 0$.

(b) In fact

$$\sum_{g \in G} U(g) p_{g,f} = \frac{1}{|G|} \sum_{h \in G} p_{h^{-1}*f,f} = \frac{1}{|G|} \sum_{h \in G} \mu(h) = \frac{1}{|G|}\,.$$

□

The probability distribution μ on G is called symmetric iff $\mu(g) = \mu(g^{-1})$ for all $g \in G$. If this is the case, then the chain is reversible. We just have to check the detailed balance equations

$$U(g) p_{g,h} = U(h) p_{h,g}$$

that is

$$\frac{1}{|G|} \mu(h * g^{-1}) = \frac{1}{|G|} \mu(g * h^{-1})\,,$$

which is true because of the assumed symmetry of μ.

6.3 Finite State Space

6.3.1 Perron–Fröbenius

Consider an HMC that is irreducible and positive recurrent. If its initial distribution is the stationary distribution, it keeps the same distribution at all times. The chain is then said to be in the stationary regime, or in equilibrium, or in steady state. A question arises naturally: What is the long-run behavior of the chain when the initial distribution μ is arbitrary? The classical form of the main result in this direction is that for arbitrary states i and j,

$$\lim_{n\uparrow\infty} p_{ij}(n) = \pi(j)\,, \tag{6.10}$$

if the chain is ergodic, according to the following definition:

Definition 6.3.1 *An irreducible positive recurrent and aperiodic* HMC *is called ergodic.*

When the state space is finite, the asymptotic behavior of the n-step transition matrix depends on the eigenstructure of the transition matrix. The Perron–Fröbenius theorem detailing the eigenstructure of non-negative matrices is therefore all that is needed, at least in theory.

The basic results of the theory of matrices relative to eigenvalues and eigenvectors are reviewed in the appendix, from which we quote the following one, relative to a square matrix A of dimension r with *distinct* eigenvalues denoted $\lambda_1, \ldots, \lambda_r$. Let $u_1, \ldots, u_r$ and $v_1, \ldots, v_r$ be the associated sequences of left and right-eigenvectors, respectively. Then, $u_1, \ldots, u_r$ form an independent collection of vectors, and so do $v_1, \ldots, v_r$. Also, $u_i^T v_j = 0$ if $i \neq j$. Since eigenvectors are determined up to multiplication by an arbitrary non-null scalar, one can choose them in such a way that $u_i^T v_i = 1$ $(1 \leq i \leq r)$. We then have the spectral decomposition

$$A^n = \sum_{i=1}^{r} \lambda_i^n v_i u_i^T. \tag{6.11}$$

EXAMPLE 6.3.2: TWO-STATE CHAIN. Consider the transition matrix on $E = \{1, 2\}$

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where $\alpha, \beta \in (0, 1)$. Its characteristic polynomial $(1-\alpha-\lambda)(1-\beta-\lambda)-\alpha\beta$ admits the roots $\lambda_1 = 1$ and $\lambda_2 = 1-\alpha-\beta$. Observe at this point that $\lambda = 1$ is always an eigenvalue of a stochastic $r \times r$ matrix $\mathbf{P}$, associated with the right-eigenvector $v = \mathbf{1}$ with all entries equal to 1, since $\mathbf{P1} = \mathbf{1}$. Also, the stationary distribution $\pi^T = \frac{1}{\alpha+\beta}(\beta, \alpha)$ is the left-eigenvector corresponding to the eigenvalue 1. In this example, the representation (6.11) takes the form

$$\mathbf{P}^n = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & +\beta \end{pmatrix},$$

and therefore, since $|1-\alpha-\beta| < 1$,

$$\lim_{n\uparrow\infty} \mathbf{P}^n = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}.$$

In particular, the result of convergence to steady state,

$$\lim_{n\uparrow\infty} \mathbf{P}^n = \mathbf{1}\pi^T,$$

is obtained for this special case in a purely algebraic way. In addition, this algebraic method gives the convergence speed, which is exponential and determined by the second-largest eigenvalue absolute value. This is a general fact, which follows from the Perron–Frobenius theory of non-negative matrices below.

A matrix $A = \{a_{ij}\}_{1\le i,j\le r}$ with real coefficients is called non-negative (resp., positive) if all its entries are non-negative (resp., positive). A non-negative matrix A is called stochastic if $\sum_{j=1}^{r} a_{ij} = 1$ for all i, and substochastic if $\sum_{j=1}^{r} a_{ij} \le 1$ $(1 \le i \le r)$, with strict inequality for at least one i.

Non-negativity (resp., positivity) of A will be denoted by $A \ge 0$ (resp., $A > 0$). If A and B are two matrices of the same dimensions with real coefficients, the notation $A \ge B$ (resp., $A > B$) means that $A - B \ge 0$ (resp., $A - B > 0$).

The communication graph of a square non-negative matrix A is the directed graph with the state space $E = \{1, \dots, r\}$ as its set of vertices and a directed edge from vertex i to vertex j if and only if $a_{ij} > 0$.

A non-negative square matrix A is called irreducible (resp., irreducible aperiodic) if it has the same communication graph as an irreducible (resp., irreducible aperiodic) stochastic matrix. It is called primitive if there exists an integer k such that $A^k > 0$.

EXAMPLE 6.3.3: A non-negative matrix is primitive if and only if it is irreducible and aperiodic (Exercise 15.3.1).

(Perron, 1907; Frobenius, 1908, 1909, 1912) Let A be a non-negative primitive $r \times r$ matrix. Then, there exists a real eigenvalue λ_1 with algebraic as well as geometric multiplicity one such that $\lambda_1 > 0$, and $\lambda_1 > |\lambda_j|$ for any other eigenvalue λ_j. Moreover, the left-eigenvector u_1 and the right-eigenvector v_1 associated with λ_1 can be chosen positive and such that $u_1^T v_1 = 1$.

Let $\lambda_2, \lambda_3, \dots, \lambda_r$ be the eigenvalues of A other than λ_1 ordered in such a way that

$$\lambda_1 > |\lambda_2| \ge \cdots \ge |\lambda_r| \,. \tag{6.12}$$

The quantity $|\lambda_2|$ is the second largest eigenvalue modulus, abbreviated as "SLEM".

We may always order the eigenvalues in such a way that if $|\lambda_2| = |\lambda_j|$ for some $j \ge 3$, then $m_2 \ge m_j$, where m_j is the algebraic multiplicity of λ_j. Then

$$A^n = \lambda_1^n v_1 u_1^T + O(n^{m_2-1}|\lambda_2|^n). \tag{6.13}$$

If in addition A is stochastic (resp., substochastic), $\lambda_1 = 1$ (resp., $\lambda_1 < 1$).

If A is stochastic and irreducible with period $d > 1$, then there are exactly d distinct eigenvalues of modulus 1, namely the d-th roots of unity, and all other eigenvalues have modulus strictly less than 1.

6.3.2 The Limit Distribution

The next result generalizes the observation in Example 6.3.2 and is a direct consequence of the Perron–Fröbenius theorem.

Theorem 6.3.4 *If* $\mathbf{P}$ *is a transition matrix on* $E = \{1, \ldots, r\}$ *that is irreducible and aperiodic, and therefore primitive, then*

$$v_1 = \mathbf{1},\ u_1 = \pi,$$

where π *is the unique stationary distribution. Therefore*

$$\mathbf{P}^n = \mathbf{1}\pi^T + O(n^{m_2-1}|\lambda_2|^n)\,. \tag{6.14}$$

Quasi-stationary Distributions

Let $\{X_n\}_{n\geq 0}$ be an HMC with finite state space E. Suppose that the set of recurrent states R and the set of transient states T are both non-empty. In the block decomposition of $\mathbf{P}$ with respect to the partition $R \cup T = E$,

$$\mathbf{P} = \begin{pmatrix} D & 0 \\ B & \mathbf{Q} \end{pmatrix}$$

the matrix $\mathbf{Q}$ is sub-stochastic, since B is not identically null (otherwise, the transient set would be closed, and therefore recurrent, being finite). We assume, in addition, that $\mathbf{Q}$ is irreducible and aperiodic. Let $\nu = \inf\{n \geq 0; X_n \in R\}$ be the entrance time into R. Recall that ν is almost surely finite, since T is a finite set. What is the distribution of X_n for large n, conditioned by the fact that X_n is still in T?

Theorem 6.3.5 *(Bartlett, 1957)*

$$\lim_{n\uparrow\infty} P_i(X_n = j \mid \nu > n) = \frac{u_1(j)}{\sum_{k\in T} u_1(k)}\,.$$

Proof. First recall that

$$\mathbf{Q}^n = \lambda_1^n v_1 u_1^T + O(n^{m_2-1}|\lambda_2|^n), \tag{6.15}$$

where λ_1, v_1, u_1, m_2, and λ_2 are as in the above statement of Perron–Fröbenius theorem, with $A = \mathbf{Q}$. In particular, $\lambda_1 \in (0,1)$ and $|\lambda_2| < \lambda_1$. For $i, j \in T$,

$$P_i(X_n = j \mid \nu > n) = \frac{P_i(X_n = j, \nu > n)}{P_i(\nu > n)} = \frac{P_i(X_n = j)}{P_i(X_n \in T)}.$$

Therefore,

$$P_i(X_n = j \mid \nu > n) = \frac{p_{ij}(n)}{\sum_{k\in T} p_{ik}(n)}.$$

In view of (6.15),

$$p_{ik}(n) = \lambda_1^n v_1(i) u_1(k) + O(n^{m_2-1}|\lambda_2|^n).$$

Therefore,

$$P_i(X_n = j \mid \nu > n) = \frac{u_1(j)}{\sum_{k\in T} u_1(k)} + O\left(n^{m_2-1}\left|\frac{\lambda_2}{\lambda_1}\right|^n\right), \tag{6.16}$$

and in particular,

$$\lim_{n\uparrow\infty} P_i(X_n = j \mid \nu > n) = \frac{u_1(j)}{\sum_{k\in T} u_1(k)}. \tag{6.17}$$

□

The probability distribution $\{u_1(i)/\sum_{k\in T} u_1(k)\}_{i\in T}$ is called the quasi-stationary distribution of the chain relative to T.

6.3.3 Spectral Densities

Let $\{Y_n\}_{n\in\mathbb{Z}}$ be a sequence of square-integrable real random variables such that the quantities $E[Y_n]$ and $E[Y_nY_{n+k}]$ $(k \in \mathbb{Z})$ are independent of n. Such sequences are called *wide-sense stationary*. Let then

$$m_Y := E[Y_n] \text{ and } R_Y(k) := E[Y_nY_{n+k}] .$$

The function R_Y is called the *covariance function* of the above stochastic sequence. This function plays a fundamental role in signal processing. We shall compute it for sequences that are functions of stationary HMC's.

More precisely, let $\{X_n\}_{n\in\mathbb{Z}}$ be an irreducible stationary discrete-time HMC with finite state space $E = \{1, 2, \ldots, r\}$, transition matrix $\mathbf{P}$ and (unique) stationary distribution π. Then, for any given function $f : E \to \mathbb{R}$, the stochastic process

$$Y_n := f(X_n) \tag{6.18}$$

is wide-sense stationary (in fact, stationary). Its mean is

$$m_Y = \pi^T f$$

where $f^T = (f(1), f(2), \ldots, f(r))$. Also, as simple calculations reveal,

$$E[Y_{k+n}Y_n] = f^T D_\pi \mathbf{P}^k f ,$$

where D_π is the diagonal matrix whose diagonal is π. Note that one may express the mean as $m_Y = f^T\pi = f^T D_\pi \mathbf{1}$ where $\mathbf{1}$ is a column vector with all entries equal to 1. In particular

$$m_Y^2 = f^T D_\pi \mathbf{1}\mathbf{1}^T D_\pi f = f^T D_\pi \Pi f$$

where $\Pi := \pi\mathbf{1}^T$ is a square matrix with all lines identical to the stationary distribution vector π. Therefore, the covariance function of $\{Y_n\}_{n\in\mathbb{Z}}$ is

$$R_Y(k) = f^T D_\pi (\mathbf{P}^k - \Pi) f.$$

Let $\{Y_n\}_{n\in\mathbb{Z}}$ be a wide-sense stationary sequence with covariance function R satisfying the absolute summability condition

$$\sum_{n\in\mathbb{Z}} |R_Y(n)| < \infty.$$

In particular, the Fourier sum

$$f_Y(\omega) = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} R_Y(n) e^{-i\omega n}$$

is a 2π-periodic bounded and continuous function. Integrating the right-hand side of the last display term by term (this is allowed because the covariance function is absolutely summable), we obtain the inversion formula,

$$R_Y(n) = \int_{-\pi}^{+\pi} f_Y(\omega) e^{+i\omega n} \mathrm{d}\omega.$$

The function $f : [-\pi, +\pi] \to \mathbb{R}$ is the *power spectral density* (PSD) of the time series. Note that

$$\int_{-\pi}^{+\pi} f_Y(\omega) d\omega = R_Y(0) < \infty\,.$$

In the example of interest, where the WSS sequence is defined by (6.18),

$$R_Y(k) = f^T D_\pi (\mathbf{P}^k - \Pi) f.$$

In order to simplify the notation, we shall suppose that $\mathbf{P}$ is diagonalizable. We then have

$$(\mathbf{P}^k - \Pi) = \sum_{j=2}^{r} v_j u_j^T \lambda_j^k$$

where u_j is the (up to a multiplicative constant) left-eigenvector and v_j is the (up to a multiplicative constant) right-eigenvector corresponding to the eigenvalue λ_j, and where the multiplicative constants are chosen in such a way that $u_j^T v_j = 1$. Therefore

$$R_Y(k) = f^T D_\pi \left(\sum_{j=2}^{r} v_j u_j^T \lambda_j^k \right) f. \qquad (\star)$$

We know that all the eigenvalues are of modulus smaller than or equal to 1. If we suppose, as we now do, that the chain is aperiodic, then, besides the eigenvalue $\lambda_1 = 1$, all eigenvalues have a modulus strictly less than 1. In particular, the covariance function is absolutely summable and there exists a power spectral density which is easily computed using the fact that, for $|\lambda| < 1$,

$$\sum_{k\in\mathbb{Z}} \lambda^k e^{-ik\omega} = \frac{1}{|e^{-i\omega} - \lambda|^2},$$

from which it follows, by (6.11), that

$$f_Y(\omega) = \frac{1}{2\pi} \sum_{j=2}^{r} (f^T D_\pi \, v_j u_j^T f) \frac{1}{|e^{-i\omega} - \lambda_j|^2}.$$

The case occurs when the covariance function of a wide-sense stationary sequence $\{Y_n\}_{n\in\mathbb{Z}}$ is not summable, and when the power spectral density is a pseudo-density, of the symbolic form (using the Dirac pseudo-function δ)

$$f_Y(\omega) = \sum_{\ell\in\mathbb{Z}} \alpha_\ell \delta(\omega - \omega_\ell)\,,$$

where for all $\ell \in \mathbb{Z}$: $\omega_\ell \in (-\pi, +\pi]$, $\alpha_\ell \in \mathbb{R}_+$, and

$$\sum_{\ell\in\mathbb{Z}} \alpha_\ell < \infty\,.$$

Continuing the example where the wide-sense stationary sequence is defined by (6.18), suppose that the chain has a period equal to $d > 1$. In this case, there are $d-1$ eigenvalues besides $\lambda_1 = 1$ with a modulus equal to 1, and these are precisely the d-th roots of unity besides $\lambda_1 = 1$:

$$\lambda_\ell = e^{+i\omega_\ell}, \quad (2 \le \ell \le d)$$

where $\omega_\ell = (\ell - 1)\frac{2\pi}{d}$. Observing that

$$e^{+ik\omega_\ell} = \int_{(-\pi,+\pi]} e^{+ik\omega} \delta(\omega - \omega_\ell),$$

we find for the complete spectral density, in the case where the eigenvalues are distinct, or more generally, the transition matrix is diagonalizable,

$$\begin{aligned} f_Y(\omega) = &\sum_{\ell=2}^{d} (f^T D_\pi \, v_\ell u_\ell^T f)\, \delta(\omega - \omega_\ell) \\ &+ \frac{1}{2\pi} \sum_{j=d+1}^{r} (f^T D_\pi \, v_j u_j^T f) \frac{1}{|e^{-i\omega} - \lambda_j|^2}\,. \end{aligned}$$

Books for Further information

[Kemeny and Snell, 1960] (finite Markov chains) and [Kemeny and Snell, 1960] (denumerable Markov chains) are elementary introductions to Markov chains. [Karlin and Taylor, 1975] has many examples, most notably in biology. For the Perron–Fröbenius theorem and other algebraic aspects of Markov chains, [Seneta, 1981] is the fundamental reference. Continuous-time Markov chains are treated in [Brémaud, 1999].

6.4 Exercises

Exercise 6.4.1. A COUNTEREXAMPLE
The Markov property does not imply that the past and the future are independent given any information concerning the present. Find a simple example of an HMC $\{X_n\}_{n\geq 0}$ with state space $E = \{1, 2, 3, 4, 5, 6\}$ such that

$$P(X_2 = 6 \,|\, X_1 \in \{3, 4\}, X_0 = 2) \neq P(X_2 = 6 \,|\, X_1 \in \{3, 4\}).$$

Exercise 6.4.2. PAST, PRESENT, FUTURE
For an HMC $\{X_n\}_{n\geq 0}$ with state space E, prove that for all $n \in \mathbb{N}$, and all states $i_0, i_1, \ldots, i_{n-1}, i, j_1, j_2, \ldots, j_k \in E$,

$$\begin{aligned} &P(X_{n+1} = j_1, \ldots, X_{n+k} = j_k \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ &\quad = P(X_{n+1} = j_1, \ldots, X_{n+k} = j_k \mid X_n = i). \end{aligned}$$

Exercise 6.4.3. ANOTHER CONDITIONAL INDEPENDENCE PROPERTY OF HMC'S
Let $\{X_n\}_{n\geq 0}$ be an HMC with state space E and transition matrix $\mathbf{P}$. Show that for all $n \geq 1$, all $k \geq 2$, X_n is conditionally independent of $X_0, \ldots, X_{n-2}, X_{n+2}, \ldots, X_{n+k}$ given X_{n-1}, X_{n+1} and compute the conditional distribution of X_n given X_{n-1}, X_{n+1}.

Exercise 6.4.4. ON THE CUBE
Consider the HMC $\{X_n\}_{n\geq 0}$ with state space $E := \{0, 1\}^N$ where N is some positive integer with a representation as in Theorem 6.1.4, with the following specifics:

$$X_{n+1} = X_n \oplus e_{Z_{n+1}}$$

where $\oplus$ is addition modulo 2 on $\{0, 1\}^N$, e_ℓ is the vector of $\{0, 1\}^N$ with all coordinates null except the ℓth one, equal to 1, and $\{Z_n\}_{n\geq 1}$ is an IID sequence of random variables uniformly distributed on $\{1, 2, \ldots, N\}$. Let now $Y_n = h(X_n)$ where for any $a = (a_1, \ldots, a_N)$, $h(a) := \sum_{k=1}^N a_i$.

Prove that $\{Y_n\}_{n\geq 0}$ is an HMC and identify it.

Exercise 6.4.5. RECORDS
Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of geometric random variables: For $k \geq 0$, $P(Z_n = k) = (1 - p)^k p$, where $p \in (0, 1)$. Let $X_n = \max(Z_1, \ldots, Z_n)$ be the record value at time n, and suppose X_0 is an $\mathbb{N}$-valued random variable independent of the sequence $\{Z_n\}_{n\geq 1}$. Show that $\{X_n\}_{n\geq 0}$ is an HMC and give its transition matrix.

Exercise 6.4.6. STREETGANGS
Three characters, A, B, and C, armed with guns, suddenly meet at the corner of a Washington D.C. street, whereupon they naturally start shooting at one another. Each street-gang kid shoots every tenth second, as long as he is still alive. The probability of a successful hit for A, B, and C are α, β, and γ respectively. A is the most hated, and therefore, as long as he is alive, B and C ignore each other and shoot at A. For historical reasons not developed here, A cannot stand B,

and therefore shoots only at B while the latter is still alive. Lucky C is shot at if and only if he is in the presence of A alone or B alone. What are the survival probabilities of A, B, and C, respectively?

Exercise 6.4.7. The gambler's ruin
(This exercise continues Example 6.1.3.) Compute the average duration of the game when $p = \frac{1}{2}$.

Exercise 6.4.8. Alternative proof of the strong Markov property
Give an alternative proof of the strong Markov property (Theorem 7.1.3) along the following lines. Start with a representation $X_{n+1} = f(X_n, Z_{n+1})$ as in Theorem 6.1.4 and consider the sequence $\{Z_{\tau+n}\}_{n\geq 1}$ defined when $\tau < \infty$.

Exercise 6.4.9. Truncated hmc
Let $\mathbf{P}$ be a transition matrix on the countable state space E, with the positive stationary distribution π. Let A be a subset of the state space, and define the truncation of $\mathbf{P}$ on A to be the transition matrix $\mathbf{Q}$ indexed by A and given by

$$\begin{aligned} q_{ij} &= p_{ij} \text{ if } i, j \in A,\ i \neq j, \\ q_{ii} &= p_{ii} + \sum_{k\in\bar{A}} p_{ik}. \end{aligned}$$

Show that if $(\mathbf{P}, \pi)$ is reversible, then so is $(\mathbf{Q}, \frac{\pi}{\pi(A)})$.

Exercise 6.4.10. Extension to negative times
Let $\{X_n\}_{n\geq 0}$ be an hmc with state space E, transition matrix $\mathbf{P}$, and suppose that there exists a stationary distribution $\pi > 0$. Suppose moreover that the initial distribution is π. Define the matrix $\mathbf{Q} = \{q_{ij}\}_{i,j\in E}$ by (6.7). Construct $\{X_{-n}\}_{n\geq 1}$, independent of $\{X_n\}_{n\geq 1}$ given X_0, as follows:

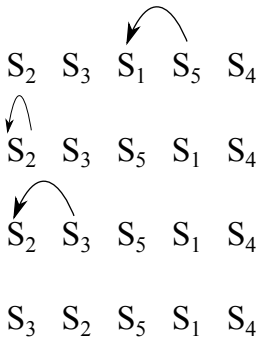
$$\begin{aligned} &P(X_{-1} = i_1, X_{-2} = i_2, \ldots, X_{-k} = i_k \,|\, X_0 = i, X_1 = j_1, \ldots, X_n = j_n) \\ &\quad = P(X_{-1} = i_1, X_{-2} = i_2, \ldots, X_{-k} = i_k \,|\, X_0 = i) = q_{ii_1} q_{i_1 i_2} \cdots q_{i_{k-1} i_k} \end{aligned}$$

for all $k \geq 1, n \geq 1, i, i_1, \ldots, i_k, j_1, \ldots, j_n \in E$. Prove that $\{X_n\}_{n\in\mathbb{Z}}$ is an hmc with transition matrix $\mathbf{P}$ and $P(X_n = i) = \pi(i)$, for all $i \in E$, all $n \in \mathbb{Z}$.

Exercise 6.4.11. Moving stones
Stones $S_1, \ldots, S_M$ are placed in line. At each time n a stone is selected at random, and this stone and the one ahead of it in the line exchange positions. If the selected stone is at the head of the line, nothing is changed. For instance, with $M = 5$: Let the current configuration be $S_2 S_3 S_1 S_5 S_4$ (S_2 is at the head of the line). If S_5 is selected, the new situation is $S_2 S_3 S_5 S_1 S_4$, whereas if S_2 is selected, the configuration is not altered. At each step, stone S_i is selected with probability $\alpha_i > 0$. Call X_n the situation at time n, for instance $X_n = S_{i_1} \cdots S_{i_M}$, meaning that stone S_{i_j} is in the jth position. Show that $\{X_n\}_{n\geq 0}$ is an irreducible hmc and that it has a stationary distribution given by the formula

$$\pi(S_{i_1} \cdots S_{i_M}) = C \alpha_{i_1}^M \alpha_{i_2}^{M-1} \cdots \alpha_{i_M},$$

for some normalizing constant C.



Exercise 6.4.12. No stationary distribution
Show that the symmetric random walk on $\mathbb{Z}$ cannot have a stationary distribution.

Exercise 6.4.13. Aperiodicity
a. Show that an irreducible transition matrix $\mathbf{P}$ with at least one state $i \in E$ such that $p_{ii} > 0$ is aperiodic.

b. Let $\mathbf{P}$ be an irreducible transition matrix on the *finite* state space E. Show that a necessary and sufficient condition for $\mathbf{P}$ to be aperiodic is the existence of an integer m such that $\mathbf{P}^m$ has all its entries positive.

c. Consider an HMC that is irreducible with period $d \geq 2$. Show that the restriction of the transition matrix to any cyclic class is irreducible. Show that the restriction of $\mathbf{P}^d$ to any cyclic class is aperiodic.

Exercise 6.4.14. A coding scheme
In certain digital communication systems, a sequence of 0s and 1s (bits) is encoded into a sequence of 0s, +1s, and −1s as follows. If the input sequence contains a 0, the output sequence contains a 0 at the same place. If the input sequence contains a 1, then the output sequence will have a −1 or a +1. The choice between −1 and +1 is made in such a way that −1s and +1s must alternate in the output sequence. The first 1 is encoded as +1. For instance, 011101 becomes $0, +1, -1, +1, 0, -1$. Find an automaton with four states +1, −1, 0_+, and 0_- for which the sequence of visited states, not counting the initial state 0_+, is exactly the encoded sequence (where 0_+ and 0_- are rewritten as 0) when it is fed by the input sequence.

Suppose that the input sequence is IID, with 0 and 1 equiprobable. The sequence of states visited by the automaton is then an HMC. Compute its transition matrix $\mathbf{P}$, its stationary distribution π, and its iterates $\mathbf{P}^n$. Call $\{Y_n\}_{n\geq 0}$ the output sequence (taking its values in $\{0, -1, +1\}$). Compute $\lim_{n\to\infty}\{E[Y_nY_{n+k}] - E[Y_n]E[Y_{n+k}]\}$ for all $k \geq 0$.

Chapter 7

Recurrence of Markov Chains

7.1 Recurrent and Transient States

7.1.1 The Strong Markov Property

In a homogeneous Markov chain, some states are visited infinitely often while others are never visited after a finite random time. These states are naturally called recurrent and transient respectively. Among the recurrent states some have the property that the mean time between successive visits to this state is finite. These are the positive recurrent states, whereas the others are called null recurrent. It turns out that for an irreducible Markov chain all states are of the same nature, transient, positive recurrent or null. We first give methods to determine if an irreducible chain is recurrent. For this we need further results concerning the distribution of a Markov chain, in particular, the strong Markov property.

The Markov property, that is, the independence of past and future given the present state, extends to the situation where the present time is a stopping time, a notion which we now introduce.

Let $\{X_n\}_{n\geq 0}$ be a stochastic process with values in the denumerable set E. For an event A, the notation $A \in \mathcal{X}_0^n$ means that there exists a function $\varphi : E^{n+1} \mapsto \{0,1\}$ such that

$$1_A(\omega) = \varphi(X_0(\omega), \ldots, X_n(\omega))\,.$$

In other terms, this event is expressible in terms of $X_0(\omega), \ldots, X_n(\omega)$. Let now τ be a random variable with values in $\overline{\mathbb{N}}$. It is called a X_0^n-stopping time if for all $m \in \mathbb{N}$, $\{\tau = m\} \in X_0^m$. In other words, it is a non-anticipative random time with respect to $\{X_n\}_{n\geq 0}$, since in order to check if $\tau = m$, it suffices to observe the process up to time m and not beyond. It is immediate to verify that if τ is a X_0^n-stopping time, then so is $\tau + n$ for all $n \geq 1$.

Example 7.1.1: Return time. Let $\{X_n\}_{n\geq 0}$ be an hmc with state space E. Define for $i \in E$ the return time to i by

$$T_i := \inf\{n \geq 1\,;\, X_n = i\}$$

using the convention $\inf \varnothing = \infty$ for the empty set of $\mathbb{N}$. This is a X_0^n-stopping time since for all $m \in \mathbb{N}$,

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_7

$$\{T_i = m\} = \{X_1 \neq i, X_2 \neq i, \ldots, X_{m-1} \neq i, X_m = i\}\,.$$

Note that $T_i \geq 1$. It is a "return" time, not to be confused with the closely related "hitting" time of i, defined as $S_i := \inf\{n \geq 0\,;\, X_n = i\}$, which is also a X_0^n-stopping time, equal to T_i if and only if $X_0 \neq i$.

EXAMPLE 7.1.2: SUCCESSIVE RETURN TIMES. This continues the previous example. Let us fix a state, conventionally labeled 0, and let T_0 be the return time to 0. We define the successive return times to 0, τ_k ($k \geq 1$) by $\tau_1 = T_0$ and for $k \geq 1$,

$$\tau_{k+1} := \inf\{n \geq \tau_k + 1\,;\, X_n = 0\}$$

with the above convention that $\inf \varnothing = \infty$. In particular, if $\tau_k = \infty$ for some k, then $\tau_{k+\ell} = \infty$ for all $\ell \geq 1$. The identity

$$\{\tau_k = m\} \equiv \left\{ \sum_{n=1}^{m-1} 1_{\{X_n = 0\}} = k - 1\,,\; X_m = 0 \right\}$$

for $m \geq 1$ shows that τ_k is a X_0^n-stopping time.

Let $\{X_n\}_{n\geq 0}$ be a stochastic process with values in the countable set E and let τ be a random time taking its values in $\overline{\mathbb{N}} := \mathbb{N} \cup \{+\infty\}$. In order to define X_τ when $\tau = \infty$, one must decide how to define X_∞. This is done by taking some arbitrary element Δ not in E, and setting

$$X_\infty = \Delta.$$

By definition, the "process after τ" is the stochastic process

$$\{S_\tau X_n\}_{n\geq 0} := \{X_{n+\tau}\}_{n\geq 0}.$$

The "process before τ," or the "process stopped at τ," is the process

$$\{X_n^\tau\}_{n\geq 0} := \{X_{n\wedge\tau}\}_{n\geq 0},$$

which freezes at time τ at the value X_τ.

Theorem 7.1.3 *Let $\{X_n\}_{n\geq 0}$ be an* HMC *with state space E and transition matrix $\mathbf{P}$. Let τ be a X_0^n-stopping time. Then for any state $i \in E$,*

(α) *Given that $X_\tau = i$, the process after τ and the process before τ are independent.*

(β) *Given that $X_\tau = i$, the process after τ is an* HMC *with transition matrix $\mathbf{P}$.*

Proof. (α) We have to show that for all times $k \geq 1, n \geq 0$, and all states $i_0, \ldots, i_n, i, j_1, \ldots, j_k$,

$$P(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \,|\, X_\tau = i, X_{\tau\wedge 0} = i_0, \ldots, X_{\tau\wedge n} = i_n) \\ = P(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \,|\, X_\tau = i).$$

We shall prove a simplified version of the above equality, namely

$$P(X_{\tau+k} = j \,|\, X_\tau = i, X_{\tau\wedge n} = i_n) = P(X_{\tau+k} = j \,|\, X_\tau = i)\,. \qquad (\star)$$

The general case is obtained by the same arguments. The left-hand side of $(\star)$ equals

$$\frac{P(X_{\tau+k} = j, X_\tau = i, X_{\tau\wedge n} = i_n)}{P(X_\tau = i, X_{\tau\wedge n} = i_n)}.$$

The numerator of the above expression can be developed as

$$\sum_{r\in\mathbb{N}} P(\tau = r, X_{r+k} = j, X_r = i, X_{r\wedge n} = i_n)\,. \qquad (\star\star)$$

(The sum is over $\mathbb{N}$ because $X_\tau = i \neq \Delta$ implies that $\tau < \infty$.) But $P(\tau = r, X_{r+k} = j, X_r = i, X_{r\wedge n} = i_n) = P(X_{r+k} = j \,|\, X_r = i,\ X_{r\wedge n} = i_n, \tau = r)$ $P(\tau = r, X_{r\wedge n} = i_n, X_r = i)$, and since $r \wedge n \leq r$ and $\{\tau = r\} \in X_0^r$, the event $B := \{X_{r\wedge n} = i_n, \tau = r\}$ is in X_0^r. Therefore, by the Markov property, $P(X_{r+k} = j \,|\, X_r = i, X_{r\wedge n} = i_n, \tau = r\} = P(X_{r+k} = j \,|\, X_r = i) = p_{ij}(k)$. Finally, expression $(\star\star)$ reduces to

$$\sum_{r\in\mathbb{N}} p_{ij}(k) P(\tau = r, X_{r\wedge n} = i_n, X_r = i) = p_{ij}(k) P(X_{\tau=i}, X_{\tau\wedge n} = i_n).$$

Therefore, the left-hand side of $(\star)$ is just $p_{ij}(k)$. Similar computations show that the right-hand side of $(\star)$ is also $p_{ij}(k)$, so that (α) is proved.

(β) We must show that for all states $i, j, k, i_{n-1}, \ldots, i_1$,

$$P(X_{\tau+n+1} = k \,|\, X_{\tau+n} = j, X_{\tau+n-1} = i_{n-1}, \ldots, X_\tau = i) \\ = P(X_{\tau+n+1} = k \,|\, X_{\tau+n} = j) = p_{jk}.$$

But the first equality follows from the fact proved in (α) that for the stopping time $\tau' = \tau + n$, the processes before and after τ' are independent given $X_{\tau'} = j$. The second equality is obtained by the same calculations as in the proof of (α). □

For an alternative and perhaps more illuminating proof, see Exercise 6.4.8.

Cycle Independence

Consider a Markov chain with a state conventionally denoted by 0 such that $P_0(T_0 < \infty) = 1$. In view of the strong Markov property, the chain starting from state 0 will return infinitely often to this state. Let $\tau_1 = T_0, \tau_2, \ldots$ be the successive return times to 0, and set $\tau_0 \equiv 0$.

By the strong Markov property, for any $k \geq 1$, the process after τ_k is independent of the process before τ_k (observe that condition $X_{\tau_k} = 0$ is always satisfied), and the process after τ_k is a Markov chain with the same transition matrix as the

original chain, and with initial state 0, by construction. Therefore, the successive times of visit to 0, the pieces of trajectory

$$\{X_{\tau_k}, X_{\tau_k+1}, \ldots, X_{\tau_{k+1}-1}\}, k \geq 0,$$

are independent and identically distributed. Such pieces are called the regenerative cycles of the chain between visits to state 0. Each random time τ_k is a regeneration time, in the sense that $\{X_{\tau_k+n}\}_{n\geq 0}$ is independent of the past $X_0, \ldots, X_{\tau_k-1}$ and has the same distribution as $\{X_n\}_{n\geq 0}$. In particular, the sequence $\{\tau_k - \tau_{k-1}\}_{k\geq 1}$ is IID.

7.1.2 The Potential Matrix Criterion of Recurrence

Consider an HMC $\{X_n\}_{n\geq 0}$ with state space E and transition matrix $\mathbf{P}$. A state $i \in E$ is called recurrent if it is visited infinitely often.

The distribution given $X_0 = j$ of $N_i = \sum_{n\geq 1} 1_{\{X_n=i\}}$, the number of visits to state i strictly after time 0, is

$$\begin{aligned} P_j(N_i = r) &= f_{ji} f_{ii}^{r-1}(1 - f_{ii}) \, (r \geq 1) \\ P_j(N_i = 0) &= 1 - f_{ji}, \end{aligned}$$

where $f_{ji} = P_j(T_i < \infty)$ and T_i is the return time to i.

Proof. An informal proof goes like this: We first go from j to i (probability f_{ji}) and then, $r-1$ times in succession, from i to i (each time with probability f_{ii}), and the last time, that is the $r+1$-st time, we leave i never to return to it (probability $1 - f_{ii}$). By the cycle independence property, all these "cycles" are independent, so that the successive probabilities multiplicate. Here is a formal proof if someone needs it.

For $r = 0$, this is just the definition of f_{ji}. Now let $r \geq 1$, and suppose that $P_j(N_i = k) = f_{ji} f_{ii}^{k-1}(1 - f_{ii})$ is true for all k, $1 \leq k \leq r$. In particular,

$$P_j(N_i > r) = f_{ji} f_{ii}^r.$$

Denoting by τ_r the rth return time to state i,

$$\begin{aligned} P_j(N_i = r+1) &= P_j(N_i = r+1, X_{\tau_{r+1}} = i) \\ &= P_j(\tau_{r+2} - \tau_{r+1} = \infty, X_{\tau_{r+1}} = i) \\ &= P_j(\tau_{r+2} - \tau_{r+1} = \infty \,|\, X_{\tau_{r+1}} = i) P_j(X_{\tau_{r+1}} = i). \end{aligned}$$

By the strong Markov property, observing that $\tau_{r+2} - \tau_{r+1}$ is the return time to i of the process after τ_{r+1},

$$P_j(\tau_{r+2} - \tau_{r+1} = \infty \,|\, X_{\tau_{r+1}} = i) = 1 - f_{ii}\,.$$

Also, $P_j(X_{\tau_{r+1}} = i) = P_j(N_i > r)$, and therefore,

$$P_j(N_i = r+1) = P_i(T_i = \infty) P_j(N_i > r) = (1 - f_{ii}) f_{ji} f_{ii}^r.$$

The result then follows by induction. □

The distribution of N_i given $X_0 = j$ and given $N_i \geq 1$ is geometric. This has two main consequences. Firstly, $P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$. In words: if starting from i the chain almost surely returns to i, and will then visit i infinitely often. Secondly,

$$E_i[N_i] = \sum_{r=1}^{\infty} r P_i(N_i = r) = \sum_{r-1}^{\infty} r f_{ii}^r (1 - f_{ii}) = \frac{f_{ii}}{1 - f_{ii}}.$$

In particular, $P_i(T_i < \infty) < 1 \iff E_i[N_i] < \infty$.

We collect these results for future reference. For any state $i \in E$,

$$P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$$

and

$$P_i(T_i < \infty) < 1 \iff P_i(N_i = \infty) = 0 \iff E_i[N_i] < \infty. \tag{7.1}$$

In particular, the event $\{N_i = \infty\}$ has P_i-probability 0 or 1.

Define the potential matrix of the transition matrix $\mathbf{P}$ to be

$$\mathbf{G} := \sum_{n \geq 0} \mathbf{P}^n.$$

Its general term

$$g_{ij} = \sum_{n=0}^{\infty} p_{ij}(n) = \sum_{n=0}^{\infty} P_i(X_n = j) = \sum_{n=0}^{\infty} E_i[1_{\{X_n = j\}}] = E_i \left[\sum_{n=0}^{\infty} 1_{\{X_n = j\}} \right]$$

is the average number of visits to state j, given that the chain starts from state i.

Recall that T_i denotes the return time to state i.

Definition 7.1.4 *The state $i \in E$ is called recurrent if*

$$P_i(T_i < \infty) = 1,$$

and otherwise it is called transient. A recurrent state $i \in E$ such that

$$E_i[T_i] < \infty$$

is called positive recurrent, and otherwise it is called null recurrent.

Although the criterion of recurrence below is of theoretical rather than practical interest, it can be helpful in a few situations, for instance in the study of recurrence of random walks, as examples will show.

Theorem 7.1.5 *The state $i \in E$ is recurrent if and only if*

$$\sum_{n=0}^{\infty} p_{ii}(n) = \infty.$$

Proof. This merely rephrases (7.1). □

EXAMPLE 7.1.6: 1-D RANDOM WALK, TAKE 2. The state space of this Markov chain is $E := \mathbb{Z}$ and the non-null terms of its transition matrix are $p_{i,i+1} = p$, $p_{i,i-1} = 1 - p$, where $p \in (0, 1)$. Since this chain is irreducible, it suffices to elucidate the nature (recurrent or transient) of any one of its states, say, 0. We have $p_{00}(2n + 1) = 0$ and

$$p_{00}(2n) = \frac{(2n)!}{n!n!} p^n (1 - p)^n.$$

By Stirling's equivalence formula $n! \sim (n/e)^n \sqrt{2\pi n}$, the above quantity is equivalent to

$$\frac{[4p(1 - p)]^n}{\sqrt{\pi n}} \qquad (\star)$$

and the nature of the series $\sum_{n=0}^{\infty} p_{00}(n)$ (convergent or divergent) is that of the series with general term $(\star)$. If $p \neq \frac{1}{2}$, in which case $4p(1-p) < 1$, the latter series converges, and if $p = \frac{1}{2}$, in which case $4p(1 - p) = 1$, it diverges. In summary, the states of the 1-D random walk are transient if $p \neq \frac{1}{2}$, recurrent if $p = \frac{1}{2}$.

A theoretical application of the potential matrix criterion is to the proof that recurrence is a (communication) class property.

Theorem 7.1.7 *If i and j communicate, they are either both recurrent or both transient.*

Proof. By definition, i and j communicate if and only if there exist integers M and N such that $p_{ij}(M) > 0$ and $p_{ji}(N) > 0$. Going from i to j in M steps, then from j to j in n steps, then from j to i in N steps, is just one way of going from i back to i in $M + n + N$ steps. Therefore, $p_{ii}(M + n + N) \geq p_{ij}(M) \times p_{jj}(n) \times p_{ji}(N)$. Similarly, $p_{jj}(N + n + M) \geq p_{ji}(N) \times p_{ii}(n) \times p_{ij}(M)$. Therefore, with $\alpha := p_{ij}(M)\, p_{ji}(N)$ (a strictly positive quantity), we have $p_{ii}(M + N + n) \geq \alpha\, p_{jj}(n)$ and $p_{jj}(M + N + n) \geq \alpha\, p_{ii}(n)$. This implies that the series $\sum_{n=0}^{\infty} p_{ii}(n)$ and $\sum_{n=0}^{\infty} p_{jj}(n)$ either both converge or both diverge. The potential matrix criterion concludes the proof. □

7.2 Positive Recurrence

7.2.1 The Stationary Distribution Criterion

We first give a necessary (yet not sufficient) condition of recurrence based on the notion of invariant measure, which extends that of a stationary distribution. It is the first step towards the stationary distribution criterion (necessary and sufficient condition) of positive recurrence.

Definition 7.2.1 *A non-trivial (that is, non-null) vector x (indexed by E) of non-negative real numbers (notation: $0 \leq x < \infty$) is called an invariant measure of the stochastic matrix $\mathbf{P}$ (indexed by E) if*

$$x^T = x^T \mathbf{P}\,. \tag{7.2}$$

Theorem 7.2.2 *Let $\mathbf{P}$ be the transition matrix of an irreducible recurrent* HMC *$\{X_n\}_{n\geq 0}$. Let 0 be an arbitrary state and let T_0 be the return time to 0. Define for all $i \in E$*

$$x_i = E_0\left[\sum_{n=1}^{T_0} 1_{\{X_n=i\}}\right]\,. \tag{7.3}$$

(For $i \neq 0$, x_i is the expected number of visits to state i before returning to 0.) Then, $0 < x < \infty$ and x is an invariant measure of $\mathbf{P}$.

Proof. We make three preliminary observations. First, it will be convenient to rewrite (7.3) as

$$x_i = E_0\left[\sum_{n\geq 1} 1_{\{X_n=i\}} 1_{\{n\leq T_0\}}\right]\,.$$

Next, when $1 \leq n \leq T_0$, $X_n = 0$ if and only if $n = T_0$. Therefore,

$$x_0 = 1.$$

Also, $\sum_{i\in E}\sum_{n\geq 1} 1_{\{X_n=i\}}1_{\{n\leq T_0\}} = \sum_{n\geq 1}\left(\sum_{i\in E} 1_{\{X_n=i\}}\right)1_{\{n\leq T_0\}} = \sum_{n\geq 1} 1_{\{n\leq T_0\}} = T_0$, and therefore

$$\sum_{i\in E} x_i = E_0[T_0]. \tag{7.4}$$

We introduce the quantity

$${}_0p_{0i}(n) := E_0[1_{\{X_n=i\}}1_{\{n\leq T_0\}}] = P_0(X_1 \neq 0, \cdots, X_{n-1} \neq 0, X_n = i).$$

This is the probability, starting from state 0, of visiting i at time n before returning to 0. From the definition of x,

$$x_i = \sum_{n\geq 1} {}_0p_{0i}(n)\,. \tag{$\dagger$}$$

We first prove (7.2). Observe that ${}_0p_{0i}(1) = p_{0i}$, and, by first-step analysis, for all $n \geq 2$, ${}_0p_{0i}(n) = \sum_{j\neq 0} {}_0p_{0j}(n-1)p_{ji}$. Summing up all the above equalities, and taking ($\dagger$) into account, we obtain

$$x_i = p_{0i} + \sum_{j\neq 0} x_j p_{ji},$$

that is, (7.2), since $x_0 = 1$.

Next we show that $x_i > 0$ for all $i \in E$. Indeed, iterating (7.2), we find $x^T = x^T \mathbf{P}^n$, that is, since $x_0 = 1$,

$$x_i = \sum_{j \in E} x_j p_{ji}(n) = p_{0i}(n) + \sum_{j \neq 0} x_j p_{ji}(n).$$

If x_i were null for some $i \in E$, $i \neq 0$, the latter equality would imply that $p_{0i}(n) = 0$ for all $n \geq 0$, which means that 0 and i do not communicate, in contradiction to the irreducibility assumption.

It remains to show that $x_i < \infty$ for all $i \in E$. As before, we find that

$$1 = x_0 = \sum_{j \in E} x_j p_{j0}(n)$$

for all $n \geq 1$, and therefore if $x_i = \infty$ for some i, necessarily $p_{i0}(n) = 0$ for all $n \geq 1$, and this also contradicts irreducibility. □

Theorem 7.2.3 *The invariant measure of an irreducible recurrent* HMC *is unique up to a multiplicative factor.*

Proof. In the proof of Theorem 7.2.2, we showed that for an invariant measure y of an irreducible chain, $y_i > 0$ for all $i \in E$, and therefore, one can define, for all $i, j \in E$, the matrix $\mathbf{Q}$ by

$$q_{ji} = \frac{y_i}{y_j} p_{ij}\,. \qquad (\star)$$

It is a transition matrix, since $\sum_{i \in E} q_{ji} = \frac{1}{y_j} \sum_{i \in E} y_i p_{ij} = \frac{y_j}{y_j} = 1$. The general term of $\mathbf{Q}^n$ is

$$q_{ji}(n) = \frac{y_i}{y_j} p_{ij}(n)\,. \qquad (\star\star)$$

Indeed, supposing $(\star\star)$ true for n,

$$\begin{aligned} q_{ji}(n+1) &= \sum_{k \in E} q_{jk} q_{ki}(n) = \sum_{k \in E} \frac{y_k}{y_j} p_{kj} \frac{y_i}{y_k} p_{ik}(n) \\ &= \frac{y_i}{y_j} \sum_{k \in E} p_{ik}(n) p_{kj} = \frac{y_i}{y_j} p_{ij}(n+1), \end{aligned}$$

and $(\star\star)$ follows by induction.

Clearly, $\mathbf{Q}$ is irreducible, since $\mathbf{P}$ is irreducible (just observe that $q_{ji}(n) > 0$ if and only if $p_{ij}(n) > 0$ in view of $(\star\star)$). Also, $p_{ii}(n) = q_{ii}(n)$, and therefore $\sum_{n \geq 0} q_{ii}(n) = \sum_{n \geq 0} p_{ii}(n)$, and therefore $\mathbf{Q}$ is recurrent by the potential matrix criterion. Call $g_{ji}(n)$ the probability, relative to the chain governed by the transition matrix $\mathbf{Q}$, of returning to state i for the first time at step n when starting from j. First-step analysis gives

$$g_{i0}(n+1) = \sum_{j \neq 0} q_{ij} g_{j0}(n)\,,$$

that is, using $(\star)$,

$$y_i g_{i0}(n+1) = \sum_{j \neq 0} (y_j g_{j0}(n)) p_{ji}.$$

Recall that ${}_0p_{0i}(n+1) = \sum_{j \neq 0} {}_0p_{0j}(n) p_{ji}$, or, equivalently,

$$y_0 \, {}_0p_{0i}(n+1) = \sum_{j \neq 0} (y_0 \, {}_0p_{0j}(n)) p_{ji}.$$

We therefore see that the sequences $\{y_0 \, {}_0p_{0i}(n)\}$ and $\{y_i g_{i0}(n)\}$ satisfy the same recurrence equation. Their first terms ($n = 1$), respectively $y_0 \, {}_0p_{0i}(1) = y_0 p_{0i}$ and $y_i g_{i0}(1) = y_i q_{i0}$, are equal in view of ($\star$). Therefore, for all $n \geq 1$,

$${}_0p_{0i}(n) = \frac{y_i}{y_0} g_{i0}(n).$$

Summing with respect to $n \geq 1$ and using $\sum_{n \geq 1} g_{i0}(n) = 1$ ($\mathbf{Q}$ is recurrent), we obtain that $x_i = \frac{y_i}{y_0}$. □

Equality (7.4) and the definition of positive recurrence give the following.

Theorem 7.2.4 *An irreducible recurrent* HMC *is positive recurrent if and only if its invariant measures x satisfy*

$$\sum_{i \in E} x_i < \infty \, .$$

An HMC may well be irreducible and possess an invariant measure, and yet not be recurrent. The simplest example is the 1-D non-symmetric random walk, which was shown to be transient and yet admits $x_i \equiv 1$ for invariant measure. It turns out, however, that the existence of a stationary probability distribution is necessary and sufficient for an irreducible chain (not a priori assumed recurrent) to be recurrent positive.

Theorem 7.2.5 *An irreducible* HMC *is positive recurrent if and only if there exists a stationary distribution. Moreover, the stationary distribution π is, when it exists, unique, and $\pi > 0$.*

Proof. The direct part follows from Theorems 7.2.2 and 7.2.4. For the converse part, assume the existence of a stationary distribution π. Iterating $\pi^T = \pi^T \mathbf{P}$, we obtain $\pi^T = \pi^T \mathbf{P}^n$, that is, for all $i \in E$, $\pi(i) = \sum_{j \in E} \pi(j) p_{ji}(n)$. If the chain were transient, then, for all states i, j,

$$\lim_{n \uparrow \infty} p_{ji}(n) = 0 \, .$$

To prove this, let T be the last time state i is visited. Since i is transient, T is a finite random variable and in particular, $\lim_n P_j(T > n) = 0$. But $X_n = i \Rightarrow T > n$ and therefore $P_j(X_n = i) \leq P_j(T > n)$.

Now, since $p_{ji}(n)$ is bounded uniformly in j and n by 1, by dominated convergence (Theorem A.1.5):

$$\pi(i) = \lim_{n\uparrow\infty} \sum_{j\in E} \pi(j) p_{ji}(n) = \sum_{j\in E} \pi(j) \left(\lim_{n\uparrow\infty} p_{ji}(n) \right) = 0.$$

This contradicts the assumption that π is a stationary distribution ($\sum_{i\in E} \pi(i) = 1$). The chain must therefore be recurrent, and by Theorem 7.2.4, it is positive recurrent.

The stationary distribution π of an irreducible positive recurrent chain is unique (use Theorem 7.2.3 and the fact that there is no choice for a multiplicative factor but 1). Also recall that $\pi(i) > 0$ for all $i \in E$ (see Theorem 7.2.2). □

Theorem 7.2.6 *Let π be the unique stationary distribution of an irreducible positive recurrent* HMC*, and let T_i be the return time to state i. Then*

$$\pi(i) E_i[T_i] = 1. \tag{7.5}$$

Proof. This equality is a direct consequence of expression (7.3) for the invariant measure. Indeed, π is obtained by normalization of x: for all $i \in E$,

$$\pi(i) = \frac{x_i}{\sum_{j\in E} x_j},$$

and in particular, for $i = 0$, recalling that $x_0 = 1$ and using (7.4),

$$\pi(0) = \frac{1}{E_0[T_0]}.$$

Since state 0 does not play a special role in the analysis, (7.5) is true for all $i \in E$. □

The situation is extremely simple when the state space is finite.

Theorem 7.2.7 *An irreducible* HMC *with finite state space is positive recurrent.*

Proof. We first show recurrence. We have

$$\sum_{j\in E} p_{ij}(n) = 1,$$

and in particular, the limit of the left-hand side is 1. If the chain were transient, then, as we saw in the proof of Theorem 7.2.5, for all $i, j \in E$,

$$\lim_{n\uparrow\infty} p_{ij}(n) = 0,$$

and therefore, since the state space is finite

$$\lim_{n\uparrow\infty} \sum_{j\in E} p_{ij}(n) = 0\,,$$

a contradiction. Therefore, the chain is recurrent. By Theorem 7.2.2 it has an invariant measure x. Since E is finite, $\sum_{i\in E} x_i < \infty$, and therefore the chain is positive recurrent, by Theorem 7.2.4. □

Birth-and-death Markov Chain

Birth-and-death process models are omnipresent in operations research and, of course, in biology. We first define the birth-and-death process with a bounded population. The state space of such a chain is $E = \{0, 1, \ldots, N\}$ and its transition matrix is

$$\mathbf{P} = \begin{pmatrix} r_0 & p_0 & & & & & & & \\ q_1 & r_1 & p_1 & & & & & & \\ & q_2 & r_2 & p_2 & & & & & \\ & & \ddots & & & & & & \\ & & & q_i & r_i & p_i & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & q_{N-1} & r_{N-1} & p_{N-1} \\ & & & & & & p_N & r_N \end{pmatrix},$$

where $p_i > 0$ for all $i \in E\backslash\{N\}$, $q_i > 0$ for all $i \in E\backslash\{0\}$, $r_i \geq 0$ for all $i \in E$, and $p_i + q_i + r_i = 1$ for all $i \in E$. The positivity conditions placed on the p_i's and q_i's guarantee that the chain is irreducible. Since the state space is finite, it is positive recurrent (Theorem 7.2.7), and it has a unique stationary distribution. Motivated by the Ehrenfest HMC which is reversible in the stationary state, we make the educated guess that the birth-and-death process considered has the same property. This will be the case if and only if there exists a probability distribution π on E satisfying the detailed balance equations, that is, such that for all $1 \leq i \leq N$, $\pi(i-1)p_{i-1} = \pi(i)q_i$. Letting $w_0 = 1$ and for all $1 \leq i \leq N$,

$$w_i = \prod_{k=1}^{i} \frac{p_{k-1}}{q_k}$$

we find that

$$\pi(i) = \frac{w_i}{\sum_{j=0}^{N} w_j} \tag{7.6}$$

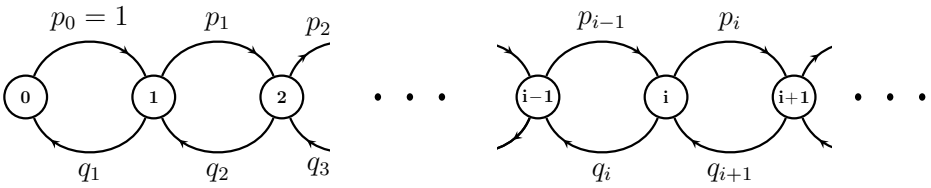
indeed satisfies the detailed balance equations and is therefore the (unique) stationary distribution of the chain.

We now consider the unbounded birth-and-death process, with state space $E = \mathbb{N}$ and transition matrix as in the previous example (only, it is "unbounded on the right"). We assume that the p_i's and q_i's are positive in order to guarantee irreducibility. The same reversibility argument as above applies with a little difference. In fact we can show that the w_i's defined above satisfy the detailed balance equations and therefore the global balance equations. Therefore the vector $\{w_i\}_{i\in E}$ is the unique, up to a multiplicative factor, invariant measure of the chain. It can be normalized to a probability distribution if and only if

$$\sum_{j=0}^{\infty} w_j < \infty\,.$$

Therefore, in this case and only in this case there exists a (unique) stationary distribution, also given by (7.6).

Note that the stationary distribution, when it exists, does not depend on the r_i's. The recurrence properties of the above unbounded birth-and-death process are therefore the same as those of the chain below, which is however not aperiodic. For aperiodicity of the original chain, it suffices to suppose at least one of the r_i's to be positive.



We now compute for the (bounded or unbounded) irreducible birth-and-death process the average time it takes to reach a state b from a state $a < b$. In fact, we shall prove that

$$E_a[T_b] = \sum_{k=a+1}^{b} \frac{1}{q_k w_k} \sum_{j=0}^{k-1} w_j \,. \tag{7.7}$$

Since obviously $E_a[T_b] = \sum_{k=a+1}^{b} E_{k-1}[T_k]$, it suffices to prove that

$$E_{k-1}[T_k] = \frac{1}{q_k w_k} \sum_{j=0}^{k-1} w_j \,. \tag{$\star$}$$

For this, consider for any given $k \in \{0, 1, \ldots, N\}$ the truncated chain which moves on the state space $\{0, 1, \ldots, k\}$ as the original chain, except in state k where it moves one step down with probability q_k and stays still with probability $p_k + r_k$. Use $\widetilde{E}$ to symbolize expectations with respect to the modified chain. The unique stationary distribution of this chain is

$$\widetilde{\pi}_\ell = \frac{w_\ell}{\sum_{j=0}^{k} w_\ell} \qquad (0 \le \ell \le k) \,.$$

First-step analysis yields $\widetilde{E}_k[T_k] = (r_k + p_k) \times 1 + q_k \left(1 + \widetilde{E}_{k-1}[T_k]\right)$, that is

$$\widetilde{E}_k[T_k] = 1 + q_k \widetilde{E}_{k-1}[T_k] \,.$$

Also

$$\widetilde{E}_k[T_k] = \frac{1}{\widetilde{\pi}_k} = \frac{1}{w_k} \sum_{j=0}^{k} w_j \,,$$

and therefore, since $\widetilde{E}_{k-1}[T_k] = E_{k-1}[T_k]$, we have $(\star)$.

EXAMPLE 7.2.8: SPECIAL CASES. In the special case where $(p_j, q_j, r_j) = (p, q, r)$ for all $j \neq 0, N$, $(p_0, q_0, r_0) = (p, q + r, 0)$ and $(p_N, q_N, r_N) = (0, p + r, q)$, we have $w_i = \left(\frac{p}{q}\right)^i$, and for $1 \le k \le N$,

$$E_{k-1}[T_k] = \frac{1}{q\left(\frac{p}{q}\right)^k} \sum_{j=0}^{k-1} \left(\frac{p}{q}\right)^j = \frac{1}{p-q}\left(1 - \left(\frac{q}{p}\right)^k\right).$$

7.2.2 The Ergodic Theorem

An important application of the strong law of large numbers is to the ergodic theorem for Markov chains giving conditions guaranteeing that empirical averages of the type

$$\frac{1}{N}\sum_{k=1}^{N} g(X_k, \ldots, X_{k+L})$$

converge to the corresponding probabilistic average computed for a stationary version of the chain. More precisely, if the chain is irreducible positive recurrent with stationary distribtion π and if $E_\pi[|g(X_0, \ldots, X_L)|] < \infty$, the above empirical average converges P_μ-almost surely to $E_\pi[g(X_0, \ldots, X_L)]$ for any initial distribution μ (Corollary 7.2.11).

We shall obtain this result as a corollary of the following proposition concerning irreducible recurrent (not necessarily positive recurrent) HMC's.

Theorem 7.2.9 *Let $\{X_n\}_{n\geq 0}$ be an irreducible recurrent* HMC*, and let x denote the canonical invariant measure associated with state $0 \in E$,*

$$x_i = E_0\left[\sum_{n\geq 1} 1_{\{X_n=i\}} 1_{\{n\leq T_0\}}\right], \tag{7.8}$$

where T_0 is the return time to 0. Define for $n \geq 1$, $\nu(n) := \sum_{k=1}^{n} 1_{\{X_k=0\}}$. Let $f : E \to \mathbb{R}$ be such that

$$\sum_{i\in E} |f(i)| x_i < \infty. \tag{7.9}$$

Then, for any initial distribution μ, P_μ-a.s.,

$$\lim_{N\uparrow\infty} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i) x_i\,. \tag{7.10}$$

Proof. Let $T_0 = \tau_1, \tau_2, \tau_3, \ldots$ be the successive return times to state 0, and define

$$U_p = \sum_{n=\tau_p+1}^{\tau_{p+1}} f(X_n)\,.$$

By the independence property of the regenerative cycles, $\{U_p\}_{p\geq 1}$ is an IID sequence. Moreover, assuming $f \geq 0$ and using the strong Markov property,

$$\begin{aligned} E[U_1] &= E_0\left[\sum_{n=1}^{T_0} f(X_n)\right] \\ &= E_0\left[\sum_{n=1}^{T_0}\sum_{i\in E} f(i)1_{\{X_n=i\}}\right] = \sum_{i\in E} f(i)E_0\left[\sum_{n=1}^{T_0} 1_{\{X_n=i\}}\right] \\ &= \sum_{i\in E} f(i)x_i\,. \end{aligned}$$

By hypothesis, this quantity is finite, and threfore the strong law of large numbers applies, to give

$$\lim_{n\uparrow\infty}\frac{1}{n}\sum_{p=1}^{n} U_p = \sum_{i\in E} f(i)x_i\,,$$

that is,

$$\lim_{n\uparrow\infty}\frac{1}{n}\sum_{k=T_0+1}^{\tau_{n+1}} f(X_k) = \sum_{i\in E} f(i)x_i\,. \tag{7.11}$$

Observing that

$$\tau_{\nu(n)} \le n < \tau_{\nu(n)+1}\,,$$

we have

$$\frac{\sum_{k=1}^{\tau_{\nu(n)}} f(X_k)}{\nu(n)} \le \frac{\sum_{k=1}^{n} f(X_k)}{\nu(n)} \le \frac{\sum_{k=1}^{\tau_{\nu(n)+1}} f(X_i)}{\nu(n)}\,.$$

Since the chain is recurrent, $\lim_{n\uparrow\infty}\nu(n) = \infty$, and therefore, from (7.11), the extreme terms of the above chain of inequality tend to $\sum_{i\in E} f(i)x_i$ as n goes to ∞, and this implies (7.10). The case of a function f of arbitrary sign is obtained by considering (7.10) written separately for $f^+ = \max(0, f)$ and $f^- = \max(0, -f)$, and then taking the difference of the two equalities obtained in this way. The difference is not an undetermined form $\infty - \infty$ due to hypothesis (7.9). □

The main result of ergodicity of Markov chains concerns the positive recurrent case.

Corollary 7.2.10 *Let $\{X_n\}_{n\ge 0}$ be an irreducible positive recurrent Markov chain with the stationary distribution π, and let $f : E \to \mathbb{R}$ be such that*

$$\sum_{i\in E} |f(i)|\pi(i) < \infty. \tag{7.12}$$

Then for any initial distribution μ, P_μ-a.s.,

$$\lim_{n\uparrow\infty}\frac{1}{N}\sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)\pi(i)\,. \tag{7.13}$$

Proof. Apply Theorem 7.2.9 to $f \equiv 1$. Condition (7.9) is satisfied, since in the positive recurrent case, $\sum_{i\in E} x_i = E_0[T_0] < \infty$. Therefore, P_μ-a.s.,

$$\lim_{N\uparrow\infty} \frac{N}{\nu(N)} = \sum_{j\in E} x_j \,.$$

Now, f satisfying (7.12) also satisfies (7.9), since x and π are proportional, and therefore, P_μ-a.s.,

$$\lim_{N\uparrow\infty} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i) x_i \,.$$

The combination of the above equalities gives, P_μ-a.s.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \lim_{N\to\infty} \frac{\nu(N)}{N} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \frac{\sum_{i\in E} f(i) x_i}{\sum_{j\in E} x_j} \,,$$

from which (7.13) follows, since π is obtained by normalization of x. □

Corollary 7.2.11 *Let $\{X_n\}_{n\geq 1}$ be an irreducible positive recurrent Markov chain with the stationary distribution π, and let $g : E^{L+1} \to \mathbb{R}$ be such that*

$$\sum_{i_0,\ldots,i_L} |g(i_0,\ldots,i_L)| \pi(i_0) p_{i_0 i_1} \cdots p_{i_{L-1} i_L} < \infty \,.$$

Then for all initial distributions μ, P_μ-a.s.

$$\lim \frac{1}{N} \sum_{k=1}^{N} g(X_k, X_{k+1}, \ldots, X_{k+L}) = \sum_{i_0,i_1,\ldots,i_L} g(i_0, i_1, \ldots, i_L) \pi(i_0) p_{i_0 i_1} \cdots p_{i_{L-1} i_L} \,.$$

Proof. Apply Corollary 7.2.10 to the "snake chain" $\{(X_n, X_{n+1}, \ldots, X_{n+L})\}_{n\geq 0}$, which is irreducible recurrent and admits the stationary distribution (see Exercise 7.5.7)

$$\pi(i_0) p_{i_0 i_1} \cdots p_{i_{L-1} i_L} \,.$$

□

Note that

$$\sum_{i_0,i_1,\ldots,i_L} g(i_0, i_1, \ldots, i_L) \pi(i_0) p_{i_0 i_1} \cdots p_{i_{L-1} i_L} = E_\pi[g(X_0, \ldots, X_L)] \,.$$

EXAMPLE 7.2.12: ERGODIC ESTIMATE OF THE TRANSITION MATRIX. Let $\{X_n\}_{n\geq 1}$ be an irreducible positive recurrent Markov chain with the stationary distribution π. Applying Corollary 7.2.11 successively with $g(i) = 1_{i_0}(i)$ and $g(i,j) = 1_{(i_0,i_1)}(i,j)$ yields

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{i_0}(X_n) = \pi(i_0)$$

and

$$\lim_{N\uparrow\infty} \frac{1}{N}\frac{1}{N}\sum_{k=1}^{N} 1_{i_0,i_1}(X_n, X_{n+1}) = \pi(i_0)p_{i_0,i_1}\,,$$

and therefore, in particular,

$$\lim_{N\uparrow\infty} \frac{1}{N}\frac{1}{N}\frac{\sum_{k=1}^{N} 1_{\{X_n=i_0, X_{n+1}=i_1\}}}{\sum_{k=1}^{N} 1_{\{X_n=i_0\}}} = p_{i_0,i_1}\,.$$

7.3 The Lyapunov Function Method

7.3.1 Foster's Condition of Positive Recurrence

The stationary distribution criterion of positive recurrence of an irreducible chain requires solving the balance equations, and this is not always feasible. Therefore one needs less general but efficient conditions guaranteeing positive recurrence.

Theorem 7.3.1 *(Foster, 1953) Let the transition matrix* $\mathbf{P}$ *on the countable state space* E *be irreducible and suppose that there exists a function* $h : E \to \mathbb{R}$ *such that* $\inf_i h(i) > -\infty$ *and*

$$\sum_{k\in E} p_{ik}h(k) < \infty \text{ for all } i \in F, \tag{7.14}$$

$$\sum_{k\in E} p_{ik}h(k) \le h(i) - \epsilon \text{ for all } i \notin F, \tag{7.15}$$

for some finite set F *and some* $\epsilon > 0$. *Then the corresponding* HMC *is positive recurrent.*

Proof. Since $\inf_i h(i) > -\infty$, one may assume without loss of generality that $h \ge 0$, by adding a constant if necessary. Call τ the return time to F, and define $Y_n = h(X_n)1_{\{n<\tau\}}$. Equality (7.15) is just $E[h(X_{n+1}) \,|\, X_n = i] \le h(i) - \epsilon$ for all $i \notin F$. For $i \notin F$,

$$\begin{aligned} E_i[Y_{n+1} \,|\, X_0^n] &= E_i[Y_{n+1}1_{\{n<\tau\}} \,|\, X_0^n] + E_i(Y_{n+1}1_{\{n\ge\tau\}} \,|\, X_0^n] \\ &= E_i[Y_{n+1}1_{\{n<\tau\}} \,|\, X_0^n] \le E_i[h(X_{n+1})1_{\{n<\tau\}} \,|\, X_0^n] \\ &= 1_{\{n<\tau\}}E_i[h(X_{n+1}) \,|\, X_0^n] = 1_{\{n<\tau\}}E_i[h(X_{n+1}) \,|\, X_n] \\ &\le 1_{\{n<\tau\}}h(X_n) - \epsilon 1_{\{n<\tau\}}, \end{aligned}$$

where the third *equality* comes from the fact that $1_{\{n<\tau\}}$ is a function of X_0^n, the fourth *equality* is the Markov property, and the last *inequality* is true because P_i-a.s., $X_n \notin F$ on $n < \tau$. Therefore, P_i-a.s., $E_i[Y_{n+1} \,|\, X_0^n] \le Y_n - \epsilon 1_{\{n<\tau\}}$, and taking expectations,

$$E_i[Y_{n+1}] \le E_i[Y_n] - \epsilon P_i(\tau > n).$$

Iterating the above equality, and observing that Y_n is non-negative, we obtain

$$0 \leq E_i[Y_{n+1}] \leq E_i[Y_0] - \epsilon \sum_{k=0}^{n} P_i(\tau > k).$$

But $Y_0 = h(i)$, P_i-a.s., and $\sum_{k=0}^{\infty} P_i(\tau > k) = E_i[\tau]$. Therefore, for all $i \notin F$,

$$E_i[\tau] \leq \epsilon^{-1} h(i).$$

For $j \in F$, first-step analysis yields

$$E_j[\tau] = 1 + \sum_{i \notin F} p_{ji} E_i[\tau].$$

Thus $E_j[\tau] \leq 1 + \epsilon^{-1} \sum_{i \notin F} p_{ji} h(i)$, and this quantity is finite in view of assumption (7.14). Therefore, the return time to F starting anywhere in F has finite expectation. Since F is a finite set, this implies positive recurrence in view of the following lemma. □

Lemma 7.3.2 *Let $\{X_n\}_{n\geq 0}$ be an irreducible* HMC*, let F be a finite subset of the state space E, and let $\tau(F)$ be the return time to F. If $E_j[\tau(F)] < \infty$ for all $j \in F$, the chain is positive recurrent.*

Proof. Select $i \in F$, and let T_i be the return time of $\{X_n\}$ to i. Let $\tau_1 = \tau(F), \tau_2, \tau_3, \ldots$ be the successive return times to F. It follows from the strong Markov property that $\{Y_n\}_{n\geq 0}$ defined by $Y_0 = X_0 = i$ and $Y_n = X_{\tau_n}$ for $n \geq 1$ is an HMC with state space F. Since $\{X_n\}$ is irreducible, so is $\{Y_n\}$. Since F is finite, $\{Y_n\}$ is positive recurrent, and in particular, $E_i[\tilde{T}_i] < \infty$, where $\tilde{T}_i$ is the return time to i of $\{Y_n\}$. Defining $S_0 = \tau_1$ and $S_k = \tau_{k+1} - \tau_k$ for $k \geq 1$, we have

$$T_i = \sum_{k=0}^{\infty} S_k 1_{\{k<\tilde{T}_i\}},$$

and therefore

$$E_i[T_i] = \sum_{k=0}^{\infty} E_i[S_k 1_{\{k<\tilde{T}_i\}}].$$

Now,

$$E_i[S_k 1_{\{k<\tilde{T}_i\}}] = \sum_{\ell \in F} E_i[S_k 1_{\{k<\tilde{T}_i\}} 1_{\{X_{\tau_k}=\ell\}}],$$

and by the strong Markov property applied to $\{X_n\}_{n\geq 0}$ and the stopping time τ_k, and the fact that the event $\{k < \tilde{T}_i\}$ belongs to the past of $\{X_n\}_{n\geq 0}$ at time τ_k,

$$\begin{aligned} E_i[S_k 1_{\{k<\tilde{T}_i\}} 1_{\{X_{\tau_k}=\ell\}}] &= E_i[S_k \,|\, k < \tilde{T}_i, X_{\tau_k} = \ell] P_i(k < \tilde{T}_i, X_{\tau_k} = \ell) \\ &= E_i[S_k \,|\, X_{\tau_k} = \ell] P_i(k < \tilde{T}_i, X_{\tau_k} = \ell). \end{aligned}$$

Observing that $E_i[S_k \,|\, X_{\tau_k} = \ell] = E_\ell[\tau(F)]$, we see that the latter expression is bounded by $(\max_{\ell \in F} E_\ell[\tau(F)])\, P_i(k < \tilde{T}_i, X_{\tau_k} = \ell)$, and therefore

$$E_i[T_i] \leq \left(\max_{\ell \in F} E_\ell(\tau(F))\right) \sum_{k=0}^{\infty} P_i(\tilde{T}_i > k) = \left(\max_{\ell \in F} E_\ell(\tau(F))\right) E_i[\tilde{T}_i] < \infty.$$

□

Remark 7.3.3 The function h in Foster's theorem is called a Lyapunov function because it plays a role similar to the Lyapunov functions in the stability theory of ordinary differential equations.

The corollary below is refered to as Pakes's lemma.

Corollary 7.3.4 *(Pakes, 1969) Let $\{X_n\}_{n\geq 0}$ be an irreducible* HMC *on $E = \mathbb{N}$ such that for all $n \geq 0$ and all $i \in E$,*

$$E[X_{n+1} - X_n \,|\, X_n = i] < \infty$$

and

$$\limsup_{i \uparrow \infty} E[X_{n+1} - X_n \,|\, X_n = i] < 0. \tag{7.16}$$

Such an HMC *is positive recurrent.*

Proof. Let -2ϵ be the left-hand side of (7.16). In particular, $\epsilon > 0$. By (7.16), for i sufficiently large, say $i > i_0$, $E[X_{n+1} - X_n \,|\, X_n = i] < -\epsilon$. We are therefore in the conditions of Foster's theorem with $h(i) = i$ and $F = \{i; i \leq i_0\}$. □

EXAMPLE 7.3.5: A RANDOM WALK ON $\mathbb{N}$. Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of integrable random variables with values in $\mathbb{Z}$ such that

$$E[Z_1] < 0,$$

and define $\{X_n\}_{n\geq 0}$, an HMC with state space $E = \mathbb{N}$, by

$$X_{n+1} = (X_n + Z_{n+1})^+,$$

where X_0 is independent of $\{Z_n\}_{n\geq 1}$. Assume irreducibility (the industrious reader will find the necessary and sufficient condition for this). Here

$$\begin{aligned} E[X_{n+1} - i \mid X_n = i] &= E[(i + Z_{n+1})^+ - i] \\ &= E[-i1_{\{Z_{n+1} \leq -i\}} + Z_{n+1}1_{\{Z_{n+1} > -i\}}] \leq E[Z_1 1_{\{Z_1 > -i\}}]. \end{aligned}$$

By dominated convergence, the limit of $E[Z_1 1_{\{Z_1 > -i\}}]$ as i tends to ∞ is $E[Z_1] < 0$ and therefore, by Pakes's lemma, the HMC is positive recurrent.

The following is a Foster-type theorem, only with a negative conclusion.

Theorem 7.3.6 *Let the transition matrix* $\mathbf{P}$ *on the countable state space* E *be irreducible and suppose that there exists a finite set* F *and a function* $h : E \to \mathbb{R}_+$ *such that*

$$\text{there exists a state } j \notin F \text{ such that } h(j) > \max_{i \in F} h(i) \tag{7.17}$$

$$\sup_{i \in E} \sum_{k \in E} p_{ik} |h(k) - h(i)| < \infty, \tag{7.18}$$

$$\sum_{k \in E} p_{ik}(h(k) - h(i)) \leq 0 \text{ for all } i \notin F. \tag{7.19}$$

Then the corresponding HMC *cannot be positive recurrent.*

Proof. Let τ be the return time to F. Observe that

$$h(X_\tau) 1_{\{\tau < \infty\}} = h(X_0) + \sum_{n=0}^{\infty} (h(X_{n+1}) - h(X_n)) \, 1_{\{\tau > n\}}.$$

Now, with $j \notin F$,

$$\begin{aligned}
&\sum_{n=0}^{\infty} E_j \left[|h(X_{n+1}) - h(X_n)| \, 1_{\{\tau > n\}} \right] \\
&\quad = \sum_{n=0}^{\infty} E_j \left[E_j \left[|h(X_{n+1}) - h(X_n)| \, | X_0^n \right] 1_{\{\tau > n\}} \right] \\
&\quad = \sum_{n=0}^{\infty} E_j \left[E_j \left[|h(X_{n+1}) - h(X_n)| \, | X_n \right] 1_{\{\tau > n\}} \right] \\
&\quad \leq K \sum_{n=0}^{\infty} P_j(\tau > n)
\end{aligned}$$

for some finite positive constant K by (7.18). Therefore, if the chain is positive recurrent, the latter bound is $K E_j[\tau] < \infty$. Therefore

$$\begin{aligned}
E_j[h(X_\tau)] &= E_j \left[h(X_\tau) 1_{\{\tau < \infty\}} \right] \\
&= h(j) + \sum_{n=0}^{\infty} E_j \left[(h(X_{n+1}) - h(X_n)) \, 1_{\{\tau > n\}} \right] > h(j),
\end{aligned}$$

by (7.19). In view of assumption (7.17), we have $h(j) > \max_{i \in F} h(i) \geq E_j[h(X_\tau)]$, hence a contradiction. The chain therefore cannot be positive recurrent. □

7.3.2 Queueing Applications

EXAMPLE 7.3.7: THE REPAIR SHOP, TAKE 4. Assuming irreducibility (see Example 6.1.17), we now seek a necessary and sufficient condition for positive recurrence. For any complex number z with modulus not larger than 1, it follows from the recurrence equation (6.5) that

$$z^{X_{n+1}+1} = \left(z^{(X_n-1)^+ +1}\right) z^{Z_{n+1}} = \left(z^{X_n} - 1_{\{X_n=0\}} + z1_{\{X_n=0\}}\right) z^{Z_{n+1}},$$

and therefore $zz^{X_{n+1}} - z^{X_n} z^{Z_{n+1}} = (z-1)1_{\{X_n=0\}} z^{Z_{n+1}}$. From the independence of X_n and Z_{n+1}, $E[z^{X_n} z^{Z_{n+1}}] = E[z^{X_n}] g_Z(z)$, and $E[1_{\{X_n=0\}} z^{Z_{n+1}}] = \pi(0) g_Z(z)$, where $\pi(0) = P(X_n = 0)$. Therefore, $zE[z^{X_{n+1}}] - g_Z(z)E[z^{X_n}] = (z-1)\pi(0)g_Z(z)$. But in steady state, $E[z^{X_{n+1}}] = E[z^{X_n}] = g_X(z)$, and therefore

$$g_X(z)\,(z - g_Z(z)) = \pi(0)(z-1)g_Z(z). \tag{7.20}$$

This gives the generating function $g_X(z) = \sum_{i=0}^{\infty} \pi(i) z^i$, as long as $\pi(0)$ is available. To obtain $\pi(0)$, differentiate (7.20):

$$g'_X(z)\,(z - g_Z(z)) + g_X(z)\,(1 - g'_Z(z)) = \pi(0)\,(g_Z(z) + (z-1)g'_Z(z))\,,$$

and let $z = 1$, to obtain, taking into account the equalities $g_X(1) = g_Z(1) = 1$ and $g'_Z(1) = E[Z]$,

$$\pi(0) = 1 - E[Z]. \tag{7.21}$$

But the stationary distribution of an irreducible HMC is positive, hence the necessary condition of positive recurrence:

$$E[Z_1] < 1.$$

We now show this condition is also sufficient for positive recurrence. This follows immediately from Pakes's lemma, since for $i \geq 1$, $E[X_{n+1} - X_n \,|\, X_n = i] = E[Z] - 1 < 0$.

From (7.20) and (7.21), we have the generating function of the stationary distribution:

$$\sum_{i=0}^{\infty} \pi(i) z^i = (1 - E[Z]) \frac{(z-1)g_Z(z)}{z - g_Z(z)}. \tag{7.22}$$

If $E[Z_1] > 1$, the chain is transient, as a simple argument based on the strong law of large numbers shows. In fact, $X_n = X_0 + \sum_{k=1}^{n} Z_k - n + \sum_{k=1}^{n} 1_{\{X_k=0\}}$, and therefore

$$X_n \geq \sum_{k=1}^{n} Z_k - n,$$

which tends to ∞ because, by the strong law of large numbers,

$$\frac{\sum_{k=1}^{n} Z_k - n}{n} \to E[Z] - 1 > 0.$$

This is of course incompatible with recurrence.

We finally examine the case $E[Z_1] = 1$, for which there are only two possibilities left: transient or null recurrent. It turns out that the chain is null recurrent in this case. The proof relies on Theorem 17.3.9. In fact, the conditions of this theorem are easily verified with $h(i) = i$ and $F = \{0\}$. Therefore, the chain is recurrent. Since it is not positive recurrent, it is null-recurrent.

EXAMPLE 7.3.8: ALOHA, TAKE 2. It turns out, as will be shown next, that the Bernoulli retransmission policy makes the ALOHA protocol *unstable*, in the sense that the chain $\{X_n\}_{n\geq 0}$ is *not positive recurrent*.

An elementary computation yields, for the ALOHA model,

$$E[X_{n+1} - X_n \,|\, X_n = i] = \lambda - b_1(i)a_0 - b_0(i)a_1. \tag{7.23}$$

Note that $b_1(i)a_0 + b_0(i)a_1$ is the probability of one successful (re-)transmission in a slot given that the backlog at the beginning of the slot is i. Equivalently, since there is at most one successful (re-)transmission in any slot, this is the average number of successful (re-)transmissions in a slot given the backlog i at the start of the slot. An elementary computation shows that $\lim_{i\uparrow\infty}(b_1(i)a_0 + b_0(i)a_1) = 0$. Therefore, outside a finite set F, the conditions of Theorem 7.3.6 are satisfied when we take h to be the identity, and remember the hypothesis that $E[A_1] < \infty$.

EXAMPLE 7.3.9: ALOHA, TAKE 3. The ALOHA protocol with a fixed retransmission probability ν is unstable, it seems natural to try a retransmission probability $\nu = \nu(k)$ depending on the number k of backlogged messages. We show that there is a choice of the function $\nu(k)$ that achieves stability of the protocol. The probability that i among the k backlogged messages at the beginning of slot n retransmit in slot n is now $\nu(k)$. The same is true for the transition probabilities. According to Pakes's lemma and using (7.23), it suffices to find a function $\nu(k)$ guaranteeing that

$$\lambda \leq \lim_{i\uparrow\infty}(b_1(i)a_0 + b_0(i)a_1) - \epsilon, \tag{7.24}$$

for some $\epsilon > 0$. We shall therefore study the function

$$g_k(\nu) = (1-\nu)^k a_1 + k\nu(1-\nu)^{k-1}a_0,$$

since condititition (7.24) is just $\lambda \leq g_i(\nu(i)) - \epsilon$. The derivative of $g_k(\nu)$ is, for $k \geq 2$,

$$g_k'(\nu) = k(1-\nu)^{k-2}[(a_0 - a_1) - \nu(ka_0 - a_1)].$$

We first assume that $a_0 > a_1$. In this case, for $k \geq 2$, the derivative is zero for

$$\nu = \nu(k) = \frac{a_0 - a_1}{ka_0 - a_1},$$

and the corresponding value of $g_k(\nu)$ is a maximum equal to

$$g_k(\nu(k)) = a_0\left(\frac{k-1}{k - a_1/a_0}\right)^{k-1}.$$

Therefore, $\lim_{k\uparrow\infty} g_k(\nu(k)) = a_0 \exp\left\{\frac{a_1}{a_0} - 1\right\}$, and we see that

$$\lambda < a_0 \exp\left\{\frac{a_1}{a_0} - 1\right\} \tag{7.25}$$

is a sufficient condition for stability of the protocol. For instance, with a Poisson distribution of arrivals

$$a_i = e^{-\lambda}\frac{\lambda^i}{i!},$$

condition (7.25) reads

$$\lambda < e^{-1}$$

(in particular, the condition $a_0 > a_1$ is satisfied a posteriori). If $a_0 \leq a_1$, the protocol can be shown to be unstable, whatever retransmission policy $\nu(k)$ is adopted (the reader is invited to check this).

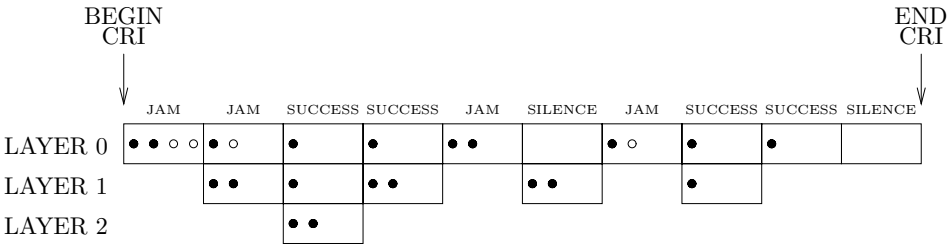
EXAMPLE 7.3.10: THE STACK ALGORITHM. (Capetanakis, 1979; Tsybakhov and Mikhailov, 1980) The slotted ALOHA protocol with constant retransmission probability was proved unstable, and it was shown that a backlog-dependent retransmission probability could restore stability . The problem then resides in the necessity for each user to know the size of the backlog in order to implement the retransmission policy. This is not practically feasible, and one must devise policies based on the actual information available by just listening to the link: collision, no transmission, or successful transmission. Such policies, which in a sense estimate the backlog, have been found that yield stability. We shall not discuss them here. Instead we shall consider a new type of *collision-resolution protocol*, called the *binary tree protocol*, or the *stack algorithm*.

In this protocol, when a collision occurs for the first time, all new requests are buffered until all the messages involved in the collision have found their way through the link. When these messages have resolved their collision problem, the buffered messages then try to retransmit. They may enter a collision, and then will try resolve their collision. Time is therefore divided into successive periods, called *collision-resolution intervals* (CRI). Let us examine the fate of the messages arriving in the first slot just after a CRI, which are the messages that arrived during the previous CRI. They all try to retransmit in the first slot of the CRI, and therefore, if there are two or more messages, a collision occurs (in the other case, the CRI has lasted just one slot, and a new CRI begins in the next slot). An unbiased coin is tossed independently for each colliding message. If it shows heads, the message joins *layer* 0 of a *stack*, whereas if it shows tails, it is placed in layer 1. In the next slot, all messages of layer 0, and only them, try the link. If there is no collision (because layer 0 was empty or contained just one message), layer 0 is eliminated, and layer 1 below pops up to become layer 0. If on the contrary there is a collision because layer 0 formed after the first slot contained two or more messages, the colliding messages again flip a coin; those with heads form the new layer 0, those with tails form the new layer 1, and the former layer 1 is pushed bottomwards to form layer 2.

In general, at each step, only layer 0 tries to retransmit. If there is no collision, layer 0 disappears, and the layers $1, 2, 3, \ldots$ become layers $0, 1, 2, \ldots$ If there is a collision, layer 0 splits into layer 0 and layer 1, and layers $1, 2, 3, \ldots$ become layer $2, 3, 4, \ldots$ It should be noted that in this protocol, each message knows at every instant

in which layer it is, just by listening to the channel that gives the information: collision or no collision. In that sense, the protocol is *distributed*, because there is no central operator broadcasting non-locally available information, such as the size of the backlog, to all users.

Once a collision is resolved, that is, when all layers have disappeared, a new CRI begins. The number of customers that are starting this CRI are those that have arrived in the CRI that just ended. The figure below gives an example of what happens in a CRI.



In the figure above, the four messages at the begining are those buffered in the previous CRI. A black dot correponds to a message with "heads", that is authorized to attempt transmission.

Since the fresh requests sequence $\{A_n\}_{n\geq 1}$ is IID, the sequence, $\{X_n\}_{n\geq 0}$, where X_n is the *length* of the n-th CRI, forms an irreducible HMC. Stability of the protocol is naturally identified with positive recurrence of this chain, which will now be proved with the help of Pakes's lemma. It suffices to show that

$$\limsup_{i\uparrow\infty} E[X_{n+1} - X_n \,|\, X_n = i] < 0 \tag{7.26}$$

and for all i,

$$E[X_{n+1} \,|\, X_n = i] < \infty. \tag{7.27}$$

For this, let Z_n be the number of fresh arrivals in the n-th CRI. We have

$$\begin{aligned} E[X_{n+1} \,|\, X_n = i] &= \sum_{k=0}^{\infty} E[X_{n+1} \,|\, X_n = i, Z_n = k] P(Z_n = k \,|\, X_n = i) \\ &= \sum_{k=0}^{\infty} E[X_{n+1} \,|\, Z_n = k] P(Z_n = k \,|\, X_n = i). \end{aligned}$$

It will be shown that for all $n \geq 0$,

$$E[X_{n+1} \,|\, Z_n = k] \leq \alpha k + 1, \tag{7.28}$$

where $\alpha = 2.886$, and therefore

$$E[X_{n+1} \,|\, X_n = i] \leq \sum_{k=0}^{\infty} (\alpha k + 1) P(Z_n = k \,|\, X_n = i) = \alpha E[Z_n \,|\, X_n = i] + 1.$$

Using Wald's lemma (Theorem 3.2 of Chapter 1), we have

$$E[Z_n \,|\, X_n = i] = \lambda i,$$

where λ is the *traffic intensity*, and therefore

$$E[X_{n+1} - X_n \,|\, X_n = i] \le 1 + i(\lambda\alpha - 1).$$

We see that condition (7.27) is always satisfied and that (7.26) is satisfied, provided that

$$\lambda < \frac{1}{\alpha} = 0.346. \tag{7.29}$$

It remains to prove (7.28). Let $E[X_{n+1} \,|\, Z_n = k] = L_k$ (it is indeed a quantity independent of n). Clearly,

$$L_0 = L_1 = 1,$$

since with zero or one packet at the beginning of a CRI, there is no collision. When $k \ge 2$, there is a collision, and the k users toss a coin, and depending on the result they split into two sets, layer 0 and layer 1. Among these k users, i obtain heads with probability

$$q_i(k) = \binom{k}{i} \left(\frac{1}{2}\right)^k.$$

The average length of the CRI given that there are $k \ge 2$ customers at the start, and given that the first layer 0 contains i messages, is

$$L_{k,i} = 1 + L_i + L_{k-i}.$$

Indeed, the first slot saw a collision; the i customers in the first layer 0 will take on average L_i slots to resolve their collision, and L_{k-i} more slots will be needed for the $k - i$ customers in the first-formed layer 1 (these customers are always at the bottom of the stack, in a layer traveling up and down until it becomes layer 0, at which time they start resolving their collision). Since

$$L_k = \sum_{i=0}^{k} q_i(k) L_{k,i}\,,$$

we have

$$L_k = 1 + \sum_{i=0}^{k} q_i(k)(L_i + L_{k-i}).$$

Solving for L_k, we obtain

$$L_k = \frac{1 + \sum_{i=0}^{k-1} [q_i(k) + q_{k-i}(k)] L_i}{1 - q_0(k) - q_k(k)}\,. \tag{7.30}$$

Suppose that for some $m \ge 2$, and α_m satisfying

$$\alpha_m \ge \sup_{j>m} \frac{\sum_{i=0}^{m-1} (L_i + 1)(q_i(j) + q_{j-i}(j))}{\sum_{i=0}^{m-1} i(q_i(j) + q_{j-i}(j))}\,, \tag{7.31}$$

it holds that $L_m \leq \alpha_m m - 1$. We shall then prove that for all $n \geq m$,

$$L_n \leq \alpha_m n - 1\,. \tag{7.32}$$

We do this by induction, supposing that (7.32) holds true for $n = m, m+1, \ldots, j-1$, and proving that it holds true for $n = j$. Equality (7.30) gives

$$\begin{aligned} L_j(1 - q_0(j) - q_j(j)) &= 1 + \sum_{i=0}^{j-1}(q_i(j) + q_{j-i}(j))L_i \\ &= 1 + \sum_{i=0}^{m-1} + \sum_{i=m}^{j-1} \\ &\leq 1 + \sum_{i=0}^{m-1} + \sum_{i=m}^{j-1}(q_i(j) + q_{j-i}(j))(\alpha_m i - 1)\,, \end{aligned}$$

where we used the induction hypothesis. The latter term equals

$$\begin{aligned} &1 + \sum_{i=0}^{m-1}(q_i(j) + q_{j-i}(j))(L_i - \alpha_m i + 1) + \sum_{i=0}^{j}(q_i(j) \\ &\qquad + q_{j-i}(j))(\alpha_m i - 1) - (q_0(j) + q_j(j))(\alpha_m j - 1) \\ &= 1 + \sum_{i=0}^{m-1}(q_i(j) + q_{j-i}(j))(L_i - \alpha_m i + 1) + \alpha_m j - 2 - (q_0(j) + q_j(j))(\alpha_m j - 1)\,, \end{aligned}$$

where we used the identities

$$\sum_{i=0}^{j} q_i(j) = 1,\ \sum_{i=0}^{j} i q_i(j) = jp,\ \sum_{i=0}^{j} i q_{j-i}(j) = j(1-p).$$

Therefore,

$$L_j \leq (\alpha_m j - 1) + \frac{\sum_{i=0}^{m-1}(q_i(j) + q_{j-i}(j))(L_i - \alpha_m i + 1)}{1 - q_0(j) - q_j(j)}.$$

Therefore, for $L_j \leq \alpha_m j - 1$ to hold, it suffices to have

$$\sum_{i=0}^{m-1}(q_i(j) + q_{j-i}(j))(L_i - \alpha_m i + 1) \leq 0.$$

We require this to be true for all $j > m$, and (7.31) guarantees this. It can be checked *numerically* that for $m = 6$ and $\alpha_6 = 0.286$, (7.31) is satisfied and that equality (7.32) is true for $n = 1, 2, 3, 4, 5, 6$, and this completes the proof.

7.4 Fundamental Matrix

7.4.1 Definition

The fundamental matrix of an ergodic HMC with finite state space $E = \{1, \ldots, r\}$ and stationary distribution π is the matrix

$$\mathbf{Z} := (I - (\mathbf{P} - \Pi))^{-1}, \tag{7.33}$$

where

$$\Pi := \mathbf{1}\pi^T = \begin{pmatrix} \pi(1) & \cdots & \pi(r) \\ \pi(1) & \cdots & \pi(r) \\ \vdots & & \vdots \\ \pi(1) & \cdots & \pi(r) \end{pmatrix}.$$

It gives access to a number of quantities such as, for instance, the mean time $E_i[T_j]$ to return to j from state i, or the variance of the ergodic estimate $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$.

Theorem 7.4.1 *For any ergodic transition matrix* $\mathbf{P}$ *on a finite state space, the right-hand side of (7.33) is well defined and*

$$\mathbf{Z} = I + \sum_{n\geq 1}(\mathbf{P}^n - \Pi). \tag{7.34}$$

In particular, $\sum_j \mathbf{Z}_{ij} = 1$.

Proof. First observe that

$$\begin{aligned} \Pi\mathbf{P} &= \Pi \text{ (since } \pi^T\mathbf{P} = \pi^T \text{ and } \Pi = \mathbf{1}\pi^T), \\ \mathbf{P}\Pi &= \Pi \text{ (since } \mathbf{P}\mathbf{1} = \mathbf{1} \text{ and } \Pi = \mathbf{1}\pi^T), \\ \Pi^2 &= \Pi \text{ (since } \Pi = \mathbf{1}\pi^T \text{ and } \pi^T\mathbf{1} = \mathbf{1}). \end{aligned}$$

In particular, for all $k \geq 1$, $\mathbf{P}\Pi^k = \Pi = \Pi^k\mathbf{P}$, and therefore,

$$\begin{aligned} (\mathbf{P} - \Pi)^n &= \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \mathbf{P}^k \Pi^{n-k} \\ &= \mathbf{P}^n + \left(\sum_{k=0}^{n-1} \binom{n}{k} (-1)^{n-k} \right) \Pi = \mathbf{P}^n - \Pi. \end{aligned}$$

Therefore, with $A = \mathbf{P} - \Pi$,

$$(I - A)(I + A + \cdots + A^{n-1}) = I - A^n = I - \mathbf{P}^n + \Pi.$$

Letting $n \to \infty$,

$$(I - A)(I + \sum_{n\geq 1} A^n) = I,$$

which shows that $I - (\mathbf{P} - \Pi)$ is invertible, with inverse

$$I + \sum_{n\geq 1}(\mathbf{P} - \Pi)^n = I + \sum_{n\geq 1}(\mathbf{P}^n - \Pi).$$

□

Remark 7.4.2 Some authors use another definition of the fundamental matrix:

$$\widetilde{\mathbf{Z}} = \sum_{n\geq 0} (\mathbf{P}^n - \Pi)\,, \tag{7.35}$$

that is $\widetilde{\mathbf{Z}} = Z - \Pi$.

Remark 7.4.3 Expression (7.34) is meaningful only if the chain is ergodic. In particular, if the chain is only recurrent positive, but periodic, the series on the right-hand side of (7.34) oscillates. This does not mean, however, that in the periodic case the inverse in (7.34) does not exist. As a matter of fact, it does exist, but it is not given by formula (7.34).

An Extension of the Fundamental Matrix

The following is an alternative description of the fundamental matrix that does not require, in principle, knowledge of the stationary distribution.

Let b be any vector such that

$$b^T\mathbf{1} \neq 0, \tag{7.36}$$

and define

$$\mathbf{Z} = \left(I - \mathbf{P} + \mathbf{1}b^T\right)^{-1}, \tag{7.37}$$

where $\mathbf{P}$ is an ergodic matrix on the finite space E, with the stationary distribution π. The matrix differs from the usual fundamental matrix in that π is replaced by b.

Theorem 7.4.4 *(Kemeny, 1991 ; Grinstead and Snell, 1997)*
The inverse matrix in (7.37) exists and

$$\pi^T = b^T\mathbf{Z}. \tag{7.38}$$

Proof. Since $\pi^T\mathbf{1} = 1$ and $\pi^T(I - \mathbf{P}) = 0$,

$$\pi^T\left(I - \mathbf{P} + \mathbf{1}b^T\right) = \pi^T\mathbf{1}b^T = b^T, \tag{7.39}$$

and therefore, for any vector x such that

$$\left(I - \mathbf{P} + \mathbf{1}b^T\right)x = 0, \tag{7.40}$$

we have

$$b^T x = 0$$

and

$$(I - \mathbf{P})\,x = 0.$$

Therefore, x must be a right eigenvector associated with the eigenvalue $\lambda_1 = 1$, and consequently, x is a multiple of $\mathbf{1}$. But this is compatible with $b^T x = 0$ and $b^T\mathbf{1} \neq 0$ only if $x = 0$. Therefore (7.40) implies $x = 0$, which implies that $\left(I - \mathbf{P} + \mathbf{1}b^T\right)$ is invertible; and (7.39) proves (7.38). □

7.4.2 Travel Times

For any square matrix B, let $d(B)$ is the diagonal matrix which has the same diagonal as B. In particular $d(\Pi)^{-1}$ is the diagonal matrix for which the (i,i)-th entry is $\pi(i)^{-1}$. Note also that $\mathbf{1}\mathbf{1}^T$ is the matrix with all entries equal to 1.

The quantity $m_{ij} := E_i[T_j]$ is the travel time from i to j, and $M := \{m_{ij}\}_{1\le i,j\le r}$ is the travel time matrix

Theorem 7.4.5 *The travel time matrix M of an ergodic* HMC *is given by the formula*

$$M = (I - \mathbf{Z} + \mathbf{1}\mathbf{1}^T d(\mathbf{Z}))d(\Pi)^{-1}\,. \tag{7.41}$$

Proof. We first observe that M has finite entries. Indeed, we already know that $m_{ii} = E_i[T_i] = 1/\pi(i)$ and that $\pi(i) > 0$. Also, when $i \neq j$, m_{ij} is the mean time to absorption in the modified chain where j is made absorbing, and this average time is finite.

By first-step analysis,

$$m_{ij} = 1 + \sum_{k;k\neq j} p_{ik} m_{kj}\,,$$

that is,

$$M = \mathbf{P}(M - d(M)) + \mathbf{1}\mathbf{1}^T\,. \tag{7.42}$$

We now prove that there is but one finite solution of the above equation in the unknown M. To do this, we first show that for any solution M, $d(M)$ is necessarily equal to $d(\Pi)^{-1}$. (We know this to be true when M is the mutual distance matrix, but not yet for a general solution of (†).) Indeed, premultiplying (7.42) by π^T yields

$$\pi^T M = \pi^T \mathbf{P}(M - d(M)) + (\pi^T\mathbf{1})\mathbf{1}^T = \pi^T(M - d(M)) + \mathbf{1}^T\,,$$

and therefore $\pi^T d(M) = \mathbf{1}^T$, which implies the announced result.

Now suppose that there exist two finite solutions M_1 and M_2. Since $d(M_1) = d(M_2)$, it follows that

$$M_1 - M_2 = \mathbf{P}(M_1 - M_2).$$

Therefore, any column v of $M_1 - M_2$ is a right-eigenvector of $\mathbf{P}$ corresponding to the eigenvalue 1. We know that the right-eigenspace R_λ and the left-eigenspace L_λ corresponding to any given eigenvalue λ have the same dimension. For $\lambda = 1$, we know that the dimension of L_λ is one. Therefore, R_λ has dimension 1 for $\lambda = 1$. Thus any right-eigenvector is a scalar multiple of 1. Therefore, $M_1 - M_2$ has columns of the type $\alpha\mathbf{1}$ for some α (α may a priori depend on the column). Since $d(M_1) = d(M_2)$, each column contains a zero, and therefore $\alpha = 0$ for all columns, that is, $M_1 - M_2 \equiv 0$.

At this point we have proved that M is the unique finite solution. It remains to show that M defined by (7.41) is a solution. In fact, from (7.41) and $d(M) = d(\Pi)^{-1}$,

$$M - d(\Pi)^{-1} = (-\mathbf{Z} + \mathbf{1}\mathbf{1}^T d(\mathbf{Z}))d(\Pi)^{-1}.$$

Therefore,

$$\begin{aligned}\mathbf{P}(M - d(\Pi)^{-1}) &= (-\mathbf{PZ} + \mathbf{P1}\mathbf{1}^T d(\mathbf{Z}))d(\Pi)^{-1} \\ &= (-\mathbf{PZ} + \mathbf{1}\mathbf{1}^T d(\mathbf{Z}))d(\Pi)^{-1} \\ &= M + (-\mathbf{PZ} - I + \mathbf{Z})d(\Pi)^{-1},\end{aligned}$$

where we have used the identity $\mathbf{P1} = \mathbf{1}$ for the second equality and (7.41) again for the third. Using now (7.33), that is, $I - \mathbf{Z} = \Pi - \mathbf{PZ}$, we see that

$$\mathbf{P}(M - d(\Pi)^{-1}) = M - \Pi d(\Pi)^{-1} = M - \mathbf{1}\mathbf{1}^T,$$

and (7.42) follows, since $d(M) = d(\Pi)^{-1}$. □

Theorem 7.4.6 *Let* $\mathbf{Z}$ *be the fundamental matrix as defined in (7.37). Then for all* $i \neq j$,

$$E_i[T_j] = \frac{z_{jj} - z_{ij}}{\pi(j)}. \tag{7.43}$$

Proof. We shall need two preliminary formulas. First,

$$\mathbf{Z1} = \theta\mathbf{1}, \tag{7.44}$$

where $\theta^{-1} = b^T\mathbf{1}$. Indeed, from the definiton of $\mathbf{Z}$,

$$\mathbf{Z}\left(I - \mathbf{P} + \mathbf{1}b^T\right)\mathbf{1} = \mathbf{1}. \tag{7.45}$$

But $(I - \mathbf{P})\,\mathbf{1} = 0$, and therefore (7.44) follows. We shall also use the formula

$$\mathbf{Z}\,(I - \mathbf{P}) = I - \theta\mathbf{1}b^T, \tag{7.46}$$

which follows from (7.37) and (7.44).

We now proceed to the main part of the proof. Call N the mutual distance matrix M in which the diagonal elements have been replaced by 0's. From (7.42), we obtain

$$(I - \mathbf{P})\,N = \mathbf{1}\mathbf{1}^T - D^{-1},$$

where $D = \text{diag}\ \{\pi(1), \ldots, \pi(n)\}$. Multiplying both sides by $\mathbf{Z}$, and using (7.44), we obtain

$$\mathbf{Z}\,(I - \mathbf{P})\,N = \theta\mathbf{1}\mathbf{1}^T - \mathbf{Z}D^{-1}.$$

Using (7.46),

$$\mathbf{Z}\,(I - \mathbf{P})\,N = N - \theta\mathbf{1}b^T N.$$

Therefore,

$$N = \theta\mathbf{1}\mathbf{1}^T - \mathbf{Z}D^{-1} + \theta\mathbf{1}b^T N.$$

Thus, for all $i, j \in E$,

$$n_{ij} = \theta - \frac{z_{ij}}{\pi(j)} + \theta\left(b^T N\right)_j.$$

For $i = j$, $n_{ij} = \theta - \frac{z_{jj}}{\pi(j)} + \theta\left(b^T N\right)_j = 0$, which gives $\left(b^T N\right)_j$. Finally, for $i \neq j$,

$$n_{ij} = \frac{z_{jj} - z_{ij}}{\pi(j)}.$$

□

EXAMPLE 7.4.7: THE TARGET TIME FORMULA. The quantity

$$E_i[S_\pi] := \sum_j \pi(j) E_i[S_j] \ ,$$

where S_j is the hitting time of j, is the expected time to a state j previously selected at random with the stationary probability π. From formula (7.43), we have

$$E_i[S_\pi] = \sum_j Z_{jj} - 1 \, .$$

Variance of Ergodic Estimates

Let $\{X_n\}_{n\geq 0}$ be an ergodic Markov chain with finite state space $E = \{1, 2, \ldots, r\}$. A function $f : E \to \mathbb{R}$ is represented by a column vector $f = (f(1), \ldots, f(r))^T$. The ergodic theorem tells us that the estimate $\frac{1}{n}\sum_{k=1}^n f(X_n)$ of $\langle f \rangle_\pi := E_\pi[f(X_0)]$ is asymptotically unbiased, in the sense that it converges to $\langle f \rangle_\pi$ as $n \to \infty$.

Theorem 7.4.8 *For $\{X_n\}_{n\geq 0}$ and $f : E \to \mathbb{R}$ as above, and for any initial distribution μ,*

$$\lim_{n\to\infty} \frac{1}{n} V_\mu \left(\sum_{k=1}^n f(X_k) \right) = 2 \langle f, \mathbf{Z} f \rangle_\pi - \langle f, (I + \Pi) f \rangle_\pi \ , \tag{7.47}$$

where the notation V_μ indicates that the variance is computed with respect to P_μ.

The quantity on the right-hand side will be denoted by $v(f, \mathbf{P}, \pi)$.

Proof. We first suppose that $\mu = \pi$, the stationary distribution. Then

$$\begin{aligned} \frac{1}{n} V_\pi \left(\sum_{k=1}^n f(X_k) \right) &= \frac{1}{n} \left\{ \sum_{k=1}^n V_\pi(f(X_k)) + 2 \sum_{\substack{k,j=1 \\ k<j}}^n \text{cov}_\pi(f(X_k), f(X_j)) \right\} \\ &= V_\pi(f(X_0)) + \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} \text{cov}_\pi(f(X_0), f(X_\ell)) \end{aligned}$$

where we have used the fact that when the initial distribution is π, the chain is stationary, and in particular, $\text{cov}_\pi(f(X_k), f(X_j)) = \text{cov}_\pi(f(X_0), f(X_{j-k}))$ for $k < j$. Now,

$$\begin{aligned} V_\pi(f(X_0)) &= E_\pi[f(X_0)^2] - E_\pi[f(X_0)]^2 \\ &= \sum_{i\in E} \pi(i) f(i)^2 - \left(\sum_{i\in E} \pi(i) f(i) \right)^2 = \langle f, f \rangle_\pi - \langle f, \Pi f \rangle_\pi \,. \end{aligned}$$

Also,

$$\begin{aligned} \text{cov}_{\,\pi}(f(X_0), f(X_\ell)) &= E_\pi[f(X_0) f(X_k)] - E_\pi[f(X_0)]^2 \\ &= \sum_{i\in E} \sum_{j\in E} \pi(i) p_{ij}(\ell) f(i) f(j) - E_\pi[f(X_0)]^2 \\ &= \left\langle f, \mathbf{P}^\ell f \right\rangle_\pi - \langle f, \Pi f \rangle_\pi = \langle f, (\mathbf{P}^\ell - \Pi) f \rangle_\pi . \end{aligned}$$

Since $\lim_{n\to\infty} \sum_{\ell=1}^n (\mathbf{P}^\ell - \Pi) = \mathbf{Z} - I$,

$$\lim_{n\to\infty} \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} (\mathbf{P}^\ell - \Pi) = \mathbf{Z} - I.$$

Indeed, by Cesaro's lemma: If $A_n = \sum_{\ell=1}^n \alpha_\ell$ tends to A as $n \to \infty$, then $\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n-1} A_\ell = A$. But $\frac{1}{n} \sum_{\ell=1}^{n-1} A_\ell = \frac{1}{n}(\alpha_1 + (\alpha_1 + \alpha_2) + \cdots + (\alpha_1 + \cdots + \alpha_{n-1})) = \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} \alpha_\ell$. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} V_\pi \left(\sum_{k=1}^n f(X_k) \right) = \langle f, f \rangle_\pi - \langle f, \Pi f \rangle_\pi + 2 \langle f, (\mathbf{Z} - I) f \rangle_\pi \,,$$

which is the announced result (for $\mu = \pi$).

To prove the result in the general case where the initial distribution is arbitrary, it suffices to show that for two chains $\{X_n^{(1)}\}_{n\geq 0}$ and $\{X_n^{(2)}\}_{n\geq 0}$ with transition matrix $\mathbf{P}$, and arbitrary initial distributions μ and ν, respectively, that couple at a time τ such that $E[\tau^2] < \infty$ (this is the case here, see Exercise 16.4.6),

$$\lim_{n\to\infty} \frac{1}{n} V \left(\sum_{k=1}^\infty f(X_k^{(1)}) \right) = \lim_{n\to\infty} \frac{1}{n} V \left(\sum_{k=1}^\infty f(X_n^{(2)}) \right).$$

But with $X_n = X_n^{(1)}$ or $X_n^{(2)}$,

$$\begin{aligned} V \left(\sum_{k=1}^n f(X_k) \right) &= E \left[\left(\sum_{k=1}^n f(X_k) \right)^2 \right] - E \left[\sum_{k=1}^n f(X_k) \right]^2 \\ &= E \left[\left(\sum_{k=1}^{\tau\wedge n} + \sum_{k=\tau+1}^n \right)^2 \right] - \left(E \left[\sum_{k=1}^{\tau\wedge n} \right] + E \left[\sum_{k=\tau+1}^n \right] \right)^2 \\ &= E \left[\left(\sum_{k=1}^{\tau\wedge n} \right)^2 \right] + E \left[\left(\sum_{k=\tau+1}^n \right)^2 + 2E \left[\left(\sum_{k=1}^{\tau\wedge n} \right) \left(\sum_{k=\tau+1}^n \right) \right] \right] \\ &\quad - E \left[\sum_{k=1}^{\tau\wedge n} \right]^2 - E \left[\sum_{k=\tau+1}^n \right]^2 - 2E \left[\sum_{k=1}^{\tau\wedge n} \right] E \left[\sum_{k=\tau+1}^n \right]. \end{aligned}$$

Since $\sum_{k=\tau+1}^{n} f(X_k^{(1)}) = \sum_{k=\tau+1}^{n} f(X_k^{(2)})$, it follows (with obvious shorthand notations) that

$$\frac{1}{n}\left\{V\left(\sum_{k=1}^{n} f(X_k^{(1)})\right) - \frac{1}{n}V\left(\sum_{k=1}^{n} f(X_k^{(2)})\right)\right\} = \frac{1}{n}A_n + \frac{2}{n}B_n - \frac{2}{n}C_n,$$

where

$$\begin{aligned}
A_n &= \left\{E\left[\left(\sum_{k=1}^{\tau\wedge n}(1)\right)^2\right] - E\left[\left(\sum_{k=1}^{\tau\wedge n}(2)\right)^2\right] - E\left[\sum_{k=1}^{\tau\wedge n}(1)\right]^2 + E\left[\sum_{k=1}^{\tau\wedge n}(2)\right]^2\right\},\\
B_n &= \left\{E\left[\left(\sum_{k=\tau+1}^{n}(1,2)\right)\left(\sum_{k=1}^{\tau\wedge n}(1) - \sum_{k=1}^{\tau\wedge n}(2)\right)\right]\right\},\\
C_n &= \left\{E\left[\sum_{k=\tau+1}^{n}(1,2)\right] E\left[\sum_{k=1}^{\tau\wedge n}(1) - \sum_{k=1}^{\tau\wedge n}(2)\right]\right\}.
\end{aligned}$$

Write

$$\frac{2}{n}B_n = 2E\left[\frac{\sum_{k=\tau+1}^{n}(1,2)}{n}\left(\sum_{k=1}^{\tau\wedge n}(1) - \sum_{k=1}^{\tau\wedge n}(2)\right)\right]$$

and observe that the quantity under the expectation converges, as $n \to \infty$, towards $E_\pi[f(X_0)]$ $(\sum_{k=1}^{\tau}(f(X_k^{(1)}) - f(X_k^{(2)})))$ and is for fixed n bounded in absolute value by $2(\sup|f|)\tau$, an integrable random variable. Therefore, by dominated convergence,

$$\lim_{n\to\infty}\frac{2}{n}B_n = 2E_\pi[f(X_0)]E\left[\sum_{k=1}^{\tau}(f(X_k^{(1)}) - f(X_k^{(2)}))\right].$$

A similar argument shows that $\frac{2}{n}C_n$ has the same limit. Therefore, $\lim_{n\to\infty}\frac{2}{n}(B_n - C_n) = 0$. As for A_n, it is bounded by $4(\sup|f|)^2E[\tau^2] < \infty$, and therefore $\lim_{n\to\infty}\frac{1}{n}A_n = 0$. □

We shall now give an expression of the asymptotic variance in terms of the eigenvalues, when $\mathbf{P}$ has r distinct eigenvalues. We have, in view of (6.11),

$$(\mathbf{P}^n - \Pi) = \sum_{i=2}^{r}\lambda_i^n v_i u_i^T,$$

and therefore

$$\mathbf{Z} = I + \sum_{n\geq 1}(\mathbf{P}^n - \Pi) = I + \sum_{i=2}^{r}\frac{\lambda_i}{1-\lambda_i}v_i u_i^T. \tag{7.48}$$

Also, from (15.5),

$$v(f,\mathbf{P},\pi) = V_\pi(f(X_0)) + 2\sum_{i=2}^{r}\frac{\lambda_i}{1-\lambda_i}\langle f, v_i\rangle_\pi (f^T u_i). \tag{7.49}$$

For a reversible pair $(\mathbf{P}, \pi)$, we have $u_i = Dv_i$, and therefore $f^T u_i = \langle f, v_i \rangle_\pi$. Using this observation and (20.5), we obtain from (15.7),

$$v(f, \mathbf{P}, \pi) = \sum_{i=2}^{r} \frac{1+\lambda_i}{1-\lambda_i} |\langle f, v_i \rangle_\pi|^2. \tag{7.50}$$

If one is interested in the speed of convergence to equilibrium, it is the second-largest eigenvalue modulus that is important. If one is interested in simulation, that is, the computation of $E_\pi[f(X_0)]$ as the ergodic mean $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k)$, all eigenvalues play a role if we measure the quality of the ergodic estimator by the asymptotic variance, as the above formulas show.

7.4.3 Hitting Times Formula

The next result extends formula (7.3).

Theorem 7.4.9 *Let $\{X_n\}_{n\geq 0}$ be a positive recurrent* HMC *with state space E and stationary distribution π. Let μ be a probability distribution on E and let $S \in \mathbb{N}$ be a stopping time of this chain such that $E_\mu[S] < \infty$ and $P\mu(X_S \in \cdot) = \mu$. Then for all $j \in E$,*

$$E_\mu \left[\sum_{k=0}^{S-1} 1_{\{X_n = j\}} \right] = E_\mu [S] \, \pi(j). \tag{7.51}$$

Proof. Let x_j denote the left-hand side of (7.51). If the vector x is an invariant measure of the chain, it must be of the form $x_i = c\pi(i)$ with $c = \sum_i x_i$, that is $c = E_\mu[S]$, which gives (7.51). For the proof that x is an invariant measure, write (using the fact that $P_\mu(X_S = k) = P_\mu(X_0 = k)$ for the second equality):

$$\begin{aligned}
x_k &= \sum_{n=0}^{\infty} P_\mu(X_n = k, n < S) \\
&= \sum_{n=0}^{\infty} P_\mu(X_{n+1} = k, n < S) \\
&= \sum_{n=0}^{\infty} \sum_j P_\mu(X_n = j, X_{n+1} = k, n < S) \\
&= \sum_{n=0}^{\infty} \sum_j P_\mu(X_n = j, n < S) p_{jk} \\
&= \sum_j \left(\sum_{n=0}^{\infty} P_\mu(X_n = j, n < S) \right) p_{jk} = \sum_j x_j p_{jk} \, .
\end{aligned}$$

□

In fact, (7.51) remains true when $E_\mu[S] = \infty$ (consider the stopping time $S^{(n)} = S \wedge \tau_n$, where τ_n is the n-return time to i, write (7.51) for $S^{(n)}$, and let $n \to \infty$, to obtain $E_\mu \left[\sum_{k=0}^{S-1} 1_{\{X_k = j\}} \right] = \infty$).

The particular case that is used in the sequel is for $\mu = \delta_i$. Let S be a stopping time and let i be any state such that $P_i(X_S = i) = 1$. Then, for all $j \in E$,

$$E_i\left[\sum_{k=0}^{S-1} 1_{\{X_n=j\}}\right] = E_i[S]\,\pi(j). \tag{7.52}$$

In the special case where $S = T_i$, the return time to i gives formula (7.3). In the next example, additional information is extracted from (7.51).

EXAMPLE 7.4.10: COMMUTE TIME FORMULAS. Let i and j be two distinct states and let S be the first time of return to i after the first visit to j. Then $E_i[S] = E_i[T_j] + E_j[T_i]$ (strong Markov property at T_j). Also,

$$E_i\left[\sum_{n=0}^{S-1} 1_{\{X_n=j\}}\right] = E_i\left[\sum_{n=T_j}^{S-1} 1_{\{X_n=j\}}\right] = E_j\left[\sum_{n=0}^{T_i-1} 1_{\{X_n=j\}}\right],$$

where the last equality is justified by the strong Markov property. Therefore, by (7.52),

$$E_j\left[\sum_{n=0}^{T_i-1} 1_{\{X_n=j\}}\right] = \pi(j)\left(E_i[T_j] + E_j[T_i]\right). \tag{$\star$}$$

Using words, the left-hand side of this equality is

$$E_j[\text{number of visits to } j \text{ before } i]. \tag{7.53}$$

The quantity $E_i[T_j] + E_j[T_i]$ is called the commute time between i and j. It is the average time needed to go from i to j and then return to j, or, in other words the average return time to i with the constraint of visiting j at least once. The quantities $E_i[T_j]$ can be computed, when the state space is finite, via the fundamental matrix, by means of formula (7.43).

Now, the probability that j is not visited between two successive visits of i is $P_i(T_j > T_i)$. Therefore, the number of visits to i (including time 0) before T_j has a geometric distribution with parameter $p = P_i(T_j > T_i)$, and the average number of such visits is $\frac{1}{P_i(T_j<T_i)}$. Therefore, by ($\star$), after exchanging the roles of i and j,

$$P_i(T_j < T_i) = \frac{1}{\pi(i)\left(E_i[T_j] + E_j[T_i]\right)}. \tag{7.54}$$

EXAMPLE 7.4.11: DIAGONAL OF THE FUNDAMENTAL MATRIX. For fixed $m \geq 1$, let

$$S = m + \inf\{k \geq 0\,;\, X_{m+k} = i\}.$$

Then, by (7.53),

$$E_i\left[\sum_{n=0}^{S-1} 1_{X_n=j}\right] = \pi(i)E_i[S],$$

that is,

$$E_i\left[\sum_{n=0}^{m-1} 1_{X_n=j}\right] = \pi(i)\left(m + E_{\nu_m}\left[S_i\right]\right),$$

where S_i is the hitting time of i, and ν_m is the distribution of the chain at time m. Therefore,

$$\sum_{n=0}^{m-1}\left(p_{ii}(n) - \pi(i)\right) = \pi(i)E_{\nu_m}\left[S_i\right].$$

But $\lim_{m\uparrow\infty}\nu_m = \pi$, and therefore $\sum_{n=0}^{\infty}\left(p_{ii}(n) - \pi(i)\right) = \pi(i)E_\pi\left[S_i\right]$, that is,

$$z_{ii} = \pi(i)E_\pi\left[T_i\right]. \tag{7.55}$$

EXAMPLE 7.4.12: PATTERNS IN COIN TOSSING. (Aldous and Fill, 2002) Let $\{Y_n\}_{n\geq 0}$ be an IID sequence of Bernoulli variables with $P\left(Y_1 = 1\right) = P\left(Y_1 = 0\right) = \frac{1}{2}$, and let $\{X_n\}_{n\geq 0}$ be the "snake chain" defined by

$$X_n = \left(Y_n, Y_{n+1}, \ldots, Y_{n+L-1}\right)$$

for some $L \geq 1$. Note that both $\{Y_n\}_{n\geq 0}$ and $\{X_n\}_{n\geq 0}$ are irreducible ergodic chains, with the stationary distribution as initial distribution. Define

$$\tilde{z}_{ij} = \sum_{n=0}^{\infty}\left(p_{ij}(n) - \pi(j)\right) \tag{7.56}$$

$(= z_{ij} - \pi(j))$, where $\mathbf{P}$ is the transition matrix of $\{X_n\}_{n\geq 0}$ and $\pi(j) = \frac{1}{2^L}$ is its stationary distribution.

For $n \geq L$, X_0 and X_n are independent, and therefore $p_{ij}(n) - \pi(j) = 0$, so that only the first L terms of (7.56) are nonzero. For $n < L$, $p_{ij}(n) > 0$ if and only if the pattern $j = (j_0, \ldots, j_{L-1})$ shifted n to the right and the pattern $i = (i_0, \ldots, i_{L-1})$ agree where they overlap (see the figure).

$$\begin{array}{cccc|cccccc|ccc} & & & & j_0 & j_1 & & \ldots & j_{L-2-n} & j_{L-1-n} & j_{L-n} & \ldots & j_{L-1} \\ & & & & \| & \| & & & \| & \| & & & \\ i_0 & i_1 & i_2 & \ldots & i_n & i_{n+1} & & \ldots & i_{L-2} & i_{L-1} & & & \end{array}$$

In this case $p_{ij}(n)$ equals $\frac{1}{2^n}$. Therefore, defining

$$c(i,j) = \sum_{n=0}^{L-1}\frac{1}{2^n}\chi(i,j,n),$$

where $\chi(i,j,n) = 1$ if and only if the situation depicted in the figure above is realized,

$$\tilde{z}_{ij} = c(i,j) - L2^{-L}.$$

In view of the result of the previous example,

$$E_\pi [S_i] = 2^L c(i,i) - L.$$

But remember that X_0 is always distributed as π, and that to generate the first pattern, L coin tosses are necessary. Therefore, $2^L c(i,i)$ is the average number of coin tosses needed to see pattern i for the first time.

To illustrate this, consider the pattern i = HTTTHT. We have $c(i,i) = 68$ (see the figure).

H	T	T	T	H	T	2^{-0}
	H	T	T	T	H	0
		H	T	T	T	0
			H	T	T	0
				H	T	2^{-4}
					H	0

$$\text{c(HTTTHT)} = 2^6 \left(1 + \tfrac{1}{2^4}\right)$$

Books for Further Information

For applications of Foster's theorem to multivariate Markov chains of interest (for instance, to queueing theory) see [Fayolle, Malishev and Menshikov, 1995].

7.5 Exercises

Exercise 7.5.1. An interpretation of invariant measure
A countable number of particles move independently of one another in the countable space E, each according to a Markov chain with the transition matrix $\mathbf{P}$. Let $A_n(i)$ be the number of particles in state $i \in E$ at time $n \geq 0$ and suppose that the random variables $A_0(i)$ $(i \in E)$ are independent Poisson random variables with respective means $\mu(i)$ $(i \in E)$, where $\mu = \{\mu(i)\}_{i \in E}$ is an invariant measure of $\mathbf{P}$. Show that for all $n \geq 1$, the random variables $A_n(i)$ $(i \in E)$ are independent Poisson random variables with respective means $\mu(i)$ $(i \in E)$.

Exercise 7.5.2. Doubly stochastic transition matrix
A stochastic matrix $\mathbf{P}$ on the state space E is called *doubly stochastic* if for all states i, $\sum_{j \in E} p_{ji} = 1$. Suppose in addition that $\mathbf{P}$ is irreducible, and that E is *infinite*.

Find the invariant measure of $\mathbf{P}$.

Show that $\mathbf{P}$ cannot be positive recurrent.

Exercise 7.5.3. RETURN TIME TO THE INITIAL STATE
Let τ be the first return time to the inital state of an irreducible positive recurrent HMC $\{X_n\}_{n\geq 0}$, that is, $\tau = \inf\{n \geq 1; X_n = X_0\}$, with $\tau = +\infty$ if $X_n \neq X_0$ for all $n \geq 1$. Compute the expectation of τ when the initial distribution is the stationary distribution π. Conclude that it is finite if and only if E is finite. When E is infinite, is this in contradiction with positive recurrence?

Exercise 7.5.4. THE LAZY MARKOV CHAIN, TAKE 2
Consider the lazy version of an irreducible positive recurrent HMC as described in Example 6.2.5. Suppose that the original HMC, with transition matrix $\mathbf{P}$, has no self-loops in its transition graph (that is, $p_{ii} = 0$ for all $i \in E$). Compare the expected return times in both chains. Since the lazy chain "takes more time to travel", explain quantitavely the apparent paradox.

Exercise 7.5.5. EXPONENTIAL TAILS OF HITTING TIMES
Let T_A be the hitting time of the set $A \subset E$ of a finite state space irreducible HMC and let $\overline{T}_A := \max_{i\in E} E_i[T_A]$. Prove that for all $n \geq 1$ and initial distribution μ,

$$P_\mu(T_A > n) \leq \left(\frac{\overline{T}_A}{k}\right)^{\lfloor \frac{n}{k} \rfloor} .$$

Exercise 7.5.6. TARGET TIMES
Consider an irreducible positive recurrenc HMC $\{X_n\}_{n\geq 0}$ with stationary distribution π. Let j be a fixed state and consider the quantity

$$h(j) = \sum_{i\in E} E_j[S_i]\pi(i) ,$$

where $S_i := \inf\{n \geq 0 ; X_n = i\}$ is the hitting (entrance) time of i. Prove that this quantity does not depend on $j \in E$.

An interpretation of $h(j)$ is as follows. A target state j is chosen. The initial distribution of the chain being the stationary one, $h(j)$ is the average time required to hit the target. The constant value T_{target} of h is called the target time. The fact that $T_{target} = \sum_{i,j\in E} \pi(i)\pi(j)E_i[S_j]$ justifies the following expression of the target time

$$E_\pi[S_\pi] := T_{target} .$$

Exercise 7.5.7. THE SNAKE CHAIN
Let $\{X_n\}_{n\geq 0}$ be an HMC with state space E and transition matrix $\mathbf{P}$. Let for $L \geq 1$ $Y_n := (X_n, X_{n+1}, \ldots, X_{n+L})$.

(a) The process $\{Y_n\}_{n\geq 0}$ takes its values in $F = E^{L+1}$. Prove that $\{Y_n\}_{n\geq 0}$ is an HMC and give the general entry of its transition matrix. (The chain $\{Y_n\}_{n\geq 0}$ is called the *snake chain* of length $L + 1$ associated with $\{X_n\}_{n\geq 0}$.)

(b) Show that if $\{X_n\}_{n\geq 0}$ is irreducible, then so is $\{Y_n\}_{n\geq 0}$ if we restrict the state space of the latter to be $F = \{(i_0, \ldots, i_L) \in E^{L+1};\ p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{L-1} i_L} > 0\}$. Show that if the original chain is irreducible aperiodic, so is the snake chain.

(c) Show that if $\{X_n\}_{n\geq 0}$ has a stationary distribution π, then $\{Y_n\}_{n\geq 0}$ also has a stationary distribution. Which one?

Exercise 7.5.8. ABBABAA!
A sequence of A's and B's is formed as follows. The first letter is chosen at random, $P(A) = P(B) = \frac{1}{2}$, as is the second letter, independently of the first one. When the first $n \geq 2$ letters have been selected, the $(n+1)$st is chosen, independently of the letters in positions $k \leq n-2$ conditionally on the pair at position $n-1$ and n, as follows:

$$P(A \mid AA) = \frac{1}{2}, P(A \mid AB) = \frac{1}{2}, P(A \mid BA) = \frac{1}{4}, P(A \mid BB) = \frac{1}{4}.$$

What is the proportion of A's and B's in a long chain?

Exercise 7.5.9. Mean time between successive visits of a set
Let $\{X_n\}_{n\geq 0}$ be an irreducible positive recurrent HMC with stationary distribution π. Let A be a subset of the state space E and let $\{\tau(k)\}_{k\geq 1}$ be the sequence of return times to A. Show that

$$\lim_{k\uparrow\infty} \frac{\tau(k)}{k} = \frac{1}{\sum_{i\in A} \pi(i)}.$$

(This extends Formula (7.5)).

Exercise 7.5.10. Fixed-age retirement policy
Let $\{U_n\}_{n\geq 1}$ be a sequence of IID random variables taking their values in $\mathbb{N}_+ = \{1, 2, \ldots, \}$. The random variable U_n is interpreted as the lifetime of some equipment, or "machine", the n-th one, which is replaced by the $(n+1)$st one upon failure. Thus at time 0, machine 1 is put in service until it breaks down at time U_1, whereupon it is immediately replaced by machine 2, which breaks down at time $U_1 + U_2$, and so on. The time to next failure of the current machine at time n is denoted by X_n. More precisely, the process $\{X_n\}_{n\geq 0}$ takes its values in $E = \mathbb{N}$, equals 0 at time $R_k = \sum_{i=1}^{k} U_i$, equals $U_{k+1} - 1$ at time $R_k + 1$, and then decreases of by unit per unit of time until it reaches the value 0 at time R_{k+1}. It is assumed that for all $k \in \mathbb{N}_+$, $P(U_1 > k) > 0$, so that the state space E is $\mathbb{N}$. Then $\{X_n\}_{n\geq 0}$ is an irreducible HMC called the forward recurrence chain. We assume positive recurrence, that is $E[U] < \infty$, where $U = U_1$.

A. Show that the chain is irreducible. Give a necessary and sufficient condition for positive recurrence. Assuming positive recurrence, what is the stationary distribution? A visit of the chain to state 0 corresponds to a breakdown of a machine. What is the empirical frequency of breakdowns?

B. Suppose that the cost of a breakdown is so important that it is better to replace a working machine during its lifetime (breakdown implies costly repairs, whereas

replacement only implies moderate maintenance costs). The fixed-age retirement policy fixes an integer $T \geq 1$ and requires that a machine having reached age T be immediately replaced. What is the empirical frequency of breakdowns (not replacements)?

Chapter 8

Random Walks on Graphs

8.1 Pure Random Walks

8.1.1 The Symmetric Random Walks on $\mathbb{Z}$ and $\mathbb{Z}^3$

A pure random walk is the motion on a graph of a particle that is not allowed to rest and that chooses equiprobably the next move among all possible ones available. This chapter opens with the classical random walks on $\mathbb{Z}$ and $\mathbb{Z}^3$, and the general pure random walk on a graph.

The one-dimensional symmetric random walk on $\mathbb{Z}$ is the simplest example of a pure random walk on a graph. Here, the nodes are the relative integers and the edges are all unordered pairs $(i, i+1)$, $(i \in \mathbb{Z})$. Let $\{X_n\}_{n\geq 0}$ be such a symmetric random walk on $\mathbb{Z}$. Example 7.1.6 showed that it is recurrent. It is in fact null recurrent.

Proof. Let $\tau_1 = T_0, \tau_2, \ldots$ be the successive return times to state 0. Observe that for $n \geq 1$,

$$P_0(X_{2n} = 0) = \sum_{k\geq 1} P_0(\tau_k = 2n),$$

and therefore, for all $z \in \mathbb{C}$ such that $|z| < 1$,

$$\sum_{n\geq 1} P_0(X_{2n} = 0) z^{2n} = \sum_{k\geq 1} \sum_{n\geq 1} P_0(\tau_k = 2n) z^{2n} = \sum_{k\geq 1} E_0[z^{\tau_k}].$$

But $\tau_k = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_k - \tau_{k-1})$ and therefore, in view of Theorem 7.1.3, and since $\tau_1 = T_0$,

$$E_0[z^{\tau_k}] = (E_0[z^{T_0}])^k.$$

In particular,

$$\sum_{n\geq 0} P_0(X_{2n} = 0) z^{2n} = \frac{1}{1 - E_0[z^{T_0}]}$$

(note that the latter sum includes the term for $n = 0$, that is, 1). Direct evaluation of the left-hand side yields

$$\sum_{n\geq 0} \frac{1}{2^{2n}} \frac{(2n)!}{n!n!} z^{2n} = \frac{1}{\sqrt{1 - z^2}}.$$

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_8

Therefore, the generating function of the return time to 0 given $X_0 = 0$ is

$$E_0[z^{T_0}] = 1 - \sqrt{1 - z^2}\,.$$

Its first derivative $\frac{z}{\sqrt{1-z^2}}$ tends to ∞ as $z \to 1$ from below via real values. Therefore, by Abel's theorem (Theorem A.1.3), $E_0[T_0] = \infty$. □

Null recurrence of the symmetric random walk on $\mathbb{Z}$ implies that the time required to reach state 0 from a given state k has infinite mean. This suggests that the probability given the initial state k that T_0 takes large values is large. The following result gives a bound on how large it is. More precisely:

Theorem 8.1.1 *For a symmetric random walk on $\mathbb{Z}$,*

$$P_k(T_0 > r) \le \frac{12|k|}{\sqrt{r}}\,. \tag{8.1}$$

The following result of independent interest, called the reflection principle, will be used in the proof of Theorem 8.1.1.

Theorem 8.1.2 *For all positive integers j, k and n,*

$$P_k(T_0 < n,\, X_n = j) = P_k(X_n = -j)\,,$$

and therefore, summing over $j > 0$,

$$P_k(T_0 < n,\, X_n > 0) = P_k(X_n < 0)\,.$$

Proof. By the strong Markov property, for $m < n$,

$$P_k(T_0 = m,\, X_n = j) = P_k(T_0 = m)P_0(X_{n-m} = j)\,.$$

Since the distribution of X_n is symmetric when the initial position is 0, the right-hand side is

$$P_k(T_0 = m)P_0(X_{n-m} = -j) = P_k(T_0 = m,\, X_n = -j)\,,$$

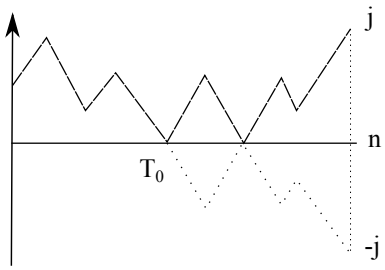
and therefore

$$P_k(T_0 = m,\, X_n = +j) = P_k(T_0 = m,\, X_n = -j)\,.$$

Summation over $m < n$ yields

$$P_k(T_0 < n,\, X_n = j) = P_k(T_0 < n,\, X_n = -j) = P_k(X_n = -j),$$

where, for the last equality, it was observed that starting from a positive position and reaching a negative position at time n implies that position 0 has been reached for the first time strictly before time n. □



Remark 8.1.3 A combinatorial interpretation of the proof is the following. There is a one-to-one correspondence between the paths that hit 0 before time n and reach position $j > 0$ at time n, and the paths that reach position $-j$ at time n. In fact, given a path that hits 0 before time n and reaches position $j > 0$, associate to it the path that is reflected with respect to position 0 after time T_0 (see the figure). This is the reflection principle.

EXAMPLE 8.1.4: THE GAMBLER'S RUIN, TAKE 2. A gambler with initial fortune 1 plays a heads and tails fair coin game with a one dollar stake at each toss. What is the distribution of the duration of the game until he is broke? In other terms, what is the distribution of the return time to 0 of a symmetric random walk starting from position 1? Note that in this case T_0 is necessarily odd. We have by the strong Markov property and the reflection principle (Theorem 15.1.6)

$$\begin{aligned}
P_1(T_0 = 2m+1) &= P_1(T_0 > 2m, X_{2m} = 1, X_{2m+1} = 0) \\
&= P_1(T_0 > 2m, X_{2m} = 1) P_1(X_{2m+1} = 0 \,|\, X_{2m} = 1) \\
&= P_1(T_0 > 2m, X_{2m} = 1) \frac{1}{2} \\
&= \frac{1}{2} \{P_1(X_{2m} = 1) - P_1(T_0 \le 2m, X_{2m} = 1)\} \\
&= \frac{1}{2} \{P_1(X_{2m} = 1) - P_1(X_{2m} = -1)\} \\
&= \frac{1}{2} \left\{ \binom{2m}{m} 2^{-2m} - \binom{2m}{m-1} 2^{-2m} \right\} = \frac{\binom{2m}{m}}{m+1} 2^{-2m-1} .
\end{aligned}$$

The way is now clear for the proof of Theorem 8.1.1.

Proof. It suffices to consider the case $k > 0$, by symmetry. The bound is an immediate consequence of the two following results:

$$P_k(T_0 > r) = P_0(-k < X_r \le +k) \qquad (\star)$$

and

$$P_0(Xr = k) < \frac{3}{\sqrt{r}} . \qquad (\dagger)$$

We start with $(\star)$:

$$P_k(T_0 > r)P_k(X_r > 0, T_0 \leq r) + P_k(X_r > 0, T_0 > r)$$
$$= P_k(X_r > 0, T_0 \leq r) + P_k(T_0 > r) = P_k(X_r < 0) + P_k(T_0 > r)\,,$$

where the last equality is the reflection principle. But by symmetry of the random walk, $P_k(X_r < 0) = P_k(X_r > 2k)$. Therefore

$$P_k(T_0 > r) = P_k(X_r > 0) - P_k(X_r > 2k)$$
$$= P_k(0 < X_r \leq 2k) = P_0(-k < X_r \leq +k)\,.$$

We now turn to the proof of $(\dagger)$. Let $k = 0, 1, \ldots, r$. Starting from state 0, the event $X_{2r} = 2k$ occurs if and only if there are $r + k$ upward moves and $r - k$ downward moves of the random walks. Therefore

$$P(X_{2r} = 2k) = \binom{2r}{r+k} 2^{-2r}\,.$$

The right-hand side is maximized for $k = 0$, and therefore

$$P(X_{2r} = 2k) \leq \binom{2r}{r} 2^{-2r} \leq \sqrt{\frac{8}{\pi}} \frac{1}{\sqrt{2r}}\,,$$

by Stirling's approximation. To obtain a bound for $P(X_{2r+1} = 2k+1)$, condition on the first move of the random walk and use the previous bound to obtain (Exercise 8.4.4)

$$P(X_{2r+1} = 2k + 1) \leq \frac{4}{\sqrt{\pi}} \frac{1}{\sqrt{2r+1}}\,.$$

□

The Symmetric Random Walk on $\mathbb{Z}^3$

(Polya, 1921) The state space of this HMC is $E = \mathbb{Z}^3$. Denoting by e_1, e_2, and e_3 the canonical basis vectors of $\mathbb{R}^3$ (respectively $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$), the non-null terms of the transition matrix of the 3-D symmetric random walk are given by

$$p_{x,x\pm e_i} = \frac{1}{6}\,.$$

The state $0 := (0,0,0)$ (and therefore all states, since the chain is irreducible) is transient.

Proof. Clearly, $p_{00}(2n+1) = 0$ for all $n \geq 0$, and (exercise)

$$p_{00}(2n) = \sum_{0 \leq i+j \leq n} \frac{(2n)!}{(i!j!(n-i-j)!)^2} \left(\frac{1}{6}\right)^{2n}\,.$$

This can be rewritten as

$$p_{00}(2n) = \sum_{0 \leq i+j \leq n} \frac{1}{2^{2n}} \binom{2n}{n} \left(\frac{n!}{i!j!(n-i-j)!}\right)^2 \left(\frac{1}{3}\right)^{2n}\,.$$

Using the trinomial formula

$$\sum_{0\le i+j\le n} \frac{n!}{i!j!(n-i-j)!}\left(\frac{1}{3}\right)^n = 1\,,$$

we obtain the bound

$$p_{00}(2n) \le K_n \frac{1}{2^{2n}}\binom{2n}{n}\left(\frac{1}{3}\right)^n,$$

where

$$K_n = \max_{0\le i+j\le n} \frac{n!}{i!j!(n-i-j)!}.$$

For large values of n, K_n is bounded as follows. Let i_0 and j_0 be the values of i, j that maximize $n!/(i!j!(n+-i-j)!)$ in the domain of interest $0 \le i+j \le n$. From the definition of i_0 and j_0, the quantities

$$\frac{n!}{(i_0-1)!j_0!(n-i_0-j_0+1)!}\,,$$

$$\frac{n!}{(i_0+1)!j_0!(n-i_0-j_0-1)!}\,,$$

$$\frac{n!}{i_0!(j_0-1)!(n-i_0-j_0+1)!}\,,$$

$$\frac{n!}{i_0!(j_0+1)!(n-i_0-j_0-1)!}$$

are bounded by

$$\frac{n!}{i_0!j_0!(n-i_0-j_0)!}.$$

The corresponding inequalities reduce to

$$n-i_0-1 \le 2j_0 \le n-i_0+1 \text{ and } n-j_0-1 \le 2i_0 \le n-j_0+1,$$

and this shows that for large n, $i_0 \sim n/3$ and $j_0 \sim n/3$. Therefore, for large n,

$$p_{00}(2n) \sim \frac{n!}{(n/3)!(n/3)!2^{2n}e^n}\binom{2n}{n}.$$

By Stirling's equivalence formula, the right-hand side of the latter equivalence is in turn equivalent to

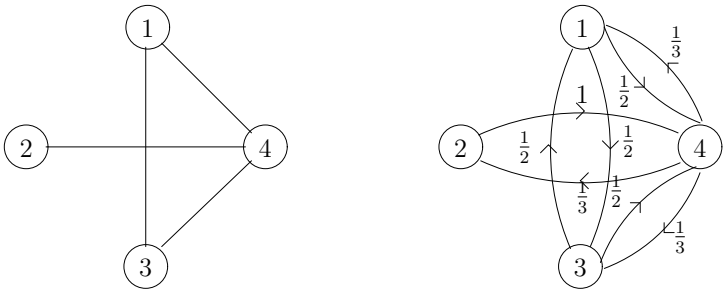
$$\frac{3\sqrt{3}}{2(\pi n)^{3/2}},$$

the general term of a convergent series. State 0 is therefore transient. □

One might wonder at this point about the symmetric random walk on $\mathbb{Z}^2$, which moves at each step northward, southward, eastward and westward equiprobably. It is in fact recurrent (Exercise 8.4.5). Exercise 8.4.6 asks you to prove that the symmetric random walk on $\mathbb{Z}^p$, $p \ge 4$ is transient.

8.1.2 Pure Random Walk on a Graph

Consider a finite non-directed connected graph $G = (V, \mathcal{E})$ where V is the set of vertices, or nodes, and $\mathcal{E}$ is the set of edges. Let d_i be the *index* of vertex i (the number of edges "adjacent" to vertex i). Since there are no isolated nodes (a consequence of the connectedness assumption), $d_i > 0$ for all $i \in V$. Transform this graph into a directed graph by splitting each edge into two directed edges of opposite directions, and make it a transition graph by associating to the directed edge from i to j the transition probability $\frac{1}{d_i}$ (see the figure below). Note that $\sum_{i \in V} d_i = 2|\mathcal{E}|$.



A random walk on a graph

The corresponding HMC with state space $E \equiv V$ is irreducible (G is connected). It therefore admits a unique stationary distribution π, that we attempt to find a stationary distribution via Theorem 6.2.9. Let i and j be connected by an edge, and therefore $p_{ij} = \frac{1}{d_i}$ and $p_{ji} = \frac{1}{d_j}$, so that the detailed balance equation between these two states is

$$\pi(i)\frac{1}{d_i} = \pi(j)\frac{1}{d_j}.$$

This gives $\pi(i) = Kd_i$, where K is obtained by normalization: $K = \left(\sum_{j \in E} d_j\right)^{-1} = (2|\mathcal{E}|)^{-1}$. Therefore

$$\pi(i) = \frac{d_i}{2|\mathcal{E}|}\,.$$

EXAMPLE 8.1.5: RANDOM WALK ON THE HYPERCUBE, TAKE 2. The random walk on the (n-dimensional) hypercube is the random walk on the graph with set of vertices $E = \{0,1\}^n$ and edges between vertices x and y that differ in just one coordivate. For instance, in three dimensions, the only possible motions of a particle performing the random walk on the cube is along its edges in both directions. Clearly, whatever the dimension $n \geq 2$, $d_i = \frac{1}{n}$, and the stationary distribution is the uniform distribution.

The lazy random walk on the graph is, by definition, the Markov chain on V with the transition probabilities $p_{ii} = \frac{1}{2}$ and for $i, j \in V$ such that i and j are connected by an edge of the graph, $p_{i,i} = \frac{1}{2d_i}$. This modified chain admits the same stationary

distribution as the original random walk. The difference is that the lazy version is always aperiodic, whereas the original version may be periodic.

8.1.3 Spanning Trees and Cover Times

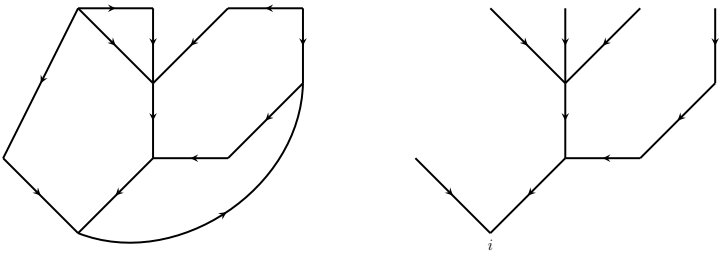
Let $\{X_n\}_{n\in\mathbb{Z}}$ be an irreducible stationary HMC with finite state space E, transition matrix $\mathbf{P}$ and stationary distribution π. Let $G = (E, \mathcal{A})$ be the associated directed graph, where $\mathcal{A}$ is the set of directed edges (arcs), that is, of ordered pairs of states (i, j) such that $p_{ij} > 0$. The weight of an arc (i, j) is p_{ij}. A rooted spanning tree of G is a directed subgraph of G with the following properties:

(i) As an undirected graph it is a connected graph with E as set of vertices.

(ii) As an undirected graph it is without cycles.

(iii) As a directed graph, each of its vertex has out degree 1, except one vertex, the root, that has out degree 0.

Denote by $\mathcal{S}$ the set of spanning trees of G, and by $\mathcal{S}_i$ the subset of $\mathcal{S}$ consisting of rooted spanning trees with vertex $i \in E$. The weight $w(S)$ of a rooted spanning tree of $S \in \mathcal{S}$ is the product of the weights of all the directed edges in S.

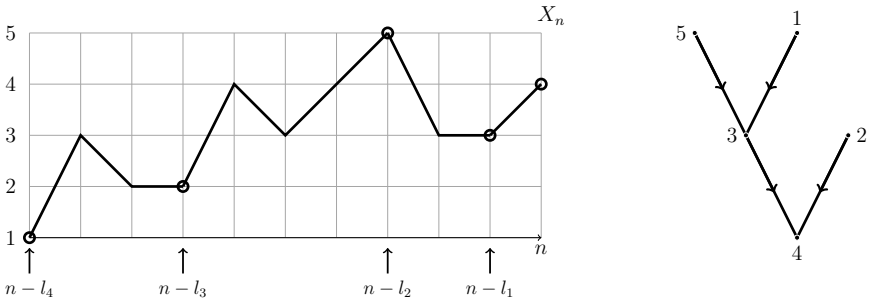


A directed graph and one of its directed spanning tree

Theorem 8.1.6 *The stationary distribution π of $\mathbf{P}$ is given by*

$$\pi(i) = \frac{\sum_{S\in\mathcal{S}_i} w(S)}{\sum_{S\in\mathcal{S}} w(S)}\,. \tag{8.2}$$

Proof. (Anantharam and Tsoucas, 1989) Define a stochastic process $\{Y_n\}_{n\in\mathbb{Z}}$ taking its values in $\mathcal{S}$ as follows. The root of Y_n is X_n, say $X_n = i$. Now, by screening the past values $X_{n-1}, X_{n-2}, \ldots$ in this order, let $X_{n-\ell_1}$ be the first value different from X_n, let $X_{n-\ell_2}$, $\ell_2 > \ell_1$, be the first value different from X_n and $X_{n-\ell_1}$, let $X_{n-\ell_3}$, $\ell_3 > \ell_2$, be the first value different from X_n, $X_{n-\ell_1}$ and $X_{n-\ell_2}$. Continue this procedure until you have exhausted the (finite) state space E. The spanning tree Y_n is the one with directed edges $(X_{n-\ell_1}, X_{n-\ell_1+1} = X_n)$, $(X_{n-\ell_2}, X_{n-\ell_2+1})$, $(X_{n-\ell_3}, X_{n-\ell_3+1}) \ldots$



Since the chain $\{X_n\}_{n\in\mathbb{Z}}$ is stationary, so is the stochastic process $\{Y_n\}_{n\in\mathbb{Z}}$. It is moreover an HMC. We denote by Q_{ST} the transition probability from S to T.

The transition from $Y_n = S \in \mathcal{S}$ with root i to $Y_{n+1} = T \in \mathcal{S}$ with root j in one step is the following:

(a) Add to S the directed (i, j), thus creating a directed spanning graph with a unique directed loop that contains i and j (this may be a self-loop at i).

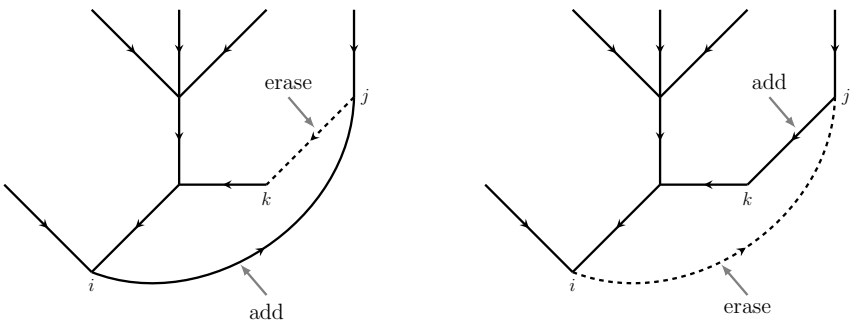
(b) Delete the unique directed edge of S out of j, say (j, k), thus breaking the loop and producing a rooted spanning tree $T \in \mathcal{S}$ with root j.

(c) Such transition occurs with probability p_{ij}.

We now describe the rooted spanning trees S that can lead to the rooted spanning tree T with root j according to the transition matrix Q. T with root j can be obtained from the spanning tree S if and only if S can be constructed from T by the following reverse procedure based on a suitable vertex k:

(α) Add to T the directed edge (j, k), thus creating a directed spanning graph with unique directed loop containing j and k (possibly a self-loop at j).

(β) Delete the unique directed edge (i, j) that lies in the loop, thus breaking the loop and producing a rooted spanning tree $T \in \mathcal{S}$ with root i.



Let k be the unique vertex used in the reverse procedure. Observing that to pass from T to S, we first added the edge (i, j) and then deleted the unique directed

edge (j,k), and that to pass from S to T, we added the directed edge (j,k) and then deleted the edge (j,i). Therefore

$$w(S)Q_{ST} = w(T)R_{TS}$$

where $R_{TS} := p_{jk}$. It follows that

$$\sum_S w(S)Q_{ST} = \sum_S w(T)R_{TS} = w(T)\,.$$

Therefore, the stationary distribution $\{\rho(S)\}_{S\in\mathcal{S}}$ of the chain is

$$\rho(S) = \frac{w(S)}{\sum_{S'} w(S')}\,,$$

and therefore,

$$\pi(i) = \sum_{T\in\mathcal{S}_i} \rho(T) = \frac{\sum_{T\in\mathcal{S}_i} w(T)}{\sum_{T\in\mathcal{S}} w(T)}\,.$$

□

Corollary 8.1.7 *Let $\{X_n\}_{n\in\mathbb{Z}}$ be the stationary random walk on the complete graph built on the finite state space E. (In particular $p_{ij} = \frac{1}{|E|-1}$ for all $j \neq i$ and the stationary distribution is the uniform distribution on E.) Let for all i $(i \in E)$ $\tau_i := \inf\{n \geq 0\,;\, X_n = i\}$. The directed graph with directed edges*

$$(X_{\tau_i}, X_{\tau_i - 1}), \qquad i \neq X_0$$

is uniformly distributed over $\mathcal{S}$.

Proof. Use the proof of Theorem 8.1.6 and the time-reversibility of the random walk. □

The cover time of an HMC is the number of steps it takes to visit all the states. We derive a bound on the maximum (with respect to the initial state) average cover time of the random walk on a graph. For this we shall first observe that the average return time to a given state $i \in E$ is $E_i[T_i] = \frac{1}{\pi(i)} = \frac{2|\mathcal{E}|}{d_i}$. By first-step analysis, denoting by N_i the set of states (vertices) adjacent to i,

$$\frac{2|\mathcal{E}|}{d_i} = E_i[T_i] = \frac{1}{d_i}\sum_{j\in N_i}(1 + E_j[T_i])$$

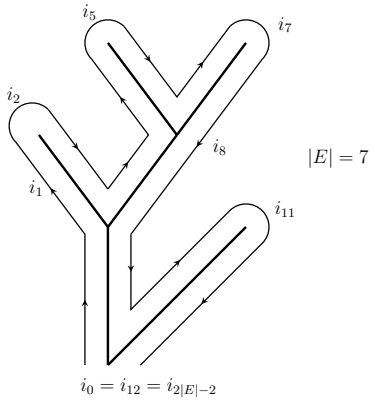
and therefore

$$2|\mathcal{E}| = \sum_{j\in N_i}(1 + E_j[T_i]) \geq 1 + E_j[T_i]\,,$$

from which we obtain the rough bound

$$E_j[T_i] \leq 2|\mathcal{E}| - 1$$

for any pair (i,j) of states. Let now i_0 be an arbitrary state and consider the spanning circuit obtained by a depth-first census of the vertices of the graph (see the figure below), say $i_0, i_1, i_{2|V|-2} = i_0$.



From any vertex i_0 it is possible to traverse the entire spanning tree exactly twice and end up in i_0. Clearly, the average cover time from i_0 is smaller than or equal to the time needed to visit all the vertices of the tree in the order $i_0, i_1, i_{2|V|-2} = i_0$. The (average) time required to go from i_0 to i_1, plus the time needed to go from i_1 to i_0, is less that the return time to i_0, which is in turn bounded by $2|\mathcal{E}| - 1$. The time required to go from i_1 to i_2, plus the time needed to go from i_2 to i_1, is less that the return time to i_1, which is less than $2|\mathcal{E}| - 1$, and so on. Therefore the cover time is bounded by $(|V| - 1) \times (2|\mathcal{E}| - 1) \leq 2|V| \times |\mathcal{E}|$.

EXAMPLE 8.1.8: COVER TIME OF THE CYCLIC RANDOM WALK. The vertices are n points uniformly distributed on the unit circle, and the n edges are those linking the neighbouring vertices. Let c_n denote the cover time for a pure random walk on this n-cycle graph. This average time does not depend on the starting vertex, say 0. Let τ be the first time at which $n - 1$ vertices have been visited. Clearly $E[\tau] = c_{n-1}$. Also at time τ, the position of the random walker is of the form $i - 1$ or $i + 1$, where i is the vertex that has not been visited yet, say $i - 1$. The random walker will visit i either by walking through the vertices $i - 2, i - 3, \ldots$ or by going directly from $i - 1$ to i. He is in the same sitution as the symmetric gambler whose initial fortune is 1 and plays against a gambler whose initial fortune is $n - 1$. The average time before a gambler is broke is $1(n - 1) = n - 1$. Therefore $c_n = c_{n-1} + n - 1$. Since $c_1 = 0$, $c_n = \frac{1}{2}n(n - 1)$. The rough bound above would have given $2n^2$.

EXAMPLE 8.1.9: COVER TIME OF THE RANDOM WALK ON THE COMPLETE GRAPH. The complete graph K_n has n vertices and all possible edges. Therefore the probability of moving in one step from a given edge to another edge is $\frac{1}{n-1}$. Consider now the modified walk with loops. From a given edge the probability of moving to another edge or of staying still is the same: $\frac{1}{n}$. Clearly the cover time in this modified model is greater than in the original model. For the modified model, the cover time is the same as the time to complete the collection of the coupon

collector of n objects. Therefore the cover time of the complete graph random walk is smaller than $(1 + o(1))n \log n$. The rough bound would have given $2n^2(n-1)$.

8.2 Symmetric Walks on a Graph

8.2.1 Reversible Chains as Symmetric Walks

The role of the graph structure is also important in the so-called symmetric walks on a graph which are in fact reversible Markov chains.

Let $G = (V, \mathcal{E})$ be a finite graph, that is, V is a finite collection of vertices, or nodes, and $\mathcal{E}$ is a subset of (unordered) pairs of vertices, denoted by $e = \langle i, j \rangle$. One then notes $i \sim j$ the fact that i and j are the end vertices of edge $\langle i, j \rangle$. This graph is assumed connected. The edge/branch $e = \langle i, j \rangle$ has a positive number $c_e = c_{ij}$ $(= c_{ji})$ attached to it. In preparation for the electrical network analogy, call c_e the conductance of edge e, and call its reciprocal $R_e = \frac{1}{c_e}$ the resistance of e. Denote by C the family $\{c_e\}_{e \in \mathcal{E}}$ and call it the conductance (parameter set) of the network. If $\langle i, j \rangle \notin \mathcal{E}$, let $c_{ij} = 0$ by convention.

Define an HMC on $E := V$ with transition matrix $\mathbf{P}$

$$p_{ij} = \frac{c_{ij}}{C_i},$$

where $C_i = \sum_{j \in V} c_{ij}$. The homogeneous Markov chain introduced in this way is called the random walk on the graph G with conductance C, or the (G, C)-random walk. The state X_n at time n is interpreted as the position on the set of vertices of a particle at time n. When on vertex i the particle chooses to move to an adjacent vertex j with a probability proportional to the conductance of the corresponding edge, that is with probability $p_{ij} = \frac{c_{ij}}{C_i}$. Note that this HMC is irreducible since the graph G is assumed connected and the conductances are positive. Moreover, if $\sum_{j \in V} C_j < \infty$ (for instance if the graph is finite, that is to say, with a finite number of vertices), its stationary probability is

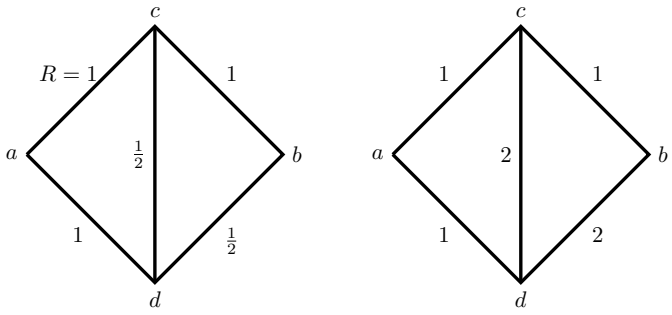
$$\pi(i) = \frac{C_i}{\sum_{j \in V} C_j} \tag{8.3}$$

and moreover, it is reversible. To prove this, it suffices to check the reversibility equations

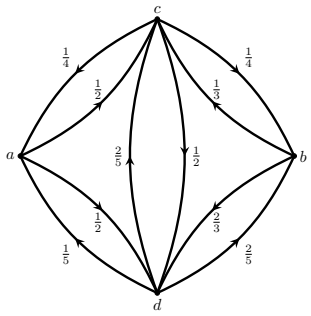
$$\pi(i) \frac{c_{ij}}{C_i} = \pi(j) \frac{c_{ji}}{C_j},$$

using the hypothesis that $c_{ij} = c_{ji} = c_e$.

EXAMPLE 8.2.1: ILLUSTRATION, TAKE 1. (Doyle and Snell, 2000) The figure on the left-hand side describes the network in terms of resistances, whereas the one on the right is in terms of conductances.



The figure below shows the transition graph of the associated reversible HMC



By (8.3), the stationary distribution is $\pi^T = \frac{1}{14}(2,3,4,5)$.

A symmetric random walk on the graph G is a particular (G,C)-random walk for which $c_e \equiv 1$ (or any constant). In this case, at any given time, the particle being in a given site chooses at random the adjacent site where it will move. The corresponding stationary probability then takes the form

$$\pi(i) = \frac{d_i}{2|\mathcal{E}|}$$

where d_i is the degree of node i (the number of nodes to which it is connected) and $|\mathcal{E}|$ is the number of edges.

The connection between random walks and reversible HMC's is in fact both ways. Given a reversible irreducible positive recurrent transition matrix $\mathbf{P} = \{p_{ij}\}_{i,j\in V}$ on V with stationary probability π, we may define the conductance of edge $e = \langle i,j\rangle$ by $c_{ij} = \pi(i)p_{ij}$ ($= c_{ji}$ by reversibility) and define in this way a random walk with the same transition matrix. In particular $C_i = \pi(i)$ and $p_{ij} = \frac{c_{ij}}{C_i}$.

The following result, called the essential edge lemma, is a useful trick for obtaining average passage times (it was already used in Subsection 7.2.1 for birth-and-death processes). Consider a (G,C)-random walk on a connected graph with the following

property. There exist an edge $e = \langle v, x\rangle$ (the essential edge) such that removal of this edge results in two disjoint components. The first one (containing v) has for set of vertices $A(v,x)$ and for set of edges $\mathcal{E}(v,x)$, the second one (containing x) has for set of vertices $A(x,v)$ and for set of edges $\mathcal{E}(x,v)$.

Lemma 8.2.2 *Under the above condition,*

$$E_v[T_x] = \frac{2\sum_{e\in\mathcal{E}(v,x)} c_e}{c_{vx}} + 1\,, \tag{8.4}$$

and

$$E_v[T_x] + E_x[T_v] = \frac{2\sum_e c_e}{c_{vx}}\,. \tag{8.5}$$

Proof. Consider the symmetric random walk with vertices $A(v,x)\cup\{x\}$ and edges $\mathcal{E}(v,x)\cup\{\langle v,x\rangle\}$. For each edge of the modified random walk, the conductance is that of the original random walk. The average time to reach x from v in this restricted graph is obviously equal to that of the original graph. Now, in the restricted random walk, by first-step analysis,

$$E_x[T_x] = 1 + E_v[T_x]$$

and $E_x[T_x]$ is the inverse of the stationary probability of x, that is

$$E_x[T_x] = \frac{2c_{vx} + 2\sum_{e\in\mathcal{E}(v,x)} c_e}{c_{vx}}\,,$$

which gives (8.4). Exchanging the roles of x and v, and combining the two results gives (8.5). □

8.2.2 The Electrical Network Analogy

For finite reversible HMC's, a quantity such as $P_i(T_a < T_b)$ can sometimes be obtained rather simply using an analogy with electrical networks (Kakutani, 1945; Kemeny, Snell and Knapp, 1960). Once the chain is identified, in a way that will be explained, to a network of resistances whose nodes are its states, the above quantity is seen to be the effective resistance between nodes a and b. This effective resistance is then computed by successive reductions of the network to a single branch between these nodes. The theory is then applied to study recurrence in reversible chains with a countable state space.

The setting and notation are those of Subsection 8.2.1. The pair (G,C) will now be interpreted as an electrical network where electricity flows along the edges of the graph (the "branches" of the electrical network). By convention, if $i \not\sim j$, $c_{ij} = 0$. To each directed pair (i,j) there is associated a potential difference Φ_{ij} and a current I_{ij} which are real numbers and satisfy the antisymmetry conditions

$$I_{ji} = -I_{ij} \text{ and } \Phi_{ji} = -\Phi_{ij}$$

for all edges $\langle i, j\rangle$. Two distinct nodes will play a particular role: the source a and the sink b. The currents and the potential differences are linked by the following fundamental laws of electricity:

Kirchoff's potential law: For any sequence of vertices $i_1, i_2, \ldots, i_{n+1}$ such that $i_{n+1} = i_1$ and $i_k \sim i_{k+1}$ for all $1 \le k \le n$,

$$\sum_{\ell=1}^{n} \Phi_{i_\ell, i_{\ell+1}} = 0\,.$$

Kirchoff's current law: For all nodes $i \in V$, $i \neq a, b$,

$$\sum_{j \in V} I_{ij} = 0\,.$$

Ohm's law: For all edges $e = \langle i, j\rangle$

$$I_{ij} = c_e \Phi_{ij}\,.$$

It readily follows from Kirchoff's potential law that there exists a function $\Phi : V \to \mathbb{R}$ determined up to an additive constant such that

$$\Phi_{ij} = \Phi(j) - \Phi(i)\,.$$

Note that, by Ohm's law, the current I_{ij} and the potential difference $\Phi(j) - \Phi(i)$ have the same sign ("currents flow in the direction of increasing potential"). Define $I = \{I_{ij}\}_{i,j \in V}$ to be the current matrix. When the three fundamental laws are satisfied, we say that (Φ, I) is a realization of the electrical network (G, C).

From Kirchoff's current law and Ohm's law, we have that for all $i \neq a, b$,

$$\sum_{i \in V} c_{ij}(\Phi(j) - \Phi(i)) = 0\,,$$

or equivalently

$$\Phi(i) = \sum_{j \in V} \frac{c_{ij}}{C_i} \Phi(j)\,.$$

Therefore,

Theorem 8.2.3 *The potential function Φ is harmonic on $V \backslash \{a, b\}$ with respect to the (G, C)-random walk.*

In particular, by Theorem 17.3.15, it is uniquely determined by its boundary values $\Phi(a)$ and $\Phi(b) = 0$.

Probabilistic Interpretation of Voltage

We shall now interpret a given realization (Φ, I) of the electrical network (G, C) in terms of the associated (G, C)-random walk. We start with the voltage.

Theorem 8.2.4 *Call Φ_1 the solution corresponding to a unit voltage at source a and a null voltage at sink b:*

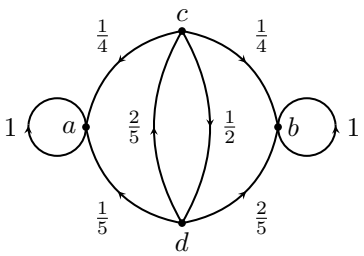
$$\Phi_1(a) = 1\,, \qquad \Phi_1(b) = 0\,.$$

Then, for all $i \in V$,

$$\Phi_1(i) = P_i\,(T_a < T_b)\ .$$

Proof. Using the one-step forward method, one shows that the function h given by $h(i) = P_i\,(T_a < T_b)$ (the probability that starting from i, a is reached before b) is harmonic on $D = V\backslash\{a, b\}$ and that $h(a) = 1$ and $h(b) = 0$. Recall that a function harmonic on $D = V\backslash\{a, b\}$ is uniquely determined by its values on $\{a, b\}$. Therefore, $\Phi_1 \equiv h$. □

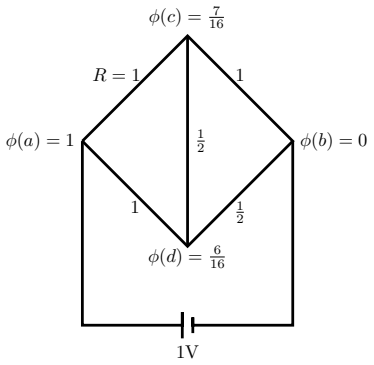
EXAMPLE 8.2.5: ILLUSTRATION, TAKE 2. Modify the HMC of the running example so as to make states (nodes) a and b absorbing. For $i \in \{a, b, c, d\}$, $h(i)$ defined above is the probability that, starting from i, this HMC is absorbed in a. (Compare with the gambler's ruin problem.)



The recurrence equations for h are

$$\begin{aligned} h(a) &= 1 \\ h(b) &= 0 \\ h(c) &= \frac{1}{4} + \frac{1}{2}h(d) \\ h(d) &= \frac{1}{5} + \frac{2}{5}h(c)\,. \end{aligned}$$

The solution is represented in the figure below, where $\Phi_1 = h$ is the voltage map corresponding to a 1 Volt battery.



Probabilistic Interpretation of Current

We now interpret the current. A particle performs the (G, C)-random walk starting from a, except that now it is supposed to leave the network once it has reached b. We show that the current I_{ij} from i to j is proportional to the expected number of passages of this particle from i to j minus the expected number of passages in the opposite direction, from j to i.

Proof. Let $u(i)$ be the expected number of visits to node i before it reaches b and leaves the network. Clearly $u(b) = 0$. Also for $i \neq a, b$, $u(i) = \sum_{j \in V} u(j) p_{ji}$. But $C_i p_{ij} = C_j p_{ji}$ so that $u(i) = \sum_{j \in V} u(j) p_{ij} \frac{C_i}{C_j}$ and finally

$$\frac{u(i)}{C_i} = \sum_{j \in V} p_{ij} \frac{u(j)}{C_j} \,.$$

Therefore the function Φ given by

$$\Phi(i) = \frac{u(i)}{C_i}$$

is harmonic on $D = V \backslash \{a, b\}$. It is the unique such function whose values at a and at b are specified by

$$\Phi(a) = \frac{u(a)}{C_a} \,, \qquad \Phi(b) = 0 \,. \tag{$\star$}$$

With such a voltage function,

$$\begin{aligned} I_{ij} &= (\Phi(i) - \Phi(j)) c_{ij} \\ &= \left(\frac{u(i)}{C_i} - \frac{u(j)}{C_j} \right) c_{ij} \\ &= u(i) \frac{c_{ij}}{C_i} - u(j) \frac{c_{ji}}{C_j} = u(i) p_{ij} - u(j) p_{ji} \,. \end{aligned}$$

But $u(i)p_{ij}$ is the expected number of crossings from i to j and $u(j)p_{ji}$ is the expected number of crossings in the opposite direction. □

Under voltage Φ determined by $(\star)$,

$$I_a := \sum_{j \in V} I_{aj} = 1$$

because, in view of the probabilistic interpretation of current in this case, the sum is equal to the expected value of the difference between the number of times the particle leaves a and the number of times it enters a, that is 1 (each time the particle enters a it leaves it immediately, except for the one time when it leaves a forever to be eventually absorbed in b).

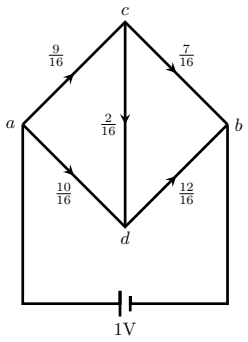
Similarly, let $I_{1,a}$ be the current out of a when the unit voltage is applied to a ($\Phi_1(a) = 1$). Since multiplication of the voltage by a factor implies multiplication of the current by the same factor, we have that

$$\frac{\Phi(a)}{I_a} = \frac{\Phi_1(a)}{I_{1,a}},$$

that is,

$$\Phi(a) = \frac{1}{I_{1,a}}. \tag{8.6}$$

EXAMPLE 8.2.6: ILLUSTRATION, TAKE 3. The figure below gives the currents, as given from the voltages by Ohm's law. The current out of a is $I_{1,a} = I_{1,ac} + I_{1,ad} = \frac{9}{16} + \frac{10}{16} = \frac{19}{16}$, by Kirchoff's law.



8.3 Effective Resistance and Escape Probability

8.3.1 Computation of the Effective Resistance

The effective resistance between a and b is defined by

$$R_{eff}(a \leftrightarrow b) = \frac{\Phi(a)}{I_a}. \tag{8.7}$$

As we saw before, this quantity does not depend on the value $\Phi(a)$. When $\Phi(a) = \Phi_1(a) = 1$, the effective conductance equals the current $I_{1,a}$ flowing out of a. But in this case

$$\begin{aligned} I_{1,a} &= \sum_{j\in V}(\Phi_1(a) - \Phi_1(j))c_{aj} = \sum_{j\in V}(1 - \Phi_1(j))c_{aj} \\ &= C_a\left(1 - \sum_{j\in V}\Phi_1(j)\frac{c_{aj}}{C_a}\right) = C_a\left(1 - \sum_{j\in V}p_{aj}\Phi_1(j)\right). \end{aligned}$$

But the quantity $\left(1 - \sum_{j\in V}p_{aj}\Phi_1(j)\right)$ is the "escape probability"

$$P_{esc} := P_a(T_b < T_a)\,,$$

that is, the probability that the particle starting from a reaches b before returning to a. Therefore

$$P_{esc} = \frac{1}{C_a R_{eff}(a \leftrightarrow b)}\,.$$

EXAMPLE 8.3.1: ILLUSTRATION, TAKE 4. The effective resistance is

$$R_{eff}(a \leftrightarrow b) = \frac{1}{I_{1,a}} = \frac{1}{\frac{19}{16}} = \frac{16}{19}\,.$$

In particular, the probability — starting from a— of returning to a before hitting b is $P_{esc} = \frac{1}{C_a R_{eff}(a\leftrightarrow b)} = \frac{1}{2\times\frac{16}{19}} = \frac{19}{32}$.

EXAMPLE 8.3.2: COMMUTE TIME AND EFFECTIVE RESISTANCE. Recalling formula (7.54):

$$P_a\left(T_b < T_a\right) = \frac{1}{\pi(a)\left(E_a\left[T_b\right] + E_b\left[T_a\right]\right)}$$

and the expression

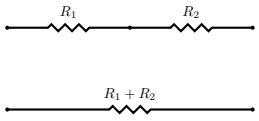
$$\pi(a) = \frac{C_a}{C}\,,$$

we obtain the following formula:

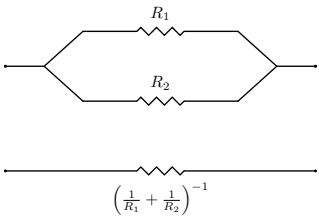
$$E_a[T_b] + E_b[T_a] = CR_{eff}(a \leftrightarrow b)\,. \tag{8.8}$$

(The left-hand side is the commute time between a and b.)

In order to compute the effective resistance, we have at our disposition the procedure used in the simplification of resistance networks, such as the following two basic rules.



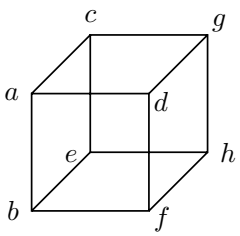
Series configuration



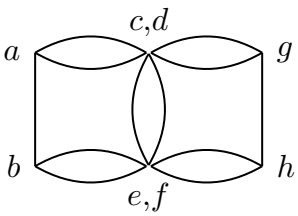
Parallel configuration

We can also merge nodes with the same voltage.

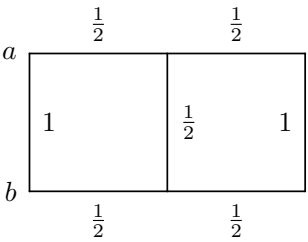
EXAMPLE 8.3.3: THE CUBIC NETWORK. All the resistances in the network below are unit resistance.



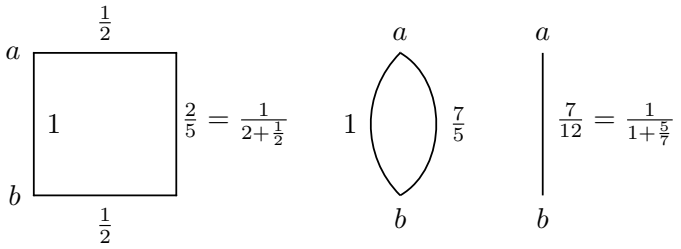
By symmetry, the nodes c and d have the same voltage, and can therefore be merged. Similarly for the nodes e and f.



One can then use the rule for resistances in parallel to further simplify the network:

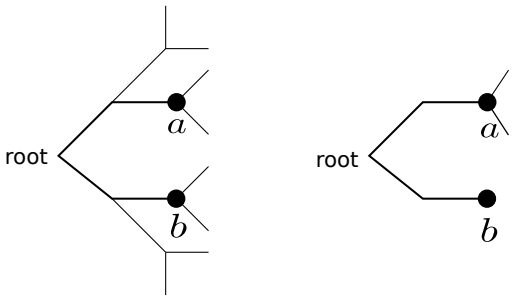


Alternating the series and parallel simplifications, we have:



Therefore the effective resistance between a and b is $R_{eff}(a \leftrightarrow b) = \frac{7}{12}$.

EXAMPLE 8.3.4: THE BINARY TREE. Consider the pure random walk on the full binary tree of depth k. Select two nodes a and b. Let $\mathcal{P}(a \leftrightarrow b)$ be the shortest path linking a and b. In view of computing $P_a(T_b < T_a)$, we make the preliminary observation that this quantity does not change if one cuts all edges that are not in $\mathcal{P}(a \leftrightarrow b)$ and have an endpoint in $\mathcal{P}(a \leftrightarrow b) \backslash \{a\}$. We are therefore left with the graph $\mathcal{P}(a \leftrightarrow b)$ plus, when a is not a leaf of the tree, the edges leading to a that do not belong to $\mathcal{P}(a \leftrightarrow b)$. Therefore $R_{eff}(a \leftrightarrow b) = d(a, b)$, the graph distance between a and b, and therefore $P_a(T_b < T_a) = \frac{1}{3d(a,b)}$ if a is not a leaf, $= \frac{1}{d(a,b)}$ if a is a leaf.



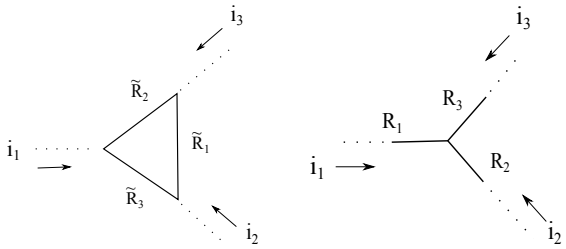
Another basic rule of reduction of electrical networks is the star-triangle transformation (Exercise 8.4.11). It states that the two following electrical network configurations are equivalent if and only if for $i = 1, 2, 3$,

$$R_i \tilde{R}_i = \delta$$

where

$$\delta = R_1 R_2 R_3 \left(R_1^{-1} + R_2^{-1} + R_3^{-1} \right) = \frac{\tilde{R}_1 \tilde{R}_2 \tilde{R}_3}{\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3} .$$

("Equivalence" means that if one network supports the currents i_1, i_2 and i_3 entering the triangle at nodes 1, 2 and 3 respectively, so does the other network.)



See Exercise 8.4.17 for an example of application.

8.3.2 Thompson's and Rayleigh's Principles

By definition, a flow on the graph G with source a and sink b is a collection of real numbers $J = \{J_{ij}\}_{i,j\in V}$, such that

(a) $J_{ij} = -J_{ji}$,

(b) $J_{ij} = 0$ if $i \not\sim j$,

(c) $\sum_{j\in V} J_{ij} = 0$ for all $i \neq a, b$.

Denote by $J_i = \sum_{j\in V} J_{ij}$ the flow out of i. A unit flow J is one for which $J_a = 1$. In general,

$$J_a = -J_b\,.$$

Indeed, since $J_i = 0$ for all $i \neq a, b$,

$$\begin{aligned} J_a + J_b &= \sum_{i\in V} J_i \\ &= \sum_{i,j\in V} J_{ij} = \frac{1}{2}\sum_{i,j\in V} (J_{ij} + J_{ji}) = 0\,. \end{aligned}$$

Also, for any function $w : V \to \mathbb{R}$,

$$(w(a) - w(b))J_a = \frac{1}{2}\sum_{i,j\in V} (w(j) - w(i))J_{ij}\,. \tag{8.9}$$

Indeed, from the properties of flows,

$$\begin{aligned} \sum_{i,j\in V} (w(i) - w(j))J_{ij} &= \sum_{i,j\in V} w(i)J_{ij} - \sum_{i,j\in V} w(j)J_{ij} \\ &= \sum_{i,j\in V} w(i)J_{ij} + \sum_{i,j\in V} w(j)J_{ji} \\ &= \sum_{i\in V} w(i)J_i + \sum_{j\in V} w(j)J_j \\ &= w(a)J_a + w(b)J(b) + w(a)J_a + w(b)J(b) \\ &= w(a)J_a - w(b)J(a) + w(a)J_a - w(b)J(a) = 2(w(a) - w(b))J_a\,. \end{aligned}$$

The energy dissipated in the network by the flow J is by definition the quantity

$$E(J) := \frac{1}{2} \sum_{i,j \in V} J_{ij}^2 R_{ij}.$$

This is a meaningful electrical quantity for the special case where the flow is a current I corresponding to a potential Φ, in which case, by Ohm's law:

$$E(I) = \frac{1}{2} \sum_{i,j \in V} I_{ij}^2 R_{ij} = \frac{1}{2} \sum_{i,j \in V} I_{ij}(\Phi(j) - \Phi(i)).$$

Theorem 8.3.5 *The effective resistance between the source a and the sink b is equal to the energy dissipated in the network when the current I_a out of a is the unit current.*

Proof. By (8.9),

$$E(I) = (\Phi(a) - \Phi(b))I_a = \Phi(a)I_a\,,$$

and by definition (8.7) of the effective resistance $R_{eff}(a \leftrightarrow b)$ between a and b,

$$E(I) = I_a^2 R_{eff}(a \leftrightarrow b)\,.$$

□

The following result is known as Thomson's principle.

Theorem 8.3.6 *The energy dissipation $E(J)$ is minimized among all unit flows J by the unit current flow I.*

Proof. Let J be a unit flow from a to b and let I be a unit current flow from a to b. Define $D = J - I$. This is a flow from a to b with $D_a = 0$. We have that

$$\begin{aligned}
\sum_{i,j \in V} J_{ij}^2 R_{ij} &= \sum_{i,j \in V} (I_{ij} + D_{ij})^2 R_{ij} \\
&= \sum_{i,j \in V} I_{ij}^2 R_{ij} + 2 \sum_{i,j \in V} I_{ij} D_{ij} R_{ij} + \sum_{i,j \in V} D_{ij}^2 R_{ij} \\
&= \sum_{i,j \in V} I_{ij}^2 R_{ij} + 2 \sum_{i,j \in V} (\Phi(j) - \Phi(i)) D_{ij} + \sum_{i,j \in V} D_{ij}^2 R_{ij}.
\end{aligned}$$

From (8.9) with $w = \Phi$ and $J = D$, the middle term equals $4(\Phi(a) - \Phi(b))D_a = 0$, so that

$$\sum_{i,j \in V} J_{ij}^2 R_{ij} = \sum_{i,j \in V} I_{ij}^2 R_{ij} + \sum_{i,j \in V} D_{ij}^2 R_{ij} \geq \sum_{i,j \in V} I_{ij}^2 R_{ij}\,.$$

□

We now state and prove Rayleigh's principle.

Theorem 8.3.7 *The effective resistance between two points can only increase as any resistance in the circuit increases.*

Proof. Change the resistances R_{ij} to $\overline{R}_{ij} \geq R_{ij}$ and let I and $\overline{I}$ be the corresponding unit current flows. Then

$$\overline{R}_{eff} = \frac{1}{2} \sum_{i,j \in V} \overline{I}_{ij}^2 \overline{R}_{ij} \geq \frac{1}{2} \sum_{i,j \in V} \overline{I}_{ij}^2 R_{ij} \,.$$

But by Thomson's principle,

$$\frac{1}{2} \sum_{i,j \in V} \overline{I}_{ij}^2 R_{ij} \geq \frac{1}{2} \sum_{i,j \in V} I_{ij}^2 R_{ij} = R_{eff}(a \leftrightarrow b) \,.$$

□

EXAMPLE 8.3.8: SHORTING AND CUTTING. Shorting consists in making some resistances null and therefore decreases the effective resistance. On the contrary, cutting (an edge), which consists in making the corresponding resistance infinite, increases the effective resistance.

8.3.3 Infinite Networks

Consider a (G, C)-random walk where now $G = (V, \mathcal{E})$ is an infinite connected graph with finite-degree vertices. Since the graph is infinite, this HMC may be transient. This subsection gives a method that is sometimes useful in assessing the recurrence or transience of this random walk. Note that once recurrence is proved, we have an invariant measure x, namely $x_i = C_i$ (a finite quantity since each vertex has finite degree). Positive recurrence is then granted if and only if $\sum_{i \in V} C_i < \infty$. (The latter condition alone guarantees the existence of an invariant measure, but remember that existence of an invariant measure does not imply recurrence.)

Some arbitrary vertex will be distinguished, henceforth called 0. Recall that the graph distance $d(i, j)$ between two vertices is the smallest number of edges to be crossed when going from i to j. For $N \geq 0$, let

$$K_N = \{i \in V;\, d(0, i) \leq N\}$$

and

$$\partial K_N = K_N - K_{N-1} = \{i \in V;\, d(0, i) = N\} \,.$$

Let G_N be the restriction of G to K_N. A graph $\overline{G}_N$ is obtained from G_N by merging the vertices of ∂K_N into a single vertex named b_N. Let $R_{eff}(N) := R_{eff}(0 \leftrightarrow b_N)$ be the effective resistance between 0 and b_N of the network $\overline{G}_N$. Since $\overline{G}_N$ is obtained from $\overline{G}_{N+1}$ by merging the vertices of $\partial K_N \cup \{b_{N+1}\}$, $R_{eff}(N) \leq R_{eff}(N+1)$. In particular the limit

$$R_{eff}(0 \leftrightarrow \infty) := \lim_{N \uparrow \infty} R_{eff}(N)$$

exists. It may be finite or infinite.

Theorem 8.3.9 *The probability of return to* 0 *of the* (G, C)*-random walk is*

$$P_0(X_n = 0 \text{ for some } n \geq 1) = 1 - \frac{1}{C_0 R_{eff}(0 \leftrightarrow \infty)}.$$

In particular this chain is recurrent if and only if $R_{eff}(0 \leftrightarrow \infty) = \infty$.

Proof. The function h_N defined by

$$h_N(i) := P(X_n \text{ hits } K_N \text{ before } 0)$$

is harmonic on $V_N \backslash \{\{0\} \cup K_N\}$ with boundary conditions $h_N(0) = 0$ and $h_N(i) = 1$ for all $i \in K_N$. Therefore, the function g_N defined by

$$g_N(i) = h_N(i) \text{ on } K_{N-1} \cup \{b_N\}$$

and $g_N(b_N) = 1$ is a potential function for the network $\overline{G}_N$ with source 0 and sink b_N. Therefore

$$\begin{aligned} P_0\left(X_n \text{ returns to 0 before reaching } \partial K_N\right) &= 1 - \sum_{j \sim 0} p_{0j} g_N(j) \\ &= 1 - \sum_{j \sim 0} \frac{c_{0j}}{C_0}(g_N(j) - g_N(0)). \end{aligned}$$

By Ohm's law, $\sum_{j \sim 0} c_{0j}(g_N(j) - g_N(0))$ is the total current $I_N(0)$ out of 0, and therefefore since the potential difference between b_N and 0 is 1, $I_N(0) = \frac{1}{R_{eff}(N)}$. Therefore

$$P_0\left(X_n \text{ returns to 0 before reaching } \partial K_N\right) = 1 - \frac{1}{C_0 R_{eff}(N)}$$

and the result follows since, by the sequential continuity property of probability,

$$P(X_n = 0 \text{ for some } n \geq 1) = \lim_{N \uparrow \infty} P_0\left(X_n \text{ returns to 0 before reaching } \partial K_N\right).$$

□

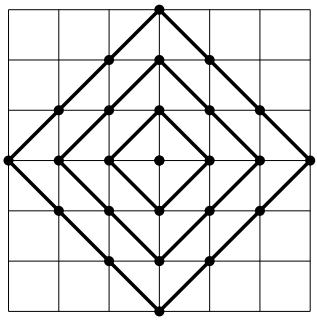
Theorem 8.3.10 *Consider two sets of conductances* C *and* $\overline{C}$ *on the same connected graph* $G = (V, \mathcal{E})$ *such that for each edge* e*,*

$$u c_e \leq \overline{c}_e \leq v c_e$$

for some constants u *and* v*,* $0 < u \leq v < \infty$*. Then the random walks* (G, C) *and* $(G, \overline{C})$ *are of the same type (either both recurrent, or both transient).*

Proof. Let C^u be the set of conductances on G defined by $c_e^u = uc_e$, and define similarly the set of conductances C^v. Observe that the random walks (G, C^u), (G, C^v) and (G, C) are the same. The rest of the proof follows from Rayleigh's monotonicity law, because (G, C) and $(G, \overline{C})$ then have effective resistances that are both finite or both infinite. □

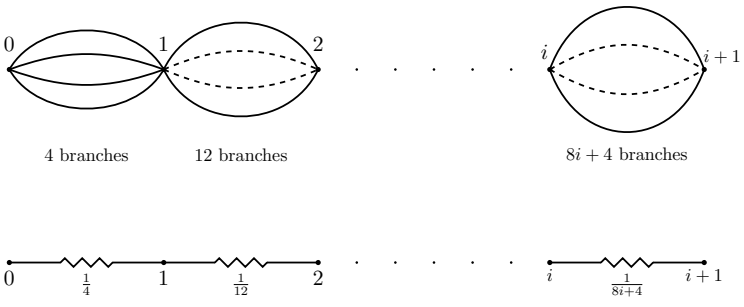
EXAMPLE 8.3.11: THE SYMMETRIC RANDOM WALK ON $\mathbb{Z}^2$. The symmetric random walk on $\mathbb{Z}^2$ corresponds in the electrical network analogy to the infinite grid where all resistances are unit resistances. The grid is actually infinite and in the figure below only "levels" up to the third one are shown. (Level 0 is the center node, level 1 consists of the 4 nodes at distance 1 of level 0, and more generally, level $i+1$ consists of the $8i+4$ nodes at distance 1 from level i.)



By Rayleigh's monotonicity law, if one shorts a set of nodes (that is, if the resistances directly linking pairs of nodes in this set are set to 0, in which case the nodes thereof have the same potential), the effective resistance between two nodes is decreased.

By shorting successively the nodes of each $i \geq 1$, we obtain for the effective resistance between node 0 and level $i+1$ in the shorted network (see the figure below),

$$\overline{R}_{eff}(i+1) = \sum_{n=0}^{i} \frac{1}{8i+4}\,.$$



$\overline{R}_{eff}(i+1)$ is, by Rayleigh's monotonicity principle, smaller than the actual effective resistance in the full grid between node 0 and any node at level $i+1$. Therefore, since $\lim_{N\uparrow\infty} \overline{R}_{eff}(N) = \infty$, the two-dimensional symmetric random walk on $\mathbb{Z}^2$ is recurrent.

Books for Further Information

[Doyle and Snell, 2000], [Kemeny, Snell and Knapp, 1960] and [Lyons and Peres, 2016]. The first reference is a pedagogical introduction, whereas the third one is more theoretical (and has more advanced material). [Klenke, 2008, 2014] has a full chapter on the electrical analogy.

8.4 Exercises

Exercise 8.4.1. PASSAGE TIMES FOR BIRTH-AND-DEATH PROCESSES
Consider the symmetric random walk on the graph $G = (V, \mathcal{E})$, where $V = \{0, \dots, N-1\}$ and $\mathcal{E} = \{\langle i-1, i\rangle ; 1 \le i \le N-1\}$. Call w_i the conductance of edge $\langle i-1, i\rangle$. Define $w := \sum_{i=1}^{N-1} w_i$. Let $a, b, c \in V$ be such that $a < b < c$. Then:

(α) $P_b(T_c < T_a) = \frac{\sum_{i=a+1}^{b} w_i^{-1}}{\sum_{i=a+1}^{c} w_i^{-1}}$,

(β) $E_b[T_c] = c - b \sum_{j=b+1}^{c} \sum_{i-1}^{j-1} w_i w_j^{-1}$,

(γ) $E_b[T_c] + E_c[T_b] = w \sum_{i=b+1}^{c} w_i^{-1}$.

Exercise 8.4.2. ON THE CIRCLE
Consider the random walk on the circle. More precisely, there are n points labeled $0, 1, 2, \dots, n-1$ orderly and equidistantly placed on a circle. A particle moves from one point to an adjacent point in the manner of a random walk on $\mathbb{Z}$. This gives rise to an HMC with the transition probabilities $p_{i,i+1} = p \in (0,1)$, $p_{i,i-1} = 1-p$, where, by the "modulo convention", $p_{0,-1} := p_{0,n-1}$ and $p_{n-1,n} := p_{n-1,0}$. Compute the average time it takes to go back to 0 when initially at 0.

Exercise 8.4.3. STREAKS OF 1'S IN A WINDOW OF FAIR COIN TOSSES
Let $\{U_n\}_{n\in\mathbb{Z}}$ be an IID sequence of equiprobable 0's and 1. Define $X_n \in \{0, 1, \dots, N\}$ by $X_n = 0$ if $U_n = 0$ and

$$X_n = k \text{ if } U_n = 1, U_{n-1} = 1, \dots, U_{n-k+1} = 1, U_{n-k} = 0\,.$$

In words, we look at the window of length N just observed at the n-th toss of a sequence of fair coin tosses, and set $X_n = k$ if the length of the last streak of 1's is k. For instance, with $N = 5$ and

$$(U_{-4}, U_{-3}, \dots, U_5) = (0110110111)$$

we have $X_0 = 1$ (the first window of size 5 is 01101 and the rightmost streak of 1's has length 1), $X_1 = 2$ (the next window of size 5 is 11011 and the rightmost

streak of 1's has length 2), $X_2 = 0$ (the next window of size 5 is 1010110 and the rightmost streak of 1's has length 0), $X_3 = 1$ (the next window of size 5 is 0101101 and the rightmost streak of 1's has length 1). The next sliding windows are 11011 and 10111 give respectively $X_3 = 2$ and $X_4 = 3$.

(a) Give the transition matrix of this HMC and its stationary distribution π.

(b) Assuming the chain stationary, give the transition matrix $\tilde{\mathbf{P}}$ of the time reversed HMC.

(c) Show that, whatever the initial state, the distribution of the reversed chain is already the stationary distribution at the N-th step.

Exercise 8.4.4.
Refer to Theorem 8.1.1. Prove that

$$P(X_{2r+1} = 2k+1) \leq \frac{4}{\sqrt{\pi}} \frac{1}{\sqrt{2r+1}}.$$

Exercise 8.4.5. NULL RECURRENCE OF THE 2-DIMENSIONAL SYMMETRIC RANDOM WALK
Show that the 2-D symmetric random walk on $\mathbb{Z}^2$ is null recurrent.

Exercise 8.4.6. TRANSIENCE OF THE 4-D SYMMETRIC RANDOM WALK
Show that the projection of the 4-D symmetric random walk on $\mathbb{Z}^3$ is a lazy symmetric random walk on $\mathbb{Z}^3$. Deduce from this that the 4-D symmetric random walk is transient. More generally, show that the symmetric random walk on $\mathbb{Z}^p$, $p \geq 5$, is transient.

Exercise 8.4.7. THE LINEAR WALK
Consider the pure random walk on the linear graph with vertices $0, 1, 2, \ldots, n$ and edges $\langle i, i+1 \rangle$ $(0 \leq i \leq n-1)$. Compute the cover time. Compare to the rough bound.

Exercise 8.4.8. THE KNIGHT RETURNS HOME
A knight moves randomly on a chessboard, making each admissible move with equal probability, and starting from a corner. What is the average time he takes to return to the corner he started from?

Exercise 8.4.9. EHRENFEST
Apply formula (8.2) to the Ehrenfest HMC.

Exercise 8.4.10. ROOTED TREES OF A GIVEN SIZE
What would you do to generate a random rooted tree with the uniform distribution on the set of rooted trees with k given vertices?

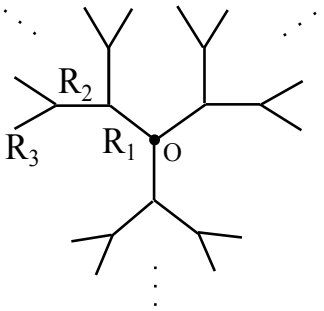
Exercise 8.4.11. THE STAR-TRIANGLE EQUIVALENCE
Show that the electrical network configurations in the figure just before Example 8.4.17 are equivalent if and only if for $i = 1, 2, 3$, $R_i \tilde{R}_i = \delta$, where

$$\delta = R_1 R_2 R_3 \left(R_1^{-1} + R_2^{-1} + R_3^{-1} \right) = \frac{\tilde{R}_1 \tilde{R}_2 \tilde{R}_3}{\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3} .$$

(Equivalence: if one network supports the currents i_1, i_2 and i_3 entering the triangle at nodes 1, 2 and 3 respectively, so does the other network.)

Exercise 8.4.12. THE URN OF EHRENFEST AS AN ELECTRICAL NETWORK
Describe the Ehrenfest HMC with $N = 2M$ particles in stationary state in terms of electrical networks. Let state M be the source and state N the sink. Compute the voltage $\Phi_1(i)$ at any state $i \in \{0, 1, \cdots, N\}$ constrained by $\Phi_1(M) = 1$ and $\Phi_1(N) = 0$. Compute $P(T_M < T_N)$.

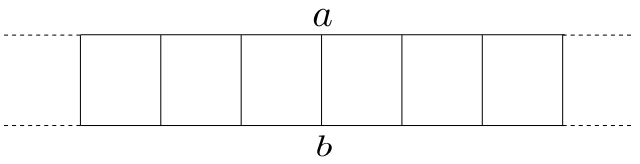
Exercise 8.4.13. THE SPHERICAL SYMMETRIC TREE
Consider the full spherical tree of degree three (see the figure) and define Γ_N to be the set of nodes at distance N from the root, called 0. Consider the symmetric random walk on the symmetric full spherical tree, where all the edges from Γ_{i-1} to Γ_i have the same resistance $R_i > 0$.



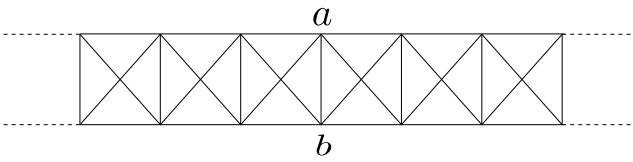
Show that a necessary and sufficient condition of recurrence of the corresponding symmetric random walk is

$$\sum_{i=1}^{\infty} \frac{R_i}{|\Gamma_i|} = \infty .$$

Exercise 8.4.14. THE LADDER
Find the effective resistance between a and b of the following infinite network of unit resistances.



Exercise 8.4.15.
Find the effective resistance between a and b of the following infinite network of unit resistances.



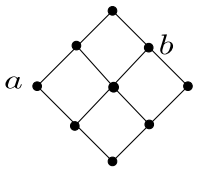
Exercise 8.4.16. $G_b(a) = C_a R_{eff}(a \leftrightarrow b)$
Consider a (G, C)-random walk, and let a and b be two distinct vertices of G. let T_b be the first return time to b. Define

$$G_b(a) := E_a \left[\sum_{n=0}^{\infty} 1_{\{X_n = a\}} 1_{\{n < T_b\}} \right]$$

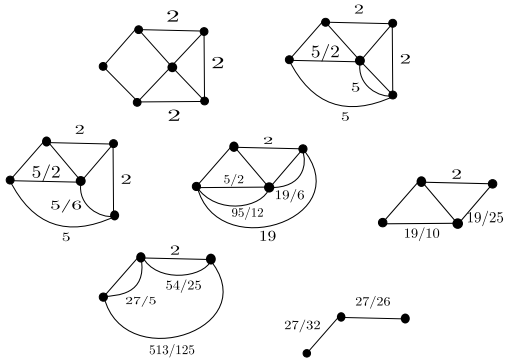
(the average number of visits to a before b is hit, given that the initial state is a). Show that

$$G_b(a) = C_a R_{eff}(a \leftrightarrow b)\,.$$

Exercise 8.4.17. Reducing the four-square network
(from [Klenke, 2014]) The goal is to find the effective resistance of the following electrical network between nodes a and b. (A resistance not appearing on a branch is conventionally taken equal to 1.)



The successive reduction operations are recorded in the sequence of figures below.



Finally the effective resistance between a and b is $\frac{27}{32} + \frac{27}{26} = \frac{629}{416}$.

(1) Give the details.

(2) What would be the result if we append to the right-most node a "square" formed by 4 unit resistances?

Exercise 8.4.18. SYMMETRIC WALK ON THE BINARY TREE
Consider the full binary tree whose root is denoted by 0. Show that

$$R_{eff}(0 \leftrightarrow \infty) = \frac{1}{2}(R_{eff}(0 \leftrightarrow \infty) + 1)$$

and

$$R_{eff}(N+1) = \frac{1}{2}(R_{eff}(N) + 1)\,.$$

Deduce from this that the symmetric random walk on the full binary tree is transient.

Chapter 9

Markov Fields on Graphs

9.1 Gibbs–Markov Equivalence

9.1.1 Local Characteristics

Markov fields are also called Gibbs fields in honour of the founder of Statistical Mechanics (Gibbs, 1902). Although they were historically of special interest to physicists, they have recently found applications in other areas, in particular in image processing.

Let $G = (V, \mathcal{E})$ be a finite graph, and let $v_1 \sim v_2$ denote the fact that $\langle v_1, v_2 \rangle$ is an edge of the graph. Such vertices are also called neighbours (one of the other). We shall also refer to vertices of V as sites. The boundary with respect to $\sim$ of a set $A \subset V$ is the set

$$\partial A := \{v \in V \backslash A\,;\, v \sim w \text{ for some } w \in A\}\,.$$

Let Λ be a finite set, the phase space. A random field on V with phases in Λ is a collection $X = \{X(v)\}_{v \in V}$ of random variables with values in Λ. A random field can be regarded as a random variable taking its values in the configuration space $E := \Lambda^V$. A configuration $x \in \Lambda^V$ is of the form $x = (x(v), v \in V)$, where $x(v) \in \Lambda$ for all $v \in V$. For a given configuration x and a given subset $A \subseteq V$, let

$$x(A) := (x(v), v \in A),$$

the restriction of x to A. If $V \backslash A$ denotes the complement of A in V, one writes $x = (x(A), x(V \backslash A))$. In particular, for fixed $v \in V$, $x = (x(v), x(V \backslash v))$, where $V \backslash v$ is a shorter way of writing $V \backslash \{v\}$.

Of special interest are the random fields characterized by local interactions. This leads to the notion of a Markov random field. The "locality" is in terms of the neighbourhood structure inherited from the graph structure. More precisely, for any $v \in V$, $N_v := \{w \in V; w \sim v\}$ is the neighborhood of v. In the following, $\widetilde{\mathcal{N}}_v$ denotes the set $\mathcal{N}_v \cup \{v\}$.

Definition 9.1.1 *The random field X is called a Markov random field* (MRF) *with respect to $\sim$ if for all sites $v \in V$, the random elements $X(v)$ and $X(V \backslash \widetilde{\mathcal{N}}_v)$ are independent given $X(\mathcal{N}_v)$.*

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_9

In mathematical symbols:

$$P(X(v) = x(v) \mid X(V\backslash v) = x(V\backslash v)) = P(X(v) = x(v) \mid X(\mathcal{N}_v) = x(\mathcal{N}_v)) \quad (9.1)$$

for all $x \in \Lambda^V$ and all $v \in V$. Property (9.1) is clearly of the Markov type: the distribution of the phase at a given site is directly influenced only by the phases of the neighboring sites.

Remark 9.1.2 Note that any random field is Markovian with respect to the trivial topology, where the neighborhood of any site is $V\backslash v$. However, the interesting Markov fields (from the point of view of modeling, simulation, and optimization) are those with relatively small neighborhoods.

Definition 9.1.3 *The local characteristic of the* MRF *at site v is the function $\pi^v : \Lambda^V \to [0,1]$ defined by*

$$\pi^v(x) := P(X(v) = x(v) \mid X(\mathcal{N}_v) = x(\mathcal{N}_v)).$$

The family $\{\pi^v\}_{v\in V}$ is called the local specification of the MRF.

One sometimes writes

$$\pi^v(x) := \pi(x(v) \mid x(\mathcal{N}_v))$$

in order to stress the role of the neighborhoods.

Theorem 9.1.4 *Two positive distributions of a random field with a finite configuration space Λ^V that have the same local specification are identical.*

Proof. Enumerate V as $\{1, 2, \ldots, K\}$. Therefore a configuration $x \in \Lambda^V$ is represented as $x = (x_1, \ldots, x_{K-1}, x_K)$ where $x_i \in \Lambda$, $1 \le i \le K$. The following identity

$$\pi(z_1, z_2, \ldots, z_k) = \prod_{i=1}^{K} \frac{\pi(z_i \mid z_1, \ldots, z_{i-1}, y_{i+1}, \ldots, y_K)}{\pi(y_i \mid z_1, \ldots, z_{i-1}, y_{i+1}, \ldots, y_K)} \pi(y_1, y_2, \ldots, y_k) \qquad (\star)$$

holds for any $z, y \in \Lambda^K$. For the proof, write

$$\pi(z) = \prod_{i=1}^{K} \frac{\pi(z_1, \ldots, z_{i-1}, z_i, y_{i+1}, \ldots, y_K)}{\pi(z_1, \ldots, z_{i-1}, y_i, y_{i+1}, \ldots, y_K)} \pi(y)$$

and use the Bayes rule to obtain for each i, $1 \le i \le K$:

$$\frac{\pi(z_1, \ldots, z_{i-1}, z_i, y_{i+1}, \ldots, y_K)}{\pi(z_1, \ldots, z_{i-1}, y_i, y_{i+1}, \ldots, y_K)} = \frac{\pi(z_i \mid z_1, \ldots, z_{i-1}, y_{i+1}, \ldots, y_K)}{\pi(y_i \mid z_1, \ldots, z_{i-1}, y_{i+1}, \ldots, y_K)}.$$

Let now π and π' be two positive probability distributions on V with the same local specification. Choose any $y \in \Lambda^V$. Identity $(\star)$ shows that for all $z \in \Lambda^V$,

$$\frac{\pi'(z)}{\pi(z)} = \frac{\pi'(y)}{\pi(y)}.$$

Therefore $\frac{\pi'(z)}{\pi(z)}$ is a constant, necessarily equal to 1 since π and π' are probability distributions. □

9.1.2 Gibbs Distributions

Consider the probability distribution

$$\pi_T(x) = \frac{1}{Z_T} e^{-\frac{1}{T}U(x)} \tag{9.2}$$

on the configuration space Λ^V, where $T > 0$ is the temperature, $U(x)$ is the energy of configuration x, and Z_T is the normalizing constant, called the partition function. Since $\pi_T(x)$ takes its values in $[0,1]$, necessarily $-\infty < U(x) \leq +\infty$. Note that $U(x) < +\infty$ if and only if $\pi_T(x) > 0$. One of the challenges associated with Gibbs models is obtaining explicit formulas for averages, considering that it is generally hard to compute the partition function. This is feasible in exceptional cases (see Exercise 9.4.4).

Such distributions are of interest to physicists when the energy is expressed in terms of a potential function describing the local interactions. The notion of a clique then plays a central role.

Definition 9.1.5 *Any singleton $\{v\} \subset V$ is a clique. A subset $C \subseteq V$ with more than one element is called a clique (with respect to $\sim$) if and only if any two distinct sites of C are mutual neighbors. A clique C is called maximal if for any site $v \notin C$, $C \cup \{v\}$ is not a clique.*

The collection of cliques will be denoted by $\mathcal{C}$.

Definition 9.1.6 *A Gibbs potential on Λ^V relative to $\sim$ is a collection $\{V_C\}_{C \subseteq V}$ of functions $V_C : \Lambda^V \to \mathbb{R} \cup \{+\infty\}$ such that*

(i) $V_C \equiv 0$ if C is not a clique, and

(ii) for all $x, x' \in \Lambda^V$ and all $C \subseteq V$,

$$x(C) = x'(C) \Rightarrow V_C(x) = V_C(x') \,.$$

The energy function U is said to derive from the potential $\{V_C\}_{C \subseteq V}$ if

$$U(x) = \sum_C V_C(x) \,.$$

The function V_C depends only on the phases at the sites inside subset C. One could write more explicitly $V_C(x(C))$ instead of $V_C(x)$, but this notation will not be used.

In this context, the distribution in (9.2) is called a Gibbs distribution (with respect to $\sim$).

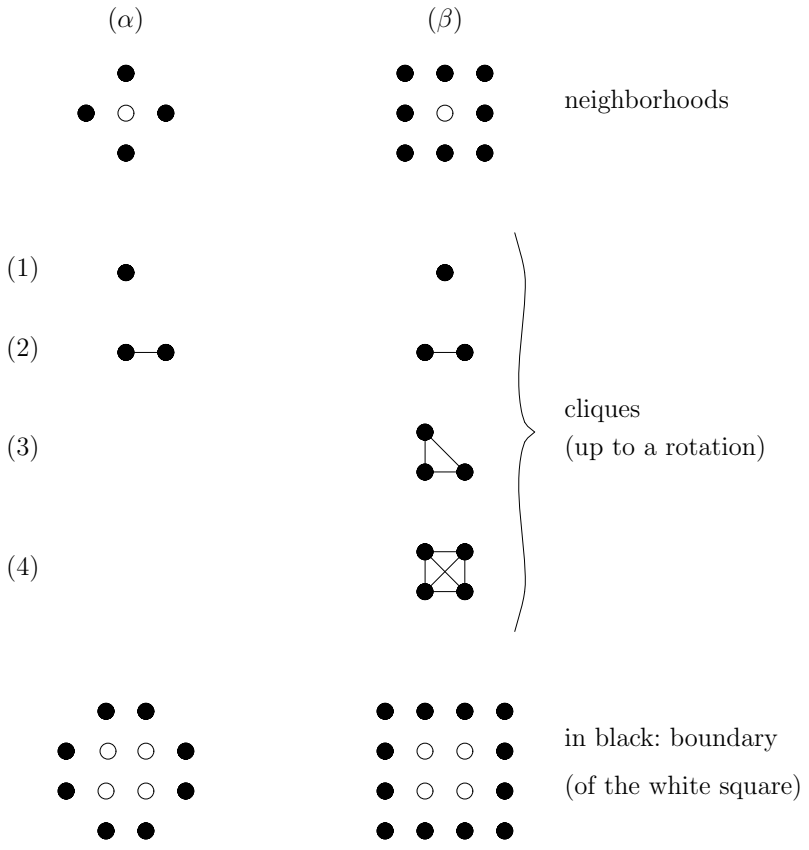
Example 9.1.7: Ising Model, take 1. (Ising, 1925) In statistical physics, the following model is regarded as a qualitatively correct idealization of a piece of ferromagnetic material. Here $V = \mathbb{Z}_m^2 = \{(i,j) \in \mathbb{Z}^2, i,j \in [1,m]\}$ and $\Lambda =$

$\{+1, -1\}$, where ± 1 is the orientation of the magnetic spin at a given site. The figure below depicts two particular neighborhood systems, their respective cliques, and the boundary of a 2×2 square for both cases. The neighborhood system in the original Ising model is as in column (α) of the figure below, and the Gibbs potential is

$$\begin{aligned} V_{\{v\}}(x) &= -\frac{H}{k} x(v), \\ V_{\langle v,w\rangle}(x) &= -\frac{J}{k} x(v)x(w), \end{aligned}$$

where $\langle v, w\rangle$ is the 2-element clique ($v \sim w$). For physicists, k is the Boltzmann constant, H is the external magnetic field, and J is the internal energy of an elementary magnetic dipole. The energy function corresponding to this potential is therefore

$$U(x) = -\frac{J}{k} \sum_{\langle v,w\rangle} x(v)x(w) - \frac{H}{k} \sum_{v \in V} x(v)\,.$$



Two examples of neighborhoods, cliques, and boundaries

The Hammersley–Clifford Theorem

Gibbs distributions with an energy deriving from a Gibbs potential relative to a neighborhood system are distributions of Markov fields relative to the same neighborhood system.

Theorem 9.1.8 *If X is a random field with a distribution π of the form $\pi(x) = \frac{1}{Z}e^{-U(x)}$, where the energy function U derives from a Gibbs potential $\{V_C\}_{C\subseteq V}$ relative to $\sim$, then X is a Markov random field with respect to $\sim$. Moreover, its local specification is given by the formula*

$$\pi^v(x) = \frac{e^{-\sum_{C\ni v} V_C(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{C\ni v} V_C(\lambda, x(V\setminus v))}}, \tag{9.3}$$

where the notation $\sum_{C\ni v}$ means that the sum extends over the sets C that contain the site v.

Proof. First observe that the right-hand side of (9.3) depends on x only through $x(v)$ and $x(\mathcal{N}_v)$. Indeed, $V_C(x)$ depends only on $(x(w), w \in C)$, and for a clique C, if $w \in C$ and $v \in C$, then either $w = v$ or $w \sim v$. Therefore, if it can be shown that $P(X(v) = x(v)|X(V\setminus v) = x(V\setminus v))$ equals the right-hand side of (9.3), then (see Exercise 2.4.21) the Markov property will be proved. By definition of conditional probability,

$$P(X(v) = x(v) \mid X(V\setminus v) = x(V\setminus v)) = \frac{\pi(x)}{\sum_{\lambda\in\Lambda}\pi(\lambda, x(V\setminus v))}\,. \tag{$\dagger$}$$

But

$$\pi(x) = \frac{1}{Z}e^{-\sum_{C\ni v} V_C(x) - \sum_{C\not\ni v} V_C(x)},$$

and similarly,

$$\pi(\lambda, x(V\setminus v)) = \frac{1}{Z}e^{-\sum_{C\ni v} V_C(\lambda, x(V\setminus v)) - \sum_{C\not\ni v} V_C(\lambda, x(V\setminus v))}.$$

If C is a clique and v is not in C, then $V_C(\lambda, x(V\setminus v)) = V_C(x)$ and is therefore independent of $\lambda \in \Lambda$. Therefore, after factoring out $\exp\left\{-\sum_{C\not\ni v} V_C(x)\right\}$, the righ-thand side of ($\dagger$) is found to be equal to the right-hand side of (9.3). □

The local energy at site v of configuration x is

$$U_v(x) = \sum_{C\ni v} V_C(x).$$

With this notation, (9.3) becomes

$$\pi^v(x) = \frac{e^{-U_v(x)}}{\sum_{\lambda\in\Lambda} e^{-U_v(\lambda, x(V\setminus v))}}.$$

EXAMPLE 9.1.9: ISING MODEL, TAKE 2. The local characteristics in the Ising model are

$$\pi_T^v(x) = \frac{e^{\frac{1}{kT}\left\{J\sum_{w;w\sim v} x(w)+H\right\}x(v)}}{e^{+\frac{1}{kT}\left\{J\sum_{w;w\sim v} x(w)+H\right\}} + e^{-\frac{1}{kT}\left\{J\sum_{w;w\sim v} x(w)+H\right\}}}.$$

Theorem 9.1.8 above is the direct part of the Gibbs–Markov equivalence theorem: A Gibbs distribution relative to a neighborhood system is the distribution of a Markov field with respect to the same neighborhood system. The converse part (Hammersley–Clifford theorem) is important from a theoretical point of view, since together with the direct part it concludes that Gibbs distributions and MRF's are essentially the same objects.

Theorem 9.1.10 *(Hammersley and Clifford, 1968) Let $\pi > 0$ be the distribution of a Markov random field with respect to $\sim$. Then*

$$\pi(x) = \frac{1}{Z}e^{-U(x)}$$

for some energy function U deriving from a Gibbs potential $\{V_C\}_{C\subseteq V}$ with respect to $\sim$.

Proof. The proof is based on the Möbius formula.

Lemma 9.1.11 *Let Φ and Ψ be two set functions defined on $\mathcal{P}(V)$, the collection of subsets of the finite set V. The two statements below are equivalent:*

$$\Phi(A) = \sum_{B\subseteq A}(-1)^{|A-B|}\Psi(B), \text{ for all } A \subseteq V, \tag{9.4}$$

$$\Psi(A) = \sum_{B\subseteq A}\Phi(B), \text{ for all } A \subseteq V, \tag{9.5}$$

where $|C|$ is the number of elements of the set C.

Proof. We first show that (9.4) implies (9.5). Write the right-hand side of (9.5) using (9.4):

$$\sum_{B\subseteq A}\Phi(B) = \sum_{B\subseteq A}\sum_{D\subseteq B}(-1)^{|B-D|}\Psi(D) = \sum_{D\subseteq A}\left(\sum_{C\subseteq A-D}(-1)^{|C|}\right)\Psi(D)\,.$$

But if $A - D = \varnothing$,

$$\sum_{C\subseteq A-D}(-1)^{|C|} = (-1)^{|\varnothing|} = (-1)^0 = 1,$$

whereas if $A - D \neq \varnothing$,

$$\begin{aligned}\sum_{C\subseteq A-D}(-1)^{|C|} &= \sum_{k=0}^{|A-D|}(-1)^k \text{ card } \{C; |C| = k, C \subseteq A - D\} \\ &= \sum_{k=0}^{|A-D|}(-1)^k \binom{|A-D|}{k} = (1-1)^{|A-D|} = 0,\end{aligned}$$

and therefore

$$\sum_{D\subseteq A}\Psi(D)\sum_{C\subseteq A-D}(-1)^{|C|} = \Psi(A).$$

We now show that (9.5) implies (9.4). Write the right-hand side of (9.4) using (9.5):

$$\begin{aligned}\sum_{B\subseteq A}(-1)^{|A-B|}\Psi(B) &= \sum_{B\subseteq A}(-1)^{|A-B|}\left(\sum_{D\subseteq B}\Phi(D)\right) \\ &= \sum_{D\subseteq B\subseteq A}(-1)^{|A-B|}\Phi(D) = \sum_{D\subseteq A}\Phi(D)\sum_{C\subseteq A-D}(-1)^{|C|}.\end{aligned}$$

By the same argument as above, the last quantity equals $\Phi(A)$. □

We now prove Theorem 9.1.10. Let 0 be a fixed element of the phase space Λ. Also, let 0 denote the configuration with all phases equal to 0. (The context will prevent confusion between $0 \in \Lambda$ and $0 \in \Lambda^V$.) Let x be a configuration, and let A be a subset of V. Let the symbol x^A represent a configuration of Λ^V coinciding with x on A, and with phase 0 outside A.

Define for $A \subseteq V, x \in \Lambda^V$,

$$V_A(x) := \sum_{B\subseteq A}(-1)^{|A-B|}\log\frac{\pi(0)}{\pi(x^B)}. \tag{9.6}$$

From the Möbius formula,

$$\log\frac{\pi(0)}{\pi(x^A)} = \sum_{B\subseteq A}V_B(x)\,,$$

and therefore, with $A = V$:

$$\pi(x) = \pi(0)e^{-\sum_{A\subseteq V}V_A(x)}\,.$$

It remains to show (a) that V_A depends only on the phases on A, and (b) that $V_A \equiv 0$ if A is not a clique with respect to $\sim$.

If $x, y \in \Lambda^V$ are such that $x(A) = y(A)$, then for any $B \subseteq A$, $x^B = y^B$, and therefore, by (9.6), $V_A(x) = V_A(y)$. This proves (a).

With t an arbitrary site in A, write (9.6) as follows:

$$\begin{aligned} V_A(x) &= \left\{ \sum_{B\subseteq A, B\not\ni t} + \sum_{B\subseteq A, B\ni t} \right\} (-1)^{|A-B|} \log \frac{\pi(0)}{\pi(x^B)} \\ &= \sum_{B\subseteq A\setminus t} (-1)^{|A-B|} \left\{ \log \frac{\pi(0)}{\pi(x^B)} - \log \frac{\pi(0)}{\pi(x^{B\cup t})} \right\} . \end{aligned}$$

That is,

$$V_A(x) = \sum_{B\subseteq A\setminus t} (-1)^{|A-B|} \log \frac{\pi(x^{B\cup t})}{\pi(x^B)} . \tag{9.7}$$

Now, if t is not in $B \subseteq A$,

$$\frac{\pi(x^{B\cup t})}{\pi(x^B)} = \frac{\pi^t(x^{B\cup t})}{\pi^t(x^B)} ,$$

and therefore

$$V_A(x) = \sum_{B\subseteq A\setminus t} (-1)^{|A-B|} \log \frac{\pi^t(x^{B\cup t})}{\pi^t(x^B)},$$

and, by the same calculations that led to (9.7),

$$V_A(x) = -\sum_{B\subseteq A} (-1)^{|A-B|} \log \pi^t(x^B) . \tag{9.8}$$

Recall that $t \in A$, and therefore, if A is not a clique, one can find $s \in A$ such that s is not a neighbor of t. Fix such an s, and split the sum in (9.8) as follows:

$$\begin{aligned} V_A(x) = &- \sum_{B\subseteq A\setminus\{s,t\}} (-1)^{|A-B|} \log \pi^t(x^B) - \sum_{B\subseteq A\setminus\{s,t\}} (-1)^{|A-(B\cup t)|} \log \pi^t(x^{B\cup t}) \\ &- \sum_{B\subseteq A\setminus\{s,t\}} (-1)^{|A-(B\cup s)|} \log \pi^t(x^{B\cup s}) - \sum_{B\subseteq A\setminus\{s,t\}} (-1)^{|A-(B\cup\{s,t\})|} \log \pi^t(x^{B\cup\{s,t\}}) \\ = &- \sum_{B\subseteq A\setminus\{s,t\}} (-1)^{|A-B|} \log \frac{\pi^t(x^B)\pi^t(x^{B\cup\{s,t\}})}{\pi^t(x^{B\cup s})\pi^t(x^{B\cup t})} . \end{aligned}$$

But since $s \neq t$ and $s \nsim t$, we have $\pi^t(x^B) = \pi^t(x^{B\cup s})$ and $\pi^t(x^{B\cup t}) = \pi^t(x^{B\cup\{s,t\}})$, and therefore $V_A(x) = 0$. □

The energy function U and the partition function are not unique, since adding a constant to the energy function is equivalent to multiplying the normalizing factor by an appropriate constant. Likewise, and more importantly, the Gibbs potential associated with π is not unique. However, uniqueness can be forced into the result if a certain property is imposed on the potential, namely normalization with respect to a fixed phase value.

Definition 9.1.12 *A Gibbs potential $\{V_C\}_{C\subseteq S}$ is said to be normalized with respect to a given phase in Λ, conventionally denoted by 0, if $V_C(x) = 0$ whenever there exists $t \in C$ such that $x(t) = 0$.*

Theorem 9.1.13 *There exists one and only one potential normalized with respect to a given phase* $0 \in \Lambda$ *corresponding to a Gibbs distribution.*

Proof. Expression (9.6) gives a normalized potential. In fact, the right-hand side of

$$V_C(x) = \sum_{B \subseteq C \setminus t} (-1)^{|C-B|} \log \frac{\pi(x^{B \cup t})}{\pi(x^B)}$$

is independent of t in the clique C, and in particular, choosing any $t \in C$ such that $x(t) = 0$, $x^{B \cup t} = x^B$ for all $B \subseteq C \setminus t$, and therefore $V_C(x) = 0$.

For the proof of uniqueness, suppose that

$$\pi(x) = \frac{1}{Z_1} e^{-U_1(x)} = \frac{1}{Z_2} e^{-U_2(x)}$$

for two energy functions U_1 and U_2 deriving from potentials V_1 and V_2, respectively, both normalized with respect to $0 \in \Lambda$. Since $U_1(0) = \sum_{C \in \mathcal{C}} V_{1,C}(0) = 0$, and similarly $U_2(0) = 0$, it follows that $Z_1 = Z_2 = \pi(0)^{-1}$, and therefore $U_1 \equiv U_2$. Suppose that $V_{1,A} = V_{2,A}$ for all $A \in \mathcal{C}$ such that $|A| \leq k$ (property $\mathcal{P}_k$). It remains to show, in view of a proof by induction, that $\mathcal{P}_k \Rightarrow \mathcal{P}_{k+1}$ and that $\mathcal{P}_1$ is true.

To prove $\mathcal{P}_k \Rightarrow \mathcal{P}_{k+1}$, fix $A \subseteq V$ with $|A| = k + 1$. To prove that $V_{1,A} \equiv V_{2,A}$ it suffices to show that $V_{1,A}(x) = V_{2,A}(x)$ for all $x \in \Lambda^V$ such that $x = x^A$. Fix such an x. Then

$$U_1(x) = \sum_C V_{1,C}(x) = \sum_{C \subseteq A} V_{1,C}(x)\,,$$

since x has phase 0 outside A and V_1 is normalized with respect to 0. Also,

$$U_1(x) = \sum_{C \subseteq A} V_{1,C}(x) = V_{1,A}(x) + \sum_{C \subseteq A, |C| \leq k} V_{1,C}(x)\,, \tag{9.9}$$

with a similar equality for $U_2(x)$. Therefore, since $U_1(x) = U_2(x)$, we obtain $V_{1,A}(x) = V_{2,A}(x)$ in view of the induction hypothesis. The root $\mathcal{P}_1$ of the induction hypothesis is true, since when $|A| = 1$, (9.9) becomes $U_1(x) = V_{1,A}(x)$, and similarly, $U_2(x) = V_{2,A}(x)$, so that $V_{1,A}(x) = V_{2,A}(x)$ is a consequence of $U_1(x) = U_2(x)$. □

In practice, the potential as well as the topology of V can be obtained directly from the expression of the energy, as the following example shows.

EXAMPLE 9.1.14: MARKOV CHAINS AS MARKOV FIELDS. Let $V = \{0, 1, \ldots N\}$ and $\Lambda = E$, a finite space. A random field X on V with phase space Λ is therefore a vector X with values in E^{N+1}. Suppose that $X_0, \ldots, X_N$ is a homogeneous Markov chain with transition matrix $\mathbf{P} = \{p_{ij}\}_{i,j \in E}$ and initial distribution $\nu = \{\nu_i\}_{i \in E}$. In particular, with $x = (x_0, \ldots, x_N)$,

$$\pi(x) = \nu_{x_0} p_{x_0 x_1} \cdots p_{x_{N-1} x_N},$$

that is,

$$\pi(x) = e^{-U(x)},$$

where

$$U(x) = -\log \nu_{x_0} - \sum_{n=0}^{N-1} (\log p_{x_n x_{n+1}}).$$

Clearly, this energy derives from a Gibbs potential associated with the nearest-neighbor topology for which the cliques are, besides the singletons, the pairs of adjacent sites. The potential functions are:

$$V_{\{0\}}(x) = -\log \nu_{x_0}, \quad V_{\{n,n+1\}}(x) = -\log p_{x_n x_{n+1}}.$$

The local characteristic at site n, $2 \le n \le N-1$, can be computed from formula (9.3), which gives

$$\pi^n(x) = \frac{\exp(\log p_{x_{n-1}x_n} + \log p_{x_n x_{n+1}})}{\sum_{y \in E} \exp(\log p_{x_{n-1} y} + \log p_{y x_{n+1}})},$$

that is,

$$\pi^n(x) = \frac{p_{x_{n-1}x_n} p_{x_n x_{n+1}}}{p^{(2)}_{x_{n-1} x_{n+1}}},$$

where $p^{(2)}_{ij}$ is the general term of the two-step transition matrix $\mathbf{P}^2$. Similar computations give $\pi^0(x)$ and $\pi^N(x)$. We note that, in view of the neighborhood structure, for $2 \le n \le N-1$, X_n is independent of $X_0, \ldots, X_{n-2}, X_{n+2}, \ldots, X_N$ given X_{n-1} and X_{n+1}.

9.1.3 Specific Models

Random Points

Let $Z := \{Z(v)\}_{v \in V}$ be a random field on V with phase space $\Lambda := \{0, 1\}$. Here $Z(v) = 1$ will be interpreted as the presence of a "point" at site v.

Recall that $\mathcal{P}(V)$ is the collection of subsets of V, and denote by $\mathbf{x}$ such a subset. A random field $Z \in \{0,1\}^V$ with distribution π being given, we associate to it the random element $\mathbf{X} \in \mathcal{P}(V)$, called a point process on V, by

$$\mathbf{X} := \{v \in V \,;\, Z(v) = 1\}.$$

Its distribution is denoted by ℓ. For any $\mathbf{x} \subseteq V$, $\ell(\mathbf{x})$ is the probability that $Z(v) = 1$ for all $v \in \mathbf{x}$ and $Z(v) = 0$ for all $v \notin \mathbf{x}$.

Let $\mathbf{X}$ be a point process on the finite set V with positive probability distribution $\{\ell(\mathbf{x})\}_{\mathbf{x} \in \mathcal{P}(V)}$. Such point process on V can be viewed as a random field on V with phase space $\Lambda \equiv \{0, 1\}$ with probability distribution $\{\pi(z)\}_{z \in \Lambda^V}$. We assume that this random field is Markov with respect to the symmetric relation $\sim$, and then say that $\mathbf{X}$ is Markov with respect to $\sim$. We have the following alternative form of the Hammersley–Clifford theorem:

Theorem 9.1.15 *(Ripley and Kelly, 1977)*[1] *The point process* $\mathbf{X}$ *on the finite set* V *with positive probability distribution* $\{\ell(\mathbf{x})\}_{\mathbf{x}\in\mathcal{P}(V)}$ *is Markov with respect to* $\sim$ *if and only if there exists a function* $\varphi : M_p^f(V) \to (0,1]$ *such that*

(i) $\varphi(\mathbf{y}) < 1$ *if and only if* $\mathbf{y}$ *is a clique for* $\sim$*, and*

(ii) for all $\mathbf{x} \in \mathcal{P}(V)$*,*

$$\ell(\mathbf{x}) = \prod_{\mathbf{y}\subseteq\mathbf{x}} \varphi(\mathbf{y})\,.$$

Proof. Necessity: The distribution π may be expressed in terms of a potential $\{V_C\}_{C\subseteq V}$ as

$$\pi(z) = \alpha e^{\sum_{C\subseteq V} V_C(z)}\,. \tag{9.10}$$

Take for a potential the (unique) one normalized with respect to phase 0. Identifying a configuration $z \in \{0,1\}^V$ with a subset $\mathbf{x}$ of V, and more generally identifying a subset C of V with a configuration $\mathbf{y} \in \mathcal{P}(V)$, the potential can be represented as a collection of functions $\{V_\mathbf{y}\}_{\mathbf{y}\in\mathcal{P}(V)}$. Note that $V_\mathbf{y}(\mathbf{x}) > 0$ if and only if $\mathbf{y}$ is a clique and $\mathbf{y} \subseteq \mathbf{x}$ (normalized potential), in which case $V_\mathbf{y}(\mathbf{x}) = V_\mathbf{y}(\mathbf{y})$. The result then follows by letting

$$\varphi(\mathbf{y}) := e^{-V_\mathbf{y}(\mathbf{y})} \qquad (\mathbf{y} \neq \varnothing)$$

and

$$\varphi(\varnothing) := \ell(\varnothing) = \alpha\,.$$

The proof of sufficiency is left for the reader as it follows the same lines as the proof of Theorem 9.1.8. □

In the case of a positive distribution ℓ of the point process $\mathbf{X}$, let

$$\lambda(u,\mathbf{x}) := \frac{\ell(\mathbf{x}\cup u)}{\ell(\mathbf{x})}$$

if $u \notin \mathbf{x}$, $= 0$ otherwise. For $u \notin \mathbf{x}$,

$$\lambda(u,\mathbf{x}) = \frac{P(Z(u)=1, \mathbf{X}\backslash u = \mathbf{x})}{P(\mathbf{X}=\mathbf{x})} = \frac{P(Z(u)=1, \mathbf{X}\backslash u = \mathbf{x})}{P(\mathbf{X}\backslash u = \mathbf{x})}\,,$$

and therefore

$$\lambda(u,\mathbf{x}) = P(Z(u)=1 \,|\, \mathbf{X}\backslash u = \mathbf{x})\,,$$

the probability that there is a point at u knowing the point process outside u. This defines the exvisible distribution (on $\{0,1\}$) at point $u \in V$.

Theorem 9.1.16 *Let* $g : V \times \mathcal{P}(V) \to \mathbb{R}$ *be a non-negative function. Then*

$$E\left[\sum_{u\in V} g(u, \mathbf{X}\backslash u)\right] = E\left[\sum_{u\in V} g(u,\mathbf{X})\lambda(u,\mathbf{X})\right]\,.$$

[1] This is the discrete version of their more general theorem concerning finite point processes on $\mathbb{R}^m$.

Proof.

$$E\left[\sum_{u\in V} g(u,\mathbf{X})\lambda(u,\mathbf{X})\right] = \sum_{\mathbf{x}\in\mathcal{P}(V)}\sum_{u\in V} g(u,\mathbf{x})\lambda(u,\mathbf{x})\ell(\mathbf{x})$$
$$= \sum_{\mathbf{x}\in\mathcal{P}(V)}\sum_{u\in V} g(u,\mathbf{x})1_{\{u\notin\mathbf{x}\}}\ell(\mathbf{x}\cup u)\,.$$

With the change of variables $\mathbf{x}\cup u\to\mathbf{y}$, the last quantity is seen to be equal to

$$\sum_{\mathbf{y}\in\mathcal{P}(V)}\sum_{u\in V} g(u,\mathbf{y}\backslash u)\ell(\mathbf{y}) = E\left[\sum_{u\in V} g(u,\mathbf{X}\backslash u)\right]\,.$$

□

The Autobinomial Texture Model

(Besag, 1974) For the purpose of image synthesis, one seeks Gibbs distributions describing pictures featuring various textures, lines separating patches with different textures (boundaries), lines per se (roads, rail tracks), randomly located objects (moon craters), etc. The corresponding model is then checked by simulation (see Chapter 19): images are drawn from the proposed Gibbs distribution, and some tuning of the parameters is done, until the images subjectively correspond to ("look like") the type of image one expects. Image synthesis is an art based on trial and error, and fortunately guided by some general principles. But these principles are difficult to formalize, and we shall mainly resort to simple examples with a pedagogical value. Note, however, that there is a domain of application where the model need not be very accurate, namely Bayesian estimation. As a matter of fact, the models proposed in this section have been devised in view of applications to Bayesian restoration of degraded pictures.

We shall begin with an all-purpose texture model that may be used to describe the texture of various materials. The set of sites is $V=\mathbb{Z}_m^2$, and the phase space is $\Lambda=\{0,1,\ldots,L\}$. In the context of image processing, a site v is a pixel (PICTure ELement), and a phase $\lambda\in\Lambda$ is a shade of grey, or a colour. The neighborhood system is

$$\mathcal{N}_v = \{w\in V; w\neq v\,;\, \|w-v\|^2\leq d\}, \tag{9.11}$$

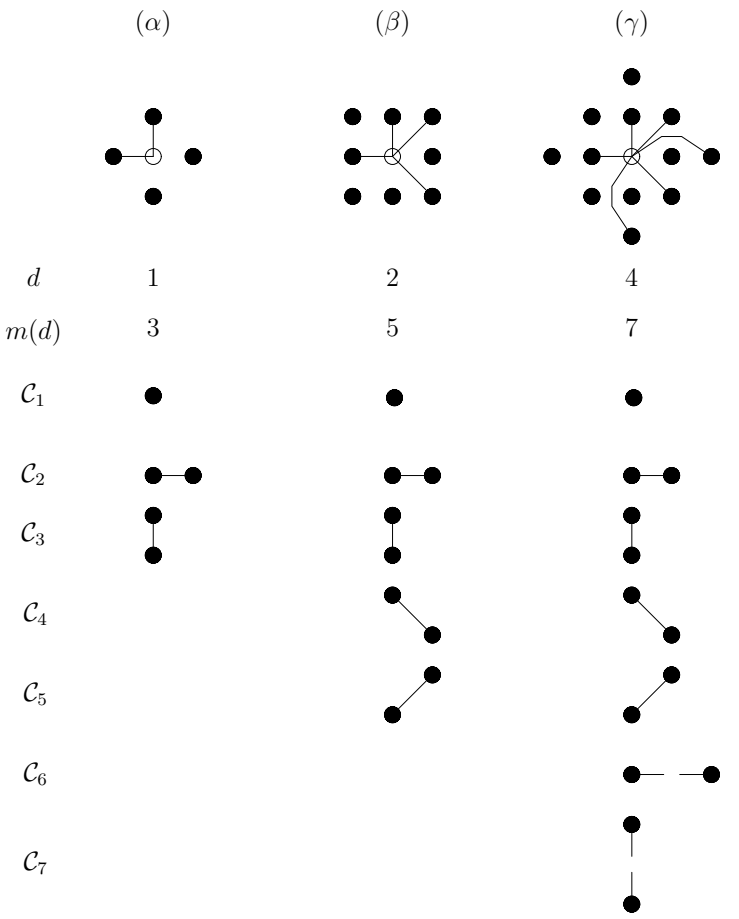
where d is a fixed positive integer and where $\|w-v\|$ is the euclidean distance between v and w. In this model the only cliques participating in the energy function are singletons and pairs of mutual neighbors. The set of cliques appearing in the energy function is a disjoint sum of collections of cliques

$$\mathcal{C} = \sum_{j=1}^{m(d)} \mathcal{C}_j,$$

where $\mathcal{C}_1$ is the collection of singletons, and all pairs $\{v,w\}$ in $\mathcal{C}_j$, $2\leq j\leq m(d)$, have the same distance $\|w-v\|$ and the same direction, as shown in the figure below. The potential is given by

$$V_C(x) = \begin{cases} -\log \binom{L}{x(v)} + \alpha_1 x(v) & \text{if } C = \{v\} \in \mathcal{C}_1, \\ \alpha_j x(v) x(w) & \text{if } C = \{v, w\} \in \mathcal{C}_j, \end{cases}$$

where $\alpha_j \in \mathbb{R}$. For any clique C not of type $\mathcal{C}_j$, $V_C \equiv 0$.



Neighborhoods and cliques of three autobinomial models

The terminology ("autobinomial") is motivated by the fact that the local system has the form

$$\pi^v(x) = \binom{L}{x(v)} \tau^{x(v)} (1-\tau)^{L-x(v)}, \tag{9.12}$$

where τ is a parameter depending on $x(\mathcal{N}_v)$ as follows:

$$\tau = \tau(\mathcal{N}_v) = \frac{e^{-\langle \alpha, b \rangle}}{1 + e^{-\langle \alpha, b \rangle}}. \tag{9.13}$$

Here $\langle \alpha, b \rangle$ is the scalar product of

$$\alpha = (\alpha_1, \ldots, \alpha_{m(d)}) \text{ and } b = (b_1, \ldots, b_{m(d)}),$$

where $b_1 = 1$, and for all j, $2 \leq j \leq m(d)$,

$$b_j = b_j(x(\mathcal{N}_v)) = x(u) + x(w),$$

where $\{v, u\}$ and $\{v, w\}$ are the two pairs in $\mathcal{C}_j$ containing v.

Proof. From the explicit formula (9.3) giving the local characteristic at site v,

$$\pi^v(x) = \frac{\exp\left\{\log\binom{L}{x(v)} - \alpha_1 x(v) - \left[\sum_{j=2}^{m(d)} \alpha_j \sum_{v;\{v,w\}\in\mathcal{C}_j} x(w)\right] x(v)\right\}}{\sum_{\lambda\in\Lambda} \exp\left\{\log\binom{L}{\lambda} - \alpha_1 \lambda - \left[\sum_{j=2}^{m(d)} \alpha_j \sum_{t;\{v,w\}\in\mathcal{C}_j} x(w)\right] \lambda\right\}}.$$

The numerator equals

$$\binom{L}{x(v)} e^{-\langle\alpha,b\rangle x(v)},$$

and the denumerator is

$$\sum_{\lambda\in\Lambda} \binom{L}{\lambda} e^{\langle -\alpha,b\rangle\lambda} = \sum_{\ell=0}^{L} \binom{L}{\ell} \left(e^{-\langle\alpha,b\rangle}\right)^{\ell} = \left(1 + e^{-\langle\alpha,b\rangle}\right)^{L}.$$

Equality (9.12) then follows. □

Expression (9.12) shows that τ is the average level of grey at site v, given $x(\mathcal{N}_v)$, and expression (9.13) shows that τ is a function of $\langle\alpha, b\rangle$. The parameter α_j controls the bond in the direction and at the distance that characterize $\mathcal{C}_j$.

Pixel-and-edge Model

(Geman and Geman, 1984) Let $X = \{X(v)\}_{v\in V}$ be a random field on V with phase space Λ, with the following structure:

$$V = V_1 + V_2\ ,\ \Lambda = \Lambda_1 \cup \Lambda_2,$$

and

$$\begin{aligned} X(v) &= Y(v_1) \in \Lambda_1 \text{ if } v = v_1 \in V_1 \\ &= Z(v_2) \in \Lambda_2 \text{ if } v = v_2 \in V_2. \end{aligned}$$

Here V_1 and V_2 are two disjoint collections of sites that can be of a different nature, or have different functions, and Λ_1 and Λ_2 are phase spaces that need not be disjoint. Define

$$Y = \{Y(v_1)\}_{v_1\in V_1},\ Z = \{Z(v_2)\}_{v_2\in V_2}.$$

The random field X may be viewed as the juxtaposition of Y and Z.

In some situations, Y is the *observed field*, and Z is the *hidden field*. Introduction of a hidden field is in principle motivated by physical considerations. From the

computational point of view, it is justified by the fact that the field Y alone usually has a Gibbsian description with large neighborhoods, whereas $X = (Y, Z)$ hopefully has small neighborhoods.

The philosophy supporting the pixel-and-edge model is the following. A digitized image can be viewed as a realization of a random field on $V^P = \mathbb{Z}_m^2$. *A* site could be, for instance, a pixel on a digital television screen, and therefore V^P will be called the set of *pixel* sites. For an observer, there is, in general, more in an image than just the colours at each pixel. For instance, an image can be perceived as a juxtaposition of zones with various textures separated by lines. However, these lines, or contours, are not seen directly on the pixels, they are inferred from them by some processing in the brain. On the other hand, if one is to sketch the picture observed on the screen, one would most likely start by drawing the lines. In any case, textures and contours are very much linked, and one should seek a description featuring the interaction between them. But as was mentioned, contours do not exist on the digital screen, they are hidden, or more accurately, they are virtual.

In this example, there is a set V^E of *edge* sites, one between each pair of adjacent pixel sites, as indicated in the figure below (a). The possible values of the phase on an edge site are blank or bar: horizontal (resp., vertical) for an edge site between two pixel sites forming a vertical (resp., horizontal) segment, as shown in the figure below (b). In this figure, not all edge sites between two pixels with a different colour have a bar, because a good model should not systematically react to what may be accidents in the global structure.

Let (i, j) denote a pixel site and (α, β) an edge site (these are the coordinates of the sites in two distinct orthogonal frames). The random field on the pixels is denoted by $X^P = \{X_{ij}^P\}_{(i,j)\in V^P}$, and that on the edge sites is $X^E = \{X_{\alpha\beta}^E\}_{(\alpha,\beta)\in V^E}$; X^P is the observed image, and X^E is the hidden, or virtual, line field. The distribution of the field $X = (X^P, X^E)$ is described by an energy function $U(x^P, x^E)$:

$$\pi(x^P, x^E) = \frac{1}{Z} e^{-U(x^P, x^E)}, \tag{9.14}$$

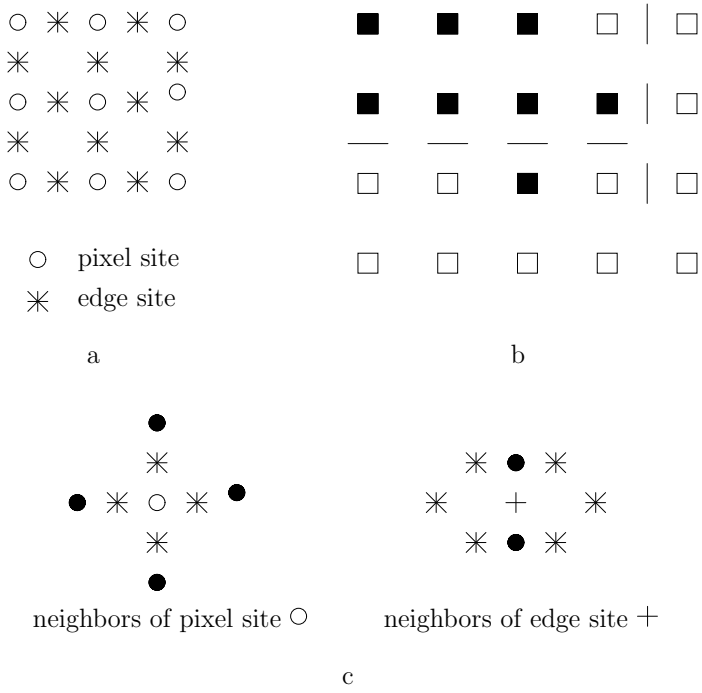
where $x^P = \{x_{ij}^P\}_{(i,j)\in V^P}$, $x^E = \{x_{\alpha\beta}^E\}_{(\alpha,\beta)\in V^E}$. The energy function derives from a potential relative to some neighborhood system, a particular choice of which is pictured in (c) of the figure. The energy function separates into two parts

$$U(x^P, x^E) = U_1(x^P, x^E) + U_2(x^E), \tag{9.15}$$

where U_1 features only cliques formed by a pair of neighboring pixels and the edge pixel in between, whereas U_2 features only the diamond cliques shown in the figure below. The energy $U_1(x^P, x^E)$ governs the relation between an edge and its two adjacent pixels. For instance, for some real constant $\alpha > 0$ and some function φ,

$$U_1(x^P, x^E) = -\alpha \sum_{\langle 1,2\rangle} \varphi(x_1^P - x_2^P)\, x_{\langle 1,2\rangle}^E,$$

where $\langle 1, 2\rangle$ represents a pair of adjacent pixel sites and $x_{\langle 1,2\rangle}^E$ is the value of the phase on the edge site between the pixel sites 1 and 2, say 0 for a blank and 1 for a bar. A possible choice of φ is



Example of a neighborhood structure for the pixel-edge model

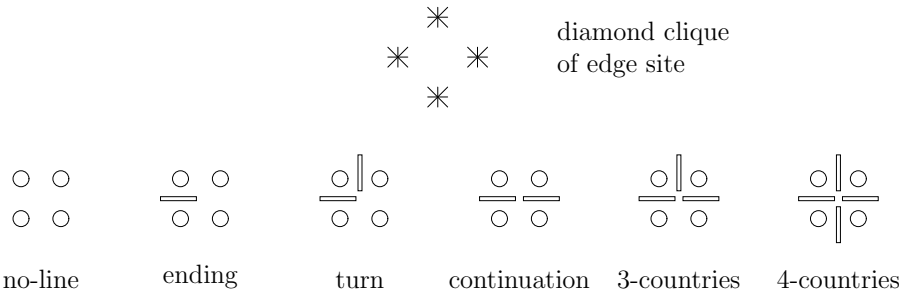
$$\varphi(x) = 1_{\{x \neq 0\}} - 1_{\{x=0\}}\,.$$

Since the most probable configurations are those with low energy, this model favors bars between two adjacent pixel sites with different colours, as is natural. More sophisticated choices of φ with a similar effect are possible. The organization of the contours is controlled by the energy

$$U_2(x^E) = \beta \sum_D w_D(x^E),$$

where $\beta > 0$ and the sum extends to all diamond cliques, and w_D is a function depending only on the phases of the four edge sites of the diamond clique D. Up to rotations of $\frac{\pi}{2}$, there are six possible values for the four-vector of phases on a given diamond clique D, as shown in the figure. If the modeler believes that for the type of images he is interested in the likelihood of the configurations shown in the figure below decreases from left to right, then the values of $w_D(x^E)$ attributed to these configurations will increase from left to right. This is generally the case because four-country border points are rare, broken lines also, and the same is true to a lesser extent for three-country border points. Also, when the picture is not a clutter of lines, the no-line configuration is the most likely.

This example admits many variations. However, too many sophisticated features could ruin the model. The purpose of the model is not so much to do image synthesis as to have a reasonable a priori model for the image F in view of Bayesian restoration of this image from a noisy version of it, as will now be explained.



Six configurations of the diamond clique

Conditional Markov Fields in Image Processing*

(∗: this subsubbsection makes use of Gaussian vectors, a notion outside the prerequisites) A number of estimation problems arising in various areas of statistics and engineering, and particularly in image processing, are solved by a method of statistical analysis known as the *maximum a posteriori* (MAP) likelihood method. The theory will not be presented, but the examples below will show its substance. These examples compute the a posteriori probability, or conditional probability, of a given MRF X with respect to another MRF Y, the observed field, say,

$$\pi(x \mid y) = P(X = x \mid Y = y),$$

and the MAP method estimates the nonobservable field X given the observed value y of Y, by $\hat{x} = \hat{x}(y)$, the value of x that maximizes $\pi(x \mid y)$:

$$\hat{x}(y) = \arg\max_x \pi(x \mid y).$$

Usually, this maximization problem is doomed by combinatorial explosion and by the complexity of standard methods of optimization. However, if

$$\pi(x \mid y) \propto e^{-U(x \mid y)} \tag{9.16}$$

(the proportionality factor depends only on y and is therefore irrelevant to maximization with respect to x) with an energy $U(x \mid y)$ that as a function of x corresponds to a topology N with small neighborhoods, then one can use a simulated annealing algorithm or a related algorithm (see Section 15.2).

EXAMPLE 9.1.17: RANDOM FLIPS OR MULTIPLICATIVE NOISE. Let X, Z be random fields on $V = \mathbb{Z}_m^2$ with phase space $\Lambda = \{-1, +1\}$, and define the field $Y = XZ$ by $y(v) = x(v)z(v), v \in V$. The field Z will be interpreted as multiplicative noise, and one can call it a random flip field, because what it does to X is to flip the phase at site v if $z(v) = -1$.

The computation below uses the fact that if $a, b, c \in \Lambda^V$, where $\Lambda = \{-1, +1\}$, then $ab = c \Leftrightarrow b = ac$:

$$\begin{aligned} P(Y = y)P(X = x \mid Y = y) &= P(X = x, Y = y) = P(X = x, ZX = y) \\ &= P(X = x, Zx = y) = P(X = x, Z = yx). \end{aligned}$$

In particular, if the noise field Z is independent of the original field X, then

$$\pi(x \mid y) \propto P(X = x)P(Z = yx).$$

The random field X has the distribution $P(X = x) \propto e^{-U(x)}$. Suppose that $(Z(v), v \in V)$ is a family of IID random variables, with $P(Z(v) = -1) = p$, $P(Z(v) = +1) = q = 1 - p$. Therefore (Exercise 9.4.1),

$$P(Z = z) = \prod_{v \in V} P(Z(v) = z(v)) \propto e^{\gamma \sum_{v \in V} z(v)},$$

where $\gamma = \frac{1}{2} \log \left(\frac{1-p}{p} \right)$. Finally, $\pi(x \mid y) \propto e^{-U(x) + \gamma \sum_{v \in V} y(v)x(v)}$.

EXAMPLE 9.1.18: IMAGE RESTORATION. This example refers to the model of Subsection 9.1.3. Recall that we have a random field $X = (X^P, X^E)$ corresponding to some energy function $U(x^P, x^E)$, which need not be made precise here (see, however, Subsection 9.1.3). The image X^P is degraded into a noisy image Y, and it is this corrupted image that is observed. Degradation combines two effects: blurring, and a possibly nonlinear interaction of the blurred image and the noise. Specifically,

$$Y_{ij} = \varphi(H(X^P)_{ij}, N_{ij}), \tag{9.17}$$

where (i, j) is a pixel site and φ, H, and N are defined as follows. First $N = \{N_{ij}\}_{(i,j) \in P_m}$ is, for instance, a family of independent centered Gaussian random variables with common variance σ^2, and is independent of (X^P, X^E). As for $H(X^P)$, it is the random field obtained by blurring X^P, that is,

$$H(X^P)_{ij} = \sum_{k,\ell} H_{k\ell} X^P_{i-k,j-\ell}, \tag{9.18}$$

where $H = \{H_{k\ell}\}_{-N \le k,\ell \le N}$ is the *blurring matrix*. In (9.18), $X^P_{i-k,j-\ell} = 0$ if $(i - k, j - \ell) \notin S^P$. A typical blurring matrix is

$$H = \begin{pmatrix} 1/80 & 1/80 & 1/80 \\ 1/80 & 9/10 & 1/80 \\ 1/80 & 1/80 & 1/80 \end{pmatrix},$$

for which $N = 1$. In this case

$$H(X^P)_{ij} = \frac{9}{10} X^P_{ij} + \frac{1}{80} \left(\sum_{\ell} X^P_{k,\ell} \right),$$

where the sum extends to the pixel sites adjacent to (i, j). As for φ, it is a function such that for fixed a, the function $b \to \varphi(a, b)$ is invertible. The inverse of this function, for fixed a, is then denoted by $b \to \psi(a, b)$. A typical example for φ is the additive noise model

$$Y_{ij} = H(X^P)_{ij} + N_{ij}. \tag{9.19}$$

To estimate X given Y, one is led to compute the a posteriori probability of image x given that the noisy image is y:

$$\pi(x^P, x^E \mid y) = P(X^P = x^P, X^E = x^E \mid Y = y).$$

Writing $\pi(y) = P(Y = y)$, we have

$$\begin{aligned}
\pi(x^P, x^E \mid y) &= \pi(y) P(X^P = x^P, X^E = x^E, Y = y) \\
&= \pi(y) P(X^P = x^P, X^E = x^E, \varphi(H(X^P), N) = y) \\
&= \pi(y) P(X^P = x^P, X^E = x^E, N = \psi(H(x^P), y)) \\
&= \pi(y) P(X^P = x^P, X^E = x^E) P(N = \psi(H(x^P), y)).
\end{aligned}$$

The reader will have noticed the abuse of notation by which the continuous character of N was ignored: The second to fourth terms are actually probability densities, and similarly for

$$P(N = \psi(H(x^P), y)) \propto e^{-\frac{1}{2\sigma^2}\|\psi(H(x^P), y)\|^2}.$$

Using the expression of the distribution of the pixel+line image in terms of the energy function, one finds

$$\pi(x^P, x^E \mid y) \propto e^{-U(x^P, x^E) - \frac{1}{2\sigma^2}\|\psi(H(x^P), y)\|^2}. \tag{9.20}$$

Therefore the *a posteriori* distribution of (X^P, X^E) given $Y = g$ is a Gibbs distribution corresponding to the energy function

$$U(x^P, x^E) = U(x^P, x^E) + \frac{1}{2\sigma^2}\|\psi(H(x^P), y)\|^2. \tag{9.21}$$

For instance, if the noise is additive, as in (9.19), then

$$U(x^P, x^E) = U(x^P, x^E) + \frac{1}{2\sigma^2}\|y - H(x^P)\|^2. \tag{9.22}$$

EXAMPLE 9.1.19: BERNOULLI–GAUSSIAN MODEL. Let $\{Y_n\}_{1 \le n \le N}$ be a real-valued stochastic process of the form

$$Y_n = \sum_{k=1}^{N} X_k h_{n-k} + Z_n,$$

where $\{Z_n\}_{1 \le n \le N}$ is a sequence of independent centered Gaussian random variables of variance σ^2, $\{X_n\}_{1 \le n \le N}$ is an IID sequence of $\{0,1\}$-valued random variables with $P(X_n = 1) = p$, and $\{h_k\}_{k \in \mathbb{Z}}$ is a deterministic function.

This is a particular case of the one-dimensional version of the model in Example 9.1.18. Here $V = \{1, \ldots, N\}$, a configuration $x \in \{0,1\}^N$ is of the form $x = (x_1, x_2, \ldots, x_N)$, X is the original image, Y is the degraded image, Z is the additive noise, and h corresponds to the blurring matrix. For this particular model, the

energy of the IID random field X is of the form $\gamma \sum_{i=1}^{N} x_i$ (Exercise 9.4.1), and the energy of the conditional field $x|y$ is

$$\gamma \sum_{i=1}^{N} x_i + \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left| y_i - \sum_{j;1 \le i-j \le N} h_j x_{i-j} \right|^2 .$$

This model is often used in problems of detection of reflectors. One says that there is a reflector at position i if $X_i = 1$. The function h is a probe signal (radar, sonar), and $\{h_{k-i}\}_{k \in \mathbb{Z}}$ is the signal reflected by the reflector at position i, if any, so that $Y_n = \sum_{k=1}^{N} X_k h_{n-k}$ is the reflected signal from which the map of reflectors X is to be recovered. The process Z is the usual additive noise of signal processing.

Of course, this model can be considerably enriched by introducing random reflection coefficients or by using a more elaborate a priori model for X, say, a Markov chain model.

Penalty Methods

Consider the simple model where the image X is additively corrupted by white Gaussian noise N of variance σ^2, and let Y be the resulting image. Calling $U(x)$ the energy function of the a priori model, the MAP estimate is

$$\hat{x} \;=\; \arg\min_x \{U(x) + \frac{1}{\sigma^2}\|y - x\|^2\}.$$

If we take an a priori model where all images are equiprobable, that is, the corresponding energy is null, then the above minimization is trivial, leading one to accept the noisy image as if it were the original image. A nontrivial a priori model introduces a penalty term $U(x)$ and forces a balance between our belief in the observed image, corresponding to a small value of $\|y - x\|^2$, and our a priori expectation as to what we should obtain, corresponding to a small value of $U(x)$. The compromise between the credibility of the observed image and the credibility of the estimate with respect to the prior distribution is embodied in the criterion $U(x) + \frac{1}{\sigma^2}\|y - x\|^2$. A non-Bayesian mind will, somehow rightly, argue that one cannot even dream of thinking that a correct a priori model is available, and that Gaussian additive white noise is at best an intellectual construction. All that he will retain from the above is the criterion

$$\lambda U(x) + \|y - x\|^2,$$

with the interpretation that the penalty term $\lambda U(x)$ corrects undesirable features of the observed image y. One of these is the usually chaotic aspect, at the fine scale. However, he does not attempt to interpret this as due to white noise. In order to correct this effect he introduces a *smoothing* penalty term $U(x)$, which is small when x is smooth, for instance

$$U(x) \;=\; \sum_{\langle s,t \rangle} \left(x(s) - x(t)\right)^2 ,$$

where the summation extends over pairs of adjacent pixels. One disadvantage of this smoothing method is that it will tend to blur the boundary between two highly contrasted regions. One must choose an edge-preserving smoothing penalty function, for instance

$$U(x) \;=\; \sum_{\langle s,t\rangle} \Psi\left(x(s) - x(t)\right),$$

where, for instance,

$$\Psi(u) = -\left(1 + \left(\frac{u}{\delta}\right)^2\right)^{-1},$$

with $\delta > 0$. This energy function favors large contrasts and therefore prevents to some extent blurring of the edges. A more sophisticated penalty function would introduce edges, as in Subsection 9.1.3, with a penalty function of the form

$$U(x^P, x^E) = U_1(x^P, x^E) + U_2(x^E),$$

where the term $U_1(x^P, x^E)$ creates the edges from the pixels, and $U_2(x^E)$ organizes the edges.

Note that the estimated image now consists of two parts: $\hat{x}^P$ solves the smoothing problem, whereas $\hat{x}^E$ extracts the boundaries. If one is really interested in boundary extraction, a sophisticated line-pixel model is desirable. If one is only interested in cleaning the picture, rough models may suffice.

The import of the Gibbs–Bayes approach with respect to the purely deterministic penalty function approach to image restoration lies in the theoretical possibility of the former to tune the penalty function by means of simulation. Indeed, if one is able to produce a typical image corresponding to the energy-penalty function, one will be able to check with the naked eye whether this penalty function respects the constraints one has in mind, and if necessary to adjust the parameters in it. The simulation issue is treated in Chapter 19. Another theoretical advantage of the Gibbs–Bayes approach is the availability of the simulated annealing algorithm (Section 15.2) to solve the minimization problem arising in MAP likelihood method or in the traditional penalty method.

9.2 Phase Transition in the Ising Model

9.2.1 Experimental Results

The first significant success of the Gibbs–Ising model was a qualitative explanation of the phase transition phenomenon in ferromagnetism (Peierls, 1936).

Consider the slightly generalized Ising model of a piece of ferromagnetic material, with spins distributed according to

$$\pi_T(x) = \frac{1}{Z_T} e^{\frac{-U(x)}{T}}. \tag{9.23}$$

The finite site space is enumerated as $V = \{1, 2, \ldots, N\}$, and therefore a configuration x is denoted by $(x(1), x(2), \ldots, x(N))$. The energy function is

$$U(x) = U_0(x) - \frac{H}{k}\sum_{i=1}^{N} x(i),$$

where the term $U_0(x)$ is assumed symmetric, that is, for any configuration x,

$$U_0(x) = U_0(-x).$$

The constant H is the external magnetic field. The *magnetic moment* of configuration x is

$$m(x) = \sum_{i=1}^{N} x(i),$$

and the *magnetization* is the average *magnetic moment* per site

$$M(H,T) = \frac{1}{N}\sum_{x\in E} \pi_T(x) m(x).$$

We have that $\frac{\partial M(H,T)}{\partial H} \geq 0$ (Exercise 9.4.13), $M(-H,T) = -M(H,T)$ and $-1 \leq M(H,T) \leq +1$. Therefore, at fixed temperature T, the magnetization $M(H,T)$ is a non-decreasing odd function of H with values in $[-1,+1]$. Also,

$$M(0,T) = 0\,, \qquad (\diamond)$$

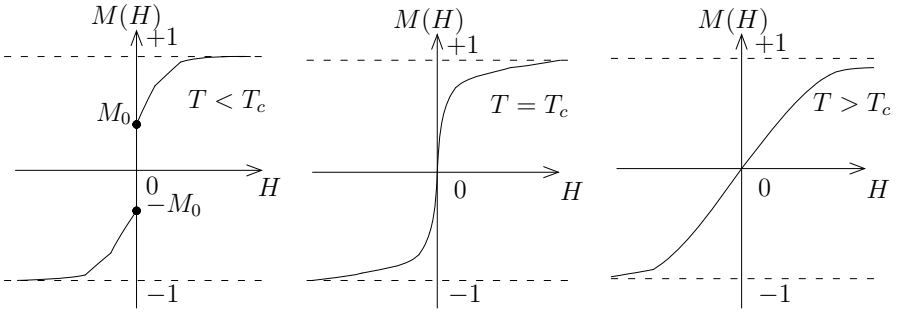
since for any configuration x, $m(-x) = -m(x)$, and therefore $\pi_T(-x) = \pi_T(x)$ when $H = 0$. Moreover, the magnetization is an analytic function of H.

However, experimental results seem to contradict the last two assertions. Indeed, if an iron bar is placed in a strong magnetic field H parallel to the axis, it is completely magnetized with magnetization $M(H,T) = +1$, and if the magnetic field is slowly decreased to 0, the magnetization decreases, but tends to a limit $M(0,T) = M_0 > 0$, in disagreement with $(\diamond)$. By symmetry, we therefore have a discontinuity of the magnetization at $H = 0$ (see the figure below (a)), in contradiction to the theoretical analyticity of the magnetization as a function of H.

This discontinuity is called a *phase transition* by physicists, by analogy with the discontinuity in density at a liquid-gas phase transition. It occurs at room temperature, and if the temperature is increased, the residual, or *spontaneous*, magnetization M_0 decreases until it reaches the value 0 at a certain temperature T_c, called the *critical temperature*. Then, for $T > T_c$, the discontinuity at 0 disappears, and the magnetization curve is smooth (Figure (c) below). At $T = T_c$, the slope at $H = 0$ is infinite, that is, the *magnetic susceptibility* is infinite (Figure (b) below).

The discrepancy between experience and theory below the critical temperature is due to the fact that the experimental results describe a situation at the *thermodynamical limit* $N = \infty$. For fixed but large N the theoretical magnetization curve is analytic, but it presents for all practical purposes the same aspect as in Figure (a) below.

To summarize the experimental results, it seems that below the critical temperature, the spontaneous magnetization has, when no external magnetic field is applied, two "choices." This phenomenon can be explained within the classical Ising model.



a b c

The DLR Problem

(Dobrushin, 1965, Lanford and Ruelle, 1969) Consider the Ising model in the absence of an external field ($H = 0$). The energy of a configuration x is of the form

$$U(x) = -J \sum_{\langle v,w \rangle} x(v)x(w),$$

where $\langle v, w \rangle$ represents an unordered pair of neighbors. When the cardinal of the site space V is infinite, the sum in the expression of the energy is not defined for all configurations, and therefore one cannot define the Gibbs distribution π_T on Λ^V by formula (9.23). However, the local specification

$$\pi_T^v(x) = \frac{e^{\beta \sum_{\langle v,w \rangle} x(v)x(w)}}{e^{\beta \sum_{\langle v,w \rangle} x(w)} + e^{-\beta \sum_{\langle v,w \rangle} x(w)}}, \tag{9.24}$$

where β is, up to a factor, the inverse temperature, is well-defined for all configurations and all sites.

In the sequel, we shall repeatedly use an abbreviated notation. For instance, if π is the distribution of a random field X under probability P, then $\pi(x(A))$ denotes $P(X(A) = x(A))$, $\pi(x(0) = +1)$ denotes $P(X(0) = +1)$, etc.

A probability distribution π_T on Λ^V is called a solution of the DLR problem if it admits the local specification (9.24).

When $V = K_N = \mathbb{Z}^2 \cap [-N, +N]^2$, we know that there exists a unique solution, given by (9.23). When $V = \mathbb{Z}^2$, one can prove (this is not done here) existence of at least one solution of the DLR problem. One way of constructing a solution is to select an arbitrary configuration z, to construct for each integer $N \geq 2$ the unique probability distribution $\pi_T^{(N)}$ on Λ^V such that

$$\pi_T^{(N)}(z(V \backslash K_{N-1})) = 1$$

(the field is frozen at the configuration z outside K_{N-1}) and such that the restriction of $\pi_T^{(N)}$ to K_{N-1} has the required local characteristics (9.24), and then let N

tend to infinity. For all configurations x and all *finite* subsets $A \subset V$, the following limit exists:

$$\pi_T(x(A)) = \lim_{N\uparrow\infty} \pi_T^{(N)}(x(A)), \tag{9.25}$$

and moreover, there exists a unique random field X with the local specification (9.24) and such that, for all configurations x and all *finite* subsets $A \subset V$,

$$P(X(A) = x(A)) = \pi_T(x(A)).$$

Note that $\pi_T^{(N)}$ depends on the configuration z only through the restriction of z to the boundary $K_N \backslash K_{N-1}$.

9.2.2 Peierls' Argument

If the DLR problem has more than one solution, one says that a phase transition occurs. The method given by Dobrushin to construct a solution suggests a way of proving phase transition when it occurs. It suffices to select two configurations z_1 and z_2, and to show that for a given finite subset $A \subset S$, the right-hand side of (9.25) is different for $z = z_1$ and $z = z_2$. In fact, for sufficiently small values of the temperature, phase transition occurs. To show this, we apply the above program with z_1 being the configuration with all spins positive and z_2 the all negative configuration, and with $A = \{0\}$, where 0 denotes the central site of $\mathbb{Z}^2$.

Denote then by $\pi_+^{(N)}$ (resp., $\pi_-^{(N)}$) the restriction to K_N of $\pi_T^{(N)}$ when $z = z_1$ (resp., $z = z_2$). We shall prove that if T is large enough, then $\pi_+^{(N)}(x(0) = -1) < \frac{1}{3}$ for all N. By symmetry, $\pi_-^{(N)}(x(0) = +1) < \frac{1}{3}$, and therefore $\pi_-^{(N)}(x(0) = -1) > \frac{2}{3}$. Passing to the limit as $N \uparrow \infty$, we see that $\pi_+(x(0) = -1) < \frac{1}{3}$ and $\pi_-(x(0) = -1) > \frac{2}{3}$, and therefore, the limiting distributions are not identical.

The above program for proving the existence of a phase transition is now carried out. For all $x \in \Lambda^{K_N}$,

$$\pi_+^{(N)}(x) = \frac{e^{-2\beta n_o(x)}}{Z_+^{(N)}}\ , \tag{9.26}$$

where $n_o(x)$ is the number of *odd bounds* in configuration x, that is, the number of cliques $\langle v, w\rangle$ such that $x(v) \neq x(w)$, and where $Z_+^{(N)}$ is the normalization factor.

Proof. It suffices to observe that

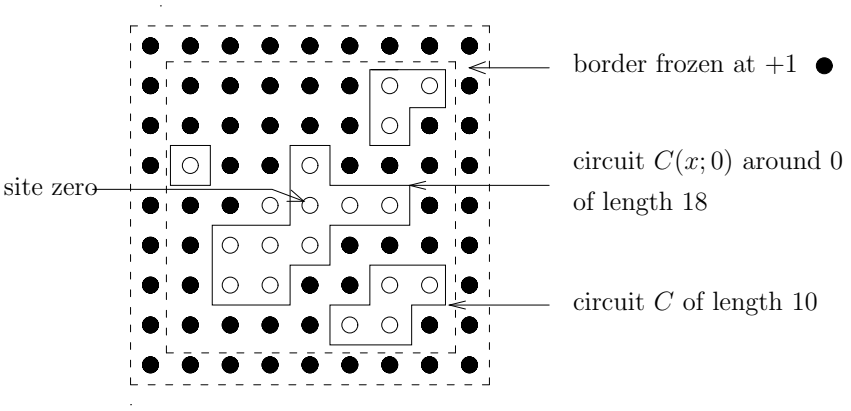
$$-\sum_{\langle v,w\rangle} x(v)x(w) = n_o(x) - n_e(x),$$

where $n_e(x)$ is the number of even bounds, and that $n_e(x) = M - n_o(x)$, where M is the total number of pair cliques. Therefore,

$$U(x) = 2\beta n_o(x) - M,$$

from which (9.26) follows. □

Before proceeding to the proof of the announced upper bound for $\pi_+^{(N)}(x(0) = -1)$, a few definitions are needed. Actually, no formal definition will be proposed; instead, the reader is referred to pictures. The figure below features *circuits* C of various lengths.



For a given configuration x, $C(x;0)$ denotes the circuit which is the boundary of the largest connected batch of sites with negative phases, containing site 0. It is a *circuit around* 0. If the phase at the central site is positive, then $C(x;0)$ is the empty set.

For a given configuration x, denote by $\widetilde{x}$ the configuration obtained by reversing all the phases inside circuit $C(x;0)$. For a given circuit C around 0,

$$\pi_+^{(N)}(C(x;0) = C) = \frac{\sum_{x\,;\,C(x;0)=C} e^{-2\beta n_o(x)}}{\sum_y e^{-2\beta n_o(y)}}.$$

But

$$\sum_z e^{-2\beta n_o(z)} \geq \sum_{y\,;\,C(y;0)=C} e^{-2\beta n_o(\widetilde{y})}$$

(one can always associate to a configuration y such that $C(y;0) = C$ the configuration $z = \widetilde{y}$, and therefore the sum on the right-hand side is a subsum of the left-hand side). Therefore,

$$\pi_+^{(N)}(C(x;0) = C) \leq \frac{\sum_{x;C(x;0)=C} e^{-2\beta n_o(x)}}{\sum_{x;C(x;0)=C} e^{-2\beta n_o(\widetilde{x})}}.$$

If x is such that $C(x;0) = C$, then $n_0(\widetilde{x}) = n_0(x) - L$, where L is the length of C, and therefore

$$\pi_+^{(N)}(C(x;0) = C) \leq e^{-2\beta L}.$$

In particular,

$$\pi_+^{(N)}(x(0) = -1) \leq \sum r(L) e^{-2\beta L},$$

where the latter summation is over all lengths L of circuits around 0, and $r(L)$ is the number of non-empty circuits around 0 of length L. The possible lengths are $4, 6, \ldots, 2f(N)$, where $f(N) \uparrow \infty$ as $N \uparrow \infty$. In order to bound $r(L)$ from above, observe that a circuit around 0 of length L must have at least one point at a distance smaller than or equal to $\frac{L}{2}$ of the central site 0. There are L^2 ways of selecting such a point, and then at most 4 ways of selecting the segment of C starting from this point, and then at most 3 ways of selecting the next connected segment, and so on, so that

$$r(L) \leq 4L^2 3^L.$$

Therefore,

$$\pi_+^{(N)}(x(0) = -1) \leq \sum_{L=4,6,\ldots} 4L^2 (3e^{-2\beta})^L.$$

Now, the series $\sum_{L=4,6,\ldots} L^2 x^L$ has a radius of convergence not less than 1, and therefore, if $3e^{-\beta}$ is small enough, or equivalently if T is large enough, $\pi_+^{(N)}(x(0) = -1) < \frac{1}{3}$ for all N.

9.3 Correlation in Random Fields

9.3.1 Increasing Events

The simplest correlation problem arising in a random field context is to determine if the spins of a given Ising model are positively correlated: is it more likely to have a positive spin at a given site when the spin at another given site is positive? More generally, given two events A and B, when can we assert that $P(A \,|\, B) \geq P(A)$? This section features a very powerful tool for this type of problem: Holley's inequality, which has for consequences two other important inequalities, Harris' inequality and the FKG inequality.

Let $E := \{0, 1\}^L$ where L is a positive integer. An element $x = (x_\ell\,,\ 1 \leq \ell \leq L) \in E$ is called a configuration. Denote by $\mathbf{0}$ the configuration with all $x_\ell = 0$ and by $\mathbf{1}$ that with all $x_\ell = 1$. Say $x \geq y$ if for all $1 \leq \ell \leq L$, $x_\ell \geq y_\ell$, with a similar definition for $x > y$ and $x = y$. The (Hamming) distance between $x \in E$ and $y \in E$ is the integer $d(x, y) := \sum_{\ell=1}^{L} 1_{\{x_\ell \neq y_\ell\}}$. For $1 \leq \ell \leq L$, let $E_\ell^0 := \{x \in E\,;\, x_\ell = 0\}$ and for any configuration $x \in E_\ell^0$, let $x + \ell$ denote the configuration y identical to x except for the ℓ-th coordinate, equal to 1. For any $x, y \in E$, call $x \wedge y$ the configuration defined by $(x \wedge y)_\ell = x_\ell \wedge y_\ell$ for all $1 \leq \ell \leq L$, with a similar definition for $x \vee y$.

The set $A \subseteq E$ is called increasing (resp., decreasing) if $x \in A$ implies that $y \in A$ for all $y \geq x$ (resp., $y \leq x$). Clearly, any non-empty increasing (resp., decreasing) set of configurations contains the configuration $\mathbf{1}$ (resp., $\mathbf{0}$).

EXAMPLE 9.3.1: BOND PERCOLATION, TAKE 1. Let L be the number of edges of some finite graph with N nodes. Edge ℓ in the configuration $x \in E$ is called open if $x_\ell = 1$. Nodes v and w are said to be connected in configuration x if there exists a path of open edges connecting them. This is denoted by $v \leftrightarrow w$. For fixed nodes v and w the set $A := \{x \in E\,;\, v \leftrightarrow w\}$ is an increasing set, and similarly for the

set B consisting of the configurations for which the number of nodes connected to v is larger than or equal to a fixed integer n.

Definition 9.3.2 *The function $f : E \to \mathbb{R}$ is called supermodular if*

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y) \text{ for all } x, y \in E\,. \qquad (\star)$$

A function $f : E \to \mathbb{R}$ is called submodular if $-f$ is supermodular.

Definition $(\star)$ is equivalent to

$$x \geq y \Longrightarrow f(x+\ell) - f(x) \geq f(y+\ell) - f(y) \text{ for all } x, y \in E_\ell^0, \text{ all } 1 \leq \ell \leq L\,. \quad (\dagger)$$

Proof. $(\star) \Rightarrow (\dagger)$: For all $1 \leq \ell \leq L$, $x \geq y$ implies $x + \ell = x \vee (y + \ell)$ and $y = x \wedge (y + \ell)$.

$(\dagger) \Rightarrow (\star)$: If $x \geq y$ the inequality is obvious. Otherwise, let $\ell_1, \ldots, \ell_n$ be an enumeration of the integers ℓ such that $x_\ell = 0$ and $y_\ell = 1$. Noting that $x \vee y = x + \ell_1 + \cdots + \ell_n$ and $x = (x \wedge y) + \ell_1 + \cdots + \ell_n$, $(\star)$ follows by n successive applications of $(\dagger)$. □

EXAMPLE 9.3.3: BOND PERCOLATION, TAKE 2. Denote by $C(x)$ the number of components in configuration x. Then C is supermodular. This is checked via $(\dagger)$. For this we note that for all $1 \leq \ell \leq L$ and all $x \in E_\ell^0$, $C(x + \ell) - C(x) = -1$ if the nodes of edge ℓ are disconnected in configuration x, and $C(x + \ell) - C(x) = 0$ otherwise. Now, for any $y \in E_\ell^0$ such that $x \geq y$, $C(x+\ell)-C(x) \geq C(y+\ell)-C(y)$, since whenever the nodes adjacent to ℓ are disconnected in configuration x, they are also disconnected in configuration y.

EXAMPLE 9.3.4: ISING MODEL, TAKE 4. Consider the Ising model with configuration space $E := \{+1, -1\}^N$ (site space $V := \{1, \ldots, N\}$, phase space $\Lambda := \{+1, -1\}$). The "spin" (phase) at site i is denoted by x_i. The energy function is $U(x) = -\sum_{i \sim j} x_i x_j$ and the corresponding Gibbs distribution is

$$\pi(x) = \frac{1}{Z(\beta)} e^{-\beta U(x)}\,,$$

where β is the inverse temperature and $Z(\beta) := \sum_{x \in E} e^{-\beta U(x)}$ is the partition function. The energy function is submodular, that is, for all x, $y \in E$,

$$U(x \vee y) + U(x \wedge y) \leq U(x) + U(y)\,.$$

Proof. Since the energy is the sum of the energies of the edges, it suffices to consider the situation where $N = 2$. The equality is obvious if the configurations

are ordered. By symmetry, the only case that remains to be checked is $x = (+1, -1)$ and $y = (-1, +1)$. Then $U(x) + U(y) = 2$. Also $x \wedge y = (-1, -1)$ and $x \vee y = (+1, +1)$, and therefore $U(x \wedge y) + U(x \vee y) = 0$. □

The magnetization function is defined by

$$M(x) := \frac{1}{N} \sum_{i=1}^{N} x_i \,.$$

The absolute magnetization $|M(x)|$ is supermodular (Exercise 9.4.16).

9.3.2 Holley's Inequality

P being a probability on E, write $P(x)$ for $P(\{x\})$.

Theorem 9.3.5 *(Holley, 1974) Let P and P' be probabilities on E, P strictly positive, such that*

$$P'(x \vee y)P(x \wedge y) \geq P'(x)P(y) \quad \textit{for all } x, y \in E \,. \tag{$\star\star$}$$

Then for any increasing set $A \subseteq E$,

$$P'(A) \geq P(A) \,.$$

Condition ($\star\star$) is equivalent to the following

$$x \geq y \Longrightarrow P'(x + \ell)P(y) \geq P'(x)P(y + \ell) \text{ for all } x, y \in E^0_\ell, \text{ all } \ell \,. \tag{$\dagger\dagger$}$$

(The proof is analogous to the proof of equivalence of ($\star$) and ($\dagger$) above.) It follows that the support of P' is an increasing set.

A proof of Holley's inequality is given in section 19.1.4.

Two important inequalities will be obtained as corollaries of Holley's inequality, namely the FKG inequality and the Harris inequality.

Definition 9.3.6 *The probability P on E is said to satisfy the lattice condition if for all $x, y \in E$,*

$$P(x \vee y)P(x \wedge y) \geq P(x)P(y) \,.$$

Corollary 9.3.7 *(Fortuin, Kasteleyn and Ginibre, 1971) Let P be a positive probability on E satisfying the lattice condition. Then for any increasing sets $A, B \subseteq E$,*

$$P(A \cap B) \geq P(A)P(B) \,. \tag{9.27}$$

Equivalently $P(A \,|\, B) \geq P(A)$. In words, the occurence of event B increases the occurence of event A.

Proof. Define P' by

$$P'(\cdot) := P(\cdot \,|\, B)\,.$$

If condition $(\star\star)$ of Theorem 9.3.5 is satisfied, then Holley's inequality reads

$$P'(A) = P(A \,|\, B) \geq P(A)\,,$$

which is (9.27). It remains to prove $(\star\star)$.

If $x \notin B$, then $P'(x) = 0$ and the inequality is trivial. If $x \in B$, $P'(x) = \frac{P(x)}{P(B)}$. Moreover, $x \vee y \in B$ since B is an increasing set, and $P'(x \vee y) = \frac{P(x \vee y)}{P(B)}$. The FKG inequality is then an immediate consequence of the lattice condition. □

EXAMPLE 9.3.8: ISING MODEL, TAKE 5. By the submodularity of the energy function, the distribution π satisfies the lattice condition. Therefore, by the FKG inequality, for any increasing sets A and B, $P(A \,|\, B) \geq P(A)$. In particular (with $A := \{x_i = 1\}$ and $B := \{x_j = 1\}$) the spins are positively correlated. In words, it is more likely to have a positive spin at site i given that the spin at site j is positive.

Definition 9.3.9 *The probability P on E is said to be of product form if for all $x \in E$*

$$P(x) = \prod_{\ell=1}^{L} p(x_\ell)\,,$$

where p is a probability distribution on $\{0, 1\}$.

Corollary 9.3.10 *(Harris, 1960) Let P be a positive probability of product form on E. Then (9.27) holds true for any increasing sets $A, B \subseteq E$.*

Proof. Harris' inequality is a consequence of the FKG inequality, as follows from the following remark: a probability P on E of product form satisfies the lattice condition. Indeed:

$$P(x \vee y)P(x \wedge y) = \left(\prod_{\{\ell\,;\, x_\ell = y_\ell\}} p(x_\ell)^2 \right) \times \left(\prod_{\{\ell\,;\, x_\ell \neq y_\ell\}} p(x_\ell)p(y_\ell) \right) = P(x)P(y)\,.$$

□

9.3.3 The Potts and Fortuin–Kasteleyn Models

The Potts model (Potts, 1952) is a natural extension of the Ising model. The set of vertices is $V = \{1, 2, \ldots, N\}$ and the phase space is $\Lambda = \{1, 2, \ldots, q\}$, each phase representing a "color". Denote by z the typical configuration. Potts's model corresponds to the probability P on Λ^V given by

$$P(z) := \frac{1}{Z(\beta)} e^{-\beta U(z)},$$

where $\beta > 0$ and

$$U(z) := \sum_{i \sim j} 1_{\{z_i \neq z_j\}}$$

(an edge being counted only once) and $Z(\beta)$ is the normalizing factor. When $q = 2$, this corresponds to the Ising model. In fact, identifying the spins $+1$ and -1 with the colours 1 and 2, the energy in the Ising model is

$$-\sum_{i \sim j} z_i z_j = \sum_{i \sim j} 1_{\{z_i \neq z_j\}} - \sum_{i \sim j} 1_{\{z_i = z_j\}} = 2U(z) - L\,.$$

Let $N_s(z) := \sum_{i=1}^{N} 1_{\{z_i = s\}}$ be the number of sites of colour s. Let

$$Q(z) := \frac{\frac{q}{N^2} \sum_{s \in \Lambda} N_s(z)^2 - 1}{q - 1}\,.$$

This index (called the Simpson index) quantifies the concentration of colour distribution in state z. We have that $0 \leq Q(z) \leq 1$, the value 0 corresponding to the uniform distribution ($N_s(z) = \frac{N}{q}$ for all colours s) and the value 1 corresponding to the monochromatic state ($N_s(z) = N$ for some colour s). Since there is no natural order among colours, the tools of comparison used previously are not directly useful. The Fortuin–Kasteleyn bound percolation model will allow us to bypass this limitation.

We now describe the Fortuin–Kasteleyn bound percolation model (Fortuin, 1972; Fortuin and Kasteleyn, 1972). Consider a graph with N nodes and L edges. The configuration space is now $E := \{0, 1\}^L$. A configuration $x = (x_1, \ldots, x_L)$ has the following interpretation: if $x_\ell = 1$, edge ℓ is called "open". Denote by

$$O(x) := \sum_{\ell=1}^{L} x_\ell$$

the number of open edges in configuration x, and by $C(x)$ the number of connected components in the graph restricted to open edges. Define for $q \in \mathbb{N}_+$ and $p \in (0, 1)$ the probability

$$P_{p,q}(x) := \frac{1}{Z_{p,q}} p^{O(x)} (1 - p)^{L - O(x)} q^{C(x)}\,, \ (x \in \{0, 1\}^L)$$

where $Z_{p,q}$ is the normalizing factor. The case $q = 1$ corresponds to independent percolation, where an edge is accepted with probability p independently of the

others. Otherwise the model is one of correlated percolation. Since $O(x \wedge y) + O(x \vee y) = O(x) + O(y)$ and since the function C is supermodular (Example 9.3.3), it follows that $P_{p,q}$ satisfies the lattice condition. Therefore, by the FKG inequality, the open edges are positively correlated (see Example 9.3.3).

Let $P = P_{p,q}$ and $P' = P_{p',q'}$. Condition (††) after the statement of Theorem 9.3.5 reads

$$\frac{p'}{1-p'}(q')^{C(x+\ell)-C(x)} \geq \frac{p}{1-p}q^{C(y+\ell)-C(y)}\,. \qquad (\star)$$

For $x, y \in E_\ell^0$ and $x \geq y$, the couple $(C(x+\ell) - C(x), C(y+\ell) - C(y))$ takes the values $(0,0)$, $(0,-1)$ and $(-1,+1)$ (if the neighbours of edge ℓ are disconnected in configuration x, they are also disconnected in configuration y). Therefore condition $(\star)$ is satisfied, and consequently the inequality

$$P_{p',q'} \geq P_{p,q}$$

holds, in the following cases:

(a) $p' \geq p$ and $q' = q$,

(b) $p' = p$ and $q' \leq q$ and

(c) $\frac{p'}{q'(1-p')} \geq \frac{p}{q(1-p)}$ and $q' \geq q$.

(Note that the two simultaneous conditions in (c) imply $p' \geq p$.)

Coupling F–K and Potts Models

By this, we mean the construction of a random field Z on the N vertices of the graph with phase space $\Lambda = \{1, \ldots, q\}$ and of a random field X on the L edges of the graph with phase space $\{0,1\}$ in such a way that Z is a Potts random field and X is a F–K random field. This of course could be done by constructing these random fields to be independent, but we shall not do this and make the construction in such a way that these fields are dependent.

For this, we start with a F–K model which we enrich by assigning independently to each connected component a colour in $\{1, \ldots, q\}$. This will have the effect of colouring each vertex of the subgraph corresponding to a given component. Let z_i be the colour received by vertex i by this procedure, and let $z = (z_1, \ldots, z_N)$ be the corresponding vertex configuration. We now have an extended configuration (x, z) in $E \times \Lambda^N = \{0,1\}^L \times \{1, \ldots, q\}^N$. The probability of an extended configuration is, since there are $q^{C(x)}$ different and equiprobable ways of colouring the connected components,

$$P(x,z) = \frac{1}{Z_{p,q}}p^{O(x)}(1-p)^{L-O(x)}q^{C(x)}\frac{1}{q^{C(x)}} = \frac{1}{Z_{p,q}}p^{O(x)}(1-p)^{L-O(x)} \qquad (9.28)$$

where $Z_{p,q}$ is the normalizing factor. Note that the configuration space is only a part of $E \times \Lambda^N$, namely the set $\mathcal{A}$ of "admissible configurations", that is configurations (x, z) such that $z_i = z_j$ for all edges $\ell = \langle i, j \rangle$ such that $z_\ell = 1$. Consequently the model (9.28) is not an independent correlation model.

Now start with a Potts random field, and define the random field on $\{0,1\}^L \times \{1,\dots,q\}^N$ as follows. Accept an edge $\ell = \langle i,j\rangle$ whose extremities are of the same colour with probability p, independently of everything else. A configuration (x,z) therefore has the probability

$$P(x,z) = \frac{1}{Z} e^{-\beta U(z)} p^{O(x)} (1-p)^{L-U(z)-O(x)} .$$

In the case $1-p = e^{-\beta}$,

$$P(x,z) = \frac{1}{Z} p^{O(x)} (1-p)^{L-O(x)} .$$

This coincides with the probability distribution (9.28). In particular

$$\begin{aligned} P(x) = \sum_{z;(x,z)\in\mathcal{A}} P(x,z) &= \frac{1}{Z} p^{O(x)} (1-p)^{L-O(x)} \sum_{z;(x,z)\in\mathcal{A}} 1 \\ &= \frac{1}{Z} p^{O(x)} (1-p)^{L-O(x)} q^{C(x)} \end{aligned}$$

is the distribution of a F–K random field on the vertices, and

$$P(z) = \sum_{x;(x,z)\in\mathcal{A}} P(x,z) = \frac{1}{Z} e^{-\beta U(z)} \sum_{x;(x,z)\in\mathcal{A}} p^{O(x)} (1-p)^{L-U(z)-O(x)} = \frac{1}{Z} e^{-\beta U(z)} ,$$

is the distribution of a Potts random field on the edges.

Let $P(z_i = z_j)$ be the probability that vertices i and j have the same colour in the Potts model, and therefore in the coupled Potts–Fortuin–Kasteleyn (P–F–K) model. Denote by $i \leftrightarrow j$ the fact that vertices i and j are in the same component. Note that i and j have the same colour if and only if one of the two disjoint events occur:

(a) $i \leftrightarrow j$ (probability $P_{p,q}(i \leftrightarrow j)$) .

(b) $i \nleftrightarrow j$ and nevertheless i and j have the same colour (probability $P_{p,q}(i \nleftrightarrow j) \times \frac{1}{q}$) .

Therefore $P(x_i = x_j) = P_{p,q}(i \leftrightarrow j) + \frac{1}{q} P_{p,q}(i \nleftrightarrow j)$, which gives

$$P(x_i = x_j) = \frac{1}{q} + \left(1 - \frac{1}{q}\right) P_{p,q}(i \leftrightarrow j) .$$

It was shown before that if $p' \geq p$, then $P_{p',q} \geq P_{p,q}$. Now, $i \leftrightarrow j$ is an increasing event of the F–K model, therefore the probability that two vertices have the same colour increases in the Potts model with p, that is with β, since the coupling was established with $p = 1 - e^{-\beta}$.

EXAMPLE 9.3.11: DECREASING AND INCREASING EVENTS IN THE POTTS MODEL. Observing that in the Potts model, the energy and the Simpson index can be written respectively as

$$U(z) = \sum_{i\sim j} 1_{\{z_i \neq z_j\}} = L - \sum_{i\sim j} 1_{\{z_i = z_j\}}$$

and

$$Q(z) = \frac{\frac{q}{N^2}\sum_{i,j} 1_{\{z_i=z_j\}} - 1}{q-1},$$

we conclude that in the Potts model the energy decreases with β whereas the Simpson index increases with β.

Books for Further Information

[Kinderman and Snell, 1980] is a pedagogical introduction to the subject. For examples of Gibbs models for which expressions of the partition function are available, see [Baxter, 1982]. [Winkler, 1995] is entirely devoted to applications in image processing. The Potts and Fortuin–Kasteleyn percolation models of Subsection 9.3.3 are treated, in the finite *and* infinite cases, in [Werner, 2009]. [Grimmett, 2010] is another important reference at the research level.

9.4 Exercises

Exercise 9.4.1. IID RANDOM FIELDS
A. Let $(Z(v)\ (v \in V)$ be a family of IID random variables with values in $\{-1, +1\}$ indexed by a finite set V, with $P(Z(v) = -1) = p \in (0, 1)$. Show that

$$P(Z = z) = Ke^{\gamma \sum_{v\in V} z(s)},$$

for some constants γ and K to be identified.

B. Do the same when the $Z(v)$s take their values in $\{0, 1\}$, with $P(Z(v) = 0) = p \in (0, 1)$.

Exercise 9.4.2. TWO-STATE HMC AS GIBBS FIELD
Consider an HMC $\{X_n\}_{n\geq 0}$ with state space $E = \{-1, 1\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \qquad (\alpha, \beta \in (0, 1))$$

and with the stationary initial distribution

$$(\nu_0, \nu_1) = \frac{1}{\alpha+\beta}(\beta, \alpha).$$

Give a representation of $(X_0, \ldots, X_N)$ as a MRF. What is the normalized potential with respect to phase 1?

Exercise 9.4.3. POISSONIAN VERSION OF BESAG'S MODEL
Consider the model of Example 9.1.3 with the following modifications. Firstly, the phase space is $\Lambda = \mathbb{N}$, and secondly, the potential is now

$$V_C(x) = \begin{cases} -\log(g(x(v))) + \alpha_1 x(v) & \text{if } C = \{v\} \in \mathcal{C}_1, \\ \alpha_j x(v)x(w) & \text{if } C = \{v, w\} \in \mathcal{C}_j, \end{cases}$$

where $\alpha_j \in \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ is strictly positive. As in the autobinomial model, for any clique C not of the type $\mathcal{C}_j$, $V_C \equiv 0$. For what function g do we have

$$\pi^s(x) = e^{-\rho} \frac{\rho^{x(v)}}{x(v)!},$$

where $\rho = e^{-\langle \alpha, b \rangle}$, and where $\langle \alpha, b \rangle$ is as in Subsection 9.1.3? (This model is the *auto-Poisson model*.)

Exercise 9.4.4. ISING ON THE TORE
(Baxter, 1965) Consider the classical Ising model of Example 9.1.7, except that the site space $V = \{1, 2, \ldots, N\}$ consists of N points arranged in this order on a circle. The neighbors of site i are $i+1$ and $i-1$, with the convention that site $N+1$ is site 1. The phase space is $\Lambda = \{+1, -1\}$. Compute the partition function. Hint: express the normalizing constant Z_N in terms of the N-th power of the matrix

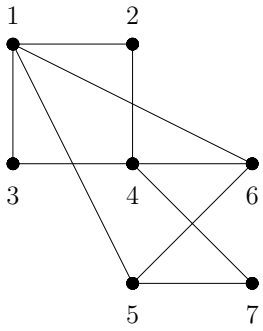
$$R = \begin{pmatrix} R(+1,+1) & R(+1,-1) \\ R(-1,+1) & R(-1,-1) \end{pmatrix} = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix},$$

where $K := \frac{J}{kT}$ and $h := \frac{H}{kT}$.

Exercise 9.4.5. CLIQUES AND BOUNDARIES
Define on $V = \mathbb{Z}^2$ the two neighborhood systems of the figure below. Describe the corresponding cliques and give the boundary of a 3×3 square for each case.

Exercise 9.4.6. JUST AN EXERCISE
Consider the nonoriented graph on $V = \{1, 2, 3, 4, 5, 6, 7\}$ in the figure below. Let the phase space be $\Lambda = \{-1, +1\}$. For a configuration $x \in \Lambda^V$, denote by $n(x)$ the number of positive *bonds*, that is, the number of edges of the graph for which the phases of the adjacent sites coincide. Define a probability distribution π on Λ^V by $\pi(x) = \frac{e^{-n(x)}}{Z}$. Give the value of the partition function Z and the local characteristics of this random field.

Exercise 9.4.7. THE MARKOV PROPERTY
Let V be a finite set of sites and Λ a finite set of phases. Let $\{X(v)\}_{v \in V}$ be a Markov field with values in Λ^V and admitting a Gibbsian description in terms



of the symmetric relation $\sim$, with Gibbs potential $\{V_C\}_{C\subset V}$. Prove that for all subsets A, B, of S such that

$$A \cap B = \varnothing$$

it holds that for all $x \in \Lambda^V$,

$$P(X(A) = x(A) \mid X(B) = x(B)) = P(X(A) = x(A) \mid X(\partial\bar{B}) = x(\partial\bar{B}))\,.$$

Exercise 9.4.8. FROZEN SITES
Let V be a finite set of sites and Λ a finite set of phases. Let $\{X(v)\}_{v\in V}$ be a Markov field with values in Λ^V and admitting a Gibbsian description in terms of the neighborhood structure $\sim$, with potential $\{V_C\}_{C\subset V}$. Let $A + B = V$ be a partition of the site. Fix $x(A) = \underline{x}(A)$ and define the distribution π_A on Λ^B by

$$\pi_A(x(B)) = \frac{e^{-U(\underline{x}(A),x(B))}}{\sum_{y(B)\in\Lambda^B} e^{-U(\underline{x}(A),y(B))}},$$

where U is the energy function associated with the potential $\{V_C\}_{C\subset V}$. Show that

$$\pi_A(x(B)) = P(X(B) = x(B) \mid X(A) = \underline{x}(A))$$

and that $\pi_A(x(B))$ is a Gibbs distribution for which you will give the neighborhood system and the corresponding cliques, as well as the local characteristics. (A Markov field with values in Λ^B and with the distribution π_A is called a version of $\{X_v\}_{v\in V}$, *frozen* on A at value $\underline{x}(A)$, or *clamped* at $\underline{x}(A)$.)

Exercise 9.4.9. HARD-CORE MODEL
Consider a random field with finite site space V and phase space $\Lambda := \{0, 1\}$ (with the interpretation that if $x(v) = 1$, the site v is "occupied" and "vacant" otherwise) evolving in time. The resulting sequence $\{X_n\}_{n\geq 0}$ is a HMC with state space F, the subset of $E = \{0, 1\}^V$ consisting of the configurations x such that for all $v \in V$, $x(v) = 1$ implies that $x(w) = 0$ for all $w \sim v$. The updating procedure is the following. If the current configuration is x, choose a site v uniformly at random,

and if no neighbour of v is occupied, make v occupied or vacant equiprobably. Show that the HMC so described is irreducible and that its stationary distribution is the uniform distribution on F.

Exercise 9.4.10. MONOTONICITY PROPERTY OF THE GIBBS SAMPLER
Let μ be an arbitrary probability measure on Λ^V and let ν be the probability measure obtained by applying the Gibbs sampler at an arbitrary site $v \in V$. Show that $d_V(\nu, \pi) \leq d_V(\mu, \pi)$.

Exercise 9.4.11. NEURAL NETWORK
The graph structure is as in the Ising model, but now the phase space is $\Lambda = \{0, 1\}$. A site v is interpreted as being a *neuron* that is *excited* if $x(v) = 1$ and *inhibited* if $x(v) = 0$. If $w \sim v$, one says that there is a *synapse* from v to w, and such a synapse has a *strength* σ_{vw}. If $\sigma_{vw} > 0$, one says that the synapse is *excitatory*; otherwise it is called *inhibitory*. The energy function is

$$U(x) = \sum_{v \in V} \sum_{w\,;\, w \sim v} \sigma_{wv} x(w) x(v) - \sum_{v \in V} h_v x(v),$$

where h_v is called the *threshold* of neuron v (we shall understand why later).

(a) Describe the corresponding Gibbs potential.

(b) Give the local characteristics.

(c) Describe the Gibbs sampling algorithm.

(d) Show that this procedure can also be described in terms of a random *threshold jitter* Σ with the cumulative distribution function

$$P(\Sigma \leq a) = \frac{e^{-a/T}}{1 + e^{-a/T}}\,, \tag{9.29}$$

the Gibbs sampler selecting phase 0 if

$$\sum_{w \in \mathcal{N}_v} (\sigma_{wv} + \sigma_{vw}) x(w) < h_v + \Sigma\,,$$

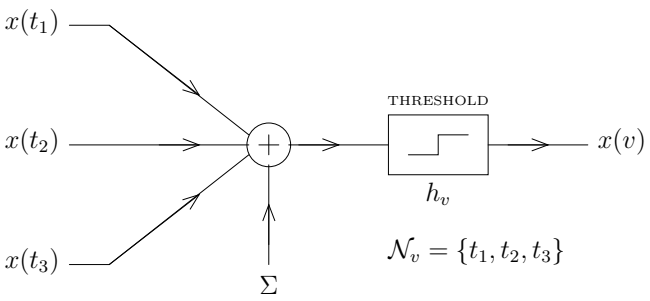
and 1 otherwise. One may interpret h_v as the *nominal threshold* at site v and $h_v + \Sigma$ as the actual (random) threshold. Also the quantity $\sum_{w \in \mathcal{N}_v} (\sigma_{wv} + \sigma_{vw}) x(w)$ is the input into neuron v. Thus the excitation of neuron v is obtained by comparing its input to a random threshold (see the figure).

Exercise 9.4.12. THERMODYNAMICS, I
Let

$$\pi_T(x) = \frac{1}{Z} e^{\frac{-E(x)}{kT}}$$

be a Gibbs distribution on the finite space $E = \Lambda^V$. Here Z is short for Z_T, and $E(x)$ is the energy of physics, differing from $U(x)$ by the Boltzmann constant k.

For any function $f : E \to \mathbb{R}$, define



Jitter sampling of a neural network

$$\langle f \rangle = \sum_{x \in E} \pi(x) f(x).$$

In particular, the *internal energy* is

$$U = \langle E \rangle = \sum_{x \in E} \pi(x) E(x).$$

The *free energy* F is defined by

$$F = -kT \ \log(Z).$$

Show that

$$U = -T^2 \frac{\partial}{\partial T} \left(\frac{F}{T} \right).$$

(This is in agreement with standard thermodynamics.)

Exercise 9.4.13. THERMODYNAMICS, II
(Continuation of Exercise 9.4.12.) For the Ising model, take

$$E(x) = E_0(x) + E_1(x),$$

where $E_0(x)$ is the interaction energy, assumed symmetric, i.e., $E_0(-x) = E_0(x)$, and

$$E_1(x) = -Hm(x),$$

where

$$m(x) = \sum_{i=1}^{N} x(i)$$

is the magnetic moment of the configuration $x = (x(1), \ldots, x(N))$ (recall that $S = \{1, 2, \ldots, N\}$), and H is the external magnetic field. The partition function, still denoted by Z, is now a function of T and H. The free energy *per site* is

$$f(H, T) = -kT \frac{1}{N} \log(Z),$$

whereas the *magnetization*

$$M(H,T) = \frac{1}{N}\langle m \rangle$$

is the average magnetic moment per site.

Show that

$$M(H,T) = -\frac{\partial}{\partial H} f(H,T)$$

and

$$\frac{\partial M}{\partial H} = \frac{1}{NkT}\left(\langle m^2 \rangle - \langle m \rangle^2\right).$$

In particular,

$$\frac{\partial M}{\partial H} \geq 0.$$

Exercise 9.4.14. THERMODYNAMICS, III
(Continuation of Exercise 9.4.13.) Compute $\lim_{N\uparrow\infty} M(H,T)$ for the Ising model on the torus (Exercise 9.4.4). Observe that this limit, as a function of H, is analytic, and null at $H = 0$. In particular, in this model, there is no phase transition.

Exercise 9.4.15. THE SPINS ARE POSITIVELY CORRELATED
Consider the Ising model with state space $E = \{-1,+1\}^N$ and energy function $U(x) := \sum_{\langle v,w \rangle} x(v)x(w)$ and probability distribution $\pi_T(x) = \frac{1}{Z(T)} e^{-\frac{1}{T}U(x)}$. Define the magnetization function $m(x) := \frac{1}{N}\sum_{v\in E} x(v)$.

(a) Prove that the energy function U is a submodular function and that the absolute magnetization function $|m|$ is a supermodular function.

(b) Show that the Gibbs distribution π satisfies the lattice condition (Definition 9.3.6). Using a famous inequality, show that "the spins are positively correlated" in the sense that for any sites v and w, given that $x(v) = +1$, $x(w) = +1$ is more likely than $x(w) = -1$.

Exercise 9.4.16. SUPER-MODULARITY OF THE ABSOLUTE MAGNETIZATION
In Example 9.3.4, prove that the absolute magnetization is supermodular, that is

$$|M(x \vee y)| + |M(x \wedge y)| \geq |M(x)| + |M(y)|.$$

Exercise 9.4.17.
Show that the Fortuin–Kasteleyn percolation model $P_{p,q}$ is bounded above and below by independent percolation models, that is for some α and β to be determined in function of p and q,

$$P_{\alpha,1} \leq P_{p,q} \leq P_{\beta,1}.$$

Exercise 9.4.18. THE LORENZ INEQUALITY
Let $h : \{1,\ldots,q\}^N \to \mathbb{R}$ be a supermodular function, and let $X_1,\ldots,X_N$ be IID random variables with values in $\{1,\ldots,q\}$. Prove the following inequality (Lorenz's inequality):

$$E\left[h(X_1,X_2,\ldots,X_N)\right] \leq E\left[h(X_1,X_1,\ldots,X_1)\right]$$

Hint: Do the case $N = 2$ first, and then proceed by induction.

Exercise 9.4.19. LORENZ AND FKG
Recall the following elementary form of the FKG inequality. Let $E \subseteq \mathbb{R}$ and let $f, g : E^n \to \mathbb{R}$ be two bounded functions that are non-decreasing in each of their arguments. Let $X_1^n := (X_1, \ldots, X_n)$ be a vector of independent variables with values in E. Then (Formula (2.22))

$$E\left[f(X_0^n)g(X_0^n)\right] \geq E\left[f(X_0^n)\right] E\left[g(X_0^n)\right] .$$

Show that this is a particular case of Lorenz's inequality of Exercise 9.4.18.

Chapter 10

Random Graphs

10.1 Branching Trees

10.1.1 Extinction and Survival

This section features what is perhaps the earliest non-trivial result concerning the evolution of a stochastic process, namely the Galton–Watson branching process. It involves a graph, here a "genealogical" tree. Francis Galton posed in 1873, in the *Educational Times*, the question of evaluating the survival probability of a given line of English peerage, and thereby initiated research in an important domain of applied probability. Branching processes have applications in numerous fields, for instance in nuclear science (because of the analogy between the growth of families and nuclear chain reactions), in chemistry (chain reactions again) and in biology (survival of a mutant gene)[1]. The results of the current section will be used in section 10.2.3 on the emergence of a giant component in Erdös–Rényi random graphs.

The recurrence equation

$$X_{n+1} = \sum_{k=1}^{X_n} Z_{n+1}^{(k)} \tag{10.1}$$

($X_{n+1} = 0$ if $X_n = 0$), where $\{Z_n^{(j)}\}_{n\geq 1, j\geq 1}$ is an IID collection of integer-valued random variables with common generating function

$$g(z) := E\left[z^Z\right] = \sum_{n\geq 0} a_n z^n$$

and independent of the integer-valued random variable X_0, defines a stochastic process $\{X_n\}_{n\geq 0}$ called a branching process. It may be interpreted as follows: X_n is the number of individuals in the n-th generation of a given population (humans, particles, etc.). Individual number k of the n-th generation gives birth to $Z_{n+1}^{(k)}$ descendents, and this accounts for (10.1). The random variable X_0 is the number of ancestors. The appelation "branching process" refers to the original preoccupation of Francis Galton in terms of a genealogical tree (see the figure). This process, also

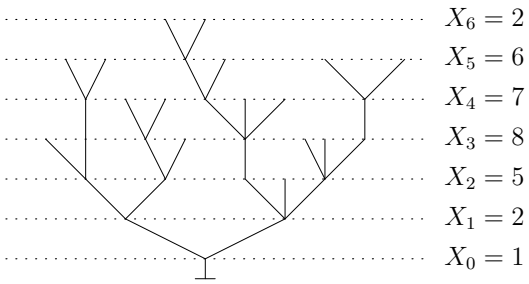
[1] See the historical remarks concerning the applications and evolution of the field in [Harris, 1963].

P. Brémaud, *Discrete Probability Models and Methods*,
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_10

called the Galton–Watson process, is in view of the recurrence equation (10.1) and the independence assumptions a homogeneous Markov chain (Theorem 6.1.4).

EXAMPLE 10.1.1: THE REPAIR SHOP, TAKE 2. The repair shop model has an interesting connection with branching processes. The first busy cycle length is the first time n at which there is no machine left in the facility. We may suppose that $Z_1 > 0$, by which it is meant that observation starts at time 1. It takes $X_1 = Z_1$ units of time before one can start the service of the X_2 machines arriving during the time these X_1 machines are repaired. Then it takes X_3 units of time before one can start the service of the machines arriving during the time these X_2 machines are repaired, and so on. This defines a sequence $\{X_n\}_{n\geq 1}$ satisfying the relation (10.1) as long as $X_n > 0$. Here $X_0 := 1$ and $\{Z_n^{(j)}\}_{n\geq 1, j\geq 1}$ is an IID collection of random variables with the same distribution as Z_1. Letting τ be the first (positive) time at which $X_n = 0$, the repair service facility is empty for the first time at time $\sum_{i=1}^{\tau-1} X_i$. Therefore, the probability of eventually having at least one "day off" for the mechanics is the probability of extinction of a branching process $\{X_n\}_{n\geq 1}$ whose typical offspring has the same distribution as Z_1.

The primary quantity of interest is the extinction probability $P(\mathcal{E})$, that is, the probability of absorption of the branching process into state 0.



Sample tree of a branching process

Theorem 10.1.2 *When there is just one ancestor ($X_0 = 1$),*

(a) $P(X_{n+1} = 0) = g(P(X_n = 0))$,

(b) $P(\mathcal{E}) = g(P(\mathcal{E}))$, *and*

(c) if $m := E[Z] < 1$, *the probability of extinction is* 1, *whereas if* $m > 1$, *the probability of extinction is* < 1 *and* > 0.

Proof.

The trivial cases $P(Z = 0) = 0$, $P(Z = 0) = 1$ and $P(Z \geq 2) = 0$ are excluded from the analysis.

(a) Let ψ_n be the generating function of X_n. Since X_n is independent of the $Z_{n+1}^{(k)}$'s, by Theorem 2.2.10,

$$\psi_{n+1}(z) = \psi_n(g(z)).$$

Iterating this equality, we obtain $\psi_{n+1}(z) = \psi_0(g^{(n+1)}(z))$, where $g^{(n)}$ is the n-th iterate of g. Since there is only *one ancestor*, $\psi_0(z) = z$, and therefore $\psi_{n+1}(z) = g^{(n+1)}(z) = g(g^{(n)}(z))$, that is,

$$\psi_{n+1}(z) = g(\psi_n(z)).$$

In particular, since $\psi_n(0) = P(X_n = 0)$, (a) is proved.

(b) An extinction occurs if and only if at some time n (and then for all subsequent times) $X_n = 0$. Therefore

$$\mathcal{E} = \cup_{n=1}^{\infty} \{X_n = 0\}.$$

Since $X_n = 0$ implies $X_{n+1} = 0$, the sequence of events $\{X_n = 0\}_{n\geq 1}$ is non-decreasing, and therefore, by monotone sequential continuity,

$$P(\mathcal{E}) = \lim_{n\uparrow\infty} P(X_n = 0).$$

The generating function g is continuous, and therefore from (a) and the last equation, the probability of extinction satisfies (b).

(c) By Theorem 2.2.8, recalling that the trivial cases where $P(Z = 0) = 1$ or $P(Z \geq 2) = 0$ have been eliminated, we have that

(α) if $E[Z] \leq 1$, the only solution of $x = g(x)$ in $[0, 1]$ is 1, and therefore $P(\mathcal{E}) = 1$. The branching process eventually becomes extinct, and

(β) if $E[Z] > 1$, there are two solutions of $x = g(x)$ in $[0, 1]$, 1 and x_0 such that $0 < x_0 < 1$. From the strict convexity and monotonicity of $g : [0, 1] \to [0, 1]$, it follows that the sequence $y_n = P(X_n = 0)$ that satisfies $y_0 = 0$ and $y_{n+1} = g(y_n)$ converges increasingly to x_0. In particular, when the mean number of descendants $E[Z]$ is strictly larger than 1, $P(\mathcal{E}) \in (0, 1)$. □

EXAMPLE 10.1.3: EXTINCTION PROBABILITY FOR A POISSON OFFSPRING. Take for the offspring distribution the Poisson distribution with mean $\lambda > 0$ whose generating function is $g(x) = e^{\lambda(x-1)}$. Suppose that $\lambda > 1$ (the supercritical case). The probability of extinction $P(\mathcal{E})$ is the unique solution in $(0, 1)$ of

$$x = e^{\lambda(x-1)}.$$

EXAMPLE 10.1.4: EXTINCTION PROBABILITY FOR A BINOMIAL OFFSPRING. Take for the offspring distribution the binomial distribution $\mathcal{B}(N, p)$, with $0 <$

$p < 1$. Its mean is $m = Np$ and its generating function is $g(x) = (px + (1-p))^N$. Suppose that $Np > 1$ (the supercritical case). The probability of extinction $P(\mathcal{E})$ is the unique solution in $(0,1)$ of

$$x = (px + (1-p))^N .$$

Example 10.1.5: Poisson branching as the limit of binomial branching. Suppose now that $p = \frac{\lambda}{N}$ with $\lambda > 1$ (therefore we are in the supercritical case) and the probability of extinction is given by the unique solution in $(0,1)$ of

$$x = \left(\frac{\lambda}{N}x + (1 - \frac{\lambda}{N})\right)^N = \left(1 - \frac{\lambda}{N}(1-x)\right)^N .$$

Letting $N \uparrow \infty$, we see that the right-hand side tends from below ($1 - x \le e^{-x}$) to the generating function of a Poisson variable with mean λ. Using this fact and the concavity of the generating functions, it follows that the probability of extinction also tends to the probability of extinction relative to the Poisson distribution.

One-by-one Exploration

The random tree corresponding to a branching process can be explored in several ways. One way is generation by generation and corresponds to the classical construction of the Galton–Watson process given above. There is an alternative way that will be useful in a few lines. At step n of the exploration, we have a set of active vertices $\mathcal{A}_n$ and a set of explored vertices $\mathcal{B}_n$. At time 0 there is one active vertex, the root of the branching tree, so that $\mathcal{A}_0 = \{\text{root}\}$, and no vertex has been explored yet: $\mathcal{B}_0 = \varnothing$. At step $n \ge 1$, one chooses a vertex v_{n-1} among the vertices active at time n (those in $\mathcal{A}_{n-1}$), and this vertex is added to the set of explored vertices, that is, $\mathcal{B}_n = \mathcal{B}_{n-1} \cup \{v_{n-1}\}$, and it is deactivated, whereas its children become active. Therefore, denoting by ξ_n the number of children of v_{n-1} and by A_n the cardinality of $\mathcal{A}_n$,

$$A_0 = 1 \text{ and } A_n = A_{n-1} - 1 + \xi_n$$

as long as $A_{n-1} > 0$. The exploration stops when there are no active vertices left, at time $Y = \inf\{n > 0;\, A_n = 0\}$, which is the size of the branching tree. By induction, as long as $A_{n-1} > 0$,

$$A_n = 1 - n + \sum_{i=1}^{n} \xi_i .$$

The one-by-one exploration procedure is summarized by the history of the branching process, that is, the random string

$$H = (\xi_1, \ldots, \xi_Y)$$

taking its values in the subset F of $\mathbb{N}^* := \left(\cup_{k\geq 1}\mathbb{N}^k\right) \cup \mathbb{N}^\infty$, determined by the following constraints. (a) If $x = (x_1, x_2, \ldots, x_k) \in \mathbb{N}^k$, $1 - \sum_{i=1}^n x_i - n > 0$ for all $n \leq k$ and $1 - \sum_{i=1}^k x_i - k = 0$, and (b) if $x = (x_1, x_2, \ldots) \in \mathbb{N}^\infty$, $1 - \sum_{i=1}^n x_i - n > 0$ for all $n \geq 1$. Finite k's correspond to histories with extinction, whereas $x \in \mathbb{N}^\infty$ represents a history without extinction.

For any sequence $(x_1, \ldots, x_k) \in F \cap \mathbb{N}^k$,

$$P(H = (x_1, \ldots, x_k)) = \prod_{i=1}^k a_{x_i}\,.$$

Conditioning by Extinction

The following question is of interest: what is the probability distribution of the history of a supercritical branching conditioned by the event that extinction occurs?

Theorem 10.1.6 *Let $\{a_k\}_{k\geq 0}$ be a supercritical offspring distribution, that is such that $\sum_{k\geq 0} k a_k > 1$. Let g_a be its generating function and $P(\mathcal{E})$ the corresponding probability of extinction, that is, the unique solution in $(0,1)$ of $P(\mathcal{E}) = g_a(P(\mathcal{E}))$. The distribution of the branching process conditioned on extinction is the same as the distribution of a subcritical branching process with offspring distribution*

$$b_k := a_k P(\mathcal{E})^{k-1} \qquad (k \geq 0). \tag{10.2}$$

Proof. We first check that (10.2) defines a probability distribution. In fact,

$$P(\mathcal{E}) = \sum_{k\geq 0} a_k P(\mathcal{E})^k = P(\mathcal{E}) \sum_{k\geq 0} b_k\,,$$

and therefore $\sum_{k\geq 0} b_k = 1$. We now check that this distribution is subcritical. Let g_b denote the generating function of $\{b_k\}_{k\geq 0}$. A simple computation reveals that $g_b(x) = P(\mathcal{E})^{-1} g_a(P(\mathcal{E})x)$ and therefore $g_b'(x) = g_a'(P(\mathcal{E})x)$, so that

$$\sum_{k\geq 0} k b_k = g_b'(1) = g_a'(P(\mathcal{E})) < 1$$

(g_a' is a strictly increasing function).

It remains to compute $P(H = (x_1, \ldots, x_k) \,|\, \text{ extinction})$ when the underlying offspring distribution is $\{a_k\}_{k\geq 0}$. For all $k \in \mathbb{N}$ and all $(x_1, \ldots, x_k) \in F$

$$\begin{aligned} P(H = (x_1, \ldots, x_k) \,|\, \text{ extinction}) &= \frac{P(H = (x_1, \ldots, x_k)\,,\ \text{extinction})}{P(\text{extinction})} \\ &= \frac{P(H = (x_1, \ldots, x_k)}{P(\text{extinction})}, \end{aligned}$$

(since the condition $(x_1, \ldots, x_k) \in F$ implies extinction at exactly time k for the history $(x_1, \ldots, x_k)$). Therefore

$$P(H = (x_1, \ldots, x_k) \mid \text{ extinction}) = \frac{1}{P(\mathcal{E})} P(H = (x_1, \ldots, x_k))$$
$$= \frac{1}{P(\mathcal{E})} \prod_{i=1}^{k} a_{x_i} = \frac{1}{P(\mathcal{E})} \prod_{i=1}^{k} b_{x_i} P(\mathcal{E})^{-(x_i - 1)}$$
$$= P(\mathcal{E})^{k-1-\sum_{i=1}^{k} x_i} \prod_{i=1}^{k} b_{x_i} = \prod_{i=1}^{k} b_{x_i} .$$

(The last equality makes use of the relation $\sum_{i=1}^{k} x_i = k-1$ when $(x_1, \ldots, x_k) \in F$.) □

EXAMPLE 10.1.7: THE POISSON CASE. For a Poisson offspring supercritical distribution with mean $\lambda > 1$,

$$b_k = e^{-\lambda} \frac{\lambda^k}{k!} P(\mathcal{E})^{k-1} = \frac{1}{P(\mathcal{E})} e^{-\lambda} \frac{(\lambda P(\mathcal{E}))^k}{k!} .$$

But in this case $P(\mathcal{E}) = g_a(P(\mathcal{E})) = e^{\lambda(P(\mathcal{E})-1)}$, or equivalently

$$\frac{1}{P(\mathcal{E})} e^{-\lambda} = e^{-\lambda P(\mathcal{E})} .$$

Therefore

$$b_k = e^{-\lambda \mathcal{P}(\mathcal{E})} \frac{(\lambda P(\mathcal{E}))^k}{k!} ,$$

which corresponds to a Poisson distribution with mean $\mu = \lambda P(\mathcal{E})$.

10.1.2 Tail Distributions

Tail of the Extinction Time

Let T be the extinction time of the Galton–Watson branching process. The distribution of T is fully described by

$$P(T \leq n) = P(X_n = 0) = \psi_n(0) \qquad (n \geq 0)$$

and $P(T = \infty) = 1 - P(\mathcal{E})$. In particular

$$\lim_{n \uparrow \infty} P(T \leq n) = P(\mathcal{E}) . \qquad (\star)$$

Theorem 10.1.8 *In the supercritical case ($m > 1$ and therefore $0 < P(\mathcal{E}) < 1$),*

$$P(\mathcal{E}) - P(T \leq n) \leq (g'(P(\mathcal{E})))^n . \qquad (10.3)$$

Proof. The probability of extinction $P(\mathcal{E})$ is the limit of the sequence $x_n = P(X_n = 0)$ satisfying the recurrence equation $x_{n+1} = g(x_n)$ with initial value $x_0 = 0$. We have that

$$0 \leq P(\mathcal{E}) - x_{n+1} = P(\mathcal{E}) - g(x_n) = g(P(\mathcal{E})) - g(x_n)\,,$$

that is,

$$\frac{P(\mathcal{E}) - x_{n+1}}{P(\mathcal{E}) - x_n} = \frac{g(P(\mathcal{E})) - g(x_n)}{P(\mathcal{E}) - x_n} \leq g'(P(\mathcal{E}))\,,$$

where we have taken into account the convexity of g and the inequality $x_n < P(\mathcal{E})$. □

EXAMPLE 10.1.9: CONVERGENCE RATE FOR THE POISSON OFFSPRING DISTRIBUTION. For a Poisson offspring with mean $m = \lambda > 1$, $g'(x) = \lambda g(x)$ and therefore $g'(P(\mathcal{E})) = \lambda P(\mathcal{E})$. Therefore

$$P(\mathcal{E}) - P(T \leq n) \leq (\lambda P(\mathcal{E}))^n\,.$$

EXAMPLE 10.1.10: CONVERGENCE RATE FOR THE BINOMIAL OFFSPRING DISTRIBUTION. For a $\mathcal{B}(N, p)$ offspring with mean $m = Np > 1$, $g'(x) = Np\frac{g(x)}{1-p(1-x)}$ and therefore

$$g'(P(\mathcal{E})) = Np\frac{P(\mathcal{E})}{1 - p(1 - P(\mathcal{E}))}\,.$$

Taking $p = \frac{\lambda}{N}$,

$$g'(P_N(\mathcal{E})) = \lambda\frac{P_N(\mathcal{E})}{1 - \frac{\lambda}{N}(1 - P_N(\mathcal{E}))}\,,$$

where the notation stresses the dependence of the extinction probability on N.

Tail of the Total Population

Theorem 10.1.11 *For a single ancestor branching process in the subcritical case (in particular, the total population size Y is finite),*

$$P(Y > n) \leq e^{-nh(1)}\,,$$

where $h(a) = \sup_{t\geq 0}\{at - \log E\left[e^{tZ}\right]\}$.

Proof. Consider the random walk $\{W_n\}_{n\geq 0}$ defined by $W_0 = 1$ and

$$W_n = W_{n-1} - 1 + \xi_n\,,$$

where the ξ_i's are IID random variables with the same distribution as Z. Then, the distributions of the sequences $\{W_n\}_{n\geq 0}$ and $\{A_n\}_{n\geq 0}$ are the same as long as they are positive. Therefore

$$\begin{aligned} P(Y > n) &= P(W_1 > 0, \ldots, W_n > 0) \leq P(W_n > 0) \\ &= P\left(1 + \sum_{i=1}^{n} \xi_i > n\right) = P\left(\sum_{i=1}^{n} \xi_i \geq n\right). \end{aligned}$$

The announced result then follows from the Chernoff bound of Theorem 3.2.3 (here we take $a = 1$ and therefore, from the discussion following the statement of the theorem and the assumption $E[Z] < 1$ for the subcritical case, $h(1) > 0$). □

10.2 The Erdös–Rényi Graph

10.2.1 Asymptotically Almost Sure Properties

Random graph theory is a vast subject to which this section is a short introduction. The random graphs considered here were introduced with the purpose of verifying if some basic property (such as the existence of cycles of a certain length, the absence of trees or connectivity) was likely to occur in large typical graphs. There is of course room for discussion concerning the qualification of Erdös–Rényi graphs as typical and other models have been proposed that more aptly fit such and such a specific application. What remains is the panoply of tools used to study these random graphs, such as the first- and second-moment methods, the probabilistic method or the Stein–Chen method. The third section features another type of random graph, the 2-dimensional grid $\mathbb{Z}^2$, where only edges between adjacent vertices are allowed. It is studied under the heading of percolation theory.

For both types of random graphs, a fundamental issue is the existence of "large" components. In the case of Erdös–Rényi graphs, a large component is said to occur if for a large set of vertices, the largest component contains a positive fraction (independent of the size) of the vertices. In the case of percolation graphs (which have an infinite number of vertices), a large component is just a component with an infinite number of vertices. The methods used to assert the existence of large components are rather different in both types of random graphs.

Suppose that the probability for a given edge to belong to a $\mathcal{G}(n,p)$ random graph depends on the number of vertices: $p = p_n$. Any function f such that $\frac{p_n}{n} \sim f(n)$ as $n \uparrow \infty$ will be called a degree growth function. Let $\mathcal{P}$ be some property that a graph may or may not have, for instance, the existence of isolated vertices. It is often of interest to evaluate the probability that $\mathcal{G}(n,p_n)$ has this property. In general, the necessary computations for a fixed size are not feasible, and one then resorts to an "asymptotic answer" by evaluating

$$\lim_{n\uparrow\infty} P(\mathcal{G}(n,p_n) \text{ satisfies property } \mathcal{P}).$$

If this limit is 1, property $\mathcal{P}$ is then said to be asymptotically almost sure (a.a.s.), or to hold with high probability (w.h.p.), for $\mathcal{G}(n,p_n)$.

Threshold Properties

The Landau notational system being often used in this chapter, the reader is directed to Section A.5 for its description.

Definition 10.2.1 *A function $\hat{p}(n)$ is called a threshold (function) for a monotone increasing property $\mathcal{P}$ if*

$$\lim_{n\uparrow\infty} P(\mathcal{G}(n,p_n)\in\mathcal{P}) = 0 \text{ or } 1$$

according to whether

$$\lim_{n\uparrow\infty} \frac{p_n}{\hat{p}(n)} = 0 \text{ or } \infty\,.$$

Theorem 10.2.2 *(Bollobás and Thomason, 1997) There exists a threshold for any non-trivial monotone property.*

Proof. Let $\mathcal{P}$ be a monotone property, say, increasing, without loss of generality. Consider the following function of $p\in[0,1]$:

$$g(p) = P(\mathcal{G}(n,p)\in\mathcal{P})\,.$$

If $p' > p$, one may construct two random graphs $\mathcal{G}(n,p')$ and $\mathcal{G}(n,p)$ such that $\mathcal{G}(n,p')\supseteq\mathcal{G}(n,p)$ (see section 16.1.1), and therefore since the property $\mathcal{P}$ is monotone increasing, $P(\mathcal{G}(n,p')\in\mathcal{P})\geq P(\mathcal{G}(n,p)\in\mathcal{P})$. Function g is therefore monotone non-decreasing. From the expression

$$P(\mathcal{G}(n,p)\in\mathcal{P}) = \sum_{G\in\mathcal{P}} p^{|\mathcal{E}(G)|}p^{\binom{n}{2}-|\mathcal{E}(G)|}\,,$$

it is a polynomial in p increasing from 0 to 1.

Therefore, for each $n\geq 1$, there exists some $\hat{p}(n)$ such that

$$P(\mathcal{G}(n,\hat{p}(n))\in\mathcal{P}) = \frac{1}{2}\,, \tag{10.4}$$

which we now show to be a threshold for $\mathcal{P}$.

Let $\mathcal{G}_1$, $\mathcal{G}_2$, …, $\mathcal{G}_k$ be k independent copies of $\mathcal{G}(n,p)$. The union of these copies is a $\mathcal{G}(n,1-(1-p)^k)$ (Exercise 10.4.8). By coupling, since $1-(1-p)^k\leq kp$,

$$\mathcal{G}(n,1-(1-p)^k)\subseteq\mathcal{G}(n,kp)\,.$$

Therefore, $\mathcal{G}(n,kp)\notin\mathcal{P}$ implies that $\mathcal{G}_1$, $\mathcal{G}_2$, …, $\mathcal{G}_k\notin\mathcal{P}$, and in particular

$$P(\mathcal{G}(n,kp)\notin\mathcal{P})\leq P(\mathcal{G}(n,p)\notin\mathcal{P})^k\,.$$

Let $\omega(n)$ be a function that increases to ∞ arbitrarily slowly as $n\uparrow\infty$. Letting $p=\hat{p}(n)$ and $k=\omega(n)$ in the last inequality, and taking (10.4) into account,

$$P(\mathcal{G}(n,\omega(n)\hat{p}(n)) \notin \mathcal{P}) \leq 2^{-\omega(n)} \to 0\,.$$

On the other hand, with $p = \frac{\hat{p}(n)}{\omega(n)}$ and $k = \omega(n)$,

$$\frac{1}{2} = P(\mathcal{G}(n,\hat{p}(n)) \notin \mathcal{P}) \leq P(\mathcal{G}(n,\hat{p}(n)\omega(n)^{-1}) \notin \mathcal{P})^{\omega(n)}\,,$$

and therefore

$$P(\mathcal{G}(n,\hat{p}(n)\omega(n)^{-1}) \notin \mathcal{P}) \geq 2^{-\frac{1}{\omega(n)}} \to 1\,.$$

□

Definition 10.2.3 *A function $\hat{p}(n)$ is called a sharp threshold (function) for a monotone increasing property $\mathcal{P}$ if for all $\varepsilon > 0$*

$$\lim_{n\uparrow\infty} P(\mathcal{G}(n,p_n) \in \mathcal{P}) = 0 \text{ or } 1$$

according to whether

$$\frac{p_n}{\hat{p}(n)} \leq 1-\varepsilon \text{ or } \geq 1+\varepsilon\,.$$

Theorem 10.2.4 *(Luczak, 1990) Let $\mathcal{P}$ be a non-decreasing graph property. Let $p = p_n$ and $m = m_n$ be such that $p_n = \frac{m_n}{\binom{n}{2}}$, and suppose that*

$$m_n = \binom{n}{2}p_n \to \infty, \qquad \frac{\binom{n}{2} - m_n}{m_n^{\frac{1}{2}}} = \frac{\binom{n}{2}(1-p_n)}{(\binom{n}{2}p_n)^{\frac{1}{2}}} \to \infty.$$

Then, for large n

$$P(\mathcal{G}_{n,m_n} \in \mathcal{P}) \leq 3P(\mathcal{G}(n,p_n) \in \mathcal{P})\,.$$

Proof. To simplify the notation, the subscript n is omitted and we write N for $\binom{n}{2}$. We have that

$$\begin{aligned} P(\mathcal{G}(n,p) \in \mathcal{P}) &= \sum_{k=0}^{N} P(\mathcal{G}_{n,k} \in \mathcal{P})P(|\mathcal{E}_{n,p}| = k) \\ &\geq \sum_{k=m}^{N} P(\mathcal{G}_{n,k} \in \mathcal{P})P(|\mathcal{E}_{n,p}| = k)\,. \end{aligned}$$

By coupling,

$$k \geq m \Rightarrow P(\mathcal{G}_{n,k} \in \mathcal{P}) \geq P(\mathcal{G}_{n,m} \in \mathcal{P})\,.$$

Therefore

$$P(\mathcal{G}(n,p) \in \mathcal{P}) \geq P(\mathcal{G}_{n,m} \in \mathcal{P}) \sum_{k=m}^{N} P(|\mathcal{E}_{n,p}| = k)\,,$$

that is,

$$P(\mathcal{G}(n,p) \in \mathcal{P}) \geq P(\mathcal{G}_{n,m} \in \mathcal{P}) \sum_{k=m}^{N} a_k\,, \qquad (\star)$$

where

$$a_k = \binom{N}{k} p^k (1-p)^{N-k} .$$

By Stirling's equivalence formula

$$a_k = (1+o(1)) \frac{N^N p^k (1-p)^{N-k}}{k^k (N-k)^{N-k} (2\pi k)^{\frac{1}{2}}} = (1+o(1))(2\pi k)^{-\frac{1}{2}} .$$

Let $k = m + t$ and $0 \le t \le m^{\frac{1}{2}}$. Then

$$\frac{a_{k+1}}{a_k} = \frac{(N-k)p}{(k+1)(1-p)} = \frac{1 - \frac{t}{N-m}}{1 + \frac{t+1}{m}} .$$

Since $\frac{t}{N-m} \le \frac{m^{\frac{1}{2}}}{N-m} \to 0$ and therefore is strictly lesser than 1 for large n, we have (using inequalities $1 + x \le e^x$ $(x \in \mathbb{R})$ and $1 - x \ge e^{-\frac{x}{1-x}}$ $(0 \le x < 1)$, and the hypotheses)

$$\frac{a_{k+1}}{a_k} \ge \exp\left\{ -\frac{t}{N-m-t} - \frac{t+1}{m} \right\} = 1 + o(1) .$$

In particular

$$\sum_{k=m}^{N} a_k \ge \sum_{k=m}^{m+m^{\frac{1}{2}}} a_k \ge \frac{1-o(1)}{(2\pi)^{\frac{1}{2}}} ,$$

and therefore, by $(\star)$, $P(\mathcal{G}(n,p) \in \mathcal{P}) \ge P(\mathcal{G}_{n,m} \in \mathcal{P})(2\pi)^{-\frac{1}{2}}$. □

Remark 10.2.5 There exists a version of Theorem 10.2.4 without the assumption of monotonicity of property $\mathcal{P}$ (Exercise 10.4.9).

The First- and Second-moment Method

The first- and second-moment bounds of Theorems 3.2.9 and 3.2.10 are basic tools for the asymptotic analysis of random graphs. The following examples illustrate the method.

EXAMPLE 10.2.6: DIAMETER LARGER THAN 2. The diameter $D(G)$ of a graph G is, by definition, the maximal length of the shortest path between two distinct vertices. It turns out that the (monotone increasing) property

$$\mathcal{P} : D(G) > 2$$

admits the threshold

$$\hat{p}(n) = \left(\frac{2 \log n}{n} \right)^{\frac{1}{2}} .$$

Proof. The property $D(G) > 2$ is equivalent to the existence of a non-adjacent pair $\{v, w\}$ of distinct vertices such that no other vertex $u \notin \{v, w\}$ is adjacent to

both v and w. Such a pair $\{v, w\}$ will be called "bad". Let for any pair $\{v, w\}$ of distinct vertices X_{vw} be the indicator function of the event "$\{v, w\}$ is bad". Let

$$X := \sum_{\{v,w\}} X_{vw} .$$

The random graph has a diameter ≤ 2 if and only if it has no bad pair, that is, if $X = 0$.

In $\mathcal{G}(n, p)$, the probability that a given vertex $u \notin \{v, w\}$ is adjacent to both v and w is p^2, or, equivalently, the probability that it is not adjacent to both v and w is $1-p^2$. The probability that no vertex is adjacent to $\{v, w\}$ is therefore $(1-p^2)^{n-2}$. The probability that v and w are not adjacent is $1-p$. Since there are $\binom{n}{2}$ distinct pairs $\{v, w\}$, the expected number of bad pairs is

$$E[X] = \binom{n}{2}(1-p^2)^{n-2}(1-p) .$$

With $p_n = c\left(\frac{\log n}{n}\right)^{\frac{1}{2}}$,

$$E[X_n] \sim \frac{1}{2}n^{2-c^2} .$$

Therefore, if $c > \sqrt{2}$, $E[X_n] \to 0$. In particular, $\mathcal{G}(n, p_n)$ has w.h.p. a diameter ≤ 2.

If $c < \sqrt{2}$, $E[X_n] \to \infty$. But this is not enough to guarantee that $\mathcal{G}(n, p_n)$ has w.h.p. a diameter > 2. We try the second-moment method, and compute

$$E[X^2] = \sum E\left[X_{vw}X_{v'w'}\right] + \sum E\left[X_{vw}X_{vw'}\right] + \sum E\left[X_{vw}X_{vw}\right]$$

where the 4 vertices in the first summation are different, the 3 vertices in the second summation are different, and the 2 vertices in the third summation are different. First observe that $X_{vw}X_{vw} = X_{vw}$, and therefore the third summation is $E[X] \sim \frac{1}{2}n^{2-c^2}$. By independence, the first sum equals

$$\sum E\left[X_{vw}\right] E\left[X_{v'w'}\right] = E\left[\sum X_{vw}\right] E\left[\sum X_{v'w'}\right] = E[X]^2 \sim n^{4-2c^2} .$$

For the second summation, observe that if $X_{vw}X_{vw'} = 1$, both pairs $\{v, w\}$ and $\{v'w\}$ are bad, which in turn implies that for every vertex $u \notin \{v, w, v'\}$, either there is no edge between u and v, or there is an edge between u and v and no edge between w and w'. For fixed u this occurs with probability

$$1 - p + p(1-p)^2 .$$

Therefore the event $X_{vw}X_{vw'} = 1$ implies an event of probability

$$\left(1 - p_n + p_n(1-p_n)^2\right)^n .$$

With $p_n = c\left(\frac{\log n}{n}\right)^{\frac{1}{2}}$, this probability is $\sim n^{-2c^2}$. Therefore $E\left[X^2\right] \leq n^{4-2c^2} + n^{3-2c^2} + n^{2-2c^2}$ and $E\left[X\right] \sim n^{2-2c^2}$,

$$E\left[X^2\right] \leq E\left[X\right](1+o(1)).$$

By the second-moment argument, $\lim_{n\uparrow\infty} P(X_n > 0) = 1$. □

EXAMPLE 10.2.7: ISOLATED VERTICES IN LARGE RANDOM GRAPHS. Let X_n be the number of isolated vertices in the Erdös–Rényi graph $\mathcal{G}(n, p_n)$ with $p_n = c\frac{\log n}{n}$. Then

(i) If $c > 1$, $P(X_n \neq 0) \to 0$ as $n \uparrow +\infty$.

(ii) If $c < 1$, $P(X_n \neq 0) \to 1$ as $n \uparrow +\infty$.

Therefore $\hat{p}(n) = \frac{\log n}{n}$ is a sharp threshold for the existence of isolated vertices.

Proof. The proof of (i) is based on Lemma 3.2.9. For each vertex u, let Z_u be the indicator function of the event "u is isolated". In particular, $X_n = \sum_{u\in V} Z_u$. Also, $E[Z_u] = (1-p_n)^{n-1}$ and therefore

$$E[X_n] = \sum_{u\in V} E[Z_u] = n(1-p_n)^{n-1} = n\left(1 - c\frac{\log n}{n}\right)^{n-1}.$$

It suffices therefore, according to Lemma 3.2.9, to prove that the latter quantity tends to 0 or, equivalently, that its logarithm tends to $-\infty$. This follows from:

$$\log n + (n-1)\log\left(1 - c\frac{\log n}{n}\right) \leq \log n\left(1 - c\left(1 - \frac{1}{n}\right)\right) \to -\infty,$$

where we have used the inequality $1 - x \leq e^{-x}$ and the hypothesis $c > 1$.

For the proof of (ii), first observe that when $c < 1$,

$$E[X_n] = n\left(1 - c\frac{\log n}{n}\right)^{n-1} \to +\infty.$$

In fact, taking logarithms, the last quantity is

$$\begin{aligned}\log n + (n-1)\log\left(1 - c\frac{\log n}{n}\right) &= \log n + (n-1)\left(-c\frac{\log n}{n} + o\left(\frac{\log n}{n}\right)\right)\\ &= \log n\left(1 - c\left(1-\frac{1}{n}\right) + \left(1-\frac{1}{n}\right)\frac{o\left(\frac{\log n}{n}\right)}{\frac{\log n}{n}}\right)\\ &= (1-c)\log n + \alpha(n),\end{aligned}$$

where $\alpha(n) \to 0$.

However, for a sequence of integer-valued random variables $\{X_n\}_{n\geq 1}$, the fact that $E[X_n] \to +\infty$ is not sufficient to guarantee that $P(X_n \neq 0) \to 1$ (see Exercise 3.3.1). One has to go beyond the first moment and use Lemma 3.2.10. According to this lemma, $P(X_n = 0) \leq \frac{\mathrm{Var}(X_n)}{E[X_n]^2}$ and therefore, it suffices to prove that

$$\mathrm{Var}(X_n) = o(E[X_n]^2)$$

to obtain that $P(X_n = 0) \to 0$. We proceed to do this.

First, we compute the variance of X_n:

$$\mathrm{Var}(X_n) = \mathrm{Var}\left(\sum_{u\in V} Z_u\right) = \sum_{u\in V}(\mathrm{Var}(Z_u)) + \sum_{u,v\in V, u\neq v} \mathrm{cov}\,(Z_u, Z_v).$$

Since the variables Z_u take the values 0 or 1, $Z_u^2 = Z_u$ and therefore

$$\mathrm{Var}(Z_u) = E[Z_u^2] - E[Z_u]^2 = E[Z_u] - E[Z_u]^2 \leq E[Z_u] = (1-p_n)^{n-1}\ .$$

Also, for $u \neq v$,

$$E[Z_u Z_v] = P(u \text{ isolated}, v \text{ isolated}) = (1-p_n)(1-p_n)^{n-2}(1-p_n)^{n-2}$$

($1-p_n$ is the probability that $\langle u, v\rangle$ is not an edge of $\mathcal{G}(n, p_n)$, and $(1-p_n)^{n-2}$ is the probability that u is not connected to a vertex in $V\backslash\{v\}$, and also the probability that v is not connected to a vertex in $V\backslash\{u\}$). Therefore

$$\mathrm{cov}\,(Z_u, Z_v) \leq (1-p_n)^{2n-3} - (1-p_n)^{2n-2} = p_n(1-p_n)^{2n-3}\,,$$

so that

$$\mathrm{Var}(X_n) \leq n(1-p_n)^{n-1} + n(n-1)p_n(1-p_n)^{2n-3}\,.$$

The first term of the right-hand side of the above inequality, which is equal to $E[X_n]$, tends to infinity. It is therefore a $o(E[X_n]^2)$. The second term also, since

$$\frac{n(n-1)p_n(1-p_n)^{2n-3}}{E[X_n]^2} = \frac{n(n-1)p_n(1-p_n)^{2n-3}}{n^2(1-p_n)^{2n-2}} = \frac{n-1}{n}\frac{p_n}{1-p_n}$$

and $p_n = c\frac{\log n}{n} \to 0$. □

EXAMPLE 10.2.8: CLIQUES IN LARGE ERDÖS–RÉNYI RANDOM GRAPHS. Let $\mathcal{G}(n, p_n)$ be an E-R graph with n vertices and probability p_n of existence of a given vertex. Define a clique to be any set of vertices C such that any pair of vertices in C are linked by an edge of the graph. Let X_n be the number of cliques of size 4 in $\mathcal{G}(n, p_n)$.

(i) If $\lim_{n\uparrow\infty} p_n n^{\frac{2}{3}} = 0$, $P(X_n \neq 0) \to 0$ (the probability of having a 4-clique tends to 0).

(ii) If $\lim_{n\uparrow\infty} p_n n^{\frac{2}{3}} = +\infty$, $P(X_n \neq 0) \to 1$.

Proof. Let $A_1, A_2, \ldots, A_{\binom{n}{4}}$ be an enumeration of the sets of four vertices. Let Z_i be the indicator function of the event that A_i is a clique of $\mathcal{G}(n, p_n)$. Then $X_n = \sum_{i=1}^{\binom{n}{4}} Z_i$. The proof now follows the lines of Example 10.2.7. First observe that $E[Z_i] = p_n^6$ (there are 6 edges in a 4-clique) and therefore $E[X_n] = \binom{n}{4} p_n^6$.

(i) It suffices to show that $E[X_n] \to 0$. This is the case since $E[X_n] = n^4 o\left(\left(n^{-\frac{2}{3}}\right)^6\right) = n^4 o(n^{-4})$.

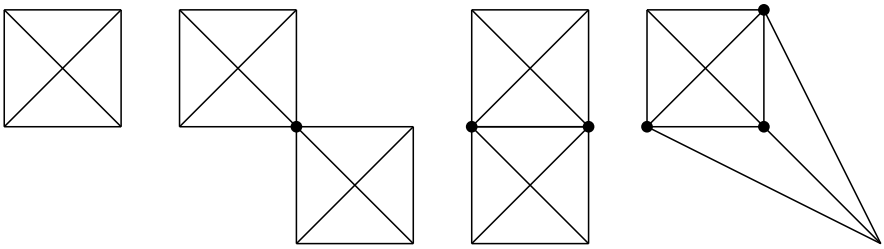
(ii) It suffices to show that $E[X_n] \to +\infty$ and $\frac{\mathrm{Var}(X_n)}{E[X_n]^2} \to 0$. First

$$E[X_n] = \binom{n}{4} p_n^6 = n^4 \omega\left(\left(n^{-\frac{2}{3}}\right)^6\right) = n^4 \omega(n^{-4}) \to +\infty.$$

For the variance of X_n, we have, as in Example 10.2.7,

$$\mathrm{Var}(X_n) \le E[X_n] + \sum_{1 \le i \ne j \le \binom{n}{4}} \mathrm{cov}\,(Z_i, Z_j)\,.$$

If $|A_i \cap A_j| = 0$ or 1, Z_i and Z_j are independent, depending only on the existence of 2 disjoint sets of edges.



A 4-clique, and three pairs of 4-cliques sharing 1, 2 and 3 vertices

If $|A_i \cap A_j| = 2$, $\mathrm{cov}\,(Z_i, Z_j) \le E[Z_i Z_j] = p_n^{11}$ since the total number of different edges of the two 4-cliques based on A_i and A_j is 11. We now count the number of pairs (A_i, A_j) sharing exactly 2 vertices. There are $\binom{n}{6}$ sextuples of vertices, and for each sextuple, there are $\binom{6}{2;2;2}$ ways to split it into two 4-tuples sharing 2 vertices (2 for $A_i \cap A_j$, 2 for A_i alone, 2 for A_j alone).

If $|A_i \cap A_j| = 3$, $\mathrm{cov}\,(Z_i, Z_j) \le E[Z_i Z_j] = p_n^9$ since the total number of different edges of the two 4-cliques based on A_i and A_j is now 9. We now count the number of pairs (A_i, A_j) sharing exactly 3 vertices. There are $\binom{n}{5}$ quintuples of vertices, and for each quintuple, there are $\binom{5}{3;1;1}$ ways to split it into two 4-tuples sharing 3 vertices.

Therefore

$$\mathrm{Var}(X_n) \le E[X_n] + \binom{n}{6}\binom{6}{2;2;2} p_n^{11} + \binom{n}{5}\binom{5}{3;1;1} p_n^9 \quad = E[X_n] + b_n + c_n.$$

The first term of the right-hand side, $E[X_n]$, tends to infinity, and is therefore a $o(E[X_n]^2)$. The second term also, because $b_n = O(n^6 p_n^{11})$ and $c_n = O(n^5 p_n^9)$ and therefore

$$\frac{b_n + c_n}{E[X_n]^2} = O\left(\frac{1}{n^2 p_n}\right) + O\left(\frac{1}{n^3 p_n^3}\right)$$

tends to 0 as $n \uparrow +\infty$ when $p_n n^{\frac{2}{3}} \to \infty$ □

10.2.2 The Evolution of Connectivity

Let $V = \{1, \ldots, n\}$ be fixed and consider the sequence of random graphs

$$G_0 := (V, \varnothing), G_1, \ldots, G_m, \ldots, G_{\binom{n}{2}} = K_n$$

each graph in the sequence being obtained by adding an edge randomly chosen among the free edges as explained just after Definition 2.1.55. In fact, $G_m \equiv \mathcal{G}_{n,m}$. As m grows from 0 to $\binom{n}{2}$, the connectivity increases. One expects that for very small values of m, isolated vertices prevail, then, as m increases, the graph will contain mostly isolated vertices and isolated edges, then mostly trees, etc. The components will become larger and larger, until one observes a giant component, that is one with a number of edges that is a fraction α (independent of n) of the number of vertices, all the other components being "much smaller". This analysis is true for "large" graphs, and can be formalized in asymptotic terms as will be done in the sequel.

The results will be stated in terms of Erdös–Rényi's graphs ($\mathcal{G}_{n,m}$) but the proofs will provide analogous results in terms of Gilbert's graphs ($\mathcal{G}_(n,p)$), where connectivity increases as p grows from 0 to 1. The corresponding statements are given as corollaries.

Theorem 10.2.9 *(a) If $m(n) \gg n^{\frac{1}{2}}$, $G_{m(n)}$ contains w.h.p a path of length 2.*

(b) If $m(n) \ll n^{\frac{1}{2}}$, $G_{m(n)}$ contains w.h.p. only isolated vertices and edges.

Proof. (a) By definition of the symbol $\gg$, $m(n) = \omega(n)n^{\frac{1}{2}}$ where $\omega(n)$ tends arbitrarily slowly to ∞ with n. Since

$$p_n = \frac{\omega(n)n^{\frac{1}{2}}}{N(n)} = \omega(n)n^{-\frac{3}{2}},$$

we also have $p_n \gg n^{-\frac{3}{2}}$.

For the proof, it suffices to show that there is a.a.s. an *isolated* path of length 2. Let P_2 be the collection of paths of length 2 of the complete graph K_n and let Y_n denote the number of isolated paths of length 2 in $\mathcal{G}(n, p_n)$, that is,

$$Y_n := \sum_{\pi \in P_2} 1_{\{\pi \text{ isolated in } \mathcal{G}(n,p_n)\}}$$

We have that (Exercise 10.4.7)

$$E[Y_n] = 3\binom{n}{3} p_n^2 (1 - p_n)^{3(n-3)+1}.$$

It holds that this quantity tends to ∞ if $m(n) = o(n)$ (and therefore $np_n = o(1)$), and *a fortiori* for $m(n) = \omega(n)n^{\frac{1}{2}}$, using coupling and the fact that the property of having a path of length 2 is a monotone increasing property. This does not suffice

to prove that $P(Y_n > 0) \to 1$. We will apply the second moment method and compute the expectation of

$$Y_n^2 = \sum_{\pi \in P_2} \sum_{\rho \in P_2} 1_{\{\pi \text{ isolated in } \mathcal{G}(n,p_n)\}} 1_{\{\rho \text{ isolated in } \mathcal{G}(n,p_n)\}} .$$

Note that the sum can be restricted to pairs π and ρ that are either identical or share no vertex. (In fact, two different paths of P_2 sharing at least one vertex could not be both isolated in $\mathcal{G}(n,p_n)$.) Now

$$E\left[Y_n^2\right] = \sum_{\pi \in P_2} \left\{ \sum_{\rho \in P_2} P(\rho \text{ is isolated in } \mathcal{G}(n,p_n)) \mid \pi \text{ isolated in } \mathcal{G}(n,p_n)) \right\} \times \cdots$$
$$\cdots P(\pi \text{ isolated in } \mathcal{G}(n,p_n)) .$$

The expression inside brackets is independent of π and therefore, with the particular choice $\pi = abc$,

$$E\left[Y_n^2\right] = E\left[Y_n\right] \times \cdots$$
$$\cdots \left(1 + \sum_{\rho \in P_2 \,;\, \rho \cap abc = \varnothing} P(\rho \text{ isolated in } \mathcal{G}(n,p_n)) \mid abc \text{ isolated in } \mathcal{G}(n,p_n)) \right) .$$

If path abc is isolated in $\mathcal{G}(n,p_n)$, then ρ such that $\rho \cap abc = \varnothing$ is isolated in $\mathcal{G}(n,p_n)$ if and only if the following edges are missing:

(a) the $3(n-3)$ edges from each one of the vertices of ρ from the $n-3$ potential edges not leading to a vertex of ρ, and

(b) the 9 edges from a vertex of ρ to a vertex of abc.

Hence

$$E\left[Y_n^2\right] \le E\left[Y_n\right] \left(1 + 3\binom{n}{3} p_n^2 (1-p_n)^{3(n-3)+1} (1-p_n)^9 \right)$$
$$= E\left[Y_n\right] \left(1 + E\left[Y_n\right] (1-p_n)^9 \right) .$$

Therefore, by Lemma 3.2.11,

$$P(Y_n > 0) \ge \frac{E\left[Y_n\right]^2}{E\left[Y_n^2\right]} \ge \frac{E\left[Y_n\right]^2}{E\left[Y_n\right] \left(1 + E\left[Y_n\right] (1-p_n)^9\right)}$$
$$= \frac{1}{E\left[Y_n\right]^{-1} + (1-p_n)^9} ,$$

a quantity that tends to 1 since $p_n \to 0$ and $E\left[Y_n\right] \to \infty$. Therefore the probability that there is an isolated path of length 2 in $\mathcal{G}(n,p_n)$ tends to 1.

Since the property $\mathcal{P}$ that there exists at least a path of length 2 is a monotone increasing property, we also have by Theorem 10.2.4 that

$$P(\mathcal{G}_{n,m(n)} \text{ has a path of length 2 }) \to 1 .$$

(b) Exercise 10.4.11. □

Corollary 10.2.10 *(a) If $p_n \gg n^{-\frac{3}{2}}$, $G(n, p_n)$ contains w.h.p. a path of length 2. (b) If $p_n \ll n^{-\frac{3}{2}}$, $G(n, p_n)$ contains w.h.p. only isolated vertices and edges.*

Theorem 10.2.11 *If $m(n) \ll n$, $G_{m(n)}$ is a forest w.h.p.*

Proof. $m(n) \ll n$ means that $m(n) = \frac{n}{\omega(n)}$, where $\omega(n)$ tends arbitrarily slowly to ∞ with n. In particular

$$p_n = \frac{m(n)}{N(n)} = \frac{2n}{n(n-1)\omega(n)} = \frac{2}{n\omega(n)}\frac{n}{n-1} \leq \frac{3}{n\omega(n)}\,.$$

If X_n denotes the number of cycles in $\mathcal{G}(n, p_n)$,

$$E\left[X_n\right] = \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p_n^k = \sum_{k=3}^{n} \frac{1}{2} n(n-1)\cdots(n-k+1) p_n^k \leq \sum_{k=3}^{n} \frac{n^k}{2k} p_n^k$$
$$\leq \sum_{k=3}^{n} \frac{n^k}{2k} \frac{1}{\omega(n)^k n^k} = \sum_{k=3}^{n} \frac{1}{\omega(n)^k} = \frac{1}{\omega(n)^3}\,.$$

Therefore

$$P(\mathcal{G}_(n, p_n) \text{ is not a forest}) = P(X_n \geq 1) \leq E\left[X_n\right] \to 0\,.$$

Since the property $\mathcal{P}$ that there is not a forest is a monotone increasing property, by Theorem 10.2.4 we also have that

$$P(\mathcal{G}_{n,m(n)} \text{ is not a forest}) \to 0\,.$$

Therefore both $\mathcal{G}_{n,m(n)}$ and $\mathcal{G}(n, p_n)$ are forests w.h.p. □

Corollary 10.2.12 *If $p_n \ll n^{-1}$, $\mathcal{G}(n, p_n)$ is w.h.p. a forest.*

10.2.3 The Giant Component

This subsection features the emblematic result of random graph theory, the emergence of a giant component.

A connected component of a graph $(V, \mathcal{E})$ is a connected subset $C \subseteq V$ (any two distinct vertices of C are linked by a chain of edges). The components $C_1, C_2, \ldots$ of the graph will be ordered by decreasing size: $|C_1| \geq |C_2| \geq \cdots$.

Consider the Erdös–Rényi random graph $\mathcal{G}(n, p)$ where $n \geq 2$ and $p \in (0, 1)$. We shall study the size of the largest component of $\mathcal{G}(n, \frac{\lambda}{n})$ as $n \uparrow \infty$. Note that the average degree of this graph is roughly λ.

It turns out, as we shall see, that the value $\lambda = 1$ is critical, in the sense that the behaviour of the Erdös–Rényi graph $\mathcal{G}(n, \frac{\lambda}{n})$ as $n \uparrow \infty$ is different according to whether $\lambda < 1$ (the subcritical case) or $\lambda > 1$ (the supercritical case).

Subcritical Case

In the subcritical case the components are a.a.s. of size at most of the order of $\log n$. More precisely:

Theorem 10.2.13 *(Erdös and Rényi, 1959) If $\lambda < 1$, there exists a finite number $a = a(\lambda)$ such that*

$$\lim_{n\uparrow\infty} P(|C_1| \leq a \log(n)) = 1\,. \tag{10.5}$$

We now prepare the proof with the following description of the exploration of the component $C(v_0)$ containing the arbitrarily chosen vertex v_0.

For each $k \geq 0$ we keep track of the set of active vertices $\mathcal{A}_k$ and of a set of explored vertices $\mathcal{B}_k$ whose respective sizes are denoted by A_k and B_k. These sets are defined by their construction, as follows. The exploration is initiated with

$$\mathcal{A}_0 = \{v_0\}, \qquad \mathcal{B}_0 = \varnothing\,.$$

At step $k \geq 1$, pick arbitrarily an active vertex $v_{k-1} \in \mathcal{A}_{k-1}$, deactivate v_{k-1} and activate all the neighbours of v_{k-1} that are not in $\mathcal{A}_{k-1} \cup \mathcal{B}_{k-1}$. Denote by $\mathcal{D}_k$ the set of newly activated vertices and by D_k its size.

These sets can be interpreted in terms of epidemics, featuring three types of individuals: the infectious, the susceptibles and the removed. The active sites form, in this interpretation, the infectious population. At time k, an indivividual that is infectious ($v_{k-1} \in \mathcal{A}_{k-1}$) transmits the disease to his neighbors that have not yet been infected (those in $\mathcal{D}_k$) and he is then immediately "removed" (either cured and immune, or dead). The removed population at time k is $\mathcal{B}_k$, of size $B_k = k$. The set $\mathcal{U}_k = V\backslash(\mathcal{A}_k \cup \mathcal{B}_k)$ (whose cardinality will be denoted by U_k) represents the population that has not yet been infected (the "susceptibles"). In particular $\mathcal{U}_0 = V\backslash\{v_0\}$.

The above construction is summarized as follows

$$\begin{aligned} &v_{k-1} \in \mathcal{A}_{k-1} \\ &\mathcal{D}_k = \{v \sim v_{k-1}; v \notin \mathcal{A}_{k-1} \cup \mathcal{B}_{k-1}\} \\ &\mathcal{A}_k = (\mathcal{A}_{k-1} \cup \mathcal{D}_k)\backslash\{v_{k-1}\} \\ &\mathcal{B}_k = \mathcal{B}_{k-1} \cup \{v_{k-1}\}\,. \end{aligned}$$

In particular, $A_0 = 1$, and as long as $A_{k-1} > 0$,

$$A_k = A_{k-1} - 1 + D_k > 0$$

and therefore

$$A_k = D_1 + \cdots + D_k - k + 1\,.$$

The exploration procedure just described generates a random tree, called "the branching process" (of the exploration process). It is not a Galton–Watson process, as the offspring distribution at each step varies. The offspring distribution at step k, that is, the distribution of D_k is given by the following two lemmas.

Lemma 10.2.14 *Conditioned on $D_1, \dots, D_{k-1}$, the random variable D_k is a binomial random variable of size $n-1-(D_1+\cdots+D_{k-1}) = n-k+1-A_{k-1}$ and parameter p.*

Proof.

$$\begin{aligned} n &= U_{k-1} + A_{k-1} + B_{k-1} \\ &= U_{k-1} + A_{k-1} + k - 1 \\ &= U_{k-1} + (D_1 + \cdots + D_{k-1}) - (k-1) + 1 + k - 1 \\ &= U_{k-1} + (D_1 + \cdots + D_{k-1}) + 1\,, \end{aligned}$$

and therefore

$$\begin{aligned} U_{k-1} &= n - 1 - (D_1 + \cdots + D_{k-1}) \\ &= n - k + 1 - A_{k-1}. \end{aligned}$$

The conclusion follows since $\mathcal{D}_k$ is selected from $\mathcal{U}_{k-1}$ by independent trials with probability p of success. □

The exploration stops at time

$$Y = \inf\{k > 0;\ A_k = 0\}\,,$$

when all the vertices of the component $C(v_0)$ have been found and deactivated, and

$$|C(v_0)| = |\mathcal{B}_Y| = Y\,.$$

Lemma 10.2.15 *For $k \geq 0$, $A_k + k - 1$ is a binomial random variable of size $n-1$ and parameter $1-(1-p)^k$.*

Proof. The construction of the set $\mathcal{D}_k$ can be described in the following recursive manner. Let us draw a collection $\{Z_{k,v}\}_{k>0, v\in V}$ of IID Bernoulli variables with parameter p. A given vertex $v \neq v_0$ is included in $\mathcal{D}_k$ if and only if $Z_{k,v} = 1$ and $v \notin \cup_{\ell=1}^{k-1} \mathcal{D}_\ell$. A vertex $v \neq v_0$ is not incorporated at step k iff $Z_{1,v} = \cdots = Z_{k,v} = 0$, and this happens with probability $(1-p)^k$, independently of all such vertices v. Therefore U_k is the sum of $n-1$ independent binomial random variables of mean $(1-p)^k$ so that

$$U_k \sim \mathcal{B}(n-1, (1-p)^k)\,.$$

But $A_k + k - 1 = (n-1) - U_k$ hence the result (see Exercise 2.4.5). □

Lemma 10.2.16 *For all $\theta > 0$,*

$$P(|C(v_0)| > k) \leq \exp\{-k\left(\theta - \lambda(1-e^\theta)\right)\}\,.$$

Proof. By the previous lemma,

$$\begin{aligned} P(|C(v_0)| > k) &\le P(A_k > 0) \\ &= P(A_k + k - 1 \ge k) \\ &= P\left(\mathcal{B}(n-1, 1-(1-p)^k) \ge k\right) . \end{aligned}$$

But, as $n - 1 < n$, and since for sufficiently large n, $1 - (1-p)^k \le kp$,

$$P\left(\mathcal{B}(n-1, 1-(1-p)^k) \ge k\right) \le P\left(\mathcal{B}(n, kp) \ge k\right) .$$

Markov's inequality gives for $\theta > 0$

$$\begin{aligned} P\left(\mathcal{B}(n, kp) \ge k\right) &= P\left(e^{\theta \mathcal{B}(n,kp)} \ge e^{\theta k}\right) \\ &\le E\left[e^{\theta \mathcal{B}(n,kp)}\right] e^{-\theta k} \\ &= (1 - kp + kpe^{\theta})^n e^{-\theta k} \\ &\le \exp\{-npk(1 - e^{\theta})\} e^{-\theta k} \\ &= \exp\{-k\left(\theta + \lambda(1 - e^{\theta})\right)\} . \end{aligned}$$

□

We are now ready to conclude the proof of Theorem 10.2.13. Optimizing with respect to $\theta > 0$ or, equivalently, taking $\theta = \log \lambda$, we obtain

$$P(|C(v_0)| > k) \le e^{-\beta k}$$

where $\beta = -\log \lambda - 1 + \lambda$. Therefore, for all $\delta > 0$,

$$\begin{aligned} P(|C_1| > \beta^{-1}(1+\delta)\log n) &= P\left(\max_{v \in V} |C(v)| > \beta^{-1}(1+\delta)\log n\right) \\ &\le \sum_{v \in V} P\left(|C(v)| > \beta^{-1}(1+\delta)\log n\right) \\ &= nP\left(|C(v_0)| > \beta^{-1}(1+\delta)\log n\right) \le n^{-\delta}. \end{aligned}$$

The announced result then follows with $a = \beta^{-1}(1+\delta)$.

Supercritical Case

In the supercritical case there exists a.a.s. one and only one large component, that is a component of size of the order of n. All other components are "small", that is, of size of the order of $\log n$. More precisely, denoting by $\mathcal{P}_e(\lambda)$ the probability of extinction of a branching process whose offspring is distributed according to a Poisson distribution with mean λ (if $\lambda > 1$, this is the unique root in $(0, 1)$ of the equation $x = e^{-\lambda(1-x)}$), we have:

Theorem 10.2.17 *(Erdös and Rényi, 1959) If $\lambda > 1$, there exists a finite positive number $a = a(\lambda)$ such that for all $\delta > 0$,*

$$\lim_{n\uparrow\infty} P\left(\left|\frac{|C_1|}{n} - (1-\mathcal{P}_e(\lambda))\right| \le \delta\right) = 1 \tag{10.6}$$

and

$$\lim_{n\uparrow\infty} P\left(|C_2| \le a\log(n)\right) = 1\,. \tag{10.7}$$

Roughly speaking: in the supercritical case, "the largest component of a big graph almost surely contains a proportion $1-\mathcal{P}_e(\lambda)$ of all vertices and almost all the other components are at most of logarithmic size".

Let $k_- = a'\log n$ where $a' = \log\frac{16\lambda}{(\lambda-1)^2}$, and let $k_+ = n^{\frac{2}{3}}$. Call v_0 a bad vertex if none of the following properties is satisfied:

(i) The branching process ends before k_- steps, that is if $|C(v_0)| \le k_-$.

(ii) For all k, $k_- \le k \le k_+$, there are at least $\frac{(\lambda-1)k}{2}$ active vertices, that is $A_k \ge \frac{(\lambda-1)k}{2}$.

Lemma 10.2.18 *Any vertex v_0 is a.a.s. a good vertex.*

Proof. Either the branching process terminates in less than k_- steps or not. Vertex v_0 is a bad vertex only in the second case, that is if there exists an integer k, $k_- \le k \le k_+$, such that $A_k < \frac{(\lambda-1)k}{2}$. This means that the number of vertices visited at step k is $< k + \frac{(\lambda-1)k}{2} = \frac{(\lambda+1)k}{2}$. Let $B(v_0,k)$ be the event that there are $< \frac{(\lambda+1)k}{2}$ vertices visited at step k. Then, observing that the branching process generated by the exploration procedure is before step k_+ stochastically bounded from below by a Galton–Watson branching process with offspring distribution $\mathcal{B}\left(n - \frac{(\lambda+1)k_+}{2}, \frac{\lambda}{n}\right)$,

$$\begin{aligned} P(B(v_0,k)) &\le P\left(\sum_{i=1}^{k} \mathcal{B}\left(n - \frac{(\lambda+1)k_+}{2}, \frac{\lambda}{n}\right) \le \frac{(\lambda+1)k}{2} - 1\right) \\ &\le P\left(\mathcal{B}\left(k\left(n - \frac{(\lambda+1)k_+}{2}\right), \frac{\lambda}{n}\right) \le \frac{(\lambda+1)k}{2} - 1\right)\,. \end{aligned}$$

Let $X := \mathcal{B}\left(k\left(n - \frac{(\lambda+1)k_+}{2}\right), \frac{\lambda}{n}\right)$. We have that $E[X] = \lambda k\left(1 - \frac{(\lambda+1)k_+}{2n}\right)$. We now apply the bound D of Example 3.2.6 with δ such that $(1-\delta)\lambda k\left(1 - \frac{(\lambda+1)k_+}{2n}\right) = \frac{(\lambda+1)k}{2}$, that is

$$\delta = 1 - \frac{\lambda+1}{\lambda(2-(\lambda+1)k_+/n)}$$

(a quantity which tends to $\frac{\lambda-1}{2\lambda}$ as $n\uparrow\infty$). In particular, after some elementary computations,

$$P(B(v_0,k)) \le \exp\left\{-\left(\frac{(\lambda-1)^2}{8\lambda} + O(n^{-\frac{1}{3}})\right)k\right\}\,.$$

The probability $P(\cup_{k=k_-}^{k_+} B(v_0,k))$ that v_0 is a bad vertex is, by the union bound, bounded by

$$\begin{aligned}\sum_{k=k_-}^{k_+} P(B(v_0,k)) &\le \sum_{k=k_-}^{k_+} e^{-\left(\frac{(\lambda-1)^2}{8\lambda}+O(n^{-\frac{1}{3}})\right)k} \\ &\le k_+ e^{-\left(\frac{(\lambda-1)^2}{8\lambda}+O(n^{-\frac{1}{3}})\right)k_-} \\ &\le n^{\frac{2}{3}} e^{-\left(\frac{(\lambda-1)^2}{8\lambda}+O(n^{-\frac{1}{3}})\right)a'\log n} = n^{\frac{2}{3}} n^{-\frac{(\lambda-1)^2}{8\lambda}a'+O(n^{-\frac{1}{3}})}.\end{aligned}$$

Taking $a' = \frac{16\lambda}{(\lambda-1)^2}$, we have that the probability that v_0 is a bad vertex is less than $n^{-\frac{4}{3}}$. Therefore, by the union bound, the probability of having a bad vertex is less than $n^{-\frac{1}{3}}$. □

Therefore the branching process starting from any vertex v_0 either terminates within $k_- = O(\log n)$ steps, or goes on for at least k_+ steps. Call vertices of the first type "small", and vertices of the second type "large".

Lemma 10.2.19 *There is a.a.s. at most one component containing all the large vertices.*

Proof. In view of Lemma 10.2.18, we may suppose that all vertices are good vertices. Let v_0 and v_1 be distinct vertices. Denote their set of active vertices at stage k_+ by $\mathcal{A}(v_0)$ and $\mathcal{A}(v_1)$ respectively (in the notation introduced at the begining of the section $\mathcal{A}(v_0) = \mathcal{A}_{k_+}$, with a similar interpretation for v_1). Suppose that $C(v_0)$ and $C(v_1)$ are both large components, that is, $|\mathcal{A}(v_0)| \ge \frac{\lambda-1}{2}k_+$ and a similar inequality for v_1. If the branching process for k_+ steps for v_0 and v_1 have some vertices in common, then we are done since this means that $C(v_0)$ and $C(v_1)$ are the same. Now,

$$\begin{aligned}P(C(v_0) \ne C(v_1)) &\le P(\text{no edge between } \mathcal{A}(v_0) \text{ and } \mathcal{A}(v_1)) \\ &\le (1-p_n)^{-\left(\frac{(\lambda-1)k_+}{2}\right)^2} \\ &\le e^{\left(-p_n\frac{(\lambda-1)k_+}{2}\right)^2} \\ &\le e^{-\frac{(\lambda-1)^2\lambda}{4}n^{\frac{1}{3}}} = o(n^{-2})\end{aligned}$$

(where we have used the inequality $1-x \le e^{-x}$). The union bound over distinct pairs of vertices (v_0,v_1) then gives that the probability that for any pair of large vertices v_0 and v_1 there exists no edge between $\mathcal{A}(v_0)$ and $\mathcal{A}(v_1)$ is $o(1)$. □

Therefore, there is a.a.s. only one large component, and all other components have a size at most $O(\log n)$. It remains to determine the size of the large component.

Lemma 10.2.20 *The number of small vertices is a.a.s.* $(1+o(1))(1-\beta)n$.

Proof. Let T_- and T_+ be the sizes of the Galton–Watson branching processes with respective offsping distributions $\mathcal{B}(n-k_-,\lambda/n)$ and $\mathcal{B}(n,\lambda/n)$, and let $T=C(v_0)$. Note that $P(T\le k_-)$ is the probability that v_0 is small. We have

$$P(T_+\le k_-)\le P(T\le k_-)\le P(T_-\le k_-)\,.$$

Let $T_{Poi}(\lambda)$ be the size of a branching process with $Poi(\lambda)$ as offspring distribution. Using estimates based on the results of Examples 10.1.3 and 10.1.10, it can be shown that

$$\lim_{n\uparrow\infty} P(T_\pm\le k_-)=\mathcal{P}_e(\lambda)\,.$$

Therefore $P(v_0 \text{ is small}) = (\mathcal{P}_e(\lambda)+o(1))$. In other notation, $P(Z(v_0)=1) = (\mathcal{P}_e(\lambda)+o(1))$, where $Z(v)$ is the indicator of the event that v is a small vertex.

Therefore, letting $N=\sum_v Z(v)$ be the number of small vertices, $E[N]=(\mathcal{P}_e(\lambda)+o(1))n$. By Chebyshev's inequality,

$$P(|N-E[N]|\ge\gamma E[N])\le\frac{\text{Var}\,(N)}{\gamma^2E[N]^2}=\frac{1}{\gamma^2}\left(\frac{E[N^2]}{E[N]^2}-1\right)$$

and therefore, if

$$\frac{E[N^2]}{E[N]^2}=1+o(1)\,,\qquad(\star)$$

we have that

$$P(|N-E[N]|\ge\gamma E[N])\le\frac{1}{\gamma^2}\times o(1)=o(1)$$

for any sufficiently slowly growing function $\gamma=\gamma(n)$, and this is enough to prove that $N=(1+o(1))E[N]$.

It remains to prove $(\star)$. We have, using the fact that $Z(v)^2=Z(v)$ and letting Z be any random variable with the common distribution of the $Z(v)$'s,

$$\begin{aligned}E\left[N^2\right]&=E\left[\left(\sum_v Z(v)\right)^2\right]=\sum_v E\left[Z(v)^2\right]+\sum_{u\ne v}E\left[Z(u)Z(v)\right]\\&=\sum_v E\left[Z(v)\right]+\sum_{u\ne v}E\left[Z(u)Z(v)\right]\\&=E\left[N\right]+\sum_{u\ne v}P(\text{both } u \text{ and } v \text{ are small})\\&=E\left[N\right]+\sum_v P(v\text{ is small})\sum_{u\ne v}P(u\text{ is small}\,|\,v\text{ is small})\,.\end{aligned}$$

Now

$$\begin{aligned}&\sum_{u\ne v}P(u\text{ is small}\,|\,v\text{ is small})\\&=\sum_{u\ne v\,;\,u\text{ in the same component as }v}\;+\sum_{u\ne v;u\text{ not in the same component as }v}\\&\le k_-+n\mathcal{P}_e(\mathcal{B}(n,\lambda/n))\end{aligned}$$

where $\mathcal{P}_e(\mathcal{B}(n, \lambda/n))$ is the extinction probability of a Galton–Watson branching process with offspring distribution $\mathcal{B}(n, \lambda/n)$. Here, we have used the fact that there are at most k_- vertices in the component containing v, and that if u is not in $C(v)$,

$$P(u \text{ is small} \,|\, v \text{ is small}) = P(u \text{ is small in the graph } G\backslash\{w\,;\, w \text{ is in } C(v)\})\,,$$

a quantity which is less than $\mathcal{P}_e(\mathcal{B}(n - k_-, \lambda/n))$. Therefore

$$E\left[N^2\right] \leq E[N] + n^2\mathcal{P}_e(\mathcal{B}(n - k_-, \lambda/n))^2(1 + o(1)) = E[N] + n^2(\mathcal{P}_e(\lambda) + o(1))^2\,.$$

(Here again the last approximation will use the results of Examples 10.1.3 and 10.1.10 and is left for the reader.) Finally

$$\text{Var } N \leq E[N] + n^2(\mathcal{P}_e(\lambda) + o(1))^2 - E\left[N\right]^2 \leq E[N] + o(E\left[N\right]^2)\,.$$

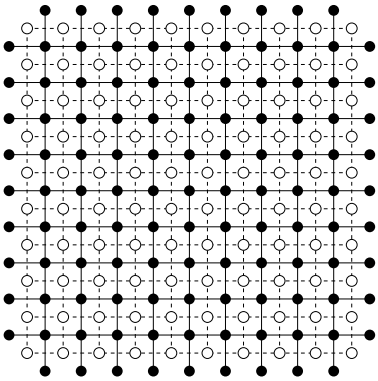
□

10.3 Percolation

10.3.1 The Basic Model

The phenomenon of physics called percolation concerns the situation of a porous material covering the whole plane and on which a drop of water falls. It refers to the possibility of a large surface to be wetted. A mathematical model will be given below. Some preliminary definitions are needed.

Consider the set $V = \mathbb{Z}^2$ (the (infinite) grid or (infinite) lattice) whose elements are called nodes or vertices. Let $\mathcal{E}_{NN}$ be the collection of nearest-neighbour potential edges, that is the collection of all unordered pairs $\langle u, v\rangle$ of mutually adjacent vertices[2]. A percolation graph on $\mathbb{Z}^2$ is, by definition, a graph $G = (V, \mathcal{E})$ where $V = \mathbb{Z}^2$ and $\mathcal{E}$ is a subset of $\mathcal{E}_{NN}$. The graph $(V, \mathcal{E}_{NN})$ is called the fully connected percolation graph. The dual grid (or lattice) in two dimensions V' is the original grid $V = \mathbb{Z}^2$ shifted by $(\frac{1}{2}, \frac{1}{2})$ (its vertices are of the form $(i', j') = (i + \frac{1}{2}, j + \frac{1}{2})$).



The grid $\mathbb{Z}^2$ and its dual

[2]Vertex $u = (i, j)$ has 4 adjacent vertices $v = (i + 1, j), (i - 1, j), (i, j + 1), (i, j - 1)$.

In a given percolation graph G, a path from vertex u to vertex v is, by definition, a sequence of vertices $v_0, v_1, \ldots, v_m$ such that $u = v_0$ and $v = v_m \neq u$, and for all i $(0 \leq i \leq m-1)$ $\langle v_i, v_{i+1}\rangle$ is an edge of G. Note that in this definition the extremities u and v must be different. Such a path is called loop-free if the vertices thereof are distinct. A loop-free path is also called a self-avoiding walk. If in addition there is an edge linking u and v, the sequence $v_0 = u, v_1, \ldots, v_m = v, u$ is called a circuit (we insist that it has to be loop-free to be so called).

The random percolation graph G_p on $\mathbb{Z}^2$, where $p \in [0,1]$, is a random element taking its values in the set of percolation graphs on $\mathbb{Z}^2$, and whose collection $\mathcal{E}_p$ of edges is randomly selected according to the following procedure. Let be given a collection $\{U_{\langle u,v\rangle}\}_{\langle u,v\rangle \in \mathcal{E}_{NN}}$ of IID random variables uniformly distributed on $[0,1]$, called the random generators. Then the potential edge $\langle u, v\rangle$ is included in $\mathcal{E}_p$ (becomes an edge of G_p) if and only if $U_{\langle u,v\rangle} \leq p$. Thus, a potential edge becomes an edge of G_p with probability p independently of all other potential edges. The specific procedure used to implement this selection allows us to construct all the random percolation graphs simultaneously, using the same collection of random generators. In particular, if $0 \leq p_1 < p_2 \leq 1$, $G_{p_1} \subseteq G_{p_2}$, by which it is meant that $\mathcal{E}_{p_1} \subseteq \mathcal{E}_{p_2}$ (see Example 16.1.1).

Two vertices u and v of a given percolation graph are said to be in the same component, or to be connected, if there exists a path of this graph connecting them. A component of the percolation graph is a set C of mutually connected vertices such that no vertex outside C is connected to a vertex in C. Its cardinality is denoted by $|C|$. Denote by $C(G, u)$ the component of the percolation graph containing vertex u.

We now introduce the notion of a dual percolation graph. The dual percolation graph of a given percolation graph G on $V = \mathbb{Z}^2$ is a percolation graph on the dual grid V' which has an edge linking two adjacent vertices u' and v' if and only if this edge does not cross an edge of G. Denote by G' such a graph. In particular G'_p is the dual random percolation graph of the random percolation graph G_p.

Percolation is said to occur in a given percolation graph if there exists an infinite component.

10.3.2 The Percolation Threshold

The fundamental result of this section is:

Theorem 10.3.1 *There exists a critical value $p_c \in [\frac{1}{3}, \frac{2}{3}]$ such that the probability that G_p percolates is null if $p < p_c$ (the subcritical case), and equal to 1 if $p > p_c$ (the supercritical case).*

The proof will be given after some preliminaries. We start with a trivial observation concerning $C(G_p, 0)$ (0 stands for $(0,0)$, the origin of $\mathbb{Z}^2$). Defining

$$\theta(p) = P(|C(G_p, 0)| = \infty),$$

the probability that the origin belongs to an unbounded component of the random percolation graph G_p, we have that $\theta(0) = 0$ and $\theta(1) = 1$. Next, $\theta : [0,1] \to [0,1]$ is non-decreasing. Indeed if $0 \leq p_1 < p_2 \leq 1$, $G_{p_1} \subseteq G_{p_2}$, and therefore $|C(G_{p_1}, 0|) = \infty$ implies $|C(G_{p_2}, 0|) = \infty$. This remark provides an opportunity to recall the notions of increasing set and increasing function defined on the set of percolation graphs.

Definition 10.3.2 *A set A of percolation graphs is called non-decreasing if for all percolation graphs $G^{(1)}, G^{(2)}$ such that $G^{(1)} \subseteq G^{(2)}$, $G^{(1)} \in A$ implies that $G^{(2)} \in A$. A function f taking its values in the set of percolation graphs on $\mathbb{Z}^2$ is called non-decreasing if $G^{(1)} \subseteq G^{(2)}$ implies that $f(G^{(1)}) \leq f(G^{(2)})$.*

In particular 1_A is a non-decreasing function whenever A is a non-decreasing set.

EXAMPLE 10.3.3: The event $\{|C(G,0)| = +\infty\}$ is a non-decreasing event. So is the event "there is a path in G from a given vertex u to a given vertex v".

In very much the same way as we proved the non-decreasingness of the function θ, one can prove the following result.

Lemma 10.3.4 *If A is a non-decreasing event, then the function $p \to P(G_p \in A)$ is non-decreasing. If f is a non-decreasing function, then the function $p \to E[f(G_p)]$ is non-decreasing.*

Theorem 10.3.1 will be obtained as a consequence of the following lemma.

Lemma 10.3.5 *There exists a critical value $p_c \in [\frac{1}{3}, \frac{2}{3}]$ such that $\theta(p) = 0$ if $p < p_c$, and $\theta(p) > 0$ if $p > p_c$.*

Proof. Part 1. We show that for $p < \frac{1}{3}$, $\theta(p) = 0$. Call $\sigma(n)$ the number of loop-free paths starting from 0 and of length n. Such a path can be constructed progressively edge by edge starting from the origin. For the first edge (from 0) there are 4 choices, and for each of the $n-1$ remaining edges there are at most 3 choices. Hence the bound

$$\sigma(n) \leq 4 \times 3^{n-1}.$$

We order these $\sigma(n)$ paths arbitrarily.

Let $N(n, G)$ be the number of paths of length n starting from 0 in a percolation graph G. If there exists in G_p an infinite path starting from 0 (or equivalently, if there exists an infinite component of G_p containing the origin) then, for each n there exists at least one path of length n starting from 0, that is,

$$\{|C(G_p, 0)| = \infty\} = \cap_{n=1}^{\infty} \{N(n, G_p) \geq 1\}$$

and therefore, for all $n \geq 1$,

$$\theta(p) \leq P(N(n, G_p) \geq 1) = P\left(\cup_{i=1}^{\sigma(n)} \{Y_i(G_p) = 1\}\right) ,$$

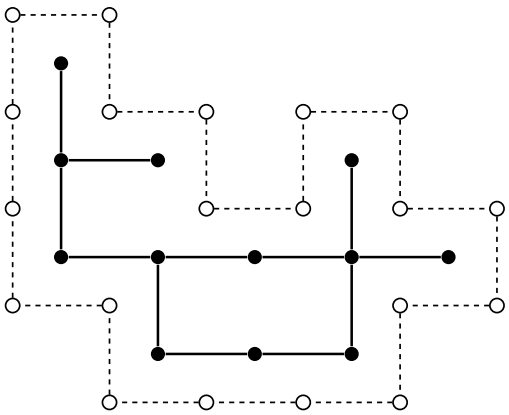
where $Y_i(G_p)$ is the indicator function for the presence in G_p of the i-th loop-free path of length n starting from 0 in the fully connected percolation graph. Therefore, by the union bound,

$$\theta(p) \leq \sum_{i=1}^{\sigma(n)} P\left(Y_i(G_p) = 1\right) \leq \sigma(n) p^n = 4p(3p)^{n-1} .$$

If $p < \frac{1}{3}$, this quantity tends to 0 as $n \uparrow +\infty$.

Part 2. We show that for $p > \frac{2}{3}$, $\theta(p) > 0$.

The statement that $|C(G_p, 0)| < \infty$ is equivalent to saying that 0 is surrounded by a circuit of G'_p.



Call $\rho(n)$ the number of circuits of length n of the fully connected dual graph that surround the origin of the original grid. We have that

$$\rho(n) \leq n\sigma(n-1) ,$$

which accounts for the fact that such circuits contain at most a path of length $n-1$ that passes through a dual vertex of the form $(\frac{1}{2}, \frac{1}{2} + i)$ for some $0 \leq i < n$.

The set $\mathcal{C}$ of circuits of the fully connected dual percolation graph that surround the origin 0 of the original grid is countable. Denote by $\mathcal{C}_k \subset \mathcal{C}$ the subset of such circuits that surround the box $B_k \subset S = \mathbb{Z}^2$ of side length k centered at the origin 0. Call $\Delta(B_k)$ the boundary of B_k. The two following statements are equivalent:

(i) There is no circuit of $\mathcal{C}_k$ that is a circuit of G'_p,

(ii) There is at least one vertex $u \in \Delta(B_k)$ with $|C(G_p, u)| = \infty$.

Therefore

$$P\left(\cup_{u\in\Delta(B_k)}\{|C(G_p,u)|=\infty\}\right) = P\left(\cap_{c\in\mathcal{C}_k}\{c \text{ is not a circuit of } G'_p\}\right)$$
$$= 1 - P\left(\cup_{c\in\mathcal{C}_k}\{c \text{ is a circuit of } G'_p\}\right)$$
$$\geq 1 - \sum_{c\in\mathcal{C}_k} P\left(\{c \text{ is a circuit of } G'_p\}\right) . \qquad (10.8)$$

A given circuit of length n occurs in the dual random percolation graph G'_p with probability $(1-p)^n$ and therefore

$$\sum_{c\in\mathcal{C}_k} P\left(\{c \text{ is a circuit of } G'_p\}\right) \leq \sum_{n=4k}^{\infty} n\sigma(n-1)(1-p)^n \leq \frac{4}{9}\sum_{n=4k}^{\infty}(3(1-p))^n n\,. \qquad (10.9)$$

If $p > \frac{2}{3}$, the series $\sum_{n=1}^{\infty}(3(1-p))^n n$ converges, and therefore, for k large enough, $\frac{4}{9}\sum_{n=4k}^{\infty}(3(1-p))^n n < 1$. From this and (10.8), we obtain that for large enough k,

$$P\left(\cup_{u\in\Delta(B_k)}\{|C(G_p,u)|=\infty\}\right) > 0$$

which implies that $P(|C(G_p,0)|=\infty) > 0$ since there is a positive probability that there exists in G_p a path from the origin to any point of the boundary of B_k. □

It remains to conclude the proof of Theorem 10.3.1.

Proof. Let A be the non-decreasing event "there exists an infinite component". The random variable $1_A(G_p)$ does not depend on any finite subset of the collection of independent variables $\{U_{\langle u,v\rangle}\}_{\langle u,v\rangle\in\mathcal{E}_{NN}}$. Therefore, by Kolmogorov's 0–1-law, $P(G_p \in A)$ can take only one of the values 0 or 1. Since on the other hand $P(G_p \in A) \geq \theta(p)$, $\theta(p) > 0$ implies $P(G_p \in A) = 1$. Also, by the union bound,

$$P(G_p \in A) \leq \sum_{u\in\mathbb{Z}^2} P(|C(G_p,u)| = +\infty)$$
$$= \sum_{u\in\mathbb{Z}^2} P(|C(G_p,0)| = +\infty) = \sum_{u\in\mathbb{Z}^2}\theta(p)\,,$$

and therefore, $\theta(p) = 0$ implies that $P(G_p \in A) = 0$. □

Books for Further Information

[Harris, 1989 (Dover ed.)] is the historical book reference on branching processes. See also [Karlin and Taylor, 1975]. [Athreya and Ney, 1972] is a point of entry to the modern theory.

The three following monographs concern mainly the Erdös–Rényi random graphs: [Bollobás, 2nd ed. 2010], [Janson, Luczak and Rućinski, 2000], [Frieze and Karonski, 2015].

[Draief and Massoulié, 2010] is a concise treatment of basically all aspects of the theory, structured around the important theme of epidemics.

[Kesten, 1982] is the historical reference on percolation. More recent references are [Bollobás and Riordan, 2006], [Grimmett, 1999] and [Grimmett, 2010]. In french, [Werner, 2009].

Applications to communications are developed in [Franceschetti and Meester, 2007].

10.4 Exercises

Exercise 10.4.1. Extinction
Compute the probability of extinction of a branching process with one ancestor when the probabilities of having $0, 1$, or 2 sons are respectively $\frac{1}{4}$, $\frac{1}{4}$, and $\frac{1}{2}$.

Exercise 10.4.2. Branching process transitions
Show that $p_{ij} := P(X_{n+1} = j \mid X_n = i)$ of the transition matrix of this chain is the coefficient of z^j in $(g(z))^i$, where $g(z)$ is the generating function of the number of descendants of a given individual.

Exercise 10.4.3. Several ancestors
Give the survival probability in the model of Section 10.1 with k ancestors, $k > 1$, in terms of the offspring generating function g_Z.

Exercise 10.4.4. Mean and variance of the branching process
Give the mean and variance of X_n in the model of Section 10.1 with one ancestor in terms of the mean m_Z and the variance σ_Z^2 of the offspring distribution.

Exercise 10.4.5. Size of the branching tree
When the probability of extinction is 1 ($m < 1$), show that the generating function g_Y of the size of the branching tree, $Y := \sum_{n\geq 0} X_n$, satisfies the equation

$$g_Y(z) = z\, g_Z\left(g_Y(z)\right) ,$$

where g_Z is the offspring generating function.

Exercise 10.4.6. Conjugate offspring distribution
Fix a probability distribution $\{b_k\}_{k\in\mathbb{N}}$ that is critical ($\sum_{k\in\mathbb{N}} k\, b_k = 1$) and such that $b_0 > 0$. For any $\lambda > 0$, define the exponentially tilted distribution $\{a_k(\lambda)\}_{k\in\mathbb{N}}$ by

$$a_k(\lambda) = b_k \frac{\lambda^k}{g_b(\lambda)} = b_k \frac{\lambda^k}{\sum_{k\geq 0} b_k \lambda^k} .$$

(a) Verify that this is indeed a probability distribution on $\mathbb{N}$ that is supercritical if $\lambda > 1$ and subcritical if $\lambda < 1$.

(b) Take for $\{b_k\}_{k\in\mathbb{N}}$ the Poisson distribution with mean 1. What is the conjugate distribution?

(c) A parameter μ is said to be a conjugate parameter of $\lambda > 0$ if

$$\frac{\lambda}{g_b(\lambda)} = \frac{\mu}{g_b(\mu)}\,.$$

Let $\lambda > 1$. Prove that there exists a unique conjugate parameter μ of λ such that $\mu \neq \lambda$ which satisfies $\mu < 1$ and is given by

$$\mu = \lambda P(\mathcal{E})(\lambda)\,,$$

where $P(\mathcal{E})(\lambda)$ is the probability of extinction relative to the offspring distribution $\{a_k(\lambda)\}_{k\in\mathbb{N}}$.

(d) Let $\lambda > 1$. Show that the distribution of the supercritical branching process history with offspring distribution $\{a_k(\lambda)\}_{k\in\mathbb{N}}$ conditioned on extinction is identical to that of the subcritical branching process history with offspring distribution $\{a_k(\mu)\}_{k\in\mathbb{N}}$ where $\mu = \lambda P(\mathcal{E})(\lambda)$ is the conjugate parameter of λ.

Exercise 10.4.7. AVERAGE CHARACTERISTICS
In the random graph $\mathcal{G}(n,p)$, compute

a. the average number of isolated vertices,

b. the average number of cycles,

c. the average number of paths of length 2, and

d. the average number of vertices of degree d.

Exercise 10.4.8. UNION OF RANDOM GRAPHS
Let $\mathcal{G}_1$, $\mathcal{G}_2$, ..., $\mathcal{G}_k$ be k independent copies of $\mathcal{G}(n,p)$. Prove that the union of these copies is a $\mathcal{G}(n, 1-(1-p)^k)$.

Exercise 10.4.9. COMPARISON OF ASYMPTOTICS
Let $\mathcal{P}$ be a graph property. Let $p = p_n$ and $m = m_n$ be such that $p_n = \frac{m_n}{\binom{n}{2}}$ and

$$\binom{n}{2} p_n \to \infty, \qquad \binom{n}{2}(1-p_n) \to \infty.$$

Show that for large n, $P(\mathcal{G}_{n,m_n} \in \mathcal{P}) \leq 10 m^{\frac{1}{2}} P(\mathcal{G}(n,p_n) \in \mathcal{P})$.

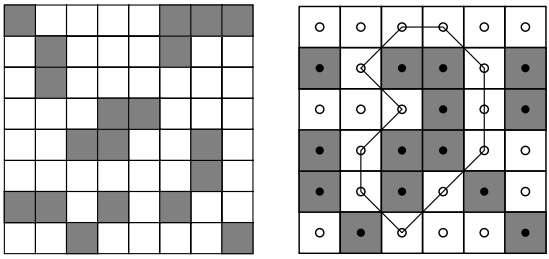
Exercise 10.4.10. FOREST GRAPH
Let ω be a function growing (arbitrarily slowly) to ∞ as $n \uparrow \infty$, say $\omega(n) = \log\log n$. Prove that if $np_n \leq \omega(n)^{-1}$, $\mathcal{G}(n,p_n)$ is a forest (contains no cycles) w.h.p.

Exercise 10.4.11. ONLY EDGES AND ISOLATED VERTICES
If $m(n) \ll n^{\frac{1}{2}}$, $\mathcal{G}_{m(n)}$ contains only isolated vertices and edges w.h.p.

Exercise 10.4.12. THE LIMIT OF THE EXPLORATION BRANCHING PROCESS
In the notation of section 10.2.3, for fixed k, what is the limit distribution as $n \uparrow \infty$ of the vector $(D_1, \ldots, D_k)$?

Exercise 10.4.13. PERCOLATION IN THE D-DIMENSIONAL GRID
Adapt the arguments for $d = 2$ to prove the existence of a percolation threshold p $(0 < p < 1)$ in the d-dimensional percolation graph.

Exercise 10.4.14. THE SITE PERCOLATION MODEL
The grid $V = \mathbb{Z}^2$ is now viewed as a paving of the plane by squares $S(v)$ of area 1 centered at vertices $v \in \mathcal{V}$. The random site percolation graph is generated by an IID collection of $\{0,1\}$-valued random variables $\{Y(v)\}_{v\in V}$, with $P(Y(v) = 1) = p \in (0,1)$. If $Y(v) = 1$, the square $S(v)$ is coloured in grey, and otherwise it is left blank. Two vertices v and w are said to be connected if and only if there exists at least one sequence of vertices $u_0 = v, u_1, \ldots, u_k = w$ such that for all i, $1 \leq i \leq k$, the squares $S(v_{i-1})$ and $S(v_i)$ share a side and are both grey. A component of the site percolation random graph is any set of vertices such that any pair of sites therein is connected. Inspired by the proof of Theorem 10.3.1, show the existence of a critical site percolation value $\tilde{p}_c$. (Since there is no concept of dual lattice in the random site percolation graph, one must find another definition for a circuit around a component. The figure below will replace a formal definition.)



Exercise 10.4.15. PERCOLATION ON THE k-ARY TREE
Consider the k-ary tree (a connected graph without cycle where each vertex has exactly k outgoing edges). A random bond percolation graph is generated by deleting edges independently with probability $1 - p$. Give the bond percolation critical probability in this case.

Exercise 10.4.16. CONNECTING THE OPPOSITE SIDES
Consider a square of the grid $\mathbb{Z}^2$ with n vertices on each side. Consider the bond percolation random graph generated by deleting edges independently with probability $\frac{1}{2}$. What is the probability that two given opposite sides are connected?

Exercise 10.4.17. TRIVIAL BOUND PERCOLATION
Give a simple example of graph for which the critical value of bound percolation is 0 (*resp.*, 1).

Chapter 11

Coding Trees

11.1 Entropy

11.1.1 The Gibbs Inequality

Entropy is an example of a physical concept that has found a new life in the engineering and computing sciences. This chapter concentrates on the aspects involving tree structures, namely source coding and the generation of discrete random variables from random numbers.

Let X be a random variable with values in a finite set $\mathcal{X}$, with probability distribution:

$$p(x) := P(X = x) \qquad (x \in \mathcal{X})\,.$$

The entropy of X is, by definition, the quantity

$$H(X) := -E\left[\log(p(X))\right] = -\sum_{x\in\mathcal{X}} p(x)\log p(x)\,. \tag{11.1}$$

(Recall the usual "log convention": $0\log 0 = 0$ and $a\log 0 = -\infty$ when $a > 0$.)

Remark 11.1.1 The notation $H(X)$ is ambiguous in that it seems to indicate a function of X, and therefore a random variable. In fact, $H(X)$ is a deterministic number, a function of the probability distribution of X. The less ambiguous notation $H(\mathbf{p})$, where $\mathbf{p} := \{p(x), x \in \mathcal{X}\}$, is also used.

The basis of the logarithm must be chosen once and for all. In base D, we shall write $H(X) = H_D(X)$. In base 2, the entropy will be expressed in bits, and in nats in base e.

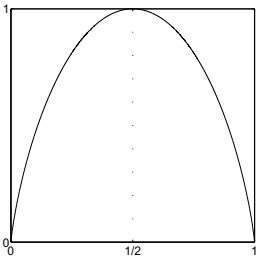
One sometimes calls $H(X)$ the quantity of information contained in X, for a reason that will be clear later on with the questionaire interpretation.

EXAMPLE 11.1.2: $\mathcal{X} = \{0, 1\}$, $P(X = 1) = p$, $P(X = 0) = 1 - p$. Then $H_2(X) = h_2(p)$, where

$$h_2(p) := -p\log_2 p - (1-p)\log_2(1-p)\,.$$

The function h_2 is concave and its maximum ($= 1$) is attained for $p = \frac{1}{2}$. It has an infinite slope at $x = 0$ and $x = 1$.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_11



The definition of entropy applies to variables taking their values in a discrete possibly infinite set $\mathcal{X}$, but it may then be an infinite quantity.

EXAMPLE 11.1.3: ENTROPY OF THE GEOMETRIC DISTRIBUTION. The corresponding entropy is

$$\begin{aligned}
&-\sum_{i\geq 1} p(1-p)^{i-1}\log p(1-p)^{i-1}\\
&\quad=-\sum_{i\geq 1} p(1-p)^{i-1}\left(\log p-\log(1-p)+i\log(1-p)\right)\\
&\quad=-\log p+\log(1-p)-\frac{1}{p}\log(1-p)=\log\frac{1-p}{p}-\frac{1}{p}\log(1-p).
\end{aligned}$$

Theorem 11.1.4 *Let $(p(x), x\in\mathcal{X})$ and $(q(x), x\in\mathcal{X})$ be two probability distributions on $\mathcal{X}$. Then (Gibbs inequality):*

$$-\sum_{x\in\mathcal{X}} p(x)\log p(x)\leq-\sum_{x\in\mathcal{X}} p(x)\log q(x)\,, \tag{11.2}$$

with equality if and only if $p(x)=q(x)$ for all $x\in\mathcal{X}$.

Proof. This inequality is a direct consequence of Jensen's inequality. In fact

$$\begin{aligned}
E\left[\log\left(\frac{q(X)}{p(X)}1_{\{p(X)>0\}}\right)\right]&\leq\log\left(E\left[\frac{q(X)}{p(X)}1_{\{p(X)>0\}}\right]\right)\\
&=\log\left(\sum_{x\in\mathcal{X}}\frac{q(x)}{p(x)}1_{\{p(x)>0\}}p(x)\right)\\
&=\log\left(\sum_{x\in\mathcal{X}}q(x)1_{\{p(x)>0\}}\right)\\
&\leq\log\left(\sum_{x\in\mathcal{X}}q(x)\right)=\log 1=0
\end{aligned}$$

and

$$E\left[\log\left(\frac{q(X)}{p(X)}1_{\{p(X)>0\}}\right)\right]$$
$$= E\left[\log q(X)1_{\{p(X)>0\}}\right] - E\left[\log p(X)1_{\{p(X)>0\}}\right]$$
$$= \sum_{x\in\mathcal{X}} p(x)\log q(x)1_{\{p(x)>0\}} - \sum_{x\in\mathcal{X}} p(x)\log p(x)1_{\{p(x)>0\}}$$
$$= \sum_{x\in\mathcal{X}} p(x)\log q(x) - \sum_{x\in\mathcal{X}} p(x)\log p(x)\,.$$

Therefore $-\sum_{x\in\mathcal{X}} p(x)\log p(x) \le -\sum_{x\in\mathcal{X}} p(x)\log q(x)$. □

Theorem 11.1.5 *Let X be a random variable with values in a finite set $\mathcal{X}$. Then, denoting by $|\mathcal{X}|$ the cardinality (number of elements) of $\mathcal{X}$,*

$$0 \le H(X) \le \log|\mathcal{X}|\,. \tag{11.3}$$

Moreover, $H(X) = 0$ if and only if X is deterministic, and $H(X) = \log|\mathcal{X}|$ if and only if X is uniformly distributed on $\mathcal{X}$.

Proof. The inequality on the left is obvious. The one on the right follows from the Gibbs inequality with $q(x) = \frac{1}{|\mathcal{X}|}$. The value 0 is possible only when for all $x \in \mathcal{X}$, $p(x)\log p(x) = 0$, that is to say when $p(x) = 0$ or $p(x) = 1$. As the $p(x)$'s sum to 1, there exists in this case one and only one $x_0 \in \mathcal{X}$ such that $p(x_0)) = 1$, that is $P(X = x_0) = 1$.
Equality $H(X) = \log|\mathcal{X}|$ is equivalent to

$$-\sum_{x\in\mathcal{X}} p(x)\log p(x) = -\sum_{x\in\mathcal{X}} p(x)\log\left(\frac{1}{|\mathcal{X}|}\right)\,,$$

which is possible (according to Theorem 11.1.4 with $q(x) = \frac{1}{|\mathcal{X}|}$) only if $p(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$. □

We have just showed that the uniform distribution maximizes entropy over all distributions on a given finite set. The next example treats a similar issue for positive integer-valued random variables.

EXAMPLE 11.1.6: GEOMETRIC DISTRIBUTION MAXIMIZES ENTROPY. Prove that the geometric distribution maximizes entropy among all positive integer-valued random variables with given finite mean μ.

Proof. The corresponding maximization problem is solved by the Lagrangian method. We find a solution to the equations

$$\frac{\partial}{\partial p_i}\left(-\sum_i p_i\log p_i\right) - \lambda\frac{\partial}{\partial p_i}\left(\sum_i ip_i\right) = 0 \qquad (i \ge 1)$$

that is

$$-\log p_i - 1 - \lambda i = 0 \qquad (i \geq 1)\,,$$

which gives $p_i = e^{-1-\lambda i}$ $(i \geq 1)$. The constraint $\sum_i p_i = 1$ finally yields

$$p_i = p(1-p)^{i-1}$$

for some $p \in (0,1)$. In fact, $p = \mu^{-1}$, where μ is the mean of the geometric distribution. □

The following result will be needed later on.

Lemma 11.1.7 *The entropy of a positive integer-valued random variable X with given finite mean μ satisfies the inequality*

$$H(X) \leq (\mu+1)\log(\mu+1) - \mu\log\mu\,.$$

Proof. The announced inequality follows from Exercise 11.1.6 according to which the entropy of any positive integer-valued random variable X with given finite mean μ is maximized by the entropy of the geometric random variable with mean μ, namely $\mu\log\mu - (\mu-1)\log(\mu-1)$. We are left to prove that

$$\mu\log\mu - (\mu-1)\log(\mu-1) \leq (\mu+1)\log(\mu+1) - \mu\log\mu$$

or equivalently

$$2\mu\log\mu \leq (\mu-1)\log(\mu-1) + (\mu+1)\log(\mu+1)\,,$$

which follows from the convexity of the function $x \to x\log x$. □

Theorem 11.1.8 *Let $X_1, \ldots, X_n$ be independent random variables, respectively with values in the finite spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$, and with entropies $H(X_1), ..., H(X_n)$. Then*

$$H(X_1, ..., X_n) = \sum_{i=1}^{n} H(X_i)\,. \tag{11.4}$$

In particular, if $X_1, ..., X_n$ are IID, *$H(X_1, ..., X_n) = nH(X_1)$.*

Proof. Let $p_i(x_i) = P(X_i = x_i)$ be the distribution of the variable $X_i \in \mathcal{X}_i$. By independence:

$$P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} p_i(x_i)\,,$$

and therefore

$$H(X_1,...,X_n) = -E\left[\log\left(\prod_{i=1}^{n} p_i(X_i)\right)\right] = -E\left[\sum_{i=1}^{n}\log p_i(X_i)\right]$$
$$= -\sum_{i=1}^{n} E\left[\log p_i(X_i)\right] = \sum_{i=1}^{n} H(X_i)\,.$$

□

Theorem 11.1.9 *Let $X_1,\ldots,X_n$ be random variables with values in the finite spaces $\mathcal{X}_1,\ldots,\mathcal{X}_n$ respectively. Let $H(X_1),...,H(X_n)$ be their respective entropies. Then,*

$$H(X_1,...,X_n) \le \sum_{i=1}^{n} H(X_i)\,, \tag{11.5}$$

with equality if and only if $X_1,...,X_n$ are independent.

Proof. By recurrence. It suffices to give the proof for two variables, X and Y, since any vector of discrete random element is a discrete random variable. We have that

$$H(X,Y) - H(X) - H(Y) = -E\left[\log p_{X,Y}(X,Y)\right] + E\left[\log\left(p_X(X)p_Y(Y)\right)\right]\,.$$

The Gibbs inequality applied to the probability distributions $p_{(X,Y)}(x,y)$ and $p_X(x)p_Y(y)$ on $\mathcal{X}\times\mathcal{Y}$ give the result. (Equality occurs if and only if the two distributions coincide, that is, if the two variables are independent.) □

Boltzmann's Interpretation of Entropy

The physicist Boltzmann made the hypothesis that the entropy (in the thermodynamical sense) of a system of n particles, each particle being in one among k microstates, is proportional to the number of undistinguishable "configurations" that the system could take. Consider for instance a system of n particles, each in one of the microstates $E_1,\ldots,E_k$. For example, the particles are the electrons of n hydrogen atoms (recall that a hydrogen atom has a single electron) and $E_1,\ldots,E_k$ are the energy levels of a given hydrogen electron. The collection of n particles is called the *system*. A macrostate of the system is a k-tuple $(n_1,\ldots,n_k)$ where n_i is the number of particles in micro-state E_i. There are $c(n,n_1,\ldots,n_k) = n!/n_1!\ldots n_k!$ configurations corresponding to the macrostate $(n_1,\ldots,n_k)$. In order to compare the number of configurations corresponding to two different macrostates $(n_1,\ldots,n_k)$ and $(\tilde{n_1},\ldots,\tilde{n_k})$, we form the ratio

$$\frac{n!}{n_1!\ldots n_k!}\Big/\frac{n!}{\tilde{n_1}!\ldots\tilde{n_k}!} = \prod_{i=1}^{k}\frac{\tilde{n_i}!}{n_i!}\,,$$

and let n tend to infinity in such a way that

$$\lim_{n\uparrow\infty}\frac{n_i}{n} = p_i \quad , \quad \lim_{n\uparrow\infty}\frac{\tilde{n_i}}{n} = \tilde{p_i}\,.$$

Stirling's equivalence formula gives

$$\prod_{i=1}^{k} \frac{\tilde{n_i}!}{n_i!} \simeq e^{n(-\sum_{i=1}^{k} p_i \log p_i + \sum_{i=1}^{k} \tilde{p_i} \log \tilde{p_i})} .$$

Therefore the macrostates $(n_1, \ldots, n_k)$ with largest entropy are the most likely. If the physics of the system do not favor any configuration satisfying the macroscopic constraints, the system will be in the macrostate that maximizes its entropy under these constraints (Boltzmann's principle).

EXAMPLE 11.1.10: THE GIBBS DISTRIBUTION. In the hydrogen example, suppose that the only constraint is that the average energy of an electron be fixed at the value E:

$$\sum_{i=1}^{k} p_i E_i = E .$$

The state of the system will be the one that maximizes $-\sum_{i=1}^{k} p_i \log p_i$ under the constraint of energy and the constraint of normalization. The Lagrange method of multipliers requires one to solve equations

$$\frac{\partial}{\partial p_i} \left(-\sum_i p_i \log p_i + \lambda(\sum_i (p_i - E_i)) + (\sum_i p_i - 1) \right) = 0$$

for $1 \leq i \leq k$. This gives $-\log p_i - 1 + \lambda E_i + \mu$, that is $p_i = K e^{-\lambda E_i}$. By normalization,

$$p_i = \frac{e^{-\lambda E_i}}{Z(\lambda)} ,$$

where

$$Z(\lambda) := \sum_{i=1}^{k} e^{-\lambda E_i}$$

(the partition function). The parameter λ is determined by the energy constraint:

$$\sum_{i=1}^{k} E_i \frac{e^{-\lambda E_i}}{Z(\lambda)} = E .$$

11.1.2 Typical Sequences

Let $\mathcal{X}$ be a finite set of cardinality D, and let $X_1, ..., X_n$ be IID random variables with values in $\mathcal{X}$ and common probability distribution $(p(x),\ x \in \mathcal{X})$. In particular, the random vector $X_1^n = (X_1, ..., X_n)$ has the probability distribution

$$p(x_1^n) := p(x_1, ..., x_n) = \prod_{i=1}^{n} p(x_i) .$$

Let

$$H_D := E\left[-\log_D p(X_1)\right]$$

be the common entropy of the above variables.

For $\varepsilon > 0$, let

$$A_\varepsilon^{(n)} := \left\{ x_1^n := (x_1, \cdots, x_n) \,;\, \left| -\frac{1}{n}\sum_{i=1}^{n} \log_D p(x_i) - H_D \right| \leq \varepsilon \right\}$$

be the set of ε-typical sequences of order n.

Theorem 11.1.11 *We then have*

$$\lim_{n\to\infty} P(X_1^n \in A_\varepsilon^{(n)}) = 1$$

and

$$\left|A_\varepsilon^{(n)}\right| \leq D^{n(H_D+\varepsilon)}\,.$$

Proof.

$$P(X_1^n \in A_\varepsilon^{(n)}) = P\left(\left| -\frac{1}{n}\sum_{i=1}^{n} \log_D p(X_i) - H_D \right| \leq \varepsilon\right)\,.$$

By the weak law of large numbers (Example 3.1.3)

$$-\frac{1}{n}\sum_{i=1}^{n} \log_D p(x_i) \xrightarrow{Pr} -E\left[\log_D p(X_1)\right] = H_D\,,$$

that is, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left| -\frac{1}{n}\sum_{i=1}^{n} \log_D p(x_i) - H_D \right| > \varepsilon\right) = 0\,,$$

and this is the first announced result. By definition of $A_\varepsilon^{(n)}$, if x_1^n belongs to this set,

$$D^{-n(H_D+\varepsilon)} \leq p(x_1^n)\,,$$

and therefore,

$$P(X_1^n \in A_\varepsilon^{(n)}) = \sum_{x_1^n \in A_\varepsilon^{(n)}} p(x_1^n) \geq D^{-n(H_D+\varepsilon)} \left|A_\varepsilon^{(n)}\right|\,.$$

The left-hand side is ≤ 1 and therefore

$$1 \geq \left|A_\varepsilon^{(n)}\right| D^{-n(H_D+\varepsilon)}\,.$$

□

The result of Theorem 11.1.11 can be used (in theory) for data compression. Construct a mapping $c : \mathcal{X}^n \to \mathcal{X}^{\lceil n(H_D+\varepsilon)\rceil+1}$ as follows. We first construct a mapping

$\tilde{c}^{(n)} : A_\varepsilon^{(n)} \to \mathcal{X}^{\lceil n(H_D+\varepsilon)\rceil}$. In view of the inequality in Theorem 11.1.11, such a mapping can be chosen to be *injective*. Define a mapping (called the compression code) $c^{(n)} : \mathcal{X}^n \to \mathcal{X}^*$ by:

$$c^{(n)}(x_1^n) = \begin{cases} \tilde{c}^{(n)}(x_1^n)1 & \text{if } x_1^n \in A_\varepsilon^{(n)} \\ 0 & \text{if } x_1^n \notin A_\varepsilon^{(n)}. \end{cases}$$

The restriction of $c^{(n)}$ to $A_\varepsilon^{(n)}$ is injective. Therefore one can always recover x_1^n from its code-word $c^{(n)}(x_1^n)$ if $X_1^n \in A_\varepsilon^{(n)}$. An error may occur only if $X_1^n \notin A_\varepsilon^{(n)}$, and the probability of such an event tends to 0 as $n \to \infty$. Decoding is therefore possible with an error as small as desired by choosing n large enough. For a sequence of symbols of $\mathcal{X}$ of length n, this coding scheme does not require n symbols, but (asymptotically) $(\lceil n(H_D+\varepsilon)\rceil + 1)$, which corresponds to a compression rate of

$$\frac{(\lceil n(H_D+\varepsilon)\rceil + 1)}{n}\,.$$

Choosing n large enough and ε small enough, this ratio can be made arbitrarily close to H_D (a quantity that is of course not greater than 1 since $H_D \le \log_D |\mathcal{X}| = \log_D D = 1$).

The following result shows that one cannot achieve a better compression rate.

Theorem 11.1.12 *Suppose that there exists a set $B^{(n)} \subseteq \mathcal{X}^n$ and a number $R > 0$ such that:*

$$\lim_{n\to\infty} P(X_1^n \in B^{(n)}) = 1$$

and

$$\left|B^{(n)}\right| \le D^{nR}\,.$$

Then, necessarily, $R \ge H_D$.

Proof. If $x_1^n \in A_\varepsilon^{(n)}$, then $p(x_1^n) \le D^{-n(H_D-\varepsilon)}$, and therefore:

$$\begin{aligned} P(X_1^n \in A_\varepsilon^{(n)} \cap B^{(n)}) &\le D^{-n(H_D-\varepsilon)} \left|A_\varepsilon^{(n)} \cap B^{(n)}\right| \\ &\le D^{-n(H_D-\varepsilon)} \left|B^{(n)}\right| \le D^{-n(H_D-\varepsilon-R)}\,. \end{aligned}$$

But, by hypothesis, $\lim_{n\to\infty} P(X_1^n \in B^{(n)}) = 1$ and (Theorem 11.1.11) $\lim_{n\to\infty} P(X_1^n \in A_\varepsilon^{(n)}) = 1$, so that

$$\lim_{n\to\infty} P(X_1^n \in A_\varepsilon^{(n)} \cap B^{(n)}) = 1\,.$$

Therefore:

$$\lim_{n\to\infty} D^{-n(H_D-\varepsilon-R)} \ge 1$$

which implies that $H_D - \varepsilon - R \le 0$, that is, $R \ge H_D - \varepsilon$. As ε is arbitrary, we obtain the announced result. □

11.1.3 Uniquely Decipherable Codes

For a given alphabet $\mathcal{A}$, we denote by $\mathcal{A}^*$ the collection of all finite chains of elements of $\mathcal{A}$ including the empty chain $\varnothing$. For instance, if $\mathcal{A} = \{0, 1\}$, the chains $y_1 = 0110$, $y_2 = 111$ and $y_3 = 0101010$ are in $\{0, 1\}^*$. Concatenating chains in $\mathcal{A}^*$ means that they are put together so as to form a chain in $\mathcal{A}^*$. In the example, 01101110101010 is obtained by concatenation of y_1, y_2 and y_3 (the resulting chain is denoted $y_1 * y_2 * y_3$ or, more simply $y_1y_2y_3$), or by concatenating y_1, $\varnothing$, y_2 and y_3. The length of a chain in $\mathcal{A}^*$ is the number of symbols that it contains (including repetitions). The length of the chain 01101110101010 is 14.

Let $\mathcal{X}$ be a finite set. A code of $\mathcal{X}$ is a function $c : \mathcal{X} \to \mathcal{A}^*$, where $\mathcal{A}$ is a finite set of cardinality D. The chain $c(x)$ is the code word associated with message $x \in \mathcal{X}$. One denotes by $l(c(x))$ the length of $c(x)$.

Definition 11.1.13 *Code c is said to be uniquely decypherable (UD) if for all integers $k \geq 1$, $l \geq 1$, and all $x_1, ..., x_k, y_1, ..., y_l \in \mathcal{X}$:*

$$c(x_1)...c(x_k) = c(y_1)...c(y_l) \Rightarrow k = l, x_1 = y_1, ..., x_k = y_k\,.$$

It is said to have the prefix property if there exists no pair $x, y \in \mathcal{X}$ $(x \neq y)$ such that $c(x)$ is a prefix of $c(y)$. Such a code is then called a prefix code.

The following result is an immediate consequence of the definition of a prefix code.

Theorem 11.1.14 *A prefix code is uniquely decipherable.*

EXAMPLE 11.1.15: Consider the following codes for $\mathcal{X} = \{1, 2, 3, 4\}$ using the binary alphabet $\mathcal{A} = \{0, 1\}$,

1. $c(1) = 00$, $c(2) = 01$, $c(3) = 10$, $c(4) = 11$.
2. $c(1) = 0$, $c(2) = 1$, $c(3) = 10$, $c(4) = 11$.
3. $c(1) = 0$, $c(2) = 10$, $c(3) = 110$, $c(4) = 111$.

Codes 1 and 3 are UD (both have the prefix property), but not that of Example 2 (for instance, $c(1) * c(2) = c(3)$).

There exist codes that are UD but do not have the prefix property (See Exercise 11.4.2).

Kraft's Inequality

Theorem 11.1.16 *(Kraft, 1949) Consider a code $c : \mathcal{X} \to \mathcal{A}^*$. Let $D := |\mathcal{A}|$.*

1. *If the code is UD, $\sum_{x \in \mathcal{X}} D^{-l(c(x))} \leq 1$ (Kraft's inequality).*

2. *If* $(l(x),\, x \in \mathcal{X})$ *is a collection of integers such that* $\sum_{x\in\mathcal{X}} D^{-l(x)} \leq 1$*, there exists a* UD *code* c *such that* $l(c(x)) = l(x)$ *for all* $x \in \mathcal{X}$*.*

Proof. (McMillan, 1956) If c is UD, define the product code of order n, $c^{(n)} : \mathcal{X}^n \to \mathcal{A}^*$, by:

$$c^{(n)}(x_1^n) = c(x_1) * ... * c(x_n)\,,$$

where $x_1^n := (x_1, ..., x_n)$. Clearly, this code is also UD. One has:

$$\left(\sum_{x\in\mathcal{X}} D^{-l(x)}\right)^n = \sum_{x_1\in\mathcal{X}} ... \sum_{x_n\in\mathcal{X}} D^{(l(x_1)+...+l(x_n))} = \sum_{x_1^n\in\mathcal{X}^n} D^{-l(x_1^n)}\,,$$

where $l(x_1^n) = l(x_1)+...+l(x_n)$ is the length of the code-word $c^{(n)}(x_1^n)$. Decompose the last sum according to each possible length k ($k \geq 1$ because in a UD code, there is no code-word of length 0). Denoting by $\alpha(k)$ the number of code-words of $c^{(n)}$ of length k and by l_{max} the maximal length of a code-word of c,

$$\left(\sum_{x\in\mathcal{X}} D^{-l(x)}\right)^n = \sum_{k\geq 1} \sum_{\substack{x_1^n\in\mathcal{X}^n \\ l(x_1^n)=k}} D^{-k} = \sum_{k\geq 1} \alpha(k) D^{-k} = \sum_{k=1}^{nl_{max}} \alpha(k) D^{-k}\,.$$

As $c^{(n)}$ is UD, there are at most D^k code-words of length k. Therefore

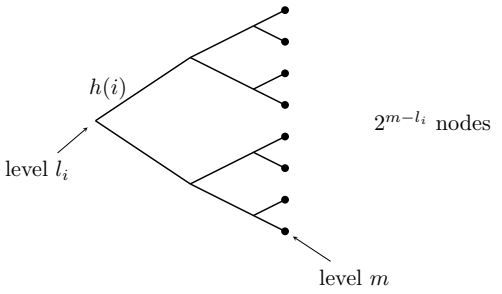
$$\left(\sum_{x\in\mathcal{X}} D^{-l(x)}\right)^n \leq \sum_{k=1}^{nl_{max}} D^k D^{-k} = nl_{max}$$

which gives

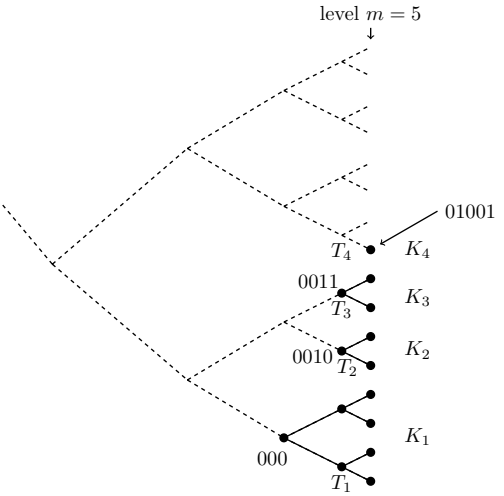
$$\sum_{x\in\mathcal{X}} D^{-l(x)} \leq (nl_{max})^{\frac{1}{n}}\,.$$

Assertion 1 follows because the right-hand side of this inequality tends to 1 as $n \to \infty$.

In order to prove assertion 2, one uses the complete D-ary tree of depth $m \geq \max_{x\in\mathcal{X}} l(x)$. (A complete D-ary tree is a connected graph without loops such that every node (vertex) has exactly $D+1$ edges stemming from it, except the root, which has exactly D outgoing edges, and selected vertices, the leaves, which have exactly one adjacent edge.) Construct a partition of the set consisting of the D^m terminal nodes (leaves) as follows. Rename the elements of $\mathcal{X}$ as $1, .., K$ in such a way that $l(1) \leq l(2) \leq ... \leq l(K)$. Assign the $D^{m-l(1)}$ first (starting from the bottom) to group 1, and then the next $D^{m-l(2)}$ to group 2, etc.



Condition $\sum_{i=1}^{K} D^{m-l(i)} \leq D^m$ guarantees that this procedure does not exhaust the terminal nodes. Then, define $c(i)$ to be the label of the root of the i-th group. This code is certainly a prefix code as the figure below shows. □



Achievable Compression Ratio

The central problem of source coding is that of finding for a given finite set $\mathcal{X}$ a code $c : \mathcal{X} \to \mathcal{A}^*$ minimizing the average code-length

$$L(c) = \sum_{x \in \mathcal{X}} p(x) l(c(x)) \,.$$

Enumerating the elements of $\mathcal{X}$ as $1, ..., K$, and denoting by $l_1, ..., l_K$ the respective lengths of the code-words $c(1), \ldots, c(K)$, one has to solve the following minimization problem:

$$\min \sum_{i=1}^{K} p_i l_i$$

under the constraint:

$$\sum_{i=1}^{K} D^{-l_i} \leq 1 \,.$$

Note that the l_i must be integers. One starts by relaxing this condition, looking for l_i's that are real non-negative. In this case the constraint is:

$$\sum_{i=1}^{K} D^{-l_i} = 1$$

because if $\sum_{i=1}^{K} D^{-l_i} < 1$, one can diminish the l_i's, and therefore the sum $\sum_{i=1}^{K} p_i l_i$, while keeping the constraint.

Lemma 11.1.17 *The solution of the problem so modified is*

$$l_i^* = -\log_D p_i \,.$$

Proof. Apply the Gibbs inequality with $q_i = D^{-l_i}$ (a probability since the modified constraint is $\sum_{i=1}^K D^{-l_i} = 1$):

$$-\sum_{i=1}^K p_i \log_D p_i \leq -\sum_{i=1}^K p_i \log_D D^{-l_i} = \sum_{i=1}^K p_i l_i \,.$$

□

Let us now return to the original optimization problem with lengths that are integers. Define

$$l_i = \lceil (-\log p_i) \rceil .$$

In particular $-\log_D p_i \leq l_i < -\log_D p_i + 1$, hence the following two remarks. Firstly, the integers just defined satisfy Kraft's constraint. Theorem 11.1.16 then guarantees the existence of a UD (in fact, prefix) code c with code-word lengths $l_1, \cdots, l_K$. Secondly,

$$H_D \leq \sum_{i=1}^K p_i l_i < H_D + 1 \,.$$

Let us now encode an IID sequence $X_1, ..., X_n$. The entropy of $(X_1, ..., X_n)$ is nH_D where H_D is the entropy of any of the elements of the sequence to be encoded, say, X_1. From the previous result, we have the existence of a code $c^{(n)} : \mathcal{X}^n \to \mathcal{A}^*$ whose average length $L(c^{(n)})$ satisfies:

$$nH_D \leq L(c^{(n)}) \leq nH_D + 1 \,.$$

Therefore the average number of letters from the alphabet $\mathcal{A}$ that are needed per symbol ($\frac{L(c^{(n)})}{n}$) satisfies:

$$H_D \leq \frac{L(c^{(n)})}{n} < H_D + \frac{1}{n}.$$

As $n \to \infty$, the quantity $\frac{L(c^{(n)})}{n}$ tends to H_D. (This is called the concatenation argument.) One cannot do better. Indeed, for any UD code with code-word lengths l_i $(1 \leq i \leq K)$, the numbers D^{-l_i} $(1 \leq i \leq K)$ define a subprobability, and therefore, by the Gibbs inequality,

$$H_D = -\sum_{i=1}^K p_i \log_D p_i \leq -\sum_{i=1}^K p_i \log_D D^{-l_i} = \sum_{i=1}^K p_i l_i \,.$$

Questionaire Interpretation of Entropy

Suppose that one among K objects, labeled 1, 2, ..., K, is chosen at random, object i with probability p_i. This object is concealed to you, and you are required to identify it. For this you are allowed to ask any number of questions whose answer is yes or no. The goal is to find a questioning strategy that minimizes the average number of questions until unambiguous identification of the object.

Each question may depend on the previous questions, and one can therefore associate with a questioning strategy a binary tree as follows. Each node of the tree is encoded in the usual way: for instance, node 001 corresponds to the path "down, down, up" when starting from the root of the tree. The root is associated with the empty string and to the first question. String (or node) 0110, for instance corresponds to the fifth question given that the answers to the first four are, in this order, no, yes, yes, no. Note that this way of coding the questions is universal and does not say anything about the nature of the questions, besides the constraint that they should have binary answers.

Now choose K nodes in the binary tree, denoted N_1, ..., N_K, with the following interpretation. If the sequence of questions (represented by a path in the tree, starting from the root) reaches node N_i, then the anwer is "i is the object chosen". One may view the binary word corresponding to node N_i as the code-word of object i. Since the questioning strategy must be admissible in the sense that it eventually leads to the correct and unambiguous decision, the binary code so defined has the prefix property. Indeed if for $i \neq j$ the code-word N_j were a prefix of N_i, the questioning strategy would produce the answer j when the object to be identified is in fact i.

The average number of questions in a strategy (identified with a set of K nodes) is the average code-word length of the prefix code so constructed, that is $H_2 = -\sum_{i=1}^{K} p_i \log_2 p_i$. This minimum is asymtotically realizable provided we accept grouping (n objects are chosen independently with the above probability in an urn with replacement, and we are to identify them simultaneously). More generally, for a discrete random variable X, the entropy $H_D(X)$ is "the minimum average number of questions with D-ary answers" needed to identify a sample of X.

11.2 Three Statistics Dependent Codes

11.2.1 The Huffman Code

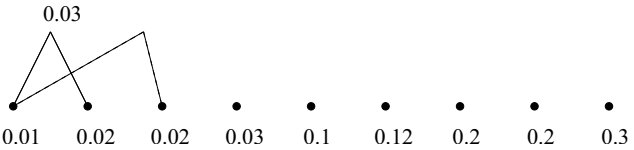
EXAMPLE 11.2.1: HUFFMAN'S ALGORITHM AT WORK ON AN EXAMPLE.

(Huffman, 1952) The probability distribution is

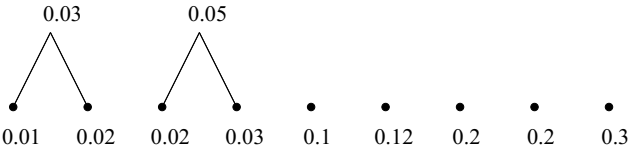
$$p = (0.01, 0.02, 0.02, 0.03, 0.1, 0.12, 0.2, 0.2, 0.3)$$

and the alphabet is $\{0, 1\}$.

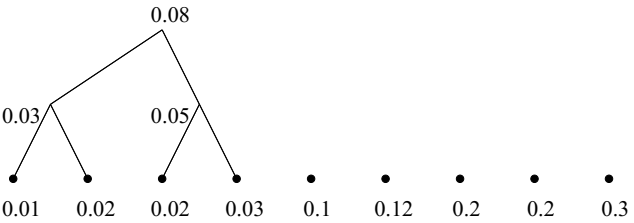
First iteration. Start by associating to each probability a point. Then merge a pair of points correspondaing to the smallest probabilities. It may occur that there are several such pairs (in the example, 2 choices). Choose one (in the example, we choose the leftmost pair).



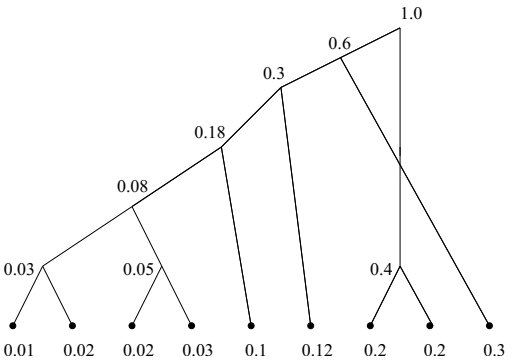
Second iteration. The two points are out of the game, they are replaced by a single point with the sum of their probabilities. Iterate. For instance, with the choice made in the first iteration:



Third iteration. For instance, with the choice made in the second iteration:



And so forth until the last iteration:
Last iteration.



Final result. Associate to each probability the corresponding string of 0's and 1's with final result

$$111111, 111110, 111101, 111100, 1110, 110, 01, 00, 10\,.$$

(Here the left branch at a node corresponds to 1, the right to 0.) The average length of the code constructed in this way (Huffman's code) is:

$$\begin{aligned} L_0 &= 6 \times (0.01 + 0.02 + 0.02 + 0.03) + 4 \times 0.1 + 3 \times 0.12 + 2 \times (0.3 + 0.2 + 0.2) \\ &= 0.48 + 0.4 = 0.36 + 1.4 = 2.64 \end{aligned}$$

The proof of the optimality of Huffman's coding algorithm rests on the following lemma.

Lemma 11.2.2 *If $n \geq 3$ and if $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ is a probability distribution such that*

$$\pi_1 \geq \pi_2 \geq \ldots \geq \pi_n > 0\,.$$

There exists an optimal code c for π such that

$$c(n) = w * 0 \quad , \quad c(n-1) = w * 1\ ,$$

for at least one binary string w, and such that code c' defined by

$$c'(i) = c(i) \quad (1 \leq i \leq n-2) \quad , \quad c'(n-1) = w$$

is optimal for the distribution $\pi' = (\pi_1, \pi_2, \ldots, \pi_{n-2}, \pi_{n-1} + \pi_n)$.

Proof. Let c be an optimal code for π with code-word lengths $\ell_1, \ldots, \ell_n$. The probabilities being ordered as indicated above, we may suppose that

$$\ell_1 \leq \ell_2 \ldots \leq \ell_n\ .$$

(In fact, if for a pair i, j such that $i < j$ we have $\ell_i > \ell_j$, the code obtained from c by exchanging the code-words $c(i)$ and $c(j)$ would have a smaller or equal average length, while keeping the prefix property.)

Moreover, $\ell_{n-1} = \ell_n$ because otherwise we could obtain a better prefix code by suppressing the last $\ell_n - \ell_{n-1}$ digits of the code-word $c(n)$.

The code-word $c(n-1)$ is therefore of the form $w * 0$ or $w * 1$, say, $w * 0$. One can then take $c(n) = w * 1$. In fact, there are only two reasons that could prevent us from doing so: either the code-word $w * 1$ is the code-word for some $c(i)$ with $i < n-1$ and one would then exchange the code-words $c(i)$ and $c(n)$, or $w * 1$ is not a code-word c. In this case one would then exchange $c(n)$ in $w * 1$ without affecting the average code length while keeping the prefix property (indeed no code-word $c(i)$ for $i \neq n$ can be a prefix of $w * 1$ because then it would be a prefix of $w * 0 = c(n-1)$, which is not possible since c is a prefix code).

Consider now the code $\tilde{c}$ induced by c on $\pi' = (\pi_1, \ldots, \pi_{n-2}, \pi_{n-1} + \pi_n)$:

$$\begin{cases} \tilde{c}(1) = c(i) \quad (1 \leq i \leq n-2) \\ \tilde{c}(n-1) = w\ . \end{cases}$$

Its average length $\tilde{L}$ is related to the average length of c by

$$L = \tilde{L} + \pi_{n-1} + \pi_n \ .$$

Let now L' be the average length of the optimal code c' for π'. Starting with c' one can define a code $\hat{c}$ for π by

$$\begin{cases} \hat{c}(1) = c'(i) \quad (1 \le i \le n-2) \\ \hat{c}(n-1) = c'(n-1) * 0 \\ \hat{c}(n) = c'(n-1) * 0 \ . \end{cases}$$

The average length $\hat{L}$ of $\hat{c}$ satisfies

$$\hat{L} = L' + \pi_{n-1} + \pi_n \ .$$

But L is the minimal length for codes of π and L' is the minimal length for codes of π'. Therefore $L' = \tilde{L}$ and $\tilde{c}$ is an optimal code for π'. □

The above lemma justifies the iterations in Huffman's coding. In fact, each iteration leads to the problem of finding the optimal code for a number of objects that has decreased by one unit, until it remains to find the optimal code for two objects which has only two code-words: 0 and 1.

11.2.2 The Shannon–Fano–Elias Code

(Elias, 1954) An advantage of this code is that it does not require the probabilities to be ordered as in the Huffman coding algorithm. Although it is not optimal, its asymptotic performance is the same as Huffman's code, as we shall see.

Since $\mathcal{X}$ has cardinality D, we may suppose that $\mathcal{X} = \{1, 2, \ldots, D\}$. Assume without loss of generality that for all $x \in \mathcal{X}$, $p(x) := P(X = x) > 0$ and let

$$F(x) := \sum_{y \le x} p(y)$$

be the cumulative distribution function of X. This is a right-continuous function with left-hand limits, with a positive jump $F(x) - F(x-) = p(x)$ at all $x \in \mathcal{X}$. Define the function

$$\overline{F}(x) := \sum_{y < x} p(y) + \frac{1}{2} p(x) \, .$$

By the positivity assumption for $p(x)$, x is uniquely determined by the knowledge of $\overline{F}(x)$. We may therefore code x by $\overline{F}(x)$, more precisely by the binary expression of it:

$$\overline{F}(x) = 0.z_1(x)z_2(x) \cdots z_n(x) \cdots$$

Since this representation may involve an infinite number of bits, we use a rounded-off value

$$\lfloor \overline{F}(x) \rfloor_{\ell(x)} := 0.z_1(x)z_2(x) \cdots z_{\ell(x)} \, .$$

Therefore

$$\overline{F}(x) - \lfloor \overline{F}(x) \rfloor_{\ell(x)} < 2^{-\ell(x)} .$$

The choice

$$\ell(x) = \lceil -\log p(x) \rceil + 1$$

ensures that

$$\frac{1}{2^{\ell(x)}} < \frac{p(x)}{2} = \overline{F}(x) - F(x)$$

and therefore that $\lfloor \overline{F}(x) \rfloor_{\ell(x)}$ lies in the interval $(F(x-), F(x))$. In particular, x is uniquely determined by its code $\lfloor \overline{F}(x) \rfloor_{\ell(x)}$. It remains to show that this code is uniquely decipherable, in fact a prefix code. Suppose, in view of contradiction, that the codewords for a and b, $a \neq b$, are respectively $z_1 z_2 \cdots z_\ell$ and $z_1 z_2 \cdots z_\ell z_{\ell+1} \cdots z_{\ell+m}$ (and therefore the prefix condition is violated). Since

$$0.z_1 z_2 \cdots z_\ell z_{\ell+1} \cdots z_{\ell+m} - 0.z_1 z_2 \cdots z_\ell < \frac{1}{2^\ell} < \frac{p(a)}{2}$$

and since $0.z_1 z_2 \cdots z_\ell$ is in the lower half of of the interval $(F(a-), F(a))$ of length $\frac{p(a)}{2}$, the code word for b would also lie in this interval, which is not the case. The non-prefix hypothesis is therefore contradicted.

The average length of the Shannon–Fano–Elias code is

$$L = \sum_{x \in \mathcal{X}} p(x) \left(\lceil \log \frac{1}{p(x)} \rceil + 1 \right) \leq H(X) + 2 .$$

The same concatenation argument as for the optimal code shows that it has the optimal asymptotic performance, with an average length per alphabet symbol $H(X)$.

11.2.3 The Tunstall Code

(Tunstall, 1967) Huffman's code transforms fixed-length messages into codewords of variable lengths. In contrast, Tunstall's code associates codewords of fixed length to variable-length messages.

A notion central to this type of code is that of "parsing". A parsing of a sequence $x_1^n := (x_1, x_2, \ldots, x_n)$ of symbols (letters) from an alphabet $\mathcal{A}$ is any sequence of strings (the "phrases") formed by successive symbols from x_1^n. For instance, for the sequence *aaabbc*, we have the parsing a, aa, b, bc, but also the trivial parsings consisting of just one phrase *aaabbc*, the sequence itself, or the parsing with all phrases of length 1, that is a, a, a, b, b, c.

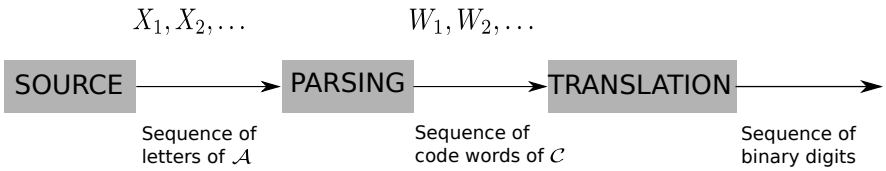
Tunstall's code consists of a parsing code and of a translation code and operates as follows. A sequence of source symbols $X_1, X_2, ...$ from an alphabet $\mathcal{A}$ of size $D \geq 2$ is parsed into a sequence $W_1, W_2, ...$ of phrases from the parsing code

$$\mathcal{C} := \{y(1), ..., y(M)\} ,$$

where for all i $(1 \leq i \leq M)$, $y(i)$ is a finite sequence of symbols from $\mathcal{A}$ of lengths $\ell_i = \ell(y(i))$:

$$y(j) = y_1(j)...y_{\ell_j}(j)\,.$$

The translation code then transforms the codeword sequence $W_1, W_2, ...$ into binary sequences of length b. This last operation must be injective, and therefore $M \leq 2^b$.



EXAMPLE 11.2.3: TUNSTALL CODING OF ABBABBBAAABABBA. In this example, the alphabet $\mathcal{A} = \{a, b\}$ ($D = 2$), the parsing code has $M = 4$ codewords

$$\begin{aligned} y(1) &= a \\ y(2) &= ba \\ y(3) &= bba \\ y(4) &= bbb \end{aligned}$$

and the translation code is

$$\begin{aligned} a &\longrightarrow 00 \\ ba &\longrightarrow 01 \\ bba &\longrightarrow 10 \\ bbb &\longrightarrow 11 \end{aligned}$$

The following sequence generated by the source

$$abbabbbaaababba$$

is parsed on-line (as soon as a codeword in the parsing code is recognized, it becomes a phrase of the parsing):

$$a|bba|bbb|a|a|a|ba|bba$$

This sequence is then translated according to the translation code:

$$00\ 10\ 11\ 00\ 00\ 00\ 01\ 10\,.$$

The code of Example 11.2.3 is a valid parsing code in the following sense:

Definition 11.2.4 *A valid parsing code $\mathcal{C}$ is one with the following properties:*

C1. Every infinite sequence of letters from $\mathcal{A}$ must have a prefix in $\mathcal{C}$.

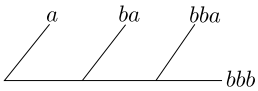
C2. It has the prefix property.

Requirement C1 is obviously necessary for encoding. If requirement C2 was not met, this would imply that some codewords of the parsing code would never be used.

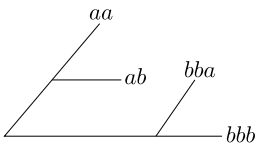
A valid parsing code is represented by a complete D-ary tree, that is, a connected graph without loops such that every node (vertex) has exactly $D+1$ edges stemming from it, except the root (which has exactly D outgoing edges) and selected vertices, the leaves (which have exactly one adjacent edge). The D edges from any given node are labeled by a letter of the alphabet $\mathcal{A}$, in a homogeneous way. For instance if the tree is represented horizontally, we label the branches from bottom to top by a_1, a_2, ..., a_D. Therefore every node of the tree is associated with a sequence of letters from the alphabet $\mathcal{A}$, this sequence being the sequence of labels read as one progresses on the tree from the root to this node.

Definition 11.2.4 implies that a valid parsing code can be represented by a complete D-ary tree for which there is a one-to-one correspondence between the leaves and the codewords, whereas no codeword of $\mathcal{C}$ corresponds to an intermediary node.

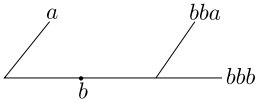
EXAMPLE 11.2.5: THREE PARSING CODES.



This parsing code is valid.



Invalid parsing code: C1 violated.



Invalid parsing code: C2 violated.

From now on, the input sequence $X_1, X_2, \ldots$ is assumed IID with distribution

$$P(X_i = x) = p(x) \qquad (x \in \mathcal{A},\, i \geq 1)\,.$$

In particular, the sequence $\{(W_i, L_i)\}_{i\geq 1}$, where $L_i := \ell(W_i)$, is IID. Let (W, L) be any random element with the common distribution of the (W_i, L_i)'s. The average length of the parsing code is $E[L]$ and

$$\frac{b}{E[L]}$$

is the average number of bits per alphabet symbol. This quantity will now be minimized by an appropriate choice of b and of the M codewords of $\mathcal{C}$.

Let $\mathcal{C}$ be a valid parsing code. Consider the associated D-ary tree. Call *intermediary* a node that is not a leaf (this includes the root).

Lemma 11.2.6

$$E[L] = \sum_{intermediary\ nodes} \Pr(nodes)$$

where $\Pr(node)$ *is the probability of the sequence of letters corresponding to the node considered (the sequence of letters read as one progresses from the root to this node).*

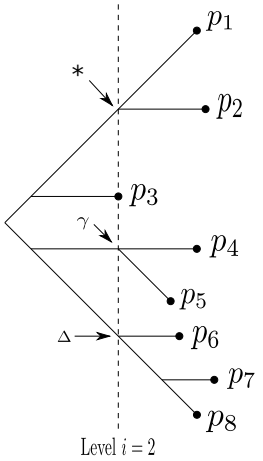
Proof. The generic parsing codeword W is of the type $W = X_1, ..., X_L$. Saying that $W = y(r)$ means that

$$X_1 = y_1(r), X_2 = y_2(r), ..., X_{\ell(r)} = y_{\ell(r)}(r), \qquad L = \ell(r)\,.$$

Therefore

$$P(W = y(r)) = \prod_{j=1}^{\ell(r)} p(y_j(r))\,.$$

In particular, the sum of the probabilities of the leaves stemming from an intermediary node is equal to the probability of this node. The sum of the probabilities of the intermediary nodes at depth i is $P(L > i)$ (see the figure below).



Here $\Pr(*) = p_1 + p_2$, $\Pr(\gamma) = p_4 + p_5$, $\Pr(\Delta) = p_6 + p_7 + p_8$,
and $P(L > 2) = \Pr(*) + \Pr(\gamma) + \Pr(\Delta) = p_1 + p_2 + p_4 + p_5 + p_6 + p_7 + p_8$.

Therefore

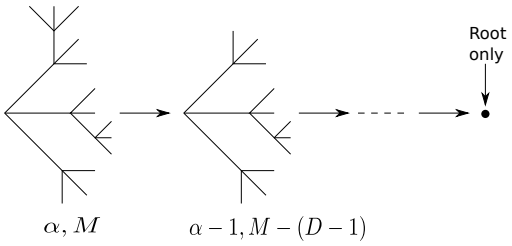
$$\sum_{\text{intermediary nodes}} \Pr(\text{nodes}) = \sum_{i=0}^{\infty} P(L > i) = E[L] .$$

□

Lemma 11.2.7 *The number α of intermediary nodes corresponding to a valid parsing code of M elements satisfies the equality*

$$M = 1 + \alpha(D-1) .$$

Proof. Consider a bunch of D leaves stemming *directly* from the same node. Such a bunch exists necessarily by condition C2 of Definition 11.2.4. By cutting the branches corresponding to these leaves and transforming the intermediary node from which they stem into a single leaf, the number of intermediary nodes is reduced by 1 and the number of leaves by $D-1$. Repeating this operation until only the root remains, we see that $1 = M - \alpha(D-1)$.



□

We now construct the optimal code. Given b, take $M \leq 2^b$. In fact, we choose α such that

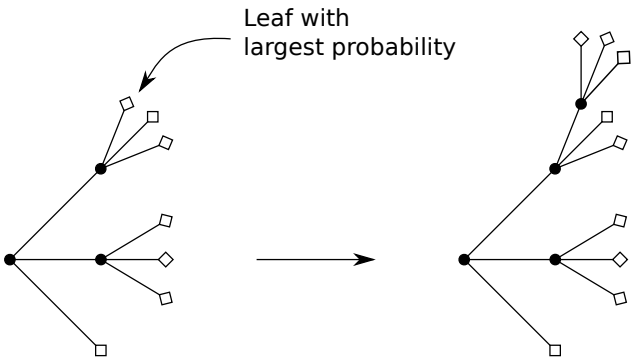
$$M \leq 2^b < M + (D-1) ,$$

that is,

$$1 + \alpha(D-1) \leq 2^b < 1 + (\alpha+1)(D-1) .$$

Having α, one must choose the intermediary nodes in order to maximize $E[L]$ or, equivalently, the sum of the probabilities of the intermediary nodes. The following algorithm does it.

Start from the root, and progress into the tree. Supose $j < \alpha$ intermediary nodes have already been selected (the black dots in the figure on the left below). Consider all the leaves stemming directly from the j already selected intermediary nodes (the white squares in the figure on the left below). Identify the leaf with highest probability. Replace this leaf by an intermediary node. Repeat the operation until α intermediary nodes have been selected.



Here $j = 3$.

Lemma 11.2.8

$$H(W) = H(X)E[L] .$$

Proof. Let $W := (X_1, ..., X_L)$ be the first phrase in the parsing of $X_1, X_2, \ldots$. We have

$$\begin{aligned}
E\left[\sum_{i=1}^{L} \log p(X_i)\right] &= E\left[\sum_{r=1}^{M}\sum_{i=1}^{L} \log p(X_i) 1_{\{W=y(r)\}}\right] \\
&= \sum_{r=1}^{M} E\left[\sum_{i=1}^{L} \log p(X_i) 1_{\{W=y(r)\}}\right] \\
&= \sum_{r=1}^{M} E\left[\sum_{i=1}^{\ell(r)} \log p(y_i(r)) 1_{\{W=y(r)\}}\right] \\
&= \sum_{r=1}^{M} \log\left(\prod_{i=1}^{\ell(r)} p(y_i(r))\right) P(W = y(r)) \\
&= \sum_{r=1}^{M} \left(\log P(W = y(r))\right) P(W = y(r)) = -H(W) .
\end{aligned}$$

Compute the same quantity in a different way:

$$\begin{aligned}
E\left[\sum_{i=1}^{L} \log p(X_i)\right] &= E\left[\sum_{i=1}^{\infty} 1_{\{i\le L\}} \log p(X_i)\right] \\
&= \sum_{i=1}^{\infty} E\left[1_{\{i\le L\}} \log p(X_i)\right] .
\end{aligned}$$

Observe that $\overline{\{L \ge i\}} = \bigcup_{k=1}^{i-1} \{L = k\}$, and that the event $\{L = k\}$ depends only on $X_1, ..., X_k$. In particular $\overline{\{L \ge i\}}$, and therefore $\{L \ge i\}$, does not depend on $X_1, ..., X_{i-1}$. By independence,

$$\sum_{i=1}^{\infty} E\left[1_{\{i\le L\}} \log p(X_i)\right] = \sum_{i=1}^{\infty} E\left[1_{\{i\le L\}}\right] E\left[\log p(X_i)\right]$$
$$= -\sum_{i=1}^{\infty} P(L \ge i) H(X) = -E\left[L\right] H(X)\,.$$

The announced equality then follows. □

Having constructed for fixed b the optimal parsing code (maximizing $E[L]$), we analyze its performance. In fact, it will be shown that, asymptotically, it optimaly compresses data, with a compression ratio equal to the entropy of the source.

Theorem 11.2.9

$$\lim_{b\to\infty} \frac{b}{E\left[L\right]} = H(X)\,.$$

Proof. Let Q be the probability of the last intermediary node selected in the above construction of the D-ary parsing tree. Then

1. each intermediary node has a probability $\ge Q$, and
2. the M leaves each have a probability $\le Q$, and
3. The M leaves each have a probability $\ge QP_{min}$,

where $P_{min} = \inf_{x\in\mathcal{A}} P(X = x)$. From conditions C1 and C2,

$$QP_{min} \le \Pr(\text{leaf}) \le Q\,. \qquad (\star)$$

Summing the first inequality on all the leaves gives $MQP_{min} \le 1$, that is

$$Q \le \frac{1}{MP_{min}}\,.$$

Therefore, by the second inequality of $(\star)$,

$$-\log \Pr(\text{leaf}) \ge \log(MP_{min})\,.$$

Summing this inequality on all the leaves:

$$H(W) \ge \log(MP_{min})\,.$$

Remembering that $2^b < M + D - 1$:

$$b < \log(M + D - 1) = \log MP_{max} + \log\left(\frac{1}{P_{min}}\right) + \log\left(1 + \frac{D-1}{M}\right)$$
$$= \log MP_{min} + c \le H(W) + c = H(X)E\left[L\right] + c\,,$$

that is

$$b < H(X)E\left[L\right] + c\,, \qquad (\dagger)$$

where

$$c := \log\left(\frac{1}{P_{min}}\right) + \log\left(1 + \frac{D-1}{M}\right).$$

Also, since $M \leq 2^b$,

$$b \geq \log M \geq H(W) = E[L]\, H(X) \qquad (\dagger\dagger)$$

(W takes M values, and therefore $H(W) \leq \log M$). From (†) and (††),

$$H(X) \leq \frac{b}{E[L]} \leq H(X) + \frac{c}{E[L]}.$$

Let $b \to \infty$. Since $2^b < M + D - 1$, M also tends to infinity and therefore:

$$c \to -\log P_{min}.$$

But, as $\log MP_{min} + c \leq H(X)E[L] + c$, we also have $E[L] \to \infty$, which shows that

$$\lim_{b\to\infty} \frac{c}{E[L]} = 0.$$

□

11.3 Discrete Distributions and Fair Coins

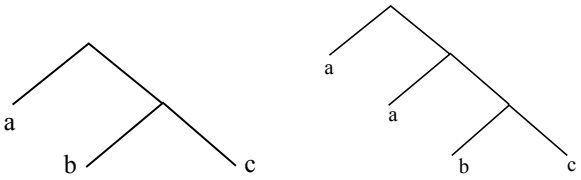
11.3.1 Representation of Discrete Distributions by Trees

The problem considered in this section is that of generating a discrete probability distribution using a fair coin. More precisely, given an IID sequence $\{U_n\}_{n\geq 1}$ of equiprobable $\{0,1\}$-valued random variable (the fair bits), can we generate a discrete random variable X with a prescribed distribution?

EXAMPLE 11.3.1: Let $\mathcal{X} = \{a, b, c\}$ and let the distribution of X be $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. The following generation algorithm is proposed. If $U_1 = 1$, set $X = a$. If $U_1 = 0, U_2 = 1$, set $X = b$. If $U_1 = 0, U_2 = 0$, set $X = c$. One readily checks that this gives the required probability distribution for X. This algorithm is best represented by the binary tree of the left in the figure below, which is explored from the root downwards according to the outcomes of the fair bits sequence. At the k-th stage of exploration, the explorator finds itself in some node at the k-th level and upon unveiling the value of the $(k+1)$-th fair bit U_{k+1} proceeds to the $(k+1)$-th level according to whether $U_{k+1} = 1$ or 0 (left if 1, right if 0). The values of X are the labels of the leaves of the binary tree. In this specific example, there is one leaf per value.

EXAMPLE 11.3.2: Let $\mathcal{X} = \{a, b, c\}$ and let the distribution of X be $(\frac{6}{8}, \frac{1}{8}, \frac{1}{8})$. Consider the tree of the figure below on the right, noting that it has two leaves for a. It is explored as in the previous example. The outcome of the algorithm is

a if $U_1 = 1$ or $U_1 = 0$ and $U_2 = 1$, and this occurs with probability $\frac{1}{2} + \frac{1}{4} = \frac{6}{8}$. Similarly, if $U_1 = 0$, $U_2 = 0$ and $U_3 = 1$, the outcome is b with probability $\frac{1}{8}$. If $U_1 = 0$, $U_2 = 0$ and $U_3 = 0$, the outcome is c with probability $\frac{1}{8}$.



Generation trees for the distributions $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{6}{8}, \frac{1}{8}, \frac{1}{8})$

11.3.2 The Knuth–Yao Tree Algorithm

(Knuth and Yao, 1976) A finite binary tree with a number of leaves equal to the number of values of X and with exactly one value of X for each leaf (as in Example 11.3.1) is called a simple generation tree. If in a possibly infinite binary tree the number of leaves is strictly larger than the number of values (and of course with exactly one value per leaf) as in Example 11.3.2, it is called a composite generation tree.

Any simple generation tree corresponds to a random variable Y, a value y being identified with a leaf at level $k = k(y)$ and having probability $2^{-k(y)}$. The average length of exploration of this tree is

$$E[T] = \sum_y k(y) 2^{-k(y)} = -P(Y = y) \log P(Y = y) = H(Y)\,.$$

The random variable Y will be called the intrinsic variable of the tree. In the case where the generation tree of X is simple, $X = Y$, and therefore the average number of fair bits needed to generate the distribution of X is $E[N] = H(Y) = H(X)$. In the case where the tree is composite, X is a function of the intrinsic variable Y. Therefore $H(X) \leq H(Y)$ and the number of fair bits needed to generate the distribution of X is

$$E[N] = H(Y) \geq H(X)\,. \qquad (\star)$$

Let henceforth $\mathcal{X} := \{1, 2, \ldots, m\}$ and $p_i := P(X = i)$ $(1 \leq i \leq m)$ be the probability distribution of X. A probability distribution corresponding to a composite binary generation tree is called a dyadic distribution. What if the distribution of X is not dyadic? The Knuth–Yao algorithm proceeds as follows. First write each p_i $(1 \leq i \leq m)$ in binary form

$$p_i = \sum_{j=1}^{\infty} p_i^{(j)}$$

where $p_i^{(j)} = \frac{1}{2^j}$ or 0. An atom of the expansion is a pair (i, j) $(1 \leq i \leq m\,,\, 1 \leq j < \infty)$ such that $p_i^{(j)} = \frac{1}{2^j}$. To such an atom associate a node at level j. An assignment such that the resulting tree has no atom at a node and exactly one atom on each

leaf is possible because the Kraft inequality is satisfied, that is, using an obvious notation,

$$\sum_{(i,j)} 2^{-\ell(i,j)} = \sum_{(i,j)} 2^{-j} = \sum_i p_i = 1\,.$$

Note however that the tree may be infinite (see Exercise 11.4.9). A leaf associated with an atom (i,j) will be assigned the value i of the variable X. Letting Y be the intrinsic variable of the tree, we have as usual that the number of fair bits needed to generate the distribution of X is $E[N] = H(Y)$.

Theorem 11.3.3

$$H(X) \le E[N] < H(X) + 2\,.$$

Proof. (Cover and Thomas, 1991) The lower bound was given above in ($\star$). We now proceed to the upper bound. Since $E[N] = H(Y)$, it suffices to show that

$$H(Y) < H(X) + 2\,. \qquad (\star\star)$$

Now

$$H(Y) = -\sum_{(i,j)} p_i^{(j)} \log p_i^{(j)} = \sum_{i=1}^{m} \sum_{j\,;\,p_i^{(j)}>0} j2^{-j} := \sum_{i=1}^{m} T_i\,.$$

To prove ($\star\star$), it is enough to show that

$$T_i < -p_i \log p_i + 2p_i \qquad (\dagger)$$

since then

$$H(Y) = \sum_{i=1}^{m} T_i < -\sum_{i=1}^{m} p_i \log p_i + 2\sum_{i=1}^{m} p_i = H(X) + 2\,.$$

Let i be fixed. There exists an integer $n = n(i)$ such that $2^{-(n+1)} > p_i \ge 2^{-n}$, that is,

$$n - 1 < -\log p_i \le n\,. \qquad (\dagger\dagger)$$

In particular, $p_i^{(j)} > 0$ only if $j \ge n$, so that we can write

$$T_i = \sum_{j\ge n\,,\,p_i^{(j)}>0} j2^{-j} \text{ and } p_i = \sum_{j\ge n\,,\,p_i^{(j)}>0} 2^{-j}\,. \qquad (\dagger\,\dagger\,\dagger)$$

In order to prove ($\dagger$), write, using ($\dagger\dagger$) and ($\dagger\,\dagger\,\dagger$),

$$\begin{aligned} T_i + p_i \log p_i - 2p_i &\le T_i - p_i(n-1) - 2p_i = T_i - (n+1)p_i \\ &= \sum_{j \ge n\,,\, p_i^{(j)} > 0} j 2^{-j} - (n+1) \sum_{j \ge n\,,\, p_i^{(j)} > 0} 2^{-j} \\ &= \sum_{j \ge n\,,\, p_i^{(j)} > 0} (j-n+1) 2^{-j} \\ &= -2^{-n} + \sum_{j \ge n+2\,,\, p_i^{(j)} > 0} (j-n+1) 2^{-j} \\ &= -2^{-n} + \sum_{k \ge 1\,,\, p_i^{(k+n+1)} > 0} k 2^{-(k+n+1)} \\ &= -2^{-n} + \sum_{k \ge 1} k 2^{-(k+n+1)} = -2^{-n} + 2^{-(n+1)} 2 = 0\,. \end{aligned}$$

□

11.3.3 Extraction Functions

We are now concerned with the inverse problem, that of generating sequences of IID variables taking the values 0 and 1 with the same probability $\frac{1}{2}$ starting from a given random variable X with known distribution. In other terms, we aim at "extracting" "random fair bits" from X. This will now be made precise.

Let $\{0,1\}^*$ denote the collection of all finite sequences of binary digits, including the void sequence. Denote by $\ell(x)$ the length of a sequence $x \in \{0,1\}^*$.

Definition 11.3.4 *Let X be a random variable with values in E. The function $\varphi : E \mapsto \{0,1\}^*$ is called an extraction function for X if whenever $P(\ell(\varphi(X)) = k) > 0$,*

$$P(\varphi(X) = x \,|\, \ell(\varphi(X)) = k) = \left(\frac{1}{2}\right)^k .$$

The variable $Y = \varphi(X)$, taking its values in $\{0,1\}^*$, is called an extraction of X. By definition of an extraction function, conditionally on $\ell(Y) = k$, the variables Y_1, Y_2, …, Y_k are IID, uniformly distributed on $\{0,1\}$. The extraction function therefore produces "independent fair bits".

EXAMPLE 11.3.5: Write (in a unique manner) the integer m as

$$m = 2^{\alpha_1} + 2^{\alpha_2} + \cdots + 2^{\alpha_k}\,,$$

where the α's are integers such that $\alpha_1 > \alpha_2 > \cdots > \alpha_k$. Therefore the list $\{0, 1, \ldots, m-1\}$ can be partitioned in k lists S_1, S_2, …, S_k. The elements of S_i contain 2^{α_i} elements and therefore can be put in bijection with the set of binary strings of length α_i, that is $\{0,1\}^{\alpha_i}$. Denote b_{α_i} this bijection. The extraction function is as follows. If $X \in S_i$, define $\varphi(X) = b_{\alpha_i}(X)$.

This is an extraction function since, conditionally on $\ell(\varphi(X)) = \alpha_i$, $\varphi(X)$ is uniformly distributed on $\{0,1\}^{\alpha_i}$.

Now, if $X \in S_i$, which occurs with probability $\frac{|S_i|}{m} = \frac{2^{\alpha_i}}{m}$, $\ell(\varphi(X)) = \alpha_i$. Therefore,

$$E\left[\ell(\varphi(X))\right] = \frac{2^{\alpha_1}}{m}\alpha_1 + \frac{2^{\alpha_2}}{m}\alpha_2 + \cdots + \frac{2^{\alpha_k}}{m}\alpha_k\,.$$

Lemma 11.3.6 *If X is uniformly distributed on $\{0,1,\ldots,m-1\}$, there exists an extraction function φ for X such that $E\left[\ell(\varphi(X))\right] \geq \lfloor \log_2 m \rfloor - 1 = \lfloor H_2(X) \rfloor - 1$.*

Proof. Proceed by induction. The result is obviously true for $m = 1$. Let now $m > 1$ be given. Suppose that the result is true for all $m' < m$. Let α_1 be as in Example 11.3.5. If $X \leq 2^{\alpha_1} - 1$ output the α_1-bit binary representation of X. Otherwise, apply to $X - 2^{\alpha_1}$, which takes its values uniformly on $\{0,1,\ldots,m-2^{\alpha_1}-1\}$, an extraction function with an average number of bits $\geq \lfloor \log_2(m-2^{\alpha_1}) \rfloor - 1$. The resulting extraction function for X therefore has an average length larger than or equal to

$$\begin{aligned}&\frac{2^{\alpha_1}}{m}\alpha_1 + \frac{2^{m-\alpha_1}}{m}\left(\lfloor \log_2(m-2^{\alpha_1}) \rfloor - 1\right)\\ &\qquad \alpha_1 + \frac{2^{m-\alpha_1}}{m}\left(\lfloor \log_2(m-2^{\alpha_1}) \rfloor - \alpha_1 - 1\right)\,.\end{aligned}$$

Now, observe that $\lfloor \log_2(m-2^{\alpha_1}) \rfloor = \alpha_2 \leq \alpha_1 - 1$. Therefore the average length of the extraction is larger than or equal to

$$\alpha_1 + \frac{2^{m-\alpha_1}}{m}\left(\alpha_2 - \alpha_1 - 1\right)\,. \qquad (\star)$$

But

$$\frac{2^{m-\alpha_1}}{m} = 1 - \frac{2^{\alpha_1}}{m} \geq 1 - \frac{2^{\alpha_1}}{2^{\alpha_1} + 2^{\alpha_2}}$$

and therefore the quantity $(\star)$ is larger than or equal to

$$\alpha_1 - \left(1 - \frac{2^{\alpha_1}}{2^{\alpha_1} + 2^{\alpha_2}}\right)(\alpha_1 - \alpha_2 + 1) \geq \alpha_1 - 1 = \lfloor \log_2 m \rfloor - 1\,.$$

□

Let $X = (X_1, X_2, \ldots, X_n) \in \{0,1\}^n$ be a vector of IID $\{0,1\}$-valued variables with $P(X_i = 1) = p \in (\frac{1}{2}, 1)$ for all i, $0 \leq i \leq n$.

Theorem 11.3.7 *(Mitzenmacher and Upfal, 2005) Let X be as above.*

(a) For any $\delta \in (0,1)$, for sufficiently large n, there exists an extraction function $\psi : \{0,1\}^n \to \{0,1\}^*$ such that $E\left[\ell(\psi(X)\right] \geq (1-\delta)nH_2(p)$.

(b) There exists an extraction function $\psi : \{0,1\}^n \to \{0,1\}^*$ such that $E\left[\ell(\psi(X)\right] \leq nH_2(p)$.

Proof. (a) Let $h(X) := \sum_{k=1}^{n} X_k$ be the Hamming weight of X. There are $\binom{n}{j}$ vectors $x \in \{0,1\}^n$ with Hamming weight j and they are equiprobable realizations of X. The realizations of Hamming weight j are encoded using an extraction function φ_j with the property indicated in Lemma 11.3.6, that is

$$E\left[\ell(\phi_j(X)) \,|\, h(X) = j\right] \geq \lfloor \log_2 \binom{n}{j} \rfloor - 1 \,.$$

The extraction function

$$\psi(X) := \sum_{j=0}^{n} \phi_j(X) 1_{\{h(X)=j\}}$$

has the average length

$$\begin{aligned} E\left[\ell(\psi(X))\right] &= \sum_{j=0}^{n} P(h(X)=j) E\left[\ell(\phi_j(X) \,|\, h(X)=j\right] \\ &\geq \sum_{j=0}^{n} P(h(X)=j)(\lfloor \log_2 \binom{n}{j} \rfloor - 1) \,. \end{aligned}$$

Let $\varepsilon < p - \frac{1}{2}$ be a constant to be chosen later. Then, for $n(p-\varepsilon) \leq j \leq n(p+\varepsilon)$, by Theorem 2.1.36,

$$\binom{n}{j} \geq \binom{n}{\lfloor n(p+\varepsilon) \rfloor} \geq \frac{2^{nH_2(p+\varepsilon)}}{n+1} \,,$$

and therefore the average length of $\psi(X)$ is larger than

$$\begin{aligned} &\sum_{j=\lfloor n(p-\varepsilon) \rfloor}^{\lceil n(p+\varepsilon) \rceil} P(h(X)=j)(\lfloor \log_2 \binom{n}{j} \rfloor - 1) \\ &\geq \left(\log_2 \frac{2^{nH_2(p+\varepsilon)}}{n+1} - 2 \right) \sum_{j=\lfloor n(p-\varepsilon) \rfloor}^{\lceil n(p+\varepsilon) \rceil} P(h(X)=j) \\ &\geq (nH_2(p+\varepsilon) - \log_2(n+1) - 1) P(|h(X) - np| \leq n\varepsilon) \,. \end{aligned}$$

By Chernoff,

$$P(|h(X) - np| \leq n\varepsilon) := P(|\sum_{k=1}^{n} X_k - np| \leq n\varepsilon) \leq 2e^{-n\varepsilon^2/3p} \,.$$

Therefore,

$$E\left[\ell(\psi(X))\right] \geq (nH_2(p+\varepsilon) - \log_2(n+1) - 1)(1 - 2e^{-n\varepsilon^2/3p}) \,.$$

Choose ε sufficiently small so that

$$nH_2(p+\varepsilon) \geq (1 - \frac{\delta}{3}) nH_2(p).$$

For $n > \frac{3p}{\varepsilon^2}\log\frac{6}{\delta}$, $1 - 2e^{-n\varepsilon^2/3p} \geq 1 - \frac{\delta}{3}$, and therefore

$$E\left[\ell(\psi(X))\right] \geq ((1-\frac{\delta}{3})nH_2(p) - \log_2(n+1) - 1)(1-\frac{\delta}{3})\,.$$

With n sufficiently large that $\frac{\delta}{3}nH_2(p) \geq \log_2(n+1)+1$, $E\left[\ell(\psi(X))\right] \geq ((1-2\frac{\delta}{3})nH_2(p))(1-\frac{\delta}{3}) \geq (1-\delta)nH_2(p)$.

(b) Since $\varphi(X)$ is a function of X, $H_2(X) \geq H_2(\varphi(X))$. But

$$\begin{aligned} H_2(\varphi(X)) &= -\sum_{x\in\mathcal{X}} P(\varphi(X)=x)\log_2 P(\varphi(X)=x) \\ &= -\sum_{x\in\mathcal{X}} P(\varphi(X)=x)\log_2 2^{-\ell(\varphi(x))} \\ &= \sum_{x\in\mathcal{X}} P(\varphi(X)=x)\ell(\varphi(x)) = E\left[\ell(\varphi(X)\right]\,. \end{aligned}$$

Therefore

$$E\left[\ell(\psi(X))\right] \leq H_2(X) = nH_2(p)\,.$$

□

Books for Further Information

A popular textbook in information theory is [Cover and Thomas, 2006]. See also [Ash, 1965], [Gallagher, 1968], [McEliece, 2002] and [MacKay, 2003]. For the generation of random variables: [Knuth, 1973].

11.4 Exercises

Exercise 11.4.1. Deterministic transformations
1. Let X be a random variable with values in the finite set $\mathcal{X}$ and let $Y = \varphi(X)$ where φ is a bijective deterministic function. Show that

$$H(Y) = H(X)\,.$$

2. Let Z be a random variable with values in the finite set $\mathcal{Z}$, and let ψ be a deterministic function (not necessarily bijective). Show that:

$$H(Z,\psi(Z)) = H(Z)\,.$$

Exercise 11.4.2. Uniquely decipherable but not prefix
Give an example of a code that is uniquely decipherable and yet not a prefix code.

Exercise 11.4.3. Extension of McMillan's theorem
The proof of Assertion 1 of Theorem 11.1.16 depends crucially on the finiteness of the code. Show that it remains true for an infinite yet denumerable code.

Exercise 11.4.4. Huffman code
Find an optimal binary code for the following probability distribution:

$$p = (0.01, 0.04, 0.05, 0.07, 0.09, 0.1, 0.14, 0.2, 0.3) \ .$$

What is its average length?

Exercise 11.4.5. Approaching entropy
Let X_i $(i \geq 1)$ be $\{0,1\}$-valued iid random variables with distribution given by $P(X_i = 1) = \frac{3}{4}$. Let H be the entropy of this variable. Find n such that the optimal code for the random element $(X_1, \ldots, X_n)$ has a length per symbol $\leq H + 10^{-2}$.

Exercise 11.4.6. Tunstall code
Find a Tunstall code for the probability distribution of Example 11.4.4 with $b = 4$, $D = 2$.

Exercise 11.4.7. The false coin
You have 15 coins, undistiguishable, except for one of them, which has a different weight. This strange coin is not known to you, but you are informed that the difference of weight is very small, say one percent lighter or heavier, but you do not know if it is lighter or heavier. All you have at disposition is a scale. Find a strategy that allows you to find the strange coin and that involves the minimal average utilization of the scale.

Exercise 11.4.8. Competitive optimality of the Shannon code
(Cover, 1991) Let $\ell(X)$ and $\ell'(X)$ be the codeword lengths of a discrete random variable X for the Shannon code and of any other uniquely decipherable code respectively. Show that for all $c \geq 1$

$$P(\ell(X) \geq \ell'(X) + c) \leq 2^{1-c} \, .$$

Exercise 11.4.9. The Knuth–Yao algorithm
Detail the Knuth–Yao algorithm for generating a random variable X taking the two values a and b with respective probabilities $\frac{2}{3}$ and $\frac{2}{3}$.

Chapter 12

Shannon's Capacity Theorem

12.1 More Information-theoretic Quantities

12.1.1 Conditional Entropy

Shannon's channel coding theorem concerns the possibility of communicating via a noisy channel with an arbitrarily small probability of error. Its proof is based on the random coding argument, perhaps the first occurence of the probabilistic method (Chapter 5). In view of defining the capacity of a channel, the central notion of Shannon's result, we need to augment our panoply of information-theoretic quantities, starting with the notion of conditional entropy.

Let X and Y be two discrete random variables with values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let p_X, p_Y and p_{XY} denote the distributions of X, Y and (X,Y) respectively. Observe that the random variables $p_X(X)$, $p_Y(Y)$ and $p_{XY}(X,Y)$ are almost surely non-null. For instance

$$P(p_X(X)=0) = E\left[1_{\{p_X(X)=0\}}\right] = \sum_{y\in\mathcal{Y}} 1_{\{p_X(x)=0\}} p_X(x) = 0\,.$$

(This observation will allow us to accept the presence of these random variables in the denominator of fractions.)

Definition 12.1.1 *The conditional entropy of X given Y is the quantity*

$$H(X|Y) := H(X,Y) - H(Y)\,.$$

Alternatively,

$$H(X|Y) = E\left[\log p_{X,Y}(X,Y)\right] - E\left[\log p_Y(Y)\right] = E\left[\log p_{X|Y}(X|Y)\right]\,,$$

where

$$p_{X|Y}(x|y) := P(X=x|Y=y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

is the conditional probability of $X=x$ given $Y=y$.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_12

Remark 12.1.2 According to the questionnaire interpretation of entropy, the entropy $H_D(X)$ is "the minimum average number of questions with D-ary answers" needed to identify a sample of X. This is a rough way of saying things that is convenient to interpret the various information-theoretic relations. An example of the kind of argument that leads to the correct result is the following. In order to identify a pair of random variables (X, Y), one can for instance start by identifying X, which requires $H(X)$ questions, and then, knowing X, to identify Y, which requires $H(Y|X)$ questions. Therefore the total number of questions necessary to identify both variables is $H(X) + H(Y|X)$ and this is $H(X, Y)$, hence the relation $H(X, Y) = H(X) + H(Y|X)$. Of course, the identities that are found in this heuristic manner must be proved by regular means, but they generally lead to the correct result.

Theorem 12.1.3 *We have the identities*

$$H(X, Y) = H(Y) + H(X|Y) = H(X) + H(Y|X) \tag{12.1}$$

and the inequality

$$H(X|Y) \leq H(X) \tag{12.2}$$

("conditioning decreases entropy").

Proof. The identities in (12.1) are direct consequences of Definition 12.1.1. From the first one, it follows that $H(X|Y) - H(X) = H(X, Y) - H(Y) - H(X)$, a negative quantity by Theorem 11.1.9. □

Let X and Y be as above, and let Z be a discrete variable with values in the finite space $\mathcal{Z}$. The following identities are recorded for future reference, but add nothing to (12.1) since a vector of discrete random variables is also a discrete variable:

$$\begin{aligned} H(X, Y|Z) &= H(Y|Z) + H(X|Y, Z) \\ &= H(X|Z) + H(Y|X, Z)\,. \end{aligned} \tag{12.3}$$

Using the heuristics of Remark 12.1.2, we expect that

$$H(X|Y, Z) \leq H(X|Y)\,. \tag{12.4}$$

(The formal proof is required in Exercise 12.3.1.)

Corollary 12.1.4 *Let $X_1, \ldots, X_n$ be random variables with values in the respective finite sets $\mathcal{X}_1, \ldots, \mathcal{X}_n$. Then*

$$H(X_1, ..., X_{n+1}) = H(X_1, ..., X_n) + H(X_{n+1}|X_1, ..., X_n).$$

In particular (sequential entropy formula, or entropy's chain rule),

$$H(X_1, ..., X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_1, ..., X_{i-1})\,.$$

Entropy of a Stationary Source

Theorem 12.1.5 *Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables taking their values in the finite set $\mathcal{X}$. Assume that it is stationary, in the sense that for all n, the distribution of the vector $(X_{1+k}, \ldots, X_{n+k})$ is independant of k. Then*

(i) $H := \lim_{n\to\infty} H(X_n|X_1, \ldots, X_{n-1})$ exists, and

(ii) $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$ exists and equals H.

Proof. (i) By inequality (12.4),

$$H(X_{n+1}|X_1, \ldots, X_n) \leq H(X_{n+1}|X_2, \ldots, X_n)\,,$$

and by stationarity,

$$H(X_{n+1}|X_2, \ldots, X_n) = H(X_n|X_1, \ldots, X_{n-1})\,.$$

Therefore

$$H(X_{n+1}|X_1, \ldots, X_n) \leq H(X_n|X_1, \ldots, X_{n-1})\,.$$

The sequence $H(X_n|X_1, \ldots, X_{n-1})$ $(n \geq 1)$ being non-increasing and bounded below by 0, converges to some $H \geq 0$.

(ii) By the chain rule,

$$\frac{1}{n} H(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1})\,,$$

and therefore, by Cesaro's theorem, $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) = H$. □

Fano's Inequality

Theorem 12.1.6 *(Fano, 1961) Let X and Y be two random variables taking their values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $\widehat{X}$ be an estimate of X based on the observation of Y, of the form $\widehat{X} = g(Y) \in \mathcal{X}$. With $P_e := P(\hat{X} \neq X)$,*

$$h(P_e) + P_e \log|\mathcal{X}| \geq H(X|Y)\,. \tag{12.5}$$

Proof. Heuristics: Suppose that you are given Y. How many questions do we need to obtain X? If we adopt a not necessarily optimal identification strategy, then the number of questions will be greater than $H(X|Y)$. Use the following strategy. First, ask if $X = Y$ or not, which requires $h(P_e)$ questions. If $X = Y$, you are done. Otherwise, with probability P_e, try to identify X, which requires at most $\log|\mathcal{X}|$ questions. Hence $h(P_e) + P_e \log|\mathcal{X}| \geq H(X|Y)$.

We now devise a regular proof, starting with the following preliminaries. Let X, Y, Z be random variables taking their values in the finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively. For a fixed $z \in \mathcal{Z}$, $H(X|Y, Z = z)$ or $H^{Z=z}(X|Y)$ is, by definition, the

conditional entropy of X given Y, computed not for the original probability, but for the conditional probability $P^{Z=z}(\cdot) = P(\cdot|Z=z)$. One can check (Exercise 12.3.2) that

$$H(X|Y,Z) = \sum_{z\in\mathcal{Z}} P(Z=z)H^{Z=z}(X|Y)\,. \tag{12.6}$$

Define the error random variable

$$E := \begin{cases} 1 & \text{if } \hat{X} \neq X \\ 0 & \text{if } \hat{X} = X\,. \end{cases}$$

By the chain rule (12.3), we have two ways of writing $H(E,X|Y)$

$$\begin{aligned} H(E,X|Y) &= H(X|Y) + H(E|X,Y) \\ &= H(E|Y) + H(X|E,Y)\,. \end{aligned}$$

Since E is a function of (X,Y), we have (see Exercise 12.3.6)

$$H(E|X,Y) = 0\,.$$

On the other hand $H(E|Y) \leq H(E)$ (conditioning decreases entropy, (12.2)) and therefore $H(E|Y) \leq h(P_e)$. Also, by (12.6),

$$\begin{aligned} H(X|E,Y) &= P_e H(X|E=1,Y) + (1-P_e)H(X|E=0,Y) \\ &= P_e H(X|E=1,Y) \end{aligned}$$

since when $E=0$, X is a function of Y ($X = \hat{X} = g(Y)$). On the other hand $H(X|E=1,Y) \leq H(X) \leq \log|\mathcal{X}|$. Combining the above observations leads to (12.5). □

If the logarithms are in base 2, $h(P_e) \leq 1$, and Fano's inequality can be weakened to $1 + P_e \log_2|\mathcal{X}| \geq H_2(X|Y)$ or

$$P_{\mathrm{e}} \geq \frac{H_2(X|Y) - 1}{\log_2|\mathcal{X}|}\,.$$

12.1.2 Mutual Information

Let X and Y be two random variables taking their values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Definition 12.1.7 *The mutual information between X and Y is the quantity*

$$I(X;Y) := E\left[\log \frac{p_{X,Y}(X,Y)}{p_X(X)p_Y(Y)}\right]\,. \tag{12.7}$$

Clearly the mutual information is symmetric in X and Y: $I(X;Y) = I(Y;X)$. Also, $I(Y;Y) = H(Y)$ (Exercise 12.3.5).

In expanded form,

$$I(X;Y) = \sum_{x,y} p_{X,Y}(x,y) \log p_{X,Y}(x,y) - \sum_{x,y} p_{X,Y}(x,y) \log q_{X,Y}(x,y)\,,$$

where $q_{X,Y}(x,y) = p_X(x)p_Y(y)$ defines a distribution on $\mathcal{X} \times \mathcal{Y}$. Therefore, by the Gibbs inequality,

$$I(X;Y) \geq 0 \tag{12.8}$$

with equality iff X and Y are independent. Also, it follows immediately from definition (12.7) that

$$I(X;Y) = H(X) + H(Y) - H(X,Y)\,.$$

Therefore, by Theorem 12.1.3,

$$I(X;Y) = H(X) - H(X|Y)\,. \tag{12.9}$$

Also, $H(X|Y) = H(X)$ if and only if $I(X;Y) = 0$, or equivalently, if and only if X and Y are independent. (Theorem 12.1.3 stated that in general, $H(X|Y) \leq H(X)$.)

Theorem 12.1.8 *Let X and Y be random variables taking their values in finite sets. Then*

$$I(X;Y) \leq I(Y;Y) = H(Y)\,,$$

with equality iff $Y = \varphi(X)$ for some deterministic function φ.

Proof. From (12.9)

$$I(X;Y) = H(Y) - H(Y|X) \leq H(Y) = I(Y;Y)\,,$$

with equality iff $H(Y|X) = 0$. But

$$\begin{aligned} H(Y|X) &= E\left[-\log p_{Y|X}(Y|X)\right] \\ &= -\sum_{x,y} p_{X,Y}(x,y)\left(\log p_{Y|X}(y|x)\right) \end{aligned}$$

is null if and only if for all x, y, $p_{Y|X}(y|x)$ is either 0 or 1. This happens if and only if $Y = \varphi(X)$ for some deterministic function φ. □

Let X, Y, Z be random variables taking their values in the finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively. We define the conditional mutual information by

$$I(X;Y|Z) := E\left[\log \frac{p_{X,Y|Z}(x,y|z)}{p_{X|Z}(x|z)p_{Y|Z}(y|z)}\right]\,.$$

Note that

$$I(X;Y|Z) = \sum_{z \in \mathcal{Z}} I(X;Y|Z=z)p_Z(z)$$

where $I(X;Y|Z=z)$ is the mutual information of X and Y considered in the probability space induced with the probability $P(\cdot|Z=z)$. In particular,

$$I(X;Y|Z) \geq 0\,, \tag{12.10}$$

with equality iff X and Y are independent given Z. Also,

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)\,. \tag{12.11}$$

Proof. Observe that

$$\begin{aligned} I(X;Y|Z) &= E\left[\log \frac{p_{X|Y,Z}(x|y,z)}{p_{X|Z}(x|z)}\right] = H(X|Z) - H(X|Y,Z) \\ &= H(X,Z) - H(Z) - H(X,Y,Z) + H(Y,Z)\,, \end{aligned}$$

and that

$$I(X;Y,Z) - I(X;Z) = H(X) + H(Y,Z) - H(X,Y,Z) - H(X) - H(Z) + H(X,Z)\,.$$

□

The Data Processing Inequality

Theorem 12.1.9 *Let X, Y, Z be random variables taking their values in the finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively. We have that*

$$I(X,Y;Z) \geq I(Y;Z)$$

with equality if and only if $X \to Y \to Z$ forms a Markov chain, that is if Z is independent of X given Y.

Proof. By (12.11) $I(X,Y;Z) - I(Y;Z) = I(Z;X|Y)$, and by (12.10), this is a non-negative quantity, null if and only if X and Z are independent given Y. □

Jointly Typical Sequences

Let X and Y be two finite valued random variables and let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of IID random variables such that $X_n \overset{\mathcal{D}}{\sim} X$ and $Y_n \overset{\mathcal{D}}{\sim} Y$. We use the following simplified notations

$$p(x_1^n) = P(X_1^n = x_1^n), \qquad p(y_1^n) = P(Y_1^n = y_1^n)$$

and

$$p(x_1^n, y_1^n) = P(X_1^n = x_1^n, Y_1^n = y_1^n)$$

(the argument x_1^n, y_1^n or (x_1^n, y_1^n) determines which function p is considered).

The following result extends Theorem 11.1.11 and gives intuitive support to the notion of mutual information. It will also be useful in the proof of Shannon's capacity theorem.

In the definition and theorem below, the logarithms used in the definition of the information-theoretic quantities are in base 2, for instance $H(X) = H_2(X)$, etc.

Definition 12.1.10 *For any $\epsilon > 0$, let $A_\epsilon^{(n)} \subseteq \mathcal{X}^n \times \mathcal{Y}^n$ be the set of ϵ-jointly typical sequences, that is, the collection of sequences (x_1^n, y_1^n) such that*

(i) $\left|-\frac{1}{n}\log_2 p_{X_1^n}(x_1^n) - H(X)\right| < \epsilon$,

(ii) $\left|-\frac{1}{n}\log_2 p_{Y_1^n}(y_1^n) - H(Y)\right| < \epsilon$, *and*

(iii) $\left|-\frac{1}{n}\log_2 p_{X_1^n, Y_1^n}(x_1^n, y_1^n) - H(X,Y)\right| < \epsilon$.

Theorem 12.1.11

(a) $P((X_1^n, Y_1^n) \in A_\epsilon^{(n)}) \underset{n\to\infty}{\to} 1$

(b) $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X,Y)+\epsilon)}$

(c) If $\tilde{X}_1^n \overset{\mathcal{D}}{\sim} X_1^n$, $\tilde{Y}_1^n \overset{\mathcal{D}}{\sim} Y_1^n$ and $\tilde{X}_1^n$ and $\tilde{Y}_1^n$ are independent, then

$$P((\tilde{X}_1^n, \tilde{Y}_1^n) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

Also, for sufficiently large n,

$$P((\tilde{X}_1^n, \tilde{Y}_1^n) \in A_\epsilon^{(n)}) \geq (1-\epsilon)\, 2^{-n(I(X;Y)+3\epsilon)}.$$

Proof. Define

$$T_1 := \left\{(x_1^n, y_1^n) \in \mathcal{X}^n \times \mathcal{Y}^n; \left|-\frac{1}{n}\log_2 p(x_1^n) - H(X)\right| < \epsilon\right\},$$

$$T_2 := \left\{(x_1^n, y_1^n) \in \mathcal{X}^n \times \mathcal{Y}^n; \left|-\frac{1}{n}\log_2 p(y_1^n) - H(Y)\right| < \epsilon\right\},$$

$$T_3 := \left\{(x_1^n, y_1^n) \in \mathcal{X}^n \times \mathcal{Y}^n; \left|-\frac{1}{n}\log_2 p(x_1^n, y_1^n) - H(X,Y)\right| < \epsilon\right\}.$$

In particular, $A_\epsilon^{(n)} = T_1 \cap T_2 \cap T_3$.

(a) By the weak law of large numbers, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|-\frac{1}{n}\log_2 p\,(X_1^n) - H(X)\right| > \epsilon\right) = 0.$$

Therefore, with each δ, one can associate an integer n_1 such that

$$n \geq n_1 \Rightarrow P\left(\left|-\frac{1}{n}\log_2 p\,(X_1^n) - H(X)\right| \geq \epsilon\right) \leq \frac{\delta}{3}.$$

Similarly, there exist integers n_2 and n_3 such that

$$n \geq n_2 \Rightarrow P\left(\left|-\frac{1}{n}\log_2 p\,(Y_1^n) - H(Y)\right| \geq \epsilon\right) \leq \frac{\delta}{3},$$

$$n \geq n_3 \Rightarrow P\left(\left|-\frac{1}{n}\log_2 p\,(X_1^n, Y_1^n) - H(X,Y)\right| \geq \epsilon\right) \leq \frac{\delta}{3}.$$

Therefore, for all $n \geq \max(n_1, n_2, n_3)$,

$$P(T_1^c \cup T_2^c \cup T_3^c) \leq P(T_1^c) + P(T_2^c) + P(T_3^c) \leq \delta\,,$$

and in particular,

$$P(A_\epsilon^{(n)}) = P\left(T_1 \cap T_2 \cap T_3\right) \geq 1 - \delta\,.$$

(b)

$$\begin{aligned}(x_1^n, y_1^n) \in A_\epsilon^{(n)} &\Rightarrow \left| -\frac{1}{n}\log_2 p\left(X_1^n, Y_1^n\right) - H(X,Y) \right| < \epsilon \\ &\Rightarrow -\frac{1}{n}\log_2 p\left(X_1^n, Y_1^n\right) < H(X,Y) + \epsilon \\ &\Leftrightarrow p(x_1^n, y_1^n) > 2^{-n(H(X,Y)+\epsilon)}\,.\end{aligned}$$

Therefore

$$1 \geq P\left(A_\epsilon^{(n)}\right) \geq \left|A_\epsilon^{(n)}\right| 2^{-n(H(X,Y)+\epsilon)}\,.$$

(c) For sufficiently large n, $P\left(A_\epsilon^{(n)}\right) \geq 1 - \epsilon$. Since

$$(x_1^n, y_1^n) \in A_\epsilon^{(n)} \Rightarrow p_{X_1^n, Y_1^n}(x_1^n, y_1^n) < 2^{-n(H(X,Y)-\epsilon)}\,,$$

we obtain

$$1 - \epsilon \leq P\left(A_\epsilon^{(n)}\right) \leq \left|A_\epsilon^{(n)}\right| 2^{-n(H(X,Y)-\epsilon)}\,.$$

On the other hand, note that

$$(x_1^n, y_1^n) \in A_\epsilon^{(n)} \Rightarrow \begin{cases} 2^{-n(H(X)+\epsilon)} < p(x_1^n) < 2^{-n(H(X)-\epsilon)} \\ 2^{-n(H(X)+\epsilon)} < p(y_1^n) < 2^{-n(H(Y)-\epsilon)} \end{cases}.$$

Then

$$\begin{aligned}P\left(\left(\tilde{X}_1^n, \tilde{Y}_1^n\right) \in A_\epsilon^{(n)}\right) &\leq \left|A_\epsilon^{(n)}\right| 2^{-n(H(X)-\epsilon)} 2^{-n(H(Y)-\epsilon)} \\ &\leq 2^{n(H(X,Y)+\epsilon)} 2^{-n(H(X)-\epsilon)} 2^{-n(H(Y)-\epsilon)} \\ &\leq 2^{(H(X)+H(Y)-H(X,Y)-3\epsilon)} = 2^{-n(I(X;Y)-3\epsilon)}\,,\end{aligned}$$

and for sufficiently large n,

$$\begin{aligned}P\left(\left(\tilde{X}_1^n, \tilde{Y}_1^n\right) \in A_\epsilon^{(n)}\right) &\geq \left|A_\epsilon^{(n)}\right| 2^{-n(H(X)+\epsilon)} 2^{-n(H(Y)+\epsilon)} \\ &\geq (1-\epsilon)\, 2^{n(H(X,Y)-\epsilon)} 2^{-n(H(X)+\epsilon)} 2^{-n(H(Y)+\epsilon)} \\ &= (1-\epsilon)\, 2^{-n(I(X;Y)+3\epsilon)}\,.\end{aligned}$$

□

12.1.3 Capacity of Noisy Channels

We introduce the notion of a communications channel, starting with one of the simplest and indeed most popular model.

EXAMPLE 12.1.12: THE BINARY SYMMETRIC CHANNEL, TAKE 1. In this type of channel, called for short the BSC channel, the inputs as well as the inputs are sequences of binary digits, and the effect of the channel is to randomly change a 0 into a 1 and vice-versa. Therefore if a sequence $x_1^n = (x_1, \ldots, x_n)$ (the input), is transmitted, the received sequence (the output) is $y_1^n = (y_1, \ldots, y_n)$, where

$$y_n = x_n \oplus B_n \,,$$

where $\oplus$ denotes addition modulo 2, and where $B_1^n = (B_1, \ldots, B_n)$ is the random noise sequence. Each noise bit B_n takes its value in $\{0, 1\}$. Therefore the value $B_n = 1$ corresponds to an error on the n-th bit transmitted. If $B_1, \ldots, B_n$ are independent random variables, identically distributed, with probability $p \in (0, 1)$ of taking the value 1, the channel considered is called a binary symmetric channel (BSC).

A channel is fed with a sequence $X_1, X_2, \ldots$, where the X_k's are random variables taking their values in a finite set $\mathcal{X}$, called the input alphabet. At the receiving end of the channel, one recovers a sequence $Y_1, Y_2, \ldots$ where the Y_k's are random variables taking their values in a finite set $\mathcal{Y}$, called the output alphabet.

Definition 12.1.13 *The channel is called memoryless if, for all $n \geq 2$,*

$$Y_n \text{ and } \left(X_1^{n-1}, Y_1^{n-1}\right) \text{ are independent given } X_n \,.$$

In other terms, for all $x \in \mathcal{X}, y \in \mathcal{Y}, x_1^{n-1} \in \mathcal{X}^{n-1}, y_1^{n-1} \in \mathcal{Y}^{n-1}$,

$$P(Y_n = y | X_n = x, X_1^{n-1} = x_1^{n-1}, Y_1^{n-1} = y_1^{n-1}) = P(Y_n = y | X_n = x) \,.$$

Definition 12.1.14 *The channel is said to be without feedback if for all $n \geq 2$, X_n and Y_1^{n-1} are independent givenX_1^{n-1}.*

Theorem 12.1.15 *Suppose that the channel is memoryless and without feedback. Then*

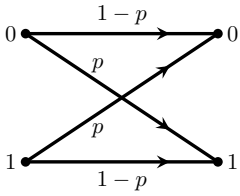
$$P(Y_1^n = y_1^n, | X_1^n = x_1^n) = \prod_{\ell=1}^{n} P(Y_\ell = y_\ell | X_\ell = x_\ell) \,.$$

Proof. Exercise 12.3.4. □

Definition 12.1.16 *The channel is called time-invariant if for all $x \in \mathcal{X}, y \in \mathcal{Y}$, the quantity $P(Y_n = y|X_n = x)$ does not depend on $n \geq 1$. Denoting it by $p(y|x)$, the matrix $\{p(y|x)\}_{x\in\mathcal{X},y\in\mathcal{Y}}$ is called the transition matrix of the channel.*

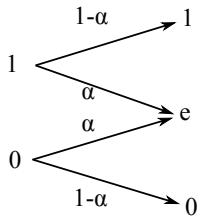
EXAMPLE 12.1.17: BINARY SYMMETRIC CHANNEL, TAKE 2. The input and output alphabets are $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. The input X and output Y are related by $Y = X \oplus B$, where $B \in \{0, 1\}$ and $P(B = 1) = p$. The channel transition matrix is

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$



EXAMPLE 12.1.18: BINARY ERASURE CHANNEL, TAKE 1. The input alphabet is $\mathcal{X} = \{0, 1\}$, the output alphabet is $\mathcal{Y} = \{0, 1, e\}$. The channel transition function is

$$\begin{pmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1-\alpha & \alpha \end{pmatrix}$$



Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, and consider a time-invariant memoryless discrete-time channel without feedback and with given transition matrix $\{p_(y|x)\}_{x\in\mathcal{X},\, y\in\mathcal{Y}}$.

Definition 12.1.19 *The capacity of the above channel is, by definition, the number*

$$C := \sup_X I(X; Y) \tag{12.12}$$

where X and Y are random variables with values in $\mathcal{X}$ and $\mathcal{Y}$ respectively, such that $P(Y = y|X = x) = p(y|x)$ and where the supremum is taken over all probability distributions p_X on $\mathcal{X}$.

More explicitly, recall that

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_X(x)p(y|x)\log_2 \frac{p(y|x)}{\sum_{z\in\mathcal{X}} p_X(z)p(y|z)}.$$

Since the supremum in the definition of capacity is over all probability distributions p_X on $\mathcal{X}$, the information capacity is a function of the channel only, through its transition matrix. Note that this supremum is achieved because the function to be optimized is concave (Exercise 12.3.3) and the set of constraints is non-empty, closed, convex and bounded.

EXAMPLE 12.1.20: BINARY SYMMETRIC CHANNEL, TAKE 3. The capacity of this channel is $C = 1 - h_2(p)$.

Proof. Note that, conditionnaly on $X = 1$, Y takes the values $\{0,1\}$ with respective probabilities $\{p, 1-p\}$. Similarily, conditionnaly on $X = 0$, Y takes the values $\{0,1\}$ with respective probabilities $\{1-p, p\}$. Therefore, for any $x \in \{0,1\}$, $H(Y|X = x) = H(p)$, and

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{x\in\{0,1\}} p_X(x)H(Y|X = x) \\ &= H(Y) - \sum_{x\in\{0,1\}} p_X(x)h_2(p) \\ &= H(Y) - h_2(p) \leq 1 - h_2(p)\,, \end{aligned}$$

where the last inequality is due to the fact that Y is equidistributed on two values. Equality holds when $P(X = 0) = \frac{1}{2}$. In fact, in this case,

$$\begin{aligned} P\,(Y = 0) &= \sum_{x\in\{0,1\}} p_X(x)P(Y = 0|X = x) \\ &= p_X(0)(1-p) + p_X(1)p = \frac{1}{2}\,. \end{aligned}$$

Therefore $H(Y) = 1$, and $C := \sup_X I(X;Y) = 1 - h_2(p)$. □

EXAMPLE 12.1.21: BINARY ERASURE CHANNEL, TAKE 2. The capacity of this channel is $C = 1 - \alpha$.

Proof. Let X be a random variable and let Y be the output of the channel corresponding to the input X. Conditionnaly on $X = 1$, Y takes the values $\{1, e\}$ with respective probabilities $\{1-\alpha, \alpha\}$. Similarily, conditionnaly on $X = 0$, it takes the values $\{0, e\}$ with respective probabilities $\{1-\alpha, \alpha\}$. Therefore, for any $x \in \{0,1\}$,

$$H(Y|X = x) = H_2(\alpha)\,.$$

Therefore

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{x\in\{0,1\}} p_X(x) H(Y|X=x) \\ &= H(Y) - \sum_{x\in\{0,1\}} p_X(x) H(\alpha) = H(Y) - H_2(\alpha)\,. \end{aligned}$$

Let $E = 1_{\{Y=e\}}$. Taking into account the fact that E is a function of Y:

$$H(Y) = H(Y,E) = H(E) + H(Y|E)\,.$$

Note that

$$\begin{aligned} P\,(E=1) &= P(Y=e) \\ &= P(X=0)P(Y=e|X=0) + P(Y=e)P(Y=e|X=1) = \alpha\,. \end{aligned}$$

Therefore $H(E) = H_2(\alpha)$. Moreover,

$$\begin{aligned} H(Y|E) &= P\,(E=0)\,H(Y|E=0) + P\,(E=1)\,H(Y\,|\,E=1) \\ &= (1-\alpha)H(X\,|\,E=0) + \alpha\times 0 = (1-\alpha)H(X)\,. \end{aligned}$$

Therefore

$$H(Y) = H(\alpha) + (1-\alpha)H(X)$$

and

$$I(X;Y) = (1-\alpha)H(X)\,.$$

Since X takes two values, $\sup_X H(X) = 1$. Therefore $C := \sup_X I(X;Y) = 1-\alpha$. □

EXAMPLE 12.1.22: THE SYMMETRIC CHANNEL. The symmetric channel is defined by the following property of its transition matrix $\{p(y\,|\,x)\}_{x\in\mathcal{X},y\in\mathcal{Y}}$. Every column is a permutation of the first column, and every line is a permutation of the first line. (In particular, the binary symmetric channel is a symmetric channel in this sense.) Denote by L the number of elements of the output alphabet $\mathcal{Y}$ and by $(q_1, q_2, \ldots, q_L)$ the first line of the channel transition matrix. The capacity of this channel is (Exercise 12.3.12)

$$C = \log L + \sum_{j=1}^{L} q_j \log q_j\,.$$

12.2 Shannon's Capacity Theorem

12.2.1 Rate versus Accuracy

In the previous chapters, the objective in terms of communications theory was to encode data for compression. Now the objective is to correct the errors introduced by the channel. For this, one has to expand rather than compress the data, thereby introducing redundancy.

EXAMPLE 12.2.1: REPETITION CODING. In the previous example, the sequence that is fed into the channel represents encoded data. The data before encoding is called the informative data. For instance, it consists of M messages, say, $M = 2^k$, so that each message may be represented as a sequence of k binary digits. This message — a binary sequence of length k — is encoded into the sequence $x_1^n = (x_1, \ldots, x_n)$ which in turn is transmitted through the binary symmetric channel. The mapping $c^{(n)} : \{1, 2, \ldots, M\} \to \{0, 1\}^n$ is called the (error-correcting) code. Its rate of transmission is the quantity $R = \frac{k}{n} = \frac{\log_2 M}{n}$. For instance, with $M = 2$, one could use a repetition code, which consists in repeating n times the binary digit to be transmitted. The rate of transmission is then $R = \frac{1}{n}$. At the receiving end of the channel, one has to decode the sequence $y_1^n = (y_1, \ldots, y_n)$. We can use for this purpose a majority decoder, deciding that the binary digit of informative data is the most frequent binary digit in the received sequence y_1^n. Assuming that n is odd, an error occurs if and only if more than $n/2$ among the noise bits B_j, $1 \leq j \leq n$ are equal to 1. Therefore the probability of error per informative bit in this coding procedure is exactly the probability that a binomial random variable $\mathcal{B}(n, p)$ exceeds the value $\frac{n}{2}$. If $p < \frac{1}{2}$ (a reasonable channel), this probability tends to 0 as n tends to infinity. The problem here is that as n tends to infinity, the rate of transmission tends to zero. A fundamental result of information theory, the celebrated Shannon capacity theorem, shows that it is always possible in theory to find error codes with asymptotically evanescent error probability as long as the rate of transmission imposed is smaller than a positive quantity depending on the channel, namely the capacity of this channel (see section 12.1.3).

We now proceed to the statement of Shannon's result. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, the alphabets used at the input and output respectively of the channel. An (error-correcting) code consists of the following items:

- A finite set of indices $\{1, ..., M\}$ (the messages)
- A coding function $c = c^{(n)} : \{1, ..., M\} \to \mathcal{X}^n$
- A decoding function $g = g^{(n)} : \mathcal{Y}^n \to \{1, ..., M\}$

We denote by $c(w) = (c_1(w), ..., c_n(w))$ the code-word for message w, and therefore the code can be represented by an $M \times n$-matrix with elements in the input alphabet $\mathcal{X}$

$$\begin{pmatrix} c_1(1) & c_2(1) & \cdots & c_n(1) \\ c_1(2) & c_2(2) & \cdots & c_n(2) \\ \vdots & \vdots & & \vdots \\ c_1(M) & c_2(M) & \cdots & c_n(M) \end{pmatrix}$$

The first important performance index of a code is its rate $R = \frac{\log_2 M}{n}$. Since $M = \lceil 2^{nR} \rceil$, such a code is called a $\left(\lceil 2^{nR} \rceil, n\right)$-code.

The channel operates as follows. A random message W enters the coder and is encoded by the sequence $c(W) = X_1^n$. The channel "corrupts" X_1^n into the sequence Y_1^n, and the latter is decoded as $\widehat{W} := g(Y_1^n)$.

Encoding and decoding

Ideally, $\widehat{W}$ should be equal to W, which it is not always. The other principal performance index is the error probability associated with the encoding-decoding procedure. Let

$$P_{e|w}(c) := \Pr(g(Y_1^n) \neq w | X_1^n = c(w))$$

be the error probability on the message w, let

$$\lambda(c) = \max_{w \in \{1,\ldots,M\}} P_{e|w}(c)$$

be the maximum error probability, and let

$$P_e(c) = \frac{1}{M} \sum_{w=1}^{M} P_{e|w}(c)$$

be the average error probability. In fact the maximum error probability is the interesting performance index in what concerns channel reliability. The average error probability will merely play an intermediary role in the calculations.

Definition 12.2.2 *Rate R is said to be achievable if there exists a sequence of $\left(\lceil 2^{nR}, n \rceil\right)$-codes such that:*

$$\lambda(c^{(n)}) \longrightarrow 0\,.$$

Theorem 12.2.3 *(Shannon, 1948) Consider a time-invariant discrete-time memoryless channel without feedback, and let C be its capacity. Any rate $R < C$ is achievable.*

This is the direct part of Shannon's theorem. In rough terms: one can transmit at rate below capacity with evanescing error probability.

The converse part of Shannon's capacity theorem consists in the proof that if one wishes to transmit information at a rate $R > C$, whatever error-correcting code that is used, the probability of error is bounded below by a positive number.

Theorem 12.2.4 *(Shannon, 1948) Consider a time-invariant discrete-time memoryless channel without feedback, and let C be its capacity.*

A. If there exists a sequence of channel codes $\left\{c^{(n)}\right\}_{n\geq 1}$ such that $\lim_{n\to\infty} P_e(c^{(n)}) = 0$, then necessarily $\limsup_{n\to\infty} R^{(n)} \leq C$.

B. Moreover, above capacity, the error probability is bounded away from 0.

12.2.2 The Random Coding Argument

We now proceed to the proof of Theorem 12.2.3.

Proof. The argument used by Shannon is the random coding argument. First one generates a random code, that is, a random matrix

$$\begin{pmatrix} C_1(1) & C_2(1) & \cdots & C_n(1) \\ C_1(2) & C_2(2) & \cdots & C_n(2) \\ \vdots & \vdots & & \vdots \\ C_1(M) & C_2(M) & \cdots & C_n(M) \end{pmatrix}$$

whose elements in $\mathcal{X}$ are independent and identically distributed according to some probability distribution q. A code is then sampled from this code distribution. Suppose that code c is drawn. We now devise a decoder (in general not optimal) $\hat{c} = \hat{c}^{(n)} : \mathcal{Y}^n \to \{1, \ldots, M\}$ as follows[1].

(α) $\hat{c}(y_1^n) = \hat{w}$ if and only if

- $(c(\hat{w}), y_1^n) \in A_\epsilon^{(n)}$ (defined in Theorem 12.1.11) and
- no other message $w \in \{1, \ldots, \lceil 2^{nR} \rceil\}$ exists such that $(c(w), y_1^n) \in A_\epsilon^{(n)}$.

(β) If no such message $\hat{w}$ exists, then an error is declared (and the receiver outputs any message).

Denote by C the random code chosen above. Taking expectation (with respect to the code randomness), we have

$$E\left[P_e(C)\right] = \frac{1}{\lceil 2^{nR} \rceil} \sum_{w=1}^{\lceil 2^{nR} \rceil} E\left[P_{e|w}(C)\right] .$$

By symmetry, $E\left[P_{e|w}(C)\right]$ is independant of w, and therefore

$$E\left[P_e(C)\right] = E\left[P_{e|1}(C)\right] . \tag{12.13}$$

We now restrict attention to the event $\{W = 1\}$. Denote by P_1 the probability P conditioned by the event $W = 1$, so that

$$E\left[P_{e|1}(C)\right] = P_1(\hat{C}(Y_1^n) \neq 1) .$$

[1]The notation $\hat{c}^{(n)}$ for $g^{(n)}$ is there to recall that the decoder is adapted to the encoder.

(The capital letter C in $\hat{C}$ tells us that the decoder is now random, since it depends on the random code C.) Let for all messages $w \in \{1, 2, \ldots, M\}$

$$E_w := \left\{(C(w), Y_1^n) \in A_\epsilon^{(n)}\right\} .$$

By the union bound,

$$\begin{aligned} E\left[P_{\mathrm{e}|1}(C)\right] &= P_1(\overline{E_1} \cup E_2 \cup \ldots \cup E_{\lceil 2^{nR} \rceil}) \\ &\leq P_1(\overline{E_1}) + \sum_{w=2}^{\lceil 2^{nR} \rceil} P_1(E_w) . \end{aligned}$$

By Theorem 12.1.11 (a), $P_1(\overline{E_1}) \to 0$. In particular, for sufficiently large n,

$$P_1(\overline{E_1}) \leq \epsilon .$$

Since for $w \neq 1$, $C(1)$ and $C(w)$ are independent, Y_1^n and $C(w)$ are independent with respect to probability P_1. In particular, by Theorem 12.1.11 (c), for sufficiently large n,

$$P_1(E_w) \leq 2^{-n(I(X;Y)-3\epsilon)} ,$$

where (X, Y) is a random vector with values in $\mathcal{X} \times \mathcal{Y}$ and distribution $p_{(X,Y)}(x, y) = q(x)\, p(x \,|\, y)$. Therefore

$$\begin{aligned} E\left[P_{e|1}(C)\right] &\leq \epsilon + \sum_{w=2}^{\lceil 2^{nR} \rceil} 2^{-n(I(X;Y)-3\epsilon)} \\ &\leq \epsilon + \left\lceil 2^{nR} \right\rceil \times 2^{-n(I(X;Y)-3\epsilon)} \\ &= \epsilon + 2^{-n(I(X;Y)-R-3\epsilon)} . \end{aligned}$$

Therefore if $R < I(X;Y) - 3\epsilon$, for sufficiently large n

$$E\left[P_{e|1}(C)\right] \leq 2\epsilon .$$

Choose $q = q^*$ so that $I(X;Y) = C$ (the capacity of the channel) (recall that the supremum in the definition of capacity is a maximum). Therefore if $R < C - 3\epsilon$, then for sufficiently large n,

$$E\left[P_e(C)\right] \leq 2\epsilon .$$

In particular (the probabilistic method argument), there exists at least a code $c^* = c^{*(n)}$ with rate $R < C$ and error probability

$$E\left[P_e(c^*)\right] \leq 2\epsilon,$$

that is

$$\frac{1}{\lceil 2^{nR} \rceil} \sum_{w=1}^{\lceil 2^{nR} \rceil} P_{\mathrm{e}|w}(c^*) \leq 2\epsilon .$$

This implies that at least half the messages w satisfy $P_{\mathrm{e}|w}(c^*) \leq 4\epsilon$. Keeping only these messages and their associated code-words $c^*(w)$ gives a new code with $M' = \left\lceil 2^{nR} \right\rceil / 2$ messages, rate

$$R' = \frac{\log_2 M'}{n} = \frac{\log_2 \left\lceil 2^{nR} \right\rceil}{n} - \frac{1}{n} ,$$

and maximal error probability $\leq 4\epsilon$. □

12.2.3 Proof of the Converse

We proceed to the proof of Theorem 12.2.4.

Proof. The following lemmas prepare the way.

Lemma 12.2.5 *Let Y_1^n be the output corresponding to the input X_1^n through a memoryless channel (not necessarily time-homogeneous, and possibly with feedback). Then*

$$I(X_1^n; Y_1^n) \leq \sum_{k=1}^{n} I(X_k; Y_k)$$

whatever the distribution of X_1^n.

Proof.

$$\begin{aligned} I(X_1^n; Y_1^n) &= H(Y_1^n) - H(Y_1^n | X_1^n) \\ &= H(Y_1^n) - \sum_{k=1}^{n} H(Y_k | Y_1^{k-1}, X_1^n) \\ &= H(Y_1^n) - \sum_{k=1}^{n} H(Y_k | X_k) \\ &\leq \sum_{k=1}^{n} H(Y_k) - \sum_{k=1}^{n} H(Y_k | X_k) = \sum_{k=1}^{n} I(X_k; Y_k)\,. \end{aligned}$$

□

Lemma 12.2.6 *Consider a discrete channel (not necessarility memoryless). Given positive integers n, M and a code $c : \{1, \ldots, M\} \to \mathcal{X}^n$, then*

$$\log M \leq \frac{1}{1 - P_e(c)} \left(I(c(W); Y_1^n) + h(P_e(c)) \right)$$

where $P(W = w) = 1/M$.

Proof. From Fano's inequality (12.5) (with $W, Y_1^n, \hat{W}$ in the roles of $X, Y, \hat{X}$ respectively),

$$H(W | Y_1^n) \leq h(P_e(c)) + P_e(c) \log M\,,$$

and therefore (Eqn. (12.1.3))

$$I\left(W; Y_1^n\right) \geq H(W) - P_e(c) \log M - h(P_e(c))\,.$$

Since W is uniformly distributed, $H(W) = \log M$, and therefore

$$I\left(W; Y_1^n\right) \geq (1 - P_e(c)) \log M - h(P_e(c))\,.$$

Moreover (Exercise 12.3.7),

$$I(c(W); Y_1^n) \geq I(W; Y_1^n)\,.$$

Combining the above inequalities yields the result. □

We are now ready to complete the proof of Theorem 12.2.4.

A. To simplify the notation, assume that $2^{nR^{(n)}}$ is an integer. For fixed n (and therefore fixed code), choose W uniformly in $\left\{1, \ldots, 2^{nR^{(n)}}\right\}$, and therefore $\log M = nR^{(n)}$. Taking Lemmas 12.2.6 and 12.2.5 into account,

$$\begin{aligned} R^{(n)} &\leq \frac{1}{1-P_e(c)}\left[\frac{1}{n}I(X_1^n; Y_1^n) + \frac{1}{n}h(P_e(c))\right] \\ &\leq \frac{1}{1-P_e(c)}\left[\frac{1}{n}I(X_1^n; Y_1^n) + \frac{1}{n}\right] \leq \frac{1}{1-P_e(c)}\left[C + \frac{1}{n}\right]\,. \end{aligned}$$

Therefore

$$R^{(n)}(1 - P_e(c)) \leq C + \frac{1}{n}\,,$$

from which it follows that $\limsup_{n\to\infty} R^{(n)} \leq C$.

B. From the last displayed inequality, we get

$$P_e(c) \geq 1 - \frac{C}{R^{(n)}} - \frac{1}{nR^{(n)}}\,.$$

This shows that if $R > C$, then for arbitrary ϵ and large enough n, $P_e(c) \geq 1 - \frac{C}{R} - \epsilon$, a positive quantity if ϵ is small enough. That is, in words: over capacity, the error probability is bounded away from 0. □

12.2.4 Feedback Does not Improve Capacity

In the absence of feedback, the channel code is a function $c : \{1, \ldots, M\} \to \mathcal{X}^n$. If feedback is allowed, the channel code is composed of a collection of functions $c : \{1, \ldots, M\} \times \mathcal{Y}^{n-1} \to \mathcal{X}$, that is

$$X_n = c(W, Y_1^{n-1})\,.$$

To prove that feedback does not improve capacity, we revisit the proof of the converse Shannon theorem. As in the case without feedback, we have

$$\begin{aligned} nR^{(n)} = H(W) &= H(W|Y_1^n) + I(W; Y_1^n) \\ &\leq H(W|Y_1^n) + I(X_1^n; Y_1^n)\,. \end{aligned}$$

Now we have to bound $I(W; Y^n)$. To this end, note that

$$\begin{aligned} I(W, Y^n) &= H(Y_1^n) - H(Y_1^n|W) = H(Y_1^n) - \sum_{k=1}^{n} H\left(Y_k|Y_1^{k-1}, W\right) \\ &= H(Y_1^n) - \sum_{k=1}^{n} H\left(Y_k|Y_1^{k-1}, W, X_k\right) \ \text{(since } X_k = f(W, Y_1^{k-1})\text{)} \\ &= H(Y^n) - \sum_{k=1}^{n} H(Y_k|X_k) \end{aligned}$$

where the third inequality is due to the fact that X_k is a function of (W, Y_1^{k-1}) and the last one is due to the fact that conditionally on the channel input X_k, the channel output Y_k is independent of W and of the past samples $Y_1, \cdots, Y_{k-1}$. Therefore

$$I(W, Y_1^n) \leq \sum_{k=1}^{n} H(Y_k) - \sum_{k=1}^{n} H(Y_k|X_k) = nI(X_k; Y_k) \leq nC\,.$$

The end of the proof is then similar to the case without feedback.

Books for Further Information

The original work of Claude Shannon was published in [Shannon and Weaver, 1949]. Otherwise, see the bibliography of Chapter 11.

12.3 Exercises

Exercise 12.3.1. $H(X|Y,Z) \leq H(X|Y)$
Prove formally the inequality (12.4).

Exercise 12.3.2. $H(X|Y,Z) = \sum_{z\in\mathcal{Z}} P(Z=z)H^{Z=z}(X|Y)$
Let X, Y, Z be random variables taking their values in the finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively. For a fixed $z \in \mathcal{Z}$, the meaning of the quantity $H(X|Y, Z=z)$ is the following. It is the conditional entropy of X given Y, computed not for the original probability, but for the conditional probability $P^{Z=z}(\cdot) = P(\cdot|Z=z)$. Prove that

$$H(X|Y,Z) = \sum_{z\in\mathcal{Z}} P(Z=z)H^{Z=z}(X|Y)\,.$$

Exercise 12.3.3. A concavity property of mutual information
Prove the following. For a fixed transition function $p(\cdot|\cdot)$, the mutual information $I(X;Y)$ is a concave function of p_X (the distribution of X).

Exercise 12.3.4. Discrete memoryless channel without feedback
Prove Theorem 12.1.15.

Exercise 12.3.5. $I(Y;Y) = H(Y)$
Prove that $I(Y;Y) = H(Y)$.

Exercise 12.3.6. $H(\varphi(X)|X) = 0$
Let X be a random variable with values in the finite set $\mathcal{X}$. Let $\varphi : \mathcal{X} \to \mathcal{Y}$ be some function with values in a finite set $\mathcal{Y}$. Prove that $H(\varphi(X)|X) = 0$.

Exercise 12.3.7. $I(\varphi(X);Y) \leq I(X;Y)$
Prove the following. For any finite-valued random variables X and Y and any deterministic function φ,

$$I(\varphi(X);Y) \leq I(X;Y)$$

with equality if and only if for all x, x', y, $\varphi(x) = \varphi(x') \Rightarrow p_{Y|X}(y|x) = p_{Y|X}(y|x')$.

Exercise 12.3.8. Optimal decoding in the BSC channel
(Continuation of Exercise 2.4.25.) Let $E = \{0,1\}^n$. The addition $\oplus$ defined on E being componentwise addition modulo 2, the observation is $X = m_\Theta \oplus Z$ where:

$$m_i \in \{0,1\}^n \quad m_i = (m_i(1), ..., m_i(n)), \quad Z = (Z_1, ..., Z_n)\,,$$

where Z and Θ are independent and the Z_i's $(1 \leq i \leq n)$ are independent and identically distributed with $\Pr(Z_i = 1) = p$. A possible interpretation is in terms of digital communications. One wishes to transmit the information Θ chosen among a finite set of "messages" which are binary strings of length n: $m_1, \ldots, m_K$. The vector Z is the "noise" inherent to all digital communications channels: if $Z_k = 1$ the k-th bit of the message Θ is flipped. This error occurs with probability p, independently for all the bits of the message.

Suppose that the hypotheses are equiprobable and that $p < \frac{1}{2}$. We have :

$$P(X = x|\Theta = i) = P(Z \oplus m_i = x) = P(Z = m_i \oplus x)\,.$$

Denote by $h(y)$ the Hamming weight of $y \in \{0,1\}^n$ (equal to the number of components of y that are equal to 1), and let

$$d(x,y) := \sum_{i=1}^{n} 1_{\{x_i \neq y_1\}} = \sum_{i=1}^{n} x_i \oplus y_i = h(x \oplus y)$$

the Hamming distance between x and y in E^n. Prove that the optimal strategy consists in choosing the hypothesis corresponding to the message closest to the observation in terms of the Hamming distance.

Exercise 12.3.9. Cascade of BSC channels
A number n of identical binary symmetric channels with error probability p are put in series. What is the capacity of the resulting channel?

Exercise 12.3.10. BSC channels in parallel
Let C_1 and C_2 be the capacities of two discrete memoryless channels, with perhaps different input (resp., output) alphabets, and with transition probabilities $p_1(y_1|x_1)$ and $p_2(y_2|x_2)$ respectively. The product channel is the discrete memoryless channel associating the output (y_1, y_2) to the input (x_1, x_2), with transition probability $p_1(y_1|x_1)p_2(y_2|x_2)$. What is the capacity of such a channel?

Exercise 12.3.11. The modulo channel
Consider the channel whose input and output both take their values in $\mathcal{X} \equiv \mathcal{Y} := \{0, 1, \cdots, L-1\}$ and are related by the equation $Y = X + Z \pmod{c}$, where Z takes its values in $\mathcal{X}$ and is independent of X. Compute the channel capacity in terms of the (arbitrary) distribution of Z.

Exercise 12.3.12. Capacity of the symmetric channel
The symmetric channel is defined by the following property of its transition matrix Every column is a permutation of the first column, and every line is a permutation of the first line. (In particular, the binary symmetric channel is a symmetric channel in this sense.) Denote by L the number of elements of the output alphabet $\mathcal{Y}$ and by $(q_1, q_2, \dots, q_L)$ the first line of the channel transition matrix. Prove that the capacity of this channel is

$$C = \log L + \sum_{j=1}^{L} q_j \log q_j \, .$$

Exercise 12.3.13. The noisy typewriter
The input symbols are the 26 letters of the alphabet. When fed with a letter, the channel gives to the receiver the said letter with probability $\frac{1}{2}$ or the next one with probability $\frac{1}{2}$ (the "next" letter of Z is A). Compute the capacity of this channel. Hint: Separate the source message in two, by sending first the odd letters A,C,..., and then the even letters B,D, ...)

Exercise 12.3.14. The asymmetric erasure channel
Compute the capacity of the channel with transition matrix

$$\begin{pmatrix} 1-\alpha-\beta & \alpha & \beta \\ \alpha & 1-\alpha-\beta & \beta \end{pmatrix} .$$

Chapter 13

The Method of Types

13.1 Divergence and Types

13.1.1 Divergence

The method of types allows one to obtain in an elementary way, in the discrete case, two fundamental results of mathematical information theory, which in the continuous case require a rather formidable technical equipment: Sanov's principle and the maximum entropy principle. The objects considered by this method are empirical averages, and the basic mathematical notion is that of divergence.

Let $\mathcal{P}$ be the collection of probability distributions on the (finite) set $\mathcal{X}$. For $p, q \in \mathcal{P}$, define

$$D(p\,;\,q) := \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{13.1}$$

(with the "log convention": $0 \log 0 := 0$ and $a \log 0 := -\infty$ when $a > 0$). This quantity is the Kullback–Leibler divergence between p and q or, more simply, the divergence between p and q.

Example 13.1.1: An interpretation in terms of source coding. Recall the Gibbs inequality,

$$-\sum_{x\in\mathcal{X}} p(x) \log p(x) \le -\sum_{x\in\mathcal{X}} p(x) \log q(x)\,,$$

with equality if and only if $p(x) = q(x)$ for all $x \in \mathcal{X}$. The heuristic interpretation of this inequality is that if we encode the elements of $\mathcal{X}$ using a code assigning to $x \in \mathcal{X}$ a codeword of length $\log p(x)$, this code has the smallest average length among all uniquely decipherable codes, where the average is computed with respect to the probability distribution p on $\mathcal{X}$. Therefore, if one believes erroneously that the probability distribution on $\mathcal{X}$ is q, and consequently chooses the corresponding optimal code with length $\log q(x)$ for $x \in \mathcal{X}$, the resulting average code length (computed with the actual probability distribution p) is $-\sum_{x\in\mathcal{X}} p(x) \log q(x)$ and is necessarily larger than the best code average length $-\sum_{x\in\mathcal{X}} p(x) \log p(x)$. Of course the discussion above is meaningful only "at the limit" (see the concatenation argument after Lemma 11.1.17). The Gibbs inequality may be rewritten as

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_13

$$D(p\,;\,q) \geq 0\,. \tag{13.2}$$

Pinsker's Theorem

The function $(p,q) \to D(p\,;\,q)$ is not a distance as the notation perhaps suggests. In fact, it is not symmetric in p and q and the triangle inequality is not available. However, D plays the role of a distance in the following sense:

$$\lim_{n\uparrow\infty} D(p_n\,;\,q) = 0 \Longrightarrow \lim_{n\uparrow\infty} d_V(p_n,q) = 0\,,$$

where $d_V(p,q) := \frac{1}{2}\sum_{x\in\mathcal{X}} |p(x) - q(x)|$ is the total variation distance between p and q. This is a consequence of the inequality below.

Theorem 13.1.2 *(Pinsker)*

$$D(p\,;\,q) \geq 2d_V(p,q)^2\,.$$

The proof is based on the following partition lemma for divergence.

Lemma 13.1.3 *Let $\mathcal{A} = (A_1,\ldots,A_k)$ be a partition of $\mathcal{X}$ and let $p \in \mathcal{P}$. Define the probability distribution $p_{\mathcal{A}}$ on $\mathcal{A}$ by*

$$p_{\mathcal{A}}(A_i) := \sum_{x\in A_i} p(x)$$

and define similarly the distribution $q_{\mathcal{A}}$ associated with the probability distribution q. Then

$$D(p\,;\,q) \geq D(p_{\mathcal{A}}\,;\,q_{\mathcal{A}})\,.$$

Proof. By the Log-Sum inequality (Exercise 13.3.4),

$$\begin{aligned} D(p\,;\,q) = \sum_{i=1}^{k}\sum_{x\in A_i} p(x)\log\frac{p(x)}{q(x)} &\geq \sum_{i=1}^{k}\left(\sum_{x\in A_i} p(x)\right)\log\frac{\sum_{x\in A_i} p(x)}{\sum_{x\in A_i} q(x)} \\ &= \sum_{i=1}^{k} p_{\mathcal{A}}(A_i)\log\frac{p_{\mathcal{A}}(A_i)}{q_{\mathcal{A}}(A_i)} = D(p_{\mathcal{A}}\,;\,q_{\mathcal{A}})\,. \end{aligned}$$

□

Proof. (of Theorem 13.1.2) Observe that if $A_1 = \{x \in \mathcal{X}\,;\,p(x) \geq q(x)\}$ and $A_2 = \{x \in \mathcal{X}\,;\,p(x) < q(x)\}$,

$$\begin{aligned} 2d_V(p,q) &= \sum_{x\in A_1}(p(x) - q(x)) - \sum_{x\in A_2}(p(x) - q(x)) \\ &= (p_{\mathcal{A}}(A_1) - q_{\mathcal{A}}(A_1)) - (p_{\mathcal{A}}(A_2) - q_{\mathcal{A}}(A_2)) \\ &= |p_{\mathcal{A}}(A_1) - q_{\mathcal{A}}(A_1)| + |p_{\mathcal{A}}(A_2) - q_{\mathcal{A}}(A_2)| = 2d_V(p_{\mathcal{A}},q_{\mathcal{A}})\,. \end{aligned}$$

Suppose the inequality proved in the case where $\mathcal{X}$ consists of just two elements. In particular

$$D(p_{\mathcal{A}}\,;\,q_{\mathcal{A}}) \geq 2d_V(p_{\mathcal{A}},q_{\mathcal{A}})^2 = 2d_V(p,q)^2\,,$$

and the result then follows from the partition lemma.

It therefore remains to verify that Pinsker's inequality holds true in the case $\mathcal{X} = \{0,1\}$. Let $p = (a, 1-a)$ and $q = (b, 1-b)$. Eliminating the trivial cases, we may suppose that $a, b \in (0,1)$. Consider the function

$$g(b) := D(p\,;\,q) - 2d_V(p,q)^2\,.$$

Since $d_V(p,q) = \frac{1}{2}\left(|a-b| + |(1-a)-(1-b)|\right) = |a-b|$,

$$g(b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b} - 2(a-b)^2\,.$$

We have $g(a) = 0$ and

$$g'(b) = (b-a)\left(\frac{1}{(1-b)b} - 4\right)\,.$$

Since $b(1-b) \leq \frac{1}{4}$, this shows that $g(b)$ has a minimum at $a = b$. Therefore,

$$D(p\,;\,q) - 2d_V(p,q)^2 \geq D(p\,;\,p) - 2d_V(p,p)^2 = 0\,.$$

□

Theorem 13.1.4 *The function $D : (p,q) \to D(p\,;\,q)$ is convex in both its arguments, that is, for all $p_1, q_1, p_2, q_2 \in \mathcal{P}$ and all $\lambda \in [0,1]$:*

$$D(\lambda p_1 + (1-\lambda)p_2\,;\,\lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1\,;\,q_1) + (1-\lambda)D(p_2\,;\,q_2)\,.$$

If q is strictly positive, the function $D : p \to D(p\,;\,q)$ is continuous.

The proof is left as an exercise (Exercise 13.3.5).

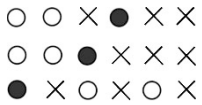
13.1.2 Empirical Averages

Let $X_1, \ldots, X_n$ be an IID sample (n-sample) of a distribution Q on the (finite) set $\mathcal{X}$, that is, a collection of n independent random variables with values in $\mathcal{X}$ and common probability distribution Q. In statistics, one is interested in the empirical distribution on $\mathcal{X}$ associated with this sample. The relevant notion is then that of the type of this sample.

Let $x_1^n := (x_1, \ldots, x_n)$ be a sequence of elements of the discrete set $\mathcal{X} = \{a_1, \ldots, a_L\}$. This sequence will also be denoted by $\mathbf{x}$. Let $h(a;\mathbf{x}) := \sum_{i=1}^n 1_{\{x_i = a\}}$ be the Hamming weight of $a \in \mathcal{X}$ in $\mathbf{x}$. The type of vector $\mathbf{x} \in \mathcal{X}^n$ is the empirical probability distribution $p_{\mathbf{x}}$ on $\mathcal{X}$ corresponding to $\mathbf{x}$:

$$p_{\mathbf{x}}(a) := \frac{h(a;\mathbf{x})}{n} \qquad (a \in \mathcal{X})\,.$$

(Note that the type of a sequence is independent of the order of its elements.)



Three sequences with the same type

In particular, for the sample $X_1^n := (X_1, \ldots, X_n)$,

$$p_{X_1^n}(a) = \frac{h(a; X_1^n)}{n}, \qquad a \in \mathcal{X}$$

is the empirical distribution on $\mathcal{X}$ associated with this sample.

For any $\mathbf{x} = x_1^n \in \mathcal{X}^n$ and any function $g : \mathcal{X} \to \mathbb{R}$, the empirical average of g on the sample $\mathbf{x}$ is

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i) = \sum_{a\in\mathcal{X}} g(a) p_{\mathbf{x}}(a)\,,$$

and in particular,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = \sum_{a\in\mathcal{X}} g(a) p_{X_1^n}(a)\,.$$

Let $\mathcal{P}_n$ denote the collection $\{p_{\mathbf{x}}\,;\, \mathbf{x} \in \mathcal{X}^n\}$ of all types. This subset of $\mathcal{P}$ consists of all the probability distributions on $\mathcal{X}$ of the form

$$p(a_1) = \frac{k_1}{n}, \ldots, p(a_L) = \frac{k_L}{n} \qquad (\sum_{i=1}^{L} k_i = n)\,.$$

Therefore, a probability $p \in \mathcal{P}_n$ is described by an integer vector $(k_1, \ldots, k_L)$ such that $k_1 + \cdots + k_L = n$. Since each one of the $|\mathcal{X}|$ components of a vector $p \in \mathcal{P}_n$ can take at most $n+1$ values,

$$|\mathcal{P}_n| \le (n+1)^{|\mathcal{X}|}\,,$$

which is a simple bound for the exact value

$$|\mathcal{P}_n| = \binom{n + |\mathcal{X}| - 1}{|\mathcal{X}| - 1}\,.$$

The type class $T_n(p)$ of $p \in \mathcal{P}_n$ is the set

$$T_n(p) := \{\mathbf{x} \in \mathcal{X}^n\,;\, p_{\mathbf{x}} = p\}\,.$$

The number of sequences $\mathbf{x} \in \mathcal{X}^n$ whose type is p, with k_i occurences of the symbol a_i $(1 \le i \le L)$, is

$$|T_n(p)| = \frac{n!}{k_1! \cdots k_L!}\,.$$

EXAMPLE 13.1.5: The sequence *bbabcccaba* has the same type as the sequence *aaabbbbccc*, namely $\left(\frac{3}{10}, \frac{4}{10}, \frac{3}{10}\right)$. The sequence *abababcbcc* belongs to $T_{10}\left(\left(\frac{3}{10}, \frac{4}{10}, \frac{3}{10}\right)\right)$.

Lemma 13.1.6 *For any type $p \in \mathcal{P}_n$,*

$$(n+1)^{-|\mathcal{X}|} 2^{-nH(p)} \leq |T_n(p)| \leq 2^{-nH(p)} .$$

Proof. Consider the multinomial expansion

$$n^n = (k_1 + \cdots + k_L)^n = \sum \frac{n!}{j_1! \cdots j_L!} k_1^{j_1} \cdots k_L^{j_L} \qquad (\star)$$

where the summation domain consists of all $(j_1, \ldots, j_L)$ such that $j_1 + \cdots + j_L = n$. The largest term in this sum is the one corresponding to $(j_1, \ldots, j_L) = (k_1, \ldots, k_L)$:

$$\frac{n!}{k_1! \cdots k_L!} k_1^{k_1} \cdots k_L^{k_L} .$$

(Indeed, if $(j_1, \ldots, j_L) \not\equiv (k_1, \ldots, k_L)$, there exists at least two indices r and s such that $j_r > k_r$ and $j_s < k_s$. Decreasing j_r by 1 and increasing j_s by 1 multiplies the corresponding term by

$$\frac{j_r}{k_r} \frac{k_s}{1 + j_s} \geq \frac{j_r}{k_r} > 1.$$

This contradicts the existence of a maximum at $(j_1, \ldots, j_L) \not\equiv (k_1, \ldots, k_L)$.)

Bounding the right-hand side sum of $(\star)$ below by its largest term, and above by the largest term times the number of terms, we obtain

$$\frac{n!}{k_1! \cdots k_L!} k_1^{k_1} \cdots k_L^{k_L} \leq n^n \leq (n+1)^{|\mathcal{X}|} \frac{n!}{k_1! \cdots k_L!} k_1^{k_1} \cdots k_L^{k_L}$$

or, equivalently,

$$\frac{n!}{k_1! \cdots k_L!} \leq \frac{n^n}{k_1^{k_1} \cdots k_L^{k_L}} \leq (n+1)^{|\mathcal{X}|} \frac{n!}{k_1! \cdots k_L!} ,$$

from which the result follows by noting that

$$\log \frac{n^n}{k_1^{k_1} \cdots k_L^{k_L}} = -\sum_{i=1}^{L} \log \frac{k_i^{k_i}}{n^{k_i}} = -n \sum_{i=1}^{L} \frac{k_i}{n} \log \frac{k_i}{n} = -nH(p) .$$

□

Let $X_1, \ldots, X_n$ be IID random variables with values in $\mathcal{X}$ and common distribution Q. Denote by Q^n the product measure $Q \times \cdots \times Q$ on $\mathcal{X}^n$, that is, the probability distribution of the random vector $(X_1, \ldots, X_n)$. Let Π be a subset of $\mathcal{P}$ (for instance Π is the set of probabilities on $\mathcal{X}$ with mean μ). By definition, with an obvious but notationally convenient abuse of notation,

$$Q^n(\Pi) := \sum_{\mathbf{x} \,;\, p_{\mathbf{x}} \in \Pi \cap \mathcal{P}_n} Q^n(\mathbf{x}) .$$

This is the probability that the sample $\{X_1, \ldots, X_n\}$ has an empirical distribution that belongs to Π. Now since $T_n(p)$ is the set of sequences $\mathbf{x} = x_1^n$ with empirical distribution $p_{\mathbf{x}} = p$,

$$Q^n(\Pi) = \sum_{p \in \Pi \cap \mathcal{P}_n} Q^n(T_n(p)) \,.$$

The next result shows in particular that the Q^n-probability of $\mathbf{x} := x_1^n$ depends only on its type $p_{\mathbf{x}}$:

Lemma 13.1.7 *(a) Let Q be a probability distribution on $\mathcal{X}$. Then*

$$Q^n(\mathbf{x}) = 2^{-n(H(p_{\mathbf{x}})+D(p_{\mathbf{x}};Q))} \,.$$

In particular, if $\mathbf{x}$ is in the type class of Q, that is, if $p_{\mathbf{x}} \equiv Q$, $Q^n(\mathbf{x}) = 2^{-nH(Q)}$.

(b) Let $p \in \mathcal{P}_n$ and let Q be any probability distribution on $\mathcal{X}$. Then

$$(n+1)^{-|\mathcal{X}|} 2^{-nD(p;Q)} \leq Q^n(T_n(p)) \leq 2^{-nD(p;Q)} \,.$$

(c) Let $p \in \mathcal{P}_n$ and let Q be any probability distribution on $\mathcal{X}$. Then, for all $\mathbf{x} \in T_n(p)$,

$$\frac{Q^n(\mathbf{x})}{p^n(\mathbf{x})} = 2^{-nD(p\,;Q)} \,.$$

Proof. (a) We have:

$$\begin{aligned} Q^n(\mathbf{x}) &= \prod_{i=1}^n Q(x_i) = \prod_{a \in \mathcal{X}} Q(a)^{h(a;\mathbf{x})} \\ &= \prod_{a \in \mathcal{X}} Q(a)^{np_{\mathbf{x}}(a)} = 2^{np_{\mathbf{x}}(a)\log Q(a)} \\ &= 2^{n(p_{\mathbf{x}}(a)\log Q(a) - p_{\mathbf{x}}(a)\log p_{\mathbf{x}}(a) + p_{\mathbf{x}}(a)\log p_{\mathbf{x}}(a))} \\ &= 2^{n\sum_{a\in\mathcal{X}}\left(-p_{\mathbf{x}}(a)\log\frac{p_{\mathbf{x}}(a)}{Q(a)} + p_{\mathbf{x}}(a)\log p_{\mathbf{x}}(a)\right)} \\ &= 2^{-n(H(p_{\mathbf{x}})+D(p_{\mathbf{x}};Q))} \,. \end{aligned}$$

(b) By the result of (a),

$$Q^n(T_n(p)) = \sum_{\mathbf{x}\in T_n(p)} Q^n(\mathbf{x}) = \sum_{\mathbf{x}\in T_n(p)} 2^{-n(D(p;Q)+H(p))} = |T_n(p)| 2^{-n(D(p;Q)+H(p))} \,.$$

The conclusion then follows from Lemma 13.1.6.

(c) Since $\mathbf{x} = x_1^n \in T_n(p)$ implies that the number of occurences of the symbol a in $\mathbf{x}$ is equal to $np(a)$,

$$\frac{Q^n(\mathbf{x})}{p^n(\mathbf{x})} = \prod_{a\in\mathcal{X}} \left(\frac{Q(a)}{p(a)}\right)^{np(a)} = 2^{n\sum_a p(a)\log\frac{Q(a)}{p(a)}} = 2^{-nD(p\,;Q)} \,.$$

□

Recall that $p_{X_1^n}$ denotes the empirical distribution of a random IID sample X_1^n.

Theorem 13.1.8 *Let X_1^n be a random* IID *sample from the probability distribution Q on $\mathcal{X}$. Then, for all $\delta > 0$,*

$$Q^n(D(p_{X_1^n}\,;\,Q) \geq \delta) \leq (n+1)^{|\mathcal{X}|}2^{-n\delta}\,.$$

Proof. The probability of the left-hand side is

$$Q^n(\{\mathbf{x}\,;\,D(p_{\mathbf{x}}\,;\,Q) \geq \delta\}) = \sum_{p\in\mathcal{P}_n;D(p\,;\,Q)\geq\delta} Q^n(T_n(p))\,.$$

But, by Lemma 13.1.7,

$$Q^n(T_n(p)) \leq 2^{-nD(p\,;\,Q)}\,.$$

Therefore

$$\sum_{p\in\mathcal{P}_n;D(p\,;\,Q)\geq\delta} Q^n(T_n(p)) \leq \sum_{p\in\mathcal{P}_n} 2^{-n\delta} = (n+1)^{|\mathcal{X}|}2^{-n\delta}\,.$$

□

In particular, it follows from Theorem 4.1.3 that

$$Pr(\lim_{n\uparrow\infty} D(p_{X_1^n}\,;\,Q) = 0) = 1\,,$$

(where Pr is the probability that makes the sequence $\{X_n\}_{n\geq 1}$ IID with common probability distribution Q) and therefore, by Pinsker's inequality,

$$Pr(\lim_{n\uparrow\infty} d_V(p_{X_1^n}, Q) = 0) = 1\,.$$

13.2 Sanov's Theorem

13.2.1 A Theorem on Large Deviations

The Chernoff bounds can be interpreted in terms of large deviations from the law of large numbers. Recall Theorem 3.2.3 and (3.4):

Let $X_1, \ldots, X_n$ be IID discrete real-valued random variables and let $a \in \mathbb{R}$. Then

$$P\left(\sum_{i=1}^n X_i \geq na\right) \leq e^{-nh^+(a)}\,,$$

where $h^+(a) := \sup_{t>0}\{at - \log E\left[e^{tX_1}\right]\}$ and

$$P\left(\sum_{i=1}^n X_i \leq na\right) \leq e^{-nh^-(a)}\,,$$

where $h^-(a) := \sup_{t<0}\{at - \log E\left[e^{tX_1}\right]\}$. Moreover, $h^+(a)$ (resp., $h^-(a)$) is positive if $a > E[X_1]$ (resp., $a < E[X_1]$).

The theme of large deviations will now be approached via the method of types. We shall be interested in certain subsets Π of probabilities on $\mathcal{X}$, for instance

$$\Pi = \{p\,;\, \sum_{a\in\mathcal{X}} g(a)p(a) > \alpha\}\,. \qquad (\star)$$

Note that, in this example, for any $\mathbf{x} \in \mathcal{X}^n$

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i) > \alpha \Longleftrightarrow \sum_{a\in\mathcal{X}} p_{\mathbf{x}}(a)g(a) > \alpha \Longleftrightarrow p_{\mathbf{x}} \in \Pi \cap \mathcal{P}_n\,.$$

Suppose now that $X_1, \ldots, X_n$ are IID random variables with values in $\mathcal{X}$ and common distribution Q. Then

$$P\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) > \alpha\right) = \sum_{\mathbf{x}\,;\, p_{\mathbf{x}}\in\Pi\cap\mathcal{P}_n} Q^n(\mathbf{x})\,.$$

Let Π be a subset of the set $\mathcal{P}$ of all probability distributions on $\mathcal{X}$. For any given distribution $Q \in \mathcal{P}$ define

$$P^* := \arg\min_{P\in\Pi} D(P\,;\, Q)\,.$$

Theorem 13.2.1 *(Sanov, 1957) Let $X_1, \ldots, X_n$ be* IID *random variables with values in $\mathcal{X}$ and common strictly positive probability distribution Q and let Π be a subset of $\mathcal{P}$. Then*

$$Q^n(\Pi \cap \mathcal{P}_n) \le (n+1)^{|\mathcal{X}|} 2^{-nD(P^*\,;\,Q)}\,. \qquad (13.3)$$

Suppose in addition that the closure of Π is the closure of its interior. Then

$$\lim_{n\uparrow\infty} \frac{1}{n} \log Q^n(\Pi \cap \mathcal{P}_n) = -D(P^*\,;\, Q)\,. \qquad (13.4)$$

In the above example for Π, (13.3) and (13.4) read, respectively

$$Pr\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) > \alpha\right) \le (n+1)^{|\mathcal{X}|} 2^{-nD(P^*\,;\,Q)}$$

and

$$\lim_{n\uparrow\infty} \frac{1}{n} \log P\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) > \alpha\right) = D(P^*\,;\, Q)\,.$$

Proof. The proof of (13.3) consists of the following chain of inequalities:

$$\begin{aligned}
Q^n(\Pi) &= \sum_{P\in\Pi\cap\mathcal{P}_n} Q^n(T_n(P)) \le \sum_{P\in\Pi\cap\mathcal{P}_n} 2^{-nD(P\,;\,Q)} \\
&\le \sum_{P\in\Pi\cap\mathcal{P}_n} \max_{P\in\Pi\cap\mathcal{P}_n} 2^{-nD(P\,;\,Q)} = \sum_{P\in\Pi\cap\mathcal{P}_n} 2^{-n\min_{P\in\Pi\cap\mathcal{P}_n} D(P\,;\,Q)} \\
&\le \sum_{P\in\Pi} 2^{-n\min_{P\in\Pi\cap\mathcal{P}_n} D(P\,;\,Q)} \\
&= \sum_{P\in\Pi} 2^{-nD(P^*\,;\,Q)} \le (n+1)^{|\mathcal{X}|} 2^{-nD(P^*\,;\,Q)}\,.
\end{aligned}$$

We now turn to the proof of (13.4). The set $\cup_n \mathcal{P}_n$ is dense in $\mathcal{P}$. Therefore the set $\Pi \cap \mathcal{P}_n$ is non-empty when n is sufficiently large. There exists a sequence of probability distributions P_n $(n \geq 1)$ such that $P_n \in \Pi \cap \mathcal{P}_n$ and $\lim_n D(P_n\,;\,Q) = D(P^*\,;\,Q)$. Therefore, for sufficiently large n,

$$Q^n(\Pi) = \sum_{P \in \Pi \cap \mathcal{P}_n} Q^n(T_n(P)) \geq Q^n(T_n(P_n)) \geq \frac{1}{(n+1)^{|\mathcal{X}|}} 2^{-nD(P_n\,;\,Q)}\,,$$

so that

$$\liminf_n \frac{1}{n} \log Q^n(\Pi \cap \mathcal{P}_n) \geq \liminf_n \left(-\frac{|\mathcal{X}| \log(n+1)}{n} - D(P_n\,;\,Q) \right) = -D(P^*\,;\,Q)\,.$$

By (13.3),

$$\frac{1}{n} \log Q^n(\Pi \cap \mathcal{P}_n) \leq |\mathcal{X}| \frac{\log(n+1)}{n} - D(P^*\,;\,Q)\,,$$

and therefore

$$\limsup_n \frac{1}{n} \log Q^n(\Pi \cap \mathcal{P}_n \leq -D(P^*\,;\,Q)\,,$$

which concludes the proof of (13.4). □

EXAMPLE 13.2.2: Let for any $p, q \in \Pi \subset \mathcal{P}$, $D(\Pi\,;\,q) := \min_{p \in \Pi} D(p\,;\,q)$. Consider the set $(\star)$, with

$$\alpha < \max_{a \in \mathcal{X}} g(a)\,.$$

This is an open set of $\mathcal{P}$, whose closure is

$$\text{cl.}\Pi = \{p\,;\, \sum_{a \in \mathcal{X}} g(a) p(a) \geq \alpha\}\,.$$

It satisfies the conditions of Sanov's theorem. By continuity of $p \to D(p\,;\,Q)$,

$$D(\Pi\,;\,Q) = D(\text{cl.}\Pi\,;\,Q) = \min D(p\,;\,Q)\,,$$

where the minimum is over the distributions p such that $\sum_a g(a)p(a) \geq \alpha$. In particular, if $\alpha > \sum_{a \in \mathcal{X}} g(a)Q(a)$, and therefore $Q \notin \text{cl.}\Pi$, we have that $D(\Pi\,;\,Q) > 0$ (use the fact that $D(p\,;\,q) > 0$ when $p \not\equiv q$). Therefore

$$\frac{1}{n} \log P\left(\frac{1}{n} \sum_{i=1}^n g(X_i) > \alpha \right) \longrightarrow 0$$

exponentially fast.

13.2.2 Computation of the Rate of Convergence

This continues Example 13.2.2. Consider the exponential family of distributions Q_t $(t \geq 0)$ associated with Q:

$$Q_t(a) := c(t)Q(a)2^{tg(a)} \text{ where } c(t) := \left(\sum_{a\in\mathcal{X}} 2^{tg(a)}\right)^{-1}.$$

The function $t \to \sum_a Q_t(a)g(a)$ is continuous and $\lim_{t\uparrow\infty} \sum_a Q_t(a)g(a) = \max_a g(a)$. Also $Q_0 = Q$. Therefore, since by assumption,

$$\sum_a Q(a)g(a) < \alpha < \max_a g(a)\,,$$

there exists some $t = t^* > 0$ such that $Q^* := Q_{t^*}$ satisfies

$$Q^*(a) = c^*Q(a)2^{t^*g(a)}\,,\ t^* > 0\,,\ \sum_a Q^*(a)g(a) = \alpha\,, \qquad (\dagger)$$

where $c^* := c(t^*)$. In particular, $Q^* \in \text{cl.}\Pi$. We show that

$$D(\Pi\,;\,Q) = D(Q^*\,;\,Q) = \log c^* + t^*\alpha\,.$$

For the first equality, it suffices to show that $D(P\,;\,Q) > D(Q^*\,;\,Q)$ for all $P \in \Pi$, that is, all P such that $\sum_a P(a)g(a) > \alpha$. Now

$$\begin{aligned} D(Q^*\,;\,Q) &= \sum_a Q^*(a)\log\frac{Q^*(a)}{Q(a)} \\ &= \sum_a Q^*(a)\left(\log c^* + t^*g(a)\right) = \log c^* + t^*\alpha \end{aligned} \qquad (13.5)$$

(by $(\dagger)$) and

$$\sum_a P(a)\log\frac{Q^*(a)}{Q(a)} = \sum_a P(a)\left(\log c^* + t^*g(a)\right) > \log c^* + t^*\alpha\,.$$

This shows that $D(Q^*\,;\,Q) < D(P\,;\,Q)$ and that

$$D(P\,;\,Q) - D(Q^*\,;\,Q) > D(P\,;\,Q) - \sum_a P(a)\log\frac{Q^*(a)}{Q(a)} = D(P\,;\,Q^*) > 0\,.$$

Replacing P in (13.5) by any Q_t, one gets

$$D(Q^*\,;\,Q_t) = \log\frac{c^*}{c} + (t^* - t)\alpha = \log c^* + t^*\alpha - (\log c + t\alpha)\,.$$

This quantity is > 0 if $Q^* \not\equiv Q_t$, which implies that

$$\log c + t\alpha = -\sum_a Q(a)2^{tg(a)} + t\alpha$$

is maximized by $t = t^*$. Therefore the large deviations exponent $D(\Pi\,;\,Q) = D(Q^*\,;\,Q)$ is equal to

$$\max_{t\geq 0}\left(\alpha t - \log\sum_a Q(a)2^{tg(a)}\right) = \max_{t\geq 0}\left(\alpha t - \log M(t)\right)\,,$$

where $M(t) = E\left[2^{tg(X_1)}\right]$.

The Chernoff bound has therefore been recovered. Exercise 13.3.8 shows that in fact the Chernoff bound gives the best exponential rate of convergence to 0 of $P\left(\frac{1}{n}\sum_{i=1}^n g(X_i) > \alpha\right)$.

13.2.3 The Maximum Entropy Principle

We begin with a useful inequality called the Pythagorean theorem for divergence.

Lemma 13.2.3 *Let $\Pi \subset \mathcal{P}$ be a closed convex set of probabilities on $\mathcal{X}$, and let $Q \notin \Pi$. Define*

$$P^* = \mathit{argmax}_{P\in\Pi} D(P\,;\,Q)\,.$$

For all $P \in \Pi$,

$$D(P\,;\,Q) \geq D(P\,;\,P^*) + D(P^*\,;\,Q)\,.$$

In particular, if $\{P_n\}_{n\geq 1}$ is a sequence of probabilities in Π such that $D(P_n\,;\,Q) \to D(P^*\,;\,Q)$, then $D(P_n\,;\,P) \to 0$.

Proof. For any $P \in \Pi$ and $\lambda \in [0,1]$, let

$$P_\lambda := \lambda P + (1-\lambda)P^*\,.$$

Since Π is convex, $P_\lambda \in \Pi$ for all $\lambda \in [0,1]$. Also $\lim_{\lambda\leftrightarrow\infty} P_\lambda = P^*$. As the function $\lambda \to D(P_\lambda\,;,Q)$ is minimized at $\lambda = 0$, the derivative of this function at $\lambda = 0$ must be non-negative. Now,

$$\frac{D(P_\lambda\,;,Q)}{d\lambda} = \sum_a\left((P(a)-P^*(a))\log\frac{P_\lambda(a)}{Q(a)} + (P(a)-P^*(a))\right)\,.$$

Therefore

$$\begin{aligned}
0 \leq \frac{D(P_\lambda\,;,Q)}{d\lambda}|_{\lambda=0} &= \sum_x (P(a)-P^*(a))\log\frac{P^*(a)}{Q(a)} \\
&= \sum_x P(a)\log\frac{P^*(a)}{Q(a)} - \sum_x P^*(a)\log\frac{P^*(a)}{Q(a)} \\
&= \sum_x P(a)\log\left(\frac{P(a)}{Q(a)}\frac{P^*(a)}{P(a)}\right) - \sum_x P^*(a)\log\frac{P(a)}{Q(a)} \\
&= D(P\,;\,Q) - D(P\,;\,P^*) - D(P^*\,;\,Q)\,.
\end{aligned}$$

□

The next result is the maximum entropy principle, also called the conditional limit theorem.

Theorem 13.2.4 *Let Π be a closed convex set of probabilities on $\mathcal{X}$, and let $Q \notin \Pi$, $Q > 0$. Define*

$$P^* = argmax_{P \in \Pi} D(P\,;\,Q)\,.$$

Let $X_1, \ldots, X_n$ be an IID *n-sample of Q. Then*

$$\lim_{n \uparrow \infty} Pr(X_1 = a \,|\, p_{X_1^n} \in \Pi) = P^*(a)\,,$$

where Pr is the probability induced by Q.

Proof. Since $q \to D(p\,;\,q)$ is a strictly convex function of p, the probability P^* such that

$$D^* := D(P^*\,;\,Q) = D(\Pi\,;\,Q) := \min_{p \in \Pi} D(p\,;\,Q)$$

is unique. Also by convexity of $q \to D(p\,;\,q)$, the sets

$$S_t := \{P \in \mathcal{P}\,;\,D(P\,;\,Q) \le t\} \qquad (t \ge 0)$$

are convex. For any $\delta > 0$, define

$$A := S_{D^*+\delta} \cap \Pi \text{ and } B := \Pi - A = \Pi - S_{D^*+\delta} \cap \Pi\,.$$

Then

$$\begin{aligned} Q^n(B) &= \sum_{P \in \Pi \cap \mathcal{P}_n\,;\,D(P\,;Q) > D^*+2\delta} Q^n(T_n(P)) \\ &\le \sum_{P \in \Pi \cap \mathcal{P}_n\,;\,D(P\,;Q) > D^*+2\delta} 2^{-nD(P\,;Q)} \\ &\le \sum_{P \in \Pi \cap \mathcal{P}_n\,;\,D(P\,;Q) > D^*+2\delta} 2^{-n(D^*+\delta)} \\ &\le (n+1)^L 2^{-n(D^*+2\delta)}\,. \end{aligned}$$

Also,

$$\begin{aligned} Q^n(A) &= \sum_{P \in \Pi \cap \mathcal{P}_n\,;\,D(P\,;Q) \le D^*+\delta} Q^n(T_n(P)) \\ &\ge \sum_{P \in \Pi \cap \mathcal{P}_n\,;\,D(P\,;Q) > D^*+\delta} \frac{1}{(n+1)^L} 2^{-nD(P\,;Q)}\,. \end{aligned}$$

For n sufficiently large, there exists at least one type in $A = S_{D^*+\delta} \cap \Pi$, and since the sum in right-hand side of the above chain of inequalities is larger that any of its terms

$$Q^n(A) \ge \frac{1}{(n+1)^L} 2^{-n(D^*+\delta)}\,.$$

In particular, for sufficiently large n,

$$\begin{aligned} Pr(p_{X_1^n} \in B \,|\, p_{X_1^n} \in \Pi) &= \frac{Q^n(B \cap \Pi)}{Q^n(\Pi)} \le \frac{Q^n(B)}{Q^n(A)} \\ &\le \frac{(n+1)^L 2^{-n(D^*+\delta)}}{(n+1)^{-L} 2^{-n(D^*+2\delta)}} = (n+1)^{2L}\, 2^{-n\delta}\,. \end{aligned}$$

Therefore, as $n \uparrow \infty$, the conditional probability of B goes to 0, and consequently, the conditional probability of A goes to 1.

Since for any $P \in A$, $D(P\,;\,Q) \leq D^* + 2\delta$, by the Pythagorean theorem,

$$D(P\,;\,P^*) + D(P^*\,;\,Q) \leq D(P\,;\,Q) \leq D^* + 2\delta = D(P^*\,;\,Q) + 2\delta\,,$$

which implies $D(P\,;\,P^*) \leq \delta$. In particular, $p_{\mathbf{x}} \in A$ implies $D(p_{\mathbf{x}}\,;\,P^*) \leq 2\delta$ and

$$Pr(D(p_{X_1^n}\,;\,P^*) \leq 2\delta \,|\, p_{X_1^n} \in \Pi) \geq Pr(p_{X_1^n} \in A \,|\, p_{X_1^n} \in \Pi)\,.$$

Therefore

$$Pr(D(p_{X_1^n}\,;\,P^*) \leq 2\delta \,|\, p_{X_1^n} \in \Pi) \to 1\,.$$

By Pinsker's theorem, this implies that, conditionally on $p_{X_1^n} \in \Pi$), $d_V(p_{X_1^n}, P^*) \to 0$ in probability. In particular, for any $a \in \mathcal{X}$, for any $\varepsilon > 0$,

$$Pr(|p_{X_1^n}(a) - P^*(a)| \geq \varepsilon \,|\, p_{X_1^n} \in \Pi) \to 1\,. \qquad (\star)$$

Since the sample $X_1, \ldots, X_n$ is IID, for any i $(1 \leq i \leq n)$,

$$Pr(X_i = a \,|\, p_{X_1^n} \in \Pi) = Pr(X_1 = a \,|\, p_{X_1^n} \in \Pi)$$

and therefore $(\star)$ implies that for all $a \in \mathcal{X}$,

$$Pr(X_1 = a \,|\, p_{X_1^n} \in \Pi) \to P^*(a)\,.$$

□

Books for Further Information

The method of types is treated in depth by its promotors in [Csiszár and Körner, 1981]. The compact survey of [Csiszár and Shields, 2004] has applications in statistics. See also the first chapters of [Dembo and Zeitouni, 2010] which, among other features, has applications to large deviations for Markov chains.

13.3 Exercises

Exercise 13.3.1. $D(p_X \,|\, q_X) \geq D(p_Y \,|\, q_Y)$
Let X and Y be two discrete random variables on the probability space $(\Omega, \mathcal{F})$ taking their values in E, and let P and Q be two probabilities on $(\Omega, \mathcal{F})$. Let $(p_X(x)\,,\, x \in E)$ and $(q_X(x)\,,\, x \in E)$ be the probability distributions of X under P and Q respectively. Similarly, let $(p_Y(x)\,,\, x \in E)$ and $(q_Y(x)\,,\, x \in E)$ be the probability distributions of Y under P and Q respectively. Let for all $x, y \in E$ $p_{Y|X}(y|x) := P(Y = y \,|\, X = x)$ and $q_{Y|X}(y|x) := Q(Y = y \,|\, X = x)$. Define in a similar way $p_{X|Y}(x|y) := P(X = x \,|\, Y = y)$ and $q_{Y|X}(y|x) := Q(X = x \,|\, Y = y)$. Assume that $q_{Y|X}(y|x) = p_{Y|X}(y|x) := r(x \,|\, y)$. Prove that

$$D(p_X \,|\, q_X) \geq D(p_Y \,|\, q_Y)$$

Exercise 13.3.2. HMC AND DIVERGENCE, TAKE 1
Let $\{X_n\}_{n\geq 0}$ be a HMC with state space E and let μ_n and μ'_n be the distributions of X_n corresponding to two different initial distributions μ_0 and μ'_0 respectively.

(a) Show that

$$D(\mu_n \,|\, \mu'_n) \geq D(\mu_{n+1} \,|\, \mu'_{n+1})\,.$$

(Use the result of Exercise 13.3.1.)

(b) Suppose that there exists a unique stationary distribution of the chain, denoted by π. Show that

$$D(\mu_n \,|\, \pi) \geq D(\mu_{n+1} \,|\, \pi)\,.$$

(The divergence between the distribution at time n and the stationary distribution decreases with n.)

Exercise 13.3.3. HMC AND DIVERGENCE, TAKE 2
Let $\{X_n\}_{n\geq 0}$ be a positive recurrent HMC with finite state space E and suppose that its stationary distribution π is the uniform distribution on E.

(a) Show that the entropy $H(X_n)$ increases with n.

(b) Give a counterexample if the stationary distribution is not uniform.

(c) Show that whatever the initial distribution, the conditional entropy $H(X_n \,|\, X_0)$ increases with n for a stationary HMC.

Exercise 13.3.4. LOG-SUM INEQUALITY
Let $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ be real non-negative numbers, and let $a := \sum_{i=1}^k a_i$ and $b := \sum_{i=1}^k b_i$. Then

$$\sum_{i=1}^{k} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b}\,,$$

with equality if and only if $a_i = cb_i$ for some constant c for all i $(1 \leq i \leq k)$.

Exercise 13.3.5. CONVEXITY OF DIVERGENCE
Prove the following:

(a) The function $D : (p, q) \to D(p\,;\, q)$ is convex in both its arguments, that is, for all $p_1, q_1, p_2, q_2 \in \mathcal{P}$, and all $\lambda \in [0, 1]$:

$$D(\lambda p_1 + (1-\lambda)p_2\,;\, \lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1\,;\, q_1) + (1-\lambda)D(p_2\,;\, q_2)\,.$$

(b) If q is strictly positive, the function $D : p \to D(p\,;\, q)$ is continuous.

Exercise 13.3.6. PARALLELOGRAM IDENTITY FOR I-DIVERGENCE
Recall the parallelogram identity where $||x - y||$ is the euclidean distance between two vectors x and y in $\mathbb{R}^m$:

$$||x - z||^2 + ||y - z||^2 = 2||\frac{1}{2}(x + y) - z||^2 + ||x - \frac{1}{2}(x + y)||^2 + ||y - \frac{1}{2}(x + y)||^2\,.$$

Prove the "analogous" identity, called the parallelogram identity for I-divergence:

$$D(p\,;\,r) + D(q\,;\,r) = 2D(\frac{1}{2}(p+q)\,;\,r) + D(p\,;\,\frac{1}{2}(p+q)) + D(q\,;\,\frac{1}{2}(p+q))\,.$$

Exercise 13.3.7. Approximate Huffman codes
You have devised a Huffman code corresponding to a long (say length 1000) IID sequence of equiprobable 0s and 1s. Find a window $[\frac{1}{2}+a, \frac{1}{2}+a]$ such that if this code is used for an IID sequence of equiprobable 0s and 1s, of the same length but with bias $p \in [\frac{1}{2}+a, \frac{1}{2}+a]$, the average length of the encoded sequence is within less than a given quantity α from the average length of a Huffman code adapted to the bias p.

Exercise 13.3.8. Chernoff gives the best exponential rate
The setting is as in Example 13.2.2. Show that for any $\varepsilon > 0$, for n sufficiently large,

$$Q^n(\Pi \cap \mathcal{P}_n) \leq 2^{-n(D(P^*\,;Q)-\varepsilon)}\,. \qquad (\star)\,.$$

Quantify "n sufficiently large" (that is, find $n_0 = n_0(\varepsilon)$ such that $(\star)$ holds for all $n \geq n_0$). Prove that $D(P^*\,;\,Q)$ is the largest constant γ such that

$$Q^n(\Pi \cap \mathcal{P}_n) \leq 2^{-n(\gamma-\varepsilon)}\,.$$

Chapter 14

Universal Source Coding

14.1 Type Encoding

14.1.1 A First Example

The source compression codes of Huffman and Shannon–Fano–Elias are adapted to specific statistics of the source. If used in a different statistical environment, they loose their optimality (see Example 13.1.1). One is therefore led to investigate the existence of codes that are less, and hopefully not, sensitive to the source statistics. These codes are called universal. The following example gives an idea of what can be expected.

Let $x_1^n := (x_1, \ldots, x_n) \in \{0,1\}^n$ be a binary sequence. It will be encoded as (k, β) where $k = k(x_1^n) := \sum_{i=1}^n x_i$ is the number of 1's in the sequence, and $\beta = \beta(x_1^n, k)$ is the lexicographical rank of the sequence among all sequences of $\{0,1\}^n$ with k ones. The number k can be encoded with $\lceil \log(n+1) \rceil$ bits, and since there are $\binom{n}{k}$ binary sequences of length n with k ones, encoding the sequence requires a total length

$$\ell(x_1^n) \le \log(n+1) + \log \binom{n}{k} + 2\,.$$

The following bounds for the binomial coefficients will be used (Exercise 14.3.1). For $p \in (0,1)$ and $n \in \mathbb{N}$ such that np is a positive integer,

$$\frac{1}{\sqrt{8np(1-p)}} \le \binom{n}{np} 2^{-nh_2(p)} \le \frac{1}{\sqrt{\pi np(1-p)}}\,.$$

In particular,

$$\log \binom{n}{np} \le nh_2(p) - \frac{1}{2}\log n - \frac{1}{2}\log\left(\pi p(1-p)\right)\,.$$

Applying this bound with $p = \frac{k}{n}$, we obtain

$$\ell(x_1^n) \le \log n + nh_2\left(\frac{k}{n}\right) - \frac{1}{2}\log n - \frac{1}{2}\log\left(\pi \frac{k}{n}\frac{n-k}{n}\right) + 3$$
$$\frac{1}{2}\log n + nh_2\left(\frac{k}{n}\right) - \frac{1}{2}\log\left(\pi \frac{k}{n}\frac{n-k}{n}\right) + 3\,.$$

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_14

Let $X_1^n := (X_1, \ldots, X_n) \in \{0,1\}^n$ be a random binary sequence such that $\lim_{n\uparrow\infty} \frac{1}{n}(X_1 + \cdots + X_n) = p$. By the preceding inequality,

$$\lim_{n\uparrow\infty} \frac{1}{n} \ell(X_1^n) = h_2(p)\,.$$

In the case where the asymptotic rate of 1's in the sequence is p, this type encoding guarantees a compression rate $h_2(p)$. In particular, the compression is optimal for a Bernoulli sequence.

14.1.2 Source Coding via Typical Sequences

We now prove the existence of a universal code with given guaranteed rate via the method of types. Let $X_1^n := (X_1, \ldots, X_n)$ be an IID sequence of random variables taking their values in the finite set $\mathcal{X}$ and with common probability distribution Q. A code of rate R consists of a sequence of encoders

$$E_n : \mathcal{X}^n \to \{1, 2, \ldots, \lceil 2^{nR} \rceil\} \qquad (n \geq 1)$$

and of a sequence of decoders

$$D_n : \{1, 2, \ldots, \lceil 2^{nR} \rceil\} \to \mathcal{X}^n \qquad (n \geq 1)\,.$$

(In the sequel, for notational convenience, we shall replace $\lceil 2^{nR} \rceil$ by 2^{nR}, leaving the fine tuning to the reader.) The sequence X_1^n is encoded as $E_n(X_1^n)$ and the compressed sequence is restituted as $D_n(E_n(X_1^n))$. The probability of error

$$P_{e,n} := Q^n(x_1^n\,;\, D_n(E_n(x_1^n) \neq x_1^n)$$

depends in general on the input distribution Q. A code is called a universal code of rate R if for any input distribution Q such that $H(Q) \leq R$, $\lim_{n\uparrow\infty} P_{e,n} = 0$.

The following result is an extension of the method of coding based on the notion of a typical sequence (Theorem 11.1.11).

Theorem 14.1.1 *There exists a universal code of rate R for all $R \geq 0$.*

Proof. (The notation is that of Chapter 13.) Let $A_n := \{\mathbf{x} \in \mathcal{X}^n\,;\, H(p_{\mathbf{x}}) \leq R_n\}$, where $R_n := R - |\mathcal{X}| \frac{\log(n+1)}{n}$. This set has at most 2^{nR} elements. Indeed, by Lemma 13.1.6,

$$\begin{aligned} |A_n| &= \sum_{p \in \mathcal{P}_n\,;\, H(p) \leq R_n} |T_n(p)| \leq \sum_{p \in \mathcal{P}_n\,;\, H(p) \leq R_n} 2^{nH(p)} \\ &\leq \sum_{p \in \mathcal{P}_n\,;\, H(p) \leq R_n} 2^{nR_n} \leq |\mathcal{P}_n| 2^{nR_n} \\ &\leq (n+1)^{|\mathcal{X}|} 2^{nR_n} = 2^{n(R_n + |\mathcal{X}| \frac{\log(n+1)}{n})} = 2^{nR}\,. \end{aligned}$$

Since $|A_n| \le 2^{nR}$ there exists a bijection f_n of A_n into a subset of $\{1, 2, \dots, 2^{nR}\}$. Define $E_n(x_1^n) = f_n(x_1^n)$ if $x_1^n \in A_n$, arbitrarily otherwise. The decoder will associate to $E_n(x_1^n)$ the correct sequence if $x_1^n \in A_n$. Therefore, if the input distribution is Q,

$$\begin{aligned} P_{e,n} = 1 - Q^n(A_n) &= \sum_{p \in \mathcal{P}_n \,;\, H(p) > R_n} Q^n(T_n(p)) \\ &\le (n+1)^{|\mathcal{X}|} \max_{p \in \mathcal{P}_n \,;\, H(p) > R_n} Q^n(T_n(p)) \\ &\le (n+1)^{|\mathcal{X}|} 2^{-n \min_{p \in \mathcal{P}_n \,;\, H(p) > R_n} D(p\,;\,Q)} . \end{aligned}$$

In particular, if

$$R_n > H(Q)\,, \tag{$\star$}$$

then, $\min_{p \in \mathcal{P}_n \,;\, H(p) > R_n} D(p\,;\,Q) \ge \min_{p \in \mathcal{P}_n \,;\, H(p) > H(Q)} D(p\,;\,Q)$ and therefore

$$P_{e,n} \le (n+1)^{|\mathcal{X}|} 2^{-n\gamma} \le 2^{-n(\gamma - |\mathcal{X}| \frac{\log(n+1)}{n})}$$

where $\gamma := \min_{p \in \mathcal{P}_n \,;\, H(p) > H(Q)} D(p\,;\,Q)$ is positive. As $R_n \uparrow R$ and $H(Q) < R$, ($\star$) holds for sufficiently large n. □

This method of encoding is not practical because it requires a list a codewords and resource consuming operations of encoding and decoding.

14.2 The Lempel–Ziv Algorithm

14.2.1 Description

(Ziv and Lempel, 1978) This kind of coding is of a nature quite different from the classical ones (Huffman, Shannon–Fano–Elias, Tunstall) in that it does not require knowledge of the statistics of the source, both for encoding and decoding. It is also different from the universal code of Subsection 14.1.2 in that it does not depend on a code, universal or not.

It transforms an infinite (input) sequence

$$x = x_1 x_2 x_3 \cdots$$

of symbols from an alphabet $\mathcal{A}$ of size D into an infinite (output) sequence of binary digits. The following example shows how this algorithm works.

Example 14.2.1: Coding aaabbc. The alphabet has three symbols: $\mathcal{A} = \{a, b, c\}$. Suppose that the initial segment of the sequence x is *aaabbc*. The algorithm features an evolutive dictionary $\mathcal{D}$ consisting of finite sequences of symbols (the dictionary words) from the alphabet $\mathcal{A}$. Each dictionary word of the current dictionary at a given step of the encoding process is encoded into a fixed-length binary sequence. Denoting by M the current size of $\mathcal{D}$, the length of the dictionary words is the minimal one, namely $\lceil \log_2 M \rceil$. The initial dictionary is

$$\mathcal{D}_1 := \{a, b, c\}\,.$$

The dictionary word-length is therefore 2 and its words are (minimally) encoded in the lexicographic order:

$$a \to 00,\ b \to 01,\ c \to 10\,.$$

We scan the input sequence for the first word in dictionary $\mathcal{D}_1$, a in the example. We therefore encode a as 00. The encoding process is now at stage

$$00, aabbc\,.$$

We then update $\mathcal{D}_1$, replacing the just recognized word (here a single letter) a by all the possible extensions: aa, ab, ac (in lexicographic order). The new dictionary is therefore

$$\mathcal{D}_2 := \{aa,\ ab,\ ac,\ b, c\}\,.$$

Its dictionary words are (minimally) encoded in lexicographic order:

$$aa \to 000,\ ab \to 001,\ ac \to 010,\ b \to 011,\ c \to 100\,.$$

We then look for the next dictionary word first encountered in the remaining sequence $aabbc$ (the not yet encoded portion of the initial sequence). We find aa, which is then encoded as 000. The encoding process is now at stage

$$00, 000, bbc\,.$$

We update $\mathcal{D}_2$ by replacing the just recognized word aa by all its possible one-letter extensions aaa, aab, aac (in lexicographic order):

$$\mathcal{D}_3 := \{aaa,\ aab,\ aac,\ ab,\ ac,\ b,\ c\},$$

which is (minimally and in lexicographic order) encoded as

$$000, 001, 010, 011, 100, 101, 110\,.$$

We then look in the remaining (not yet encoded) sequence bbc for the first word not in the dictionary. We find b which is then encoded into 101. The encoding process is now at stage

$$00, 000, 101, bc\,.$$

The next dictionary is then

$$\mathcal{D}_4 := \{aaa, aab, aac, ab, ac, ba, bb, bc, c\}$$

encoded as

$$0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000\,.$$

The remaining sequence bc turns out to be a dictionary word and it is therefore encoded into 0111. Finally, the original sequence has been encoded as

$$00, 000, 101, 0111\,.$$

14.2.2 Parsings

The Lempel–Ziv algorithm involves slicing the initial sequence into distinct strings

$$a, aa, b, bc\,. \tag{$\star$}$$

Note that all the phrases in the parsing ($\star$) are distinct, a property which is inherent to the Lempel–Ziv algorithm. A distinctive parsing is a parsing consisting of distinct phrases.

Let $\mathcal{P}(x_1^n)$ be the collection of all distinctive parsings of x_1^n and call $c(\pi(x_1^n))$ the number of phrases in the distinctive parsing $\pi(x_1^n)$, including the empty string $\varnothing$. Define

$$c(x_1^n) := \inf_{\pi(x_1^n)\in\mathcal{P}(x_1^n)} c(\pi(x_1^n))\,.$$

EXAMPLE 14.2.2: The list of all distinctive parsings of the sequence $x_1^4 = aabb$ is

$$(\varnothing, a, abb),\ (\varnothing, aab, b),\ (\varnothing, aa, bb),\ (\varnothing, a, ab, b)\,.$$

Therefore, in this example, $c(x_1^4) = 3$.

Lemma 14.2.3 *For any sequence $x = x_1x_2x_3\cdots$ of symbols in $\mathcal{A}$,*

$$n \geq c(x_1^n)\log_D \frac{c(x_1^n)}{D^3}\,, \tag{14.1}$$

and

$$c(x_1^n) = O\left(\frac{n}{\log n}\right)\,. \tag{14.2}$$

Proof. Let $c := c(x_1^n)$. This number can be written in a unique way as

$$c = \sum_{j=0}^{m-1} D^j + r, \quad 0 \leq r < D^m\,.$$

The length n of x_1^n is certainly $\geq$ than the total length of the c shortest distinct strings. Since there are D^j distinct strings of length j,

$$n \geq \sum_{j=0}^{m-1} jD^j + mr\,. \tag{$\dagger$}$$

Now

$$\sum_{j=0}^{m-1} D^j = \frac{D^m-1}{D-1} \text{ and } \sum_{j=0}^{m-1} jD^j = m\frac{D^m}{D-1} - \frac{D}{D-1}\frac{D^m-1}{D-1}\,,$$

and in particular

$$c - r = \sum_{j=0}^{m-1} D^j = \frac{D^m}{D-1} - \frac{1}{D-1}\,. \tag{$\dagger\dagger$}$$

Therefore, from (†) and (††),

$$\begin{aligned} n &\geq m\frac{D^m}{D-1} - \frac{D}{D-1}\frac{D^m-1}{D-1} + mr \\ &= m\left(c - r + \frac{1}{D-1}\right) - \frac{D}{D-1}(c-r) + mr \\ &= m\left(c + \frac{1}{D-1}\right) - \frac{D}{D-1}c + \frac{D}{D-1}r \\ &\geq mc - \frac{D}{D-1}c \geq (m-2)c\,, \end{aligned}$$

that is

$$n \geq (m-2)c\,. \qquad (\dagger\,\dagger\,\dagger)$$

On the other hand,

$$c = \sum_{j=0}^{m-1} D^j + r \leq \sum_{j=0}^{m} D^j = \frac{D^{m+1}-1}{D-1}\,,$$

so that

$$D^{m+1} \geq c(D-1) + 1 > c\,,$$

which in turn implies $\frac{c}{D^3} < D^{m-2}$ and then $m - 2 > \log_D \frac{c}{D^3}$. Therefore, from († † †),

$$n > c\log_D \frac{c}{D^3}$$

that is, (14.1).

For the proof of (14.2), define $\nu := \frac{n}{D^3}$ and $\gamma := \frac{c}{D^3}$. In particular

$$\nu > \gamma \log_D \gamma\,. \qquad (\star)$$

We must show that

$$\gamma = O\left(\frac{\nu}{\log_D \nu}\right)\,.$$

Since the result to be proved is asymptotic, we may assume that ν is large. Since $\sqrt{\nu} = o\left(\frac{\nu}{\log_D \nu}\right)$, we may suppose that $\sqrt{\nu} \leq \frac{2\nu}{\log_D \nu}$. Only two cases need to be checked:

1. $\gamma < \sqrt{\nu}$. Then, by the above assumption, $\gamma \leq \frac{2\nu}{\log_D \nu}$.

2. $\gamma \geq \sqrt{\nu}$. Then, by $(\star)$, $\gamma < \frac{\nu}{\log_D \gamma} < \frac{\nu}{\log_D \sqrt{\nu}} = \frac{2\nu}{\log_D \nu}$.

□

14.2.3 Optimality of the Lempel–Ziv Algorithm

The Lempel–Ziv algorithm has, besides the fact that its implementation does not require knowledge of the statistics of the input sequence, an interesting property: it behaves at least as well, from the compression point of view, as any information lossless finite encoder [1], a notion that we proceed to introduce.

A *finite encoder* consists of

- a finite state space Σ with s elements,
- an input alphabet $\mathcal{A}$ with $D \geq 2$ elements,
- a binary output alphabet: $\{0, 1\}$,
- an output function $f : \Sigma \times \mathcal{A} \to \{0, 1\}^*$, and
- a state transition function $g : \Sigma \times \mathcal{A} \to \Sigma$.

A finite encoder is also called in a more general framework a finite automaton. It "reads" successively the input symbols $x_1, x_2, x_3, ... \in \mathcal{A}$, and "writes" the output binary strings $y_1, y_2, y_3, y_4, ... \in \{0, 1\}^*$. Precisely: when it reads x_k while in state z_k, it writes the binary string $y_k = f(z_k, x_k)$ and moves to state $z_{k+1} = g(z_k, x_k)$. The following condensed notations summarize the above sequence of operations:

- $x_k x_{k+1} ... x_j = x_k^j$
- $f(z_k, x_k^j) = y_k^j$ (starting in state z_k and reading the input sequence x_j^k, the encoder outputs the sequence y_k^j)
- $g(z_k, x_k^j) = z_{j+1}$

The encoder is said to be *uniquely decodable* (UD) if and only if the application $f : \mathcal{A}^k \to \{0, 1\}^*$ is injective. It is called *information lossless* (IL) if and only if for all states z_k and for any distinct input sequence $x_k^j \neq \tilde{x}_k^j$,

$$f(z_k, x_k^j) = f(z_k, \tilde{x}_k^j) \Rightarrow g(z_k, x_k^j) \neq g(z_k, \tilde{x}_k^j)\,. \tag{14.3}$$

This is the minimum requirement that a decoder must satisfy, otherwise one could not distiguish x_k^j from $\tilde{x}_k^j$ given the information provided by the y_k^j's and the z_k^j's. Clearly, UD implies IL, but the converse may not be true (Exercise 14.3.7 features a counter-example).

Let E be an IL encoder. One defines the *compression ratio* for the n first input symbols by

$$\rho_E(x_1^n) := \frac{1}{n} L(y_n^1)\,,$$

where $L(y_1^n)$ is the length of the output binary sequence y_1^n. Let

[1] The proof below is borrowed from the lecture notes in information theory of Emre Telatar, IC school, EPFL.

$$\rho_s(x_1^n) := \inf_{E,|\Sigma|=s} \rho_E(x_1^n)$$

be the lower bound of this compression ratio over the encoders with s states, and define

$$\rho_s(x) := \limsup_{n\to\infty} \rho_s(x_1^n)\,.$$

Since there exists an IL encoder transforming each symbol x_k into a binary sequence of length at most $\lceil \log_2 D \rceil$, $\rho_s(x_1^n) \le \lceil \log_2 D \rceil$, and therefore $\rho_s(x) \le \lceil \log_2 D \rceil$. Clearly, $s \to \rho_s(x)$ is a non-increasing function of s (one can only improve compression with an encoder with more states). The following limit therefore exists:

$$\lim \rho_s(x) := \rho(x)\,.$$

We now proceed to the proof that the compression ratio of the Lempel–Ziv algorithm is not larger than that of any IL finite encoder.

Lemma 14.2.4 *For any* IL *encoder with s states encoding binary sequences,*

$$L(y_1^n) \ge c(x_1^n) \log_2 \left(\frac{c(x_1^n)}{8s^2} \right)\,.$$

Proof. Let x_1^n be parsed in $c = c(x_1^n)$ distinct phrases:

$$x_1^n = w_1, w_2, \ldots, w_c\,.$$

Among the w_k's, look at those for which the encoder has started in state i and finished in state j. Let c_{ij} be their number. The encoder being IL, the corresponding output strings must be distinct (by (14.3), and therefore their total length L_{ij} satisfies (14.1) with $D = 2$:

$$L_{ij} \ge c_{ij} \log_2 \left(\frac{c_{ij}}{8} \right)\,.$$

Therefore, denoting the states by $1, 2, \ldots, s$,

$$L(y_1^n) = \sum_{1\le i,j\le s} L_{ij} \ge \sum_{1\le i,j\le s} c_{ij} \log_2 \left(\frac{c_{ij}}{8} \right)\,.$$

The minimum of

$$\sum_{1\le i,j\le s} c_{ij} \log_2 \left(\frac{c_{ij}}{8} \right)$$

as a function of the c_{ij}'s, under the constraint $\sum_{i,j} c_{ij} = c := c(x_1^n))$, is attained for $c_{ij} = \frac{c}{s^2}$ for all (i,j). (In fact, under the constraint, the problem is equivalent to that of maximizing the entropy of the probability distribution $\{c_{ij}/c\,;\,(i,j) \in \Sigma^2\}$, and this is done by the uniform distribution.) Therefore

$$L(y_1^n) \ge \sum_{1\le i,j\le s} \frac{c}{s^2} \log_2 \left(\frac{c}{8s^2} \right) = c \log_2 \left(\frac{c}{8s^2} \right)\,.$$

□

Lemma 14.2.5

$$\rho(x) \geq \limsup \frac{1}{n} c(x_1^n) \log_2 c(x_1^n) \,. \tag{14.4}$$

Proof. According to Lemma 14.2.4, $L(y_1^n) \geq c(x_1^n) \log_2 \left(\frac{c(x_1^n)}{8s^2} \right)$, and therefore

$$\begin{aligned}
\rho_s(x) &\geq \limsup \frac{1}{n} c(x_1^n) \log_2 \left(\frac{c(x_1^n)}{8s^2} \right) \\
&\geq \limsup \left(\frac{1}{n} c(x_1^n) \log_2 (c(x_1^n)) - \underbrace{\frac{1}{n} c(x_1^n) \log_2 (8s^2)}_{\to 0} \right) \\
&\geq \limsup \frac{1}{n} c(x_1^n) \log_2 (c(x_1^n)) \,.
\end{aligned}$$

The announced result then follows from the fact that the right-hand side does not depend on the number of states s. □

Lemma 14.2.6

$$L_{LZ}(y_1^n) \leq c(x_1^n) \log_2 \left(c(x_1^n) \times 2D \right) \,. \tag{14.5}$$

Proof. Suppose the Lempel–Ziv algorithm has been applied to x_1^n, with the resulting distinctive parsing of the input sequence

$$x_1^n = \varnothing w_1 w_2 ... w_{c_{LZ}-1} w_{c_{LZ}} \,.$$

The first $c_{LZ} - 1$ words are by construction distinct, but nothing is known of the last word because we may be at stage n in the process of obtaining a new phrase. Nevertheless, by concatenating $w_{c_{LZ}-1}$ and $w_{c_{LZ}}$, we obtain a phrase not in $\{w_1, w_2, ... w_{c_{LZ}-2}\}$ and therefore have a parsing of x_1^n in c_{LZ} distinct phrases. Therefore, taking into account the empty string $\varnothing$,

$$c_{LZ}(x_1^n) \leq c(x_1^n) \,.$$

The size of the dictionary increases by $D - 1$ at each new phrase. Therefore, at the end of the Lempel–Ziv parsing of x_1^n, it is

$$1 + (D-1)\, c_{LZ}(x_1^n) \leq 1 + (D-1)\, c(x_1^n) \leq D c(x_1^n) \,.$$

Therefore, the words of the last dictionary have been encoded with no more than $\lceil \log_2 (D c(x_1^n)) \rceil$ binary digits. Since any phrase of the Lempel–Ziv parsing has been encoded with a smaller dictionary,

$$\begin{aligned}
L_{LZ}(y_1^n) &\leq c_{LZ}(x_1^n) \lceil \log_2 c_{LZ}(x_1^n) D \rceil \\
&\leq c_{LZ}(x_1^n) \log_2 \left(c_{LZ}(x_1^n) D \right) + 1 \\
&\leq c_{LZ}(x_1^n) \log_2 \left(c_{LZ}(x_1^n) 2D \right) \\
&\leq c(x_1^n) \log_2 \left(c(x_1^n) 2D \right) \,.
\end{aligned}$$

□

Theorem 14.2.7

$$\limsup \frac{1}{n} L_{LZ}(y_1^n) \leq \rho(x)\,.$$

Proof. From (14.5),

$$\begin{aligned}\limsup_n \frac{1}{n} L_{LZ}(y_1^n) &\leq \limsup_n \frac{1}{n} c(x_1^n) \log_2 \left(c(x_1^n) \times 2D\right) \\ &\leq \limsup_n \frac{1}{n} c(x_1^n) \left(\log_2(c(x_1^n) + \log_2(2D)\right) \\ &= \limsup_n \frac{1}{n} c(x_1^n) \log_2 c(x_1^n)\,.\end{aligned}$$

The result then follows from (14.4). □

Therefore, the Lempel–Ziv algorithm performs at least as well as any IL finite encoder such as Huffman's algorithm, which is indeed an IL finite encoder (Exercise 14.3.6).

14.2.4 Lempel–Ziv Measures Entropy

We now take another look at Lempel–Ziv's algorithm. Consider *binary* sequences without loss of generality for the method. The encoding is slightly different, but this modification facilitates the analysis while preserving the essential ideas. Suppose that the binary sequence x_1^n is parsed into $c = c(n)$ distinct phrases $y_1, \ldots, y_c$. A phrase y_i is necessarily of the form $y_j 0$ (resp., $y_j 1$) for some $j < i$. We shall encode it by $u_j 0$ (resp., $u_j 1$) where u_j is the binary expression of length $\lceil \log c \rceil$ of j.

EXAMPLE 14.2.8: The sequence 1100010111010001100000001 is parsed in

$$1, 10, 0, 01, 011, 101, 00, 0110, 000, 0001\,.$$

Here $c = 9$ and therefore the binary encoding of the z_j's require $\lceil \log 9 \rceil = 4$ bits. The encoding of the sequence is

$$(0000, 1)(0001, 0)(0000, 0)(0011, 1)(0010, 1)(0010, 1)(0011, 0)(0101, 0)(0111, 0)(1001, 1)$$

The length of the coded sequence is therefore

$$\ell(x_1^n) = c(n) \left(\lceil \log c(n) \rceil + 1\right)\,. \tag{14.6}$$

Let $\{X_n\}_{n \in \mathbb{Z}}$ be a sequence of random variables with values in the set $\mathcal{X} = \{0, 1\}$. Suppose that under probability P, the stochastic process $\{(X_{n+1}, \ldots, X_{n+k}\}_{n \geq 0}$ is a stationary ergodic HMC with state space $E := \{0, 1\}^k$. Let for $x_i \in \{0, 1\}$ $(1 \leq i \leq n)$

$$p(x_1, \ldots, x_n) := P(X_1 = x_1, \ldots, X_n = x_n)\,.$$

(This is a simplified notation for $p_{1,2,\dots,n}(x_1,\dots,x_n)$, the ambiguity of which disappears in view of the context.) In particular,

$$P(X_1^n = x_1^n \,|\, X_{-(k-1)}^0 = x_{-(k-1)}^0) = \prod_{j=1}^n p(x_j \,|\, x_{j-k}^{j-1})\,. \tag{14.7}$$

The left-hand side of (14.7) will be denoted by $p(x_1,\dots,x_n \,|\, x_{-(k-1)}^0)$. By ergodicity

$$\begin{aligned}\frac{1}{n}\log p(X_1,\dots,X_n \,|\, X_{-(k-1)}^0) &= \frac{1}{n}\sum_{j=1}^n \log p(X_j \,|\, X_{j-k}^{j-1}) \\ &\longrightarrow E\left[\log p(X_1 \,|\, X_{-(k-1)}^0)\right] = H(X_1 \,|\, X_{-(k-1)}^0)\,.\end{aligned} \tag{14.8}$$

Recall that x_1^n is parsed into c distinct phrases $y_1,\dots,y_c$. Let ν_i be the index starting y_i, that is $y_i = x_{\nu_i}^{\nu_{i+1}-1}$. Let $s_i = x_{\nu_i-k}^{\nu_i-1}$ be for $i = 1,2,\dots,c$ the k-bit chunk preceding y_i. In particular $s_1 = x_{-(k-1)}^0$.

Let $c_{\ell,s}$ be the number of phrases y_i with length ℓ and preceding k-bit chunk $s_i = s$. In particular

$$\sum_{\ell,s} c_{\ell,s} = c \tag{14.9}$$

and

$$\sum_{\ell,s} \ell c_{\ell,s} = n\,. \tag{14.10}$$

Lemma 14.2.9 *For any distinctive parsing of $x_1,\dots,x_n$,*

$$\log p(x_1,\dots,x_n \,|\, s_1) \le \sum_{\ell,s} c_{\ell,s} \log c_{\ell,s}\,. \tag{14.11}$$

This is Ziv's inequality, valid for any parsing, not only for the Lempel–Ziv parsing. Note also that the right-hand side does not depend on the probability P.

Proof. From $p_k(x_1,\dots,x_n \,|\, s_1) = \prod_{i=1}^c p(y_i \,|\, s_i)$, the concavity of the logarithm and Jensen's inequality,

$$\begin{aligned}
\log p_k(x_1,\ldots,x_n \mid s_1) &= \sum_{i=1}^{c} \log p(y_i \mid s_i) \\
&= \sum_{\ell,s} \sum_{i\,;\,|y_i|=\ell, s_i=s} \log p(y_i \mid s_i) \\
&= \sum_{\ell,s} c_{\ell,s} \sum_{i\,;\,|y_i|=\ell, s_i=s} \frac{1}{c_{\ell,s}} \log p(y_i \mid s_i) \\
&\le \sum_{\ell,s} c_{\ell,s} \log \left(\sum_{i\,;\,|y_i|=\ell, s_i=s} \frac{1}{c_{\ell,s}} p(y_i \mid s_i) \right) \\
&= \sum_{\ell,s} c_{\ell,s} \log \left(\frac{1}{c_{\ell,s}} \sum_{i\,;\,|y_i|=\ell, s_i=s} p(y_i \mid s_i) \right) .
\end{aligned}$$

Finally, since the y_i's are distinct

$$\sum_{i\,;\,|y_i|=\ell, s_i=s} p(y_i \mid s_i) \le 1 .$$

□

Theorem 14.2.10

$$\limsup_{n\uparrow\infty} \frac{c(n) \log c(n)}{n} \le H .$$

Proof. Let $\pi_{\ell,s} := \frac{c_{\ell,s}}{c}$. Then

$$\sum_{\ell,s} \pi_{\ell,s} = 1 \text{ and } \sum_{\ell,s} \ell \pi_{\ell,s} = \frac{n}{c} .$$

Define the random variables Y and Z by

$$P(Y = \ell, Z = s) = \pi_{\ell,s} .$$

Rewrite Ziv's inequality as

$$\log p(x_1,\ldots,x_n \mid s_1) \le -c \log c - \sum_{\ell,s} \pi_{\ell,s} \log \pi_{\ell,s} = -c \log c + H(Y,Z) .$$

Therefore

$$-\frac{1}{n} \log p(x_1,\ldots,x_n \mid s_1) \ge \frac{c}{n} \log c - \frac{c}{n} H(Y,Z) .$$

Now

$$H(Y,Z) \le H(Y) + H(Z)$$

and, since Z takes its values in a set of size 2^k, $H(Z) \le k$. Also, by Lemma 11.1.7,

$$\begin{aligned} H(Y) &\le (EX+1)\log(EX+1) - EX\log EX \\ &= \left(\frac{n}{c}+1\right)\log\left(\frac{n}{c}+1\right) - \frac{n}{c}\log\frac{n}{c} \\ &= \log\frac{n}{c} + \left(1+\frac{n}{c}\right)\log\left(1+\frac{c}{n}\right). \end{aligned}$$

Therefore,

$$\frac{c}{n}H(Y,Z) \le \frac{c}{n}k + \frac{c}{n}\log\frac{n}{c} + o(1).$$

Since $c \le \frac{n}{\log n}(1+o(1))$ (Lemma 14.2.3) is eventually less that $\frac{1}{e}$, and since the maximum of $\frac{c}{n}\log\frac{n}{c}$ is attained for the maximum value of c for $\frac{c}{n} \le \frac{1}{e}$,

$$\frac{c}{n}\log\frac{n}{c} \le \left(\frac{\log\log n}{\log n}\right)$$

and therefore $H(Y,Z) \to 0$ as $n \to \infty$. Finally,

$$\frac{c}{n}\log\frac{n}{c} \le -\frac{1}{n}\log p(x_1,\dots,x_n \,|\, s_1) + \varepsilon_k(n),$$

where $\lim_{n\uparrow\infty}\varepsilon_k(n) = 0$. Therefore, almost surely

$$\limsup_{n\uparrow\infty}\frac{c(n)\log c(n)}{n} \le \lim_{n\uparrow\infty} -\frac{1}{n}\log p(X_1,\dots,X_n \,|\, s_1) = H.$$

□

Theorem 14.2.11 *The length $\ell(X_1,\dots,X_n)$ of the encoded sequence satisfies*

$$\limsup_{n\uparrow\infty}\frac{1}{n}\ell(X_1,\dots,X_n) \le H.$$

Proof. Recalling (14.6)

$$\limsup_{n\uparrow\infty}\frac{1}{n}\ell(X_1,\dots,X_n) = \limsup_{n\uparrow\infty}\left(\frac{c(n)\lceil\log c(n)\rceil}{n} + \frac{c(n)}{n}\right).$$

The result follows by Theorem 14.2.10 and $\lim\frac{c(n)}{n} = 0$. □

The extension to ergodic sources is done in [Cover and Thomas, 2006]. However, the above result gives an upper bound. Heuristic arguments show that this is also a lower bound, and therefore, in this sense, the Lempel–Ziv algorithm "measures entropy".

Books for Further Information

[Cover and Thomas, 2006], [Csiszár and Shields, 2004].

14.3 Exercises

Exercise 14.3.1. Improved bounds for the binomial coefficients
For $p \in (0,1)$ and $n \in \mathbb{N}$ such that np is a positive integer,

$$\frac{1}{\sqrt{8np(1-p)}} \leq \binom{n}{np} 2^{-nh_2(p)} \leq \frac{1}{\sqrt{\pi np(1-p)}}\,.$$

Exercise 14.3.2. Run-length coding
Consider a sequence of 0's and 1's, for instance

$$00011000001111111011000\,.$$

It is encoded into a sequence of integers equal to the successive lengths of the segments of 0's and 1's. In the example

$$3257123\,.$$

(Add a symbol A or B in front to inform the receiving party that the sequence started with a 0 or a 1.) Applying the best compression algorithm to the encoded sequence, what is the overall compression ratio in the case where the input is a Bernoulli sequence of parameter $p \in (0,1)$?

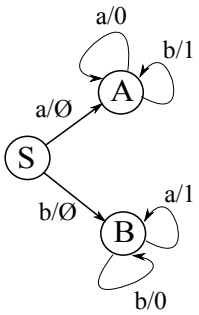
Exercise 14.3.3. Lempel–Ziv encoding
Perform the Lempel–Ziv encoding of the sequence *acabbdddaabb*.

Exercise 14.3.4. Lempel–Ziv decoding
Perform the decoding of the sequence 000001100111 obtained by Lempel–Ziv decoding of a sequence written with alphabet $\mathcal{A} = \{a, b, c\}$. (Give the details, in particular the successive dictionaries used.)

Exercise 14.3.5. Is this a Lempel–Ziv output?
Can any binary string be the output of the Ziv-Lempel encoding of an input sequence with symbols in alphabet, say, $\mathcal{A} = \{a, b, c\}$?

Exercise 14.3.6. Huffman is IL
Show that the Huffman algorithm can be implemented by an information lossless finite encoder of which you will explicit the state transition function and the output function.

Exercise 14.3.7. IL but not UD
Consider the following encoder, with state spac $\Sigma = \{A, B, C\}$.



The labels of the arrows define the input/output function. For instance, when reading b in state A, the encoder writes 1 and moves to state A (here the same state). Show that it is not UD but that it is IL.

Chapter 15

Asymptotic Behaviour of Markov Chains

15.1 Limit Distribution

15.1.1 Countable State Space

Consider an HMC that is irreducible and positive recurrent. If its initial distribution is the stationary distribution, it keeps the same distribution at all times. The chain is then said to be in the stationary regime, or in equilibrium, or in steady state. A question arises naturally: What is the long-run behavior of the chain when the initial distribution μ is arbitrary? For instance, will it converge to equilibrium? in what sense?

This question was answered in Section 6.3 for the finite state space case. In the case of infinite state space, linear algebra fails to provide the answer, and recourse to other methods is necessary. In fact, the following form of the limit theorem for Markov chains, which improves (6.10), will be proved.

Theorem 15.1.1 *Let $\{X_n\}_{n\geq 0}$ be an ergodic* HMC *on the countable state space E with transition matrix $\mathbf{P}$ and stationary distribution π, and let μ be an arbitrary initial distribution. Then*

$$\lim_{n\uparrow\infty} \sum_{i\in E} |P_\mu(X_n = i) - \pi(i)| = 0,$$

and in particular, for all $j \in E$,

$$\lim_{n\uparrow\infty} \sum_{i\in E} |p_{ji}(n) - \pi(i)| = 0.$$

Remark 15.1.2 A sequence $\{X_n\}_{n\geq 1}$ of discrete random variables with values in E is said to converge in distribution to the probability distribution π on E if for all $i \in E$, $\lim_{n\uparrow\infty} P(X_n = i) = \pi(i)$. It is said to converge in variation to this distribution if

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_15

$$\lim_{n\uparrow\infty} \sum_{i\in E} |P(X_n = i) - \pi(i)| = 0\,.$$

Thus, Theorem 15.1.1 states that the state X_n converges in stationary distribution π not only in distribution, but also in variation.

The proof of Theorem 15.1.1 will be given in Section 16.2.

EXAMPLE 15.1.3: THERMODYNAMIC IRREVERSIBILITY. The fame of the Ehrenfest diffusion model is due to the insight it provided to the once controversial issue of thermodynamic irreversibility. Indeed, according to the macroscopic theory of thermodynamics, systems progress in an orderly and irreversible manner towards equilibrium. The Ehrenfest urn is a simplified model of diffusion that captures the essential features of the phenomenon. Whatever the initial distribution of particles between the two compartments the system will settle to thermodynamical equilibrium, a macroscopic state in which the contents of A and B are both close to equality when N is "large". This means that at a fixed sufficiently large time the states that are not close to equidistribution in the two compartments are unlikely. This can be quantified as follows. By Stirling's equivalence,

$$\frac{N!}{(\alpha N)!(N-\alpha N)!} \Big/ \frac{N!}{(\beta N)!(N-\beta N)!} \sim e^{-N(h(\beta)-h(\alpha))}\,, \qquad (\star)$$

where $h(x) := -x\log x - (1-x)\log(1-x)$. The function h so defined on $[0,1]$ has a strict maximum at $x = \frac{1}{2}$, and therefore we see that all states $\alpha N \neq \frac{1}{2}N$ are very unlikely.

Boltzmann claimed that there was an arrow of time in the direction of increasing entropy, and indeed, in the diffusion experiment, equality between the thermodynamic quantities in both compartments corresponds to maximal entropy.

A controversy occured because a famous result of mechanics, Poincaré's recurrence theorem, implies that in the situation where at time 0 all molecules are in A, then whatever the time n, there will be a subsequent time $t > n$ at which all the molecules will again gather in A. This phenomenon predicted by irrefutable mathematics is, of course, never observed in daily life, where it would imply that the chunk of sugar that one patiently dissolves in one's cup of coffee could escape ingestion by reforming itself at the bottom of the cup.

Boltzmann's theory was challenged by this striking and seemingly inescapable argument. Things had to be clarified. Fortunately, Tatiana and Paul Ehrenfest came up with their Markov chain model, and in a sense saved Boltzmann's theory.

At first sight, the Ehrenfest model presents the two features that seemed incompatible: an irreversible tendency towards equilibrium, and recurrence. Here the role of Poincaré's recurrence theorem is played by the Markov chain recurrence theorem, stating that an irreducible chain with a stationary distribution visits any fixed state, say 0, infinitely often. As for the irreversible tendency towards equilibrium, one has the theorem of convergence to steady state, according to which the distribution at time n converges to the stationary distribution whatever the

initial distribution as n tends to infinity[1]. Thus, according to Markov chain theory, convergence to statistical equilibrium and recurrence are not antagonistic, and we are here at the epicenter of the refutation.

The average times between two successive occurences of a given state i being the reciprocal of the stationary probability of this state, we have from $(\star)$ that, taking into account the fact that $h(0) = 0$, and with $M := \frac{N}{2}$ (assuming N even)

$$\frac{E_M[T_M]}{E_0[T_0]} \sim e^{-Nh(\frac{1}{2})}, \qquad (\dagger).$$

This strongly suggests that the system very quickly returns to a near equilibrium state, and is very reluctant to return to a state where compartment A is nearly empty. In fact, recurrence is *not observable* for states far from $\frac{N}{2}$. For instance[2], the average time to reach 0 from state M is

$$\frac{1}{2M} 2^{2M}(1 + O(M)),$$

whereas the average time to reach state M from state 0 is less than

$$M + M \log M + O(1).$$

With $M = 10^6$ and one unit of mathematical time equal to 10^{-5} seconds, the return time to equilibrium when compartment A is initially empty is of the order of a second, whereas it would take of the order of

$$\frac{1}{2 \cdot 10^{11}} \times 2^{2^{10^6}} \text{ seconds}$$

to go from M to empty, which is an astronomical time. These numbers teach us not to spend too much time stirring the coffee, or hurry to swallow it for fear of recrystallization. From a mathematical point of view, being in the steady state at a given time does not prevent the chain from being in a rare state, only it is there rarely. The rarity of the state is equivalent to long recurrence times, so long that when there are more than a few particles in the boxes, it would take an astronomical time to witness the effects of Poincaré's recurrence theorem. Note that Boltzmann rightly argued that the recurrence times in Poincaré's theorem are extremely long, but his heuristic arguments failed to convince.

15.1.2 Absorption

We now consider the absorption problem for HMC's when the transition matrix is not necessarily assumed irreducible. The state space E is then decomposable as $E = T + \sum_j R_j$, where $R_1, R_2, \ldots$ are the disjoint recurrent classes and T is the

[1] *Stricto sensu*, this statement is not true, due to the periodicity of the chain. However, such periodicity is an artefact created by the discretization of time which is absent in the continuous-time model, or in a slight modification of the discrete-time model.

[2] These estimates can be obtained from formula (7.7) giving the average passage time from one state to another.

collection of transient states. (Note that the number of recurrent classes as well as the number of transient states may be infinite.)

What is the probability of being absorbed by a given recurrent class when starting from a given transient state? This kind of problem was already addressed when the first-step analysis method was introduced. It led to systems of linear equations with boundary conditions, for which the solution was unique, due to the finiteness of the state space. With an infinite state space, the uniqueness issue cannot be overlooked, and the absorption problem will be reconsidered with this in mind, and also with the intention of finding general matrix-algebraic expressions for the solutions. Another phenomenon not manifesting itself in the finite case is the possibility, when the set of transient states is infinite, of never being absorbed by the recurrent set. We shall consider this problem first, and then proceed to derive the distribution of the time to absorption by the recurrent set, and the probability of being absorbed by a given recurrent class.

The transition matrix can be block-partitioned as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{P}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B(1) & B(2) & \cdots & \mathbf{Q} \end{pmatrix}$$

or in condensed notation,

$$\mathbf{P} = \begin{pmatrix} D & 0 \\ B & \mathbf{Q} \end{pmatrix}. \tag{15.1}$$

This structure of the transition matrix accounts for the fact that one cannot go from a state in a given recurrent class to any state not belonging to this recurrent class. In other words, a recurrent class is closed.

Before Absorption

Let A be a subset of the state space E (typically the set of transient states, but not necessarily). We aim at computing for any initial state $i \in A$ the probability of remaining forever in A,

$$v(i) = P_i(X_r \in A;\ r \geq 0).$$

Defining $v_n(i) := P_i(X_1 \in A, \ldots, X_n \in A)$, we have, by monotone sequential continuity,

$$\lim_{n\uparrow\infty} \downarrow v_n(i) = v(i).$$

But for $j \in A$, $P_i(X_1 \in A, \ldots, X_{n-1} \in A, X_n = j) = \sum_{i_1 \in A} \cdots \sum_{i_{n-1} \in A} p_{ii_1} \cdots p_{i_{n-1}j}$ is the general term $q_{ij}(n)$ of the n-th iterate of the restriction $\mathbf{Q}$ of $\mathbf{P}$ to the set A. Therefore $v_n(i) = \sum_{j \in A} q_{ij}(n)$, that is, in vector notation,

$$v_n = \mathbf{Q}^n \mathbf{1}_A,$$

where $\mathbf{1}_A$ is the column vector indexed by A with all entries equal to 1. From this equality we obtain

$$v_{n+1} = \mathbf{Q}v_n,$$

and by dominated convergence $v = \mathbf{Q}v$. Moreover, $\mathbf{0}_A \leq v \leq \mathbf{1}_A$, where $\mathbf{0}_A$ is the column vector indexed by A with all entries equal to 0. The above result can be refined as follows:

Theorem 15.1.4 *The vector v is the maximal solution of*

$$v = \mathbf{Q}v, \mathbf{0}_A \leq v \leq \mathbf{1}_A.$$

Moreover, either $v = \mathbf{0}_A$ or $sup_{i\in A} v(i) = 1$. In the case of a finite transient set T, the probability of infinite sojourn in T is null.

Proof. Only maximality and the last statement remain to be proved. To prove maximality consider a vector u indexed by A such that $u = \mathbf{Q}u$ and $\mathbf{0}_A \leq u \leq \mathbf{1}_A$. Iteration of $u = \mathbf{Q}u$ yields $u = \mathbf{Q}^n u$, and $u \leq \mathbf{1}_A$ implies that $\mathbf{Q}^n u \leq \mathbf{Q}^n \mathbf{1}_A = v_n$. Therefore, $u \leq v_n$, which gives $u \leq v$ by passage to the limit.

To prove the last statement of the theorem, let $c = \sup_{i\in A} v(i)$. From $v \leq c\mathbf{1}_A$, we obtain $v \leq cv_n$ as above, and therefore, at the limit, $v \leq cv$. This implies either $v = \mathbf{0}_A$ or $c = 1$.

When the set T is *finite*, the probability of infinite sojourn in T is null, because otherwise at least one transient state would be visited infinitely often. □

Equation $v = \mathbf{Q}v$ reads

$$v(i) = \sum_{j\in A} p_{ij} v(j) \ (i \in A) \, .$$

First-step analysis gives this equality as a *necessary* condition. However, it does not help to determine which solution to choose, in case there are several.

EXAMPLE 15.1.5: THE REPAIR SHOP, TAKE 5. We shall prove in a different way a result already obtained in Example 7.3.7, that is: the repair shop chain is recurrent if and only if $\rho \leq 1$,. Observe that the restriction of $\mathbf{P}$ to $A_i := \{i+1, i+2, \ldots\}$, namely

$$\mathbf{Q} = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ & a_0 & a_1 & \cdots \\ & & & \cdots \end{pmatrix},$$

does not depend on $i \geq 0$. In particular, the maximal solution of $v = \mathbf{Q}v$, $\mathbf{0}_A \leq v \leq \mathbf{1}_A$ when $A \equiv A_i$ has, in view of Theorem 15.1.4, the following two interpretations. Firstly, for $i \geq 1$, $1 - v(i)$ is the probability of visiting 0 when starting from $i \geq 1$. Secondly, $(1 - v(1))$ is the probability of visiting $\{0, 1, \ldots, i\}$ when starting from $i + 1$. But when starting from $i + 1$, the chain visits $\{0, 1, \ldots, i\}$ if and only if it visits i, and therefore $(1 - v(1))$ is also the probability of visiting i when starting from $i + 1$. The probability of visiting 0 when starting from $i + 1$ is

$$1 - v(i+1) = (1 - v(1))(1 - v(i)),$$

because in order to go from $i+1$ to 0 one must first reach i, and then go to 0. Therefore, for all $i \geq 1$,

$$v(i) = 1 - \beta^i \,,$$

where $\beta = 1 - v(1)$. To determine β, write the first equality of $v = \mathbf{Q}v$:

$$v(1) = a_1 v(1) + a_2 v(2) + \cdots \,,$$

that is,

$$(1 - \beta) = a_1(1 - \beta) + a_2(1 - \beta^2) + \cdots \,.$$

Since $\sum_{i \geq 0} a_i = 1$, this reduces to

$$\beta = g(\beta) \,, \qquad (\star)$$

where g is the generating function of the probability distribution $(a_k, k \geq 0)$. Also, all other equations of $v = \mathbf{Q}v$ reduce to $(\star)$.

Under the irreduciblity assumptions $a_0 > 0$, $a_0 + a_1 < 1$, $(\star)$ has only one solution in $[0,1]$, namely $\beta = 1$ if $\rho \leq 1$, whereas if $\rho > 1$, it has two solutions in $[0,1]$, this probability is $\beta = 1$ and $\beta = \beta_0 \in (0,1)$. We must take the smallest solution. Therefore, if $\rho > 1$, the probability of visiting state 0 when starting from state $i \geq 1$ is $1 - v(i) = \beta_0^i < 1$, and therefore the chain is transient. If $\rho \leq 1$, the latter probability is $1 - v(i) = 1$, and therefore the chain is recurrent.

EXAMPLE 15.1.6: 1-D RANDOM WALK, TAKE 3. The transition matrix of the random walk on $\mathbb{N}$ with a reflecting barrier at 0,

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & & & & \\ q & 0 & p & & & \\ & q & 0 & p & & \\ & & q & 0 & p & \\ & & & \cdot & \cdot & \cdot \end{pmatrix},$$

where $p \in (0,1)$, is clearly irreducible. Intuitively, if $p > q$, there is a drift to the right, and one expects the chain to be transient. This will be proven formally by showing that the probability $v(i)$ of never visiting state 0 when starting from state $i \geq 1$ is strictly positive. In order to apply Theorem 15.1.4 with $A = \mathbb{N} - \{0\}$, we must find the general solution of $u = \mathbf{Q}u$. This equation reads

$$\begin{aligned} u(1) &= pu(2), \\ u(2) &= qu(1) + pu(3), \\ u(3) &= qu(2) + pu(4), \\ &\cdots \end{aligned}$$

and its general solution is $u(i) = u(1) \sum_{j=0}^{i-1} \left(\frac{q}{p}\right)^j$. The largest value of $u(1)$ respecting the constraint $u(i) \in [0,1]$ is $u(1) = 1 - \left(\frac{q}{p}\right)$. The solution $v(i)$ is therefore

$$v(i) = 1 - \left(\frac{q}{p}\right)^i .$$

Time to Absorption

We now turn to the determination of the distribution of τ, the time of exit from the transient set T. Theorem 15.1.4 tells us that $v = \{v(i)\}_{i\in T}$, where $v(i) = P_i(\tau = \infty)$, is the largest solution of $v = \mathbf{Q}v$ subject to the constraints $\mathbf{0}_T \leq v \leq \mathbf{1}_T$, where $\mathbf{Q}$ is the restriction of $\mathbf{P}$ to the transient set T. The probability distribution of τ when the initial state is $i \in T$ is readily computed starting from the identity

$$P_i(\tau = n) = P_i(\tau \geq n) - P_i(\tau \geq n+1)$$

and the observation that for $n \geq 1$ $\{\tau \geq n\} = \{X_{n-1} \in T\}$, from which we obtain, for $n \geq 1$,

$$P_i(\tau = n) = P_i(X_{n-1} \in T) - P(X_n \in T) = \sum_{j\in T}(p_{ij}(n-1) - p_{ij}(n)).$$

Now, $p_{ij}(n)$ is, for $i, j \in T$, the general term of $\mathbf{Q}^n$, and therefore:

Theorem 15.1.7

$$P_i(\tau = n) = \{(\mathbf{Q}^{n-1} - \mathbf{Q}^n)\mathbf{1}_T\}_i . \tag{15.2}$$

In particular, if $P_i(\tau = \infty) = 0$,

$$P_i(\tau > n) = \{\mathbf{Q}^n\mathbf{1}_T\}_i .$$

Proof. Only the last statement remains to be proved. From (15.2),

$$\begin{aligned} P_i(n < \tau \leq n+m) &= \sum_{j=0}^{m-1}\{(\mathbf{Q}^{n+j} - \mathbf{Q}^{n+j-1})\mathbf{1}_T\}_i \\ &= \left\{\left(\mathbf{Q}^n - \mathbf{Q}^{n+m}\right)\mathbf{1}_T\right\}_i , \end{aligned}$$

and therefore, if $P_i(\tau = \infty) = 0$, we obtain (15.2) by letting $m \uparrow \infty$. □

Absorbing Destinations

We seek to compute the probability of absorption by a given recurrent class when starting from a given transient state. As we shall see later, it suffices for the theory to treat the case where the recurrent classes are singletons. We therefore suppose that the transition matrix has the form

$$\mathbf{P} = \begin{pmatrix} I & 0 \\ B & \mathbf{Q} \end{pmatrix}. \tag{15.3}$$

Let f_{ij} be the probability of absorption by recurrent class $R_j = \{j\}$ when starting from the transient state i. We have

$$\mathbf{P}^n = \begin{pmatrix} I & 0 \\ L_n & \mathbf{Q}^n \end{pmatrix},$$

where $L_n = (I + \mathbf{Q} + \cdots + \mathbf{Q}^n)B$. Therefore, $\lim_{n\uparrow\infty} L_n = SB$. For $i \in T$, the (i,j) term of L_n is

$$L_n(i,j) = P(X_n = j | X_0 = i).$$

Now, if T_{R_j} is the first time of visit to R_j after time 0, then

$$L_n(i,j) = P_i(T_{R_j} \leq n),$$

since R_j is a closed state. Letting n go to ∞ gives the following:

Theorem 15.1.8 *For an* HMC *with transition matrix* $\mathbf{P}$ *of the form (15.3), the probability of absorption by recurrent class* $R_j = \{j\}$ *starting from transient state* i *is*

$$P_i(T_{R_j} < \infty) = (SB)_{i,R_j}.$$

The general case, where the recurrence classes are not necessarily singletons, can be reduced to the singleton case as follows. Let $\mathbf{P}^*$ be the matrix obtained from the transition matrix $\mathbf{P}$, by grouping for each j the states of recurrent class R_j into a single state $\hat{\jmath}$:

$$\mathbf{P}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ b_{\hat{1}} & b_{\hat{2}} & \cdots & \mathbf{Q} \end{pmatrix} \tag{15.4}$$

where $b_{\hat{\jmath}} = B(j)\mathbf{1}_T$ is obtained by summation of the columns of $B(j)$, the matrix consisting of the columns $i \in R_j$ of B. The probability f_{iR_j} of absorption by class R_j when starting from $i \in T$ equals $\hat{f}_{i\hat{\jmath}}$, the probability of ever visiting $\hat{\jmath}$ when starting from i, computed for the chain with transition matrix $\mathbf{P}^*$.

15.1.3 Variance of Ergodic Estimates

Let $\{X_n\}_{n\geq 0}$ be an ergodic Markov chain with finite state space $E = \{1, 2, \ldots, r\}$. A function $f : E \to \mathbb{R}$ is represented by a column vector $f = (f(1), \ldots, f(r))^T$. The ergodic theorem tells that the estimate $\frac{1}{n}\sum_{k=1}^n f(X_n)$ of $\langle f \rangle_\pi := E_\pi[f(X_0)]$ is asymptotically unbiased, in the sense that it converges to $\langle f \rangle_\pi$ as $n \to \infty$. However, the variance of this estimate increases indefinitely as $n \uparrow \infty$. The next result gives the asymptotic rate of increase.

Theorem 15.1.9 *For* $\{X_n\}_{n\geq 0}$ *and* $f : E \to \mathbb{R}$ *as above, and for any initial distribution* μ,

$$\lim_{n\to\infty} \frac{1}{n} V_\mu \left(\sum_{k=1}^n f(X_k) \right) = 2 \langle f, \mathbf{Z} f \rangle_\pi - \langle f, (I + \Pi) f \rangle_\pi , \tag{15.5}$$

where the notation V_μ *indicates that the variance is computed with respect to* P_μ.

The quantity on the right-hand side will be denoted by $v(f, \mathbf{P}, \pi)$.

Proof. We first suppose that $\mu = \pi$, the stationary distribution. Then

$$\begin{aligned}\frac{1}{n} V_\pi \left(\sum_{k=1}^{n} f(X_k) \right) &= \frac{1}{n} \left\{ \sum_{k=1}^{n} V_\pi(f(X_k)) + 2 \sum_{\substack{k,j=1 \\ k<j}}^{n} \text{cov}_{\,\pi}(f(X_k), f(X_j)) \right\} \\ &= V_\pi(f(X_0)) + \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} \text{cov}_{\,\pi}(f(X_0), f(X_\ell)) \,,\end{aligned}$$

where we have used the fact that when the initial distribution is π, the chain is stationary, and in particular, $\text{cov}_{\,\pi}(f(X_k), f(X_j)) = \text{cov}_{\,\pi}(f(X_0), f(X_{j-k}))$ for $k < j$. Now,

$$\begin{aligned}V_\pi(f(X_0)) &= E_\pi[f(X_0)^2] - E_\pi[f(X_0)]^2 \\ &= \sum_{i \in E} \pi(i) f(i)^2 - \left(\sum_{i \in E} \pi(i) f(i) \right)^2 = \langle f, f \rangle_\pi - \langle f, \Pi f \rangle_\pi \,.\end{aligned}$$

Also,

$$\begin{aligned}\text{cov}_{\,\pi}(f(X_0), f(X_\ell)) &= E_\pi[f(X_0) f(X_k)] - E_\pi[f(X_0)]^2 \\ &= \sum_{i \in E} \sum_{j \in E} \pi(i) p_{ij}(\ell) f(i) f(j) - E_\pi[f(X_0)]^2 \\ &= \left\langle f, \mathbf{P}^\ell f \right\rangle_\pi - \langle f, \Pi f \rangle_\pi = \langle f, (\mathbf{P}^\ell - \Pi) f \rangle_\pi .\end{aligned}$$

Since $\lim_{n\to\infty} \sum_{\ell=1}^{n} (\mathbf{P}^\ell - \Pi) = \mathbf{Z} - I$,

$$\lim_{n\to\infty} \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} (\mathbf{P}^\ell - \Pi) = \mathbf{Z} - I.$$

Indeed, by Cesaro's lemma: If $A_n = \sum_{\ell=1}^{n} \alpha_\ell$ tends to A as $n \to \infty$, then $\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n-1} A_\ell = A$. But $\frac{1}{n} \sum_{\ell=1}^{n-1} A_\ell = \frac{1}{n}(\alpha_1 + (\alpha_1 + \alpha_2) + \cdots + (\alpha_1 + \cdots + \alpha_{n-1})) = \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} \alpha_\ell$. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} V_\pi \left(\sum_{k=1}^{n} f(X_k) \right) = \langle f, f \rangle_\pi - \langle f, \Pi f \rangle_\pi + 2 \langle f, (\mathbf{Z} - I) f \rangle_\pi \,,$$

which is the announced result (for $\mu = \pi$).

To prove the result in the general case where the initial distribution is arbitrary, it suffices to show that for two chains $\{X_n^{(1)}\}_{n\geq 0}$ and $\{X_n^{(2)}\}_{n\geq 0}$ with transition matrix $\mathbf{P}$, and arbitrary initial distributions μ and ν, respectively, that couple at a time τ such that $E[\tau^2] < \infty$ (this is the case here, see Exercise 16.4.6),

$$\lim_{n\to\infty} \frac{1}{n} V \left(\sum_{k=1}^{\infty} f(X_k^{(1)}) \right) = \lim_{n\to\infty} \frac{1}{n} V \left(\sum_{k=1}^{\infty} f(X_n^{(2)}) \right).$$

But with $X_n = X_n^{(1)}$ or $X_n^{(2)}$,

$$\begin{aligned} V\left(\sum_{k=1}^{n} f(X_k)\right) &= E\left[\left(\sum_{k=1}^{n} f(X_k)\right)^2\right] - E\left[\sum_{k=1}^{n} f(X_k)\right]^2 \\ &= E\left[\left(\sum_{k=1}^{\tau\wedge n} + \sum_{k=\tau+1}^{n}\right)^2\right] - \left(E\left[\sum_{k=1}^{\tau\wedge n}\right] + E\left[\sum_{k=\tau+1}^{n}\right]\right)^2 \\ &= E\left[\left(\sum_{k=1}^{\tau\wedge n}\right)^2\right] + E\left[\left(\sum_{k=\tau+1}^{n}\right)^2 + 2E\left[\left(\sum_{k=1}^{\tau\wedge n}\right)\left(\sum_{k=\tau+1}^{n}\right)\right]\right] \\ &\quad -E\left[\sum_{k=1}^{\tau\wedge n}\right]^2 - E\left[\sum_{k=\tau+1}^{n}\right]^2 - 2E\left[\sum_{k=1}^{\tau\wedge n}\right] E\left[\sum_{k=\tau+1}^{n}\right]. \end{aligned}$$

Since $\sum_{k=\tau+1}^{n} f(X_k^{(1)}) = \sum_{k=\tau+1}^{n} f(X_k^{(2)})$, it follows that

$$\frac{1}{n}\left\{V\left(\sum_{k=1}^{n} f(X_k^{(1)})\right) - \frac{1}{n}V\left(\sum_{k=1}^{n} f(X_k^{(2)})\right)\right\} = \frac{1}{n}A_n + \frac{2}{n}B_n - \frac{2}{n}C_n,$$

where

$$\begin{aligned} A_n &= \left\{E\left[\left(\sum_{k=1}^{\tau\wedge n}(1)\right)^2\right] - E\left[\left(\sum_{k=1}^{\tau\wedge n}(2)\right)^2\right] - E\left[\sum_{k=1}^{\tau\wedge n}(1)\right]^2 + E\left[\sum_{k=1}^{\tau\wedge n}(2)\right]^2\right\}, \\ B_n &= \left\{E\left[\left(\sum_{k=\tau+1}^{n}(1,2)\right)\left(\sum_{k=1}^{\tau\wedge n}(1) - \sum_{k=1}^{\tau\wedge n}(2)\right)\right]\right\}, \\ C_n &= \left\{E\left[\sum_{k=\tau+1}^{n}(1,2)\right] E\left[\sum_{k=1}^{\tau\wedge n}(1) - \sum_{k=1}^{\tau\wedge n}(2)\right]\right\}. \end{aligned}$$

Write

$$\frac{2}{n}B_n = 2E\left[\frac{\sum_{k=\tau+1}^{n}(1,2)}{n}\left(\sum_{k=1}^{\tau\wedge n}(1) - \sum_{k=1}^{\tau\wedge n}(2)\right)\right]$$

and observe that the quantity under the expectation converges, as $n \to \infty$, towards $E_\pi[f(X_0)]$ $(\sum_{k=1}^{\tau}(f(X_k^{(1)}) - f(X_k^{(2)})))$ and is for fixed n bounded in absolute value by $2(\sup|f|)\tau$, an integrable random variable. Therefore, by dominated convergence,

$$\lim_{n\to\infty}\frac{2}{n}B_n = 2E_\pi[f(X_0)]E\left[\sum_{k=1}^{\tau}(f(X_k^{(1)}) - f(X_k^{(2)}))\right].$$

A similar argument shows that $\frac{2}{n}C_n$ has the same limit. Therefore, $\lim_{n\to\infty}\frac{2}{n}(B_n - C_n) = 0$. As for A_n, it is bounded by $4(\sup|f|)^2E[\tau^2] < \infty$, and therefore $\lim_{n\to\infty}\frac{1}{n}A_n = 0$. □

We shall now give an expression for the asymptotic variance in terms of the eigenvalues, when $\mathbf{P}$ has r distinct eigenvalues. We have, in view of (6.11),

$$(\mathbf{P}^n - \Pi) = \sum_{i=2}^{r} \lambda_i^n v_i u_i^T,$$

and therefore

$$\mathbf{Z} = I + \sum_{n\geq 1}(\mathbf{P}^n - \Pi) = I + \sum_{i=2}^{r} \frac{\lambda_i}{1-\lambda_i} v_i u_i^T. \tag{15.6}$$

Also, from (15.5),

$$v(f, \mathbf{P}, \pi) = V_\pi(f(X_0)) + 2\sum_{i=2}^{r} \frac{\lambda_i}{1-\lambda_i} \langle f, v_i\rangle_\pi \, (f^T u_i). \tag{15.7}$$

For a reversible pair $(\mathbf{P}, \pi)$, we have $u_i = Dv_i$, and therefore $f^T u_i = \langle f, v_i\rangle_\pi$. Using this observation and (20.5), we obtain from (15.7),

$$v(f, \mathbf{P}, \pi) = \sum_{i=2}^{r} \frac{1+\lambda_i}{1-\lambda_i} |\, \langle f, v_i\rangle_\pi \,|^2. \tag{15.8}$$

If one is interested in the speed of convergence to equilibrium, it is the second-largest eigenvalue modulus that is important. However, if one is interested in simulation, that is, the computation of $E_\pi[f(X_0)]$ as the ergodic mean $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k)$, all eigenvalues play a role if we measure the quality of the ergodic estimator by the asymptotic variance, as the above formulas show.

15.2 Non-homogeneous Markov Chains

15.2.1 Dobrushin's Ergodic Coefficient

(Dobrushin, 1956) For non-homogeneous Markov chains (NHMC), the Markov property is retained but the transition probabilities may depend on time. This section gives conditions guaranteeing the existence of a limit in variation of such chains, with their application to simulated annealing in view. When the state space E is finite and the chain is ergodic, Dobrushin's ergodic coefficient helps provide a computable geometric rate of convergence to steady state and will be especially useful in the next subsection to obtain a necessary and sufficient condition of weak ergodicity (yet to be defined) of non-homogeneous Markov chains.

Let E, F, G denote countable sets, finite when so indicated. A stochastic matrix indexed by $F \times E$ is a matrix whose rows (indexed by F) are probability distributions on E.

Definition 15.2.1 *Let* $\mathbf{Q}$ *be a stochastic matrix indexed by* $F \times E$. *Its ergodic coefficient is*

$$\begin{aligned}\delta(\mathbf{Q}) &:= \frac{1}{2}\sup_{i,j\in F}\sum_{k\in E}|q_{ik}-q_{jk}| \\ &= \sup_{i,j\in F} d_V(q_{i\cdot}, q_{j\cdot}) = \sup_{i,j\in F}\sup_{A\subseteq E}(q_{iA}-q_{jA})\,.\end{aligned}$$

Observe that $0 \le \delta(\mathbf{Q}) \le 1$ and that, by the result of Exercise 16.4.10,

$$\delta(\mathbf{Q}) = 1 - \inf_{i,j\in F}\sum_{k\in E} q_{ik}\wedge q_{jk}\,. \tag{15.9}$$

The ergodic coefficient is in general useless if F is infinite. In particular, if the stochastic matrix $\mathbf{Q}$ has two orthogonal rows (that is, rows i, j such that $q_{ik}q_{jk} = 0$ for all $k \in E$, and this is the most frequent case with an infinite state space), then $\delta(\mathbf{Q}) = 1$. However, for finite state spaces, this notion becomes very powerful.

EXAMPLE 15.2.2: TWO-ROW MATRIX. The ergodic coefficient of the stochastic matrix

$$\mathbf{Q} = \begin{pmatrix}\mu^T \\ \nu^T\end{pmatrix} = \begin{pmatrix}\mu_1 & \mu_2 & \mu_3 & \cdots \\ \nu_1 & \nu_2 & \nu_3 & \cdots\end{pmatrix}$$

is the distance in variation between the two rows: $\delta(\mathbf{Q}) = d_V(\mu, \nu)$.

EXAMPLE 15.2.3: COUPLING AND THE ERGODIC COEFFICIENT. We are going to construct two HMC's $\{X_n^{(1)}\}_{n\ge0}$ and $\{X_n^{(2)}\}_{n\ge0}$ with the same transition matrix $\mathbf{P}$, assumed irreducible, and with a strictly positive ergodic coefficient, in such a way that they couple at a time τ stochastically smaller than a geometric random variable with parameter $p = 1 - \delta(\mathbf{P})$.

The construction is as follows. Let $\alpha_i(j) := p_{ij}$. Suppose that at time n, $X_n^{(1)} = i_1$ and $X_n^{(2)} = i_2$. Then, $X_{n+1}^{(1)}$ and $X_{n+1}^{(2)}$ will be distributed according to the distributions α_{i_1} and α_{i_2} respectively. According to Theorem 16.1.4, this can be done in such a way that

$$P\left(X_{n+1}^{(1)} = X_{n+1}^{(2)} \,|\, X_n^{(1)} = i_1, X_n^{(2)} = i_2\right) = d_V\left(\alpha_{i_1}, \alpha_{i_2}\right)\,.$$

The latter quantity is $\ge 1 - \delta(\mathbf{P}) > 0$. Therefore the two chains will meet for the first time (and from then on be identical) at a time τ that is stochastically smaller that a geometric random variable with parameter $p = 1 - \delta(\mathbf{P})$, that is,

The following sub-multiplicativity property of the ergodic coefficient will be useful.

Theorem 15.2.4 *Let $\mathbf{Q}_1$ and $\mathbf{Q}_2$ be two stochastic matrices indexed by $F \times G$ and $G \times E$, respectively. Then*

$$\delta(\mathbf{Q}_1\mathbf{Q}_2) \le \delta(\mathbf{Q}_1)\delta(\mathbf{Q}_2)\,.$$

Proof. From Lemma 16.1.3, for any stochastic matrix $\mathbf{Q} = \{q_{ij}\}$ indexed by $E \times F$,

$$\frac{1}{2}\sum_{k\in E}|q_{ik} - q_{jk}| = \sup_{A\subseteq E}\sum_{k\in A}(q_{ik} - q_{jk})\,,$$

and in particular, with $\mathbf{Q}_1 = \{a_{ij}\}$ and $\mathbf{Q}_2 = \{b_{ij}\}$,

$$\delta(\mathbf{Q}_1\mathbf{Q}_2) = \frac{1}{2}\sup_{i,j\in F}\sup_{A\subseteq E}\sum_{k\in A}\left(\sum_{\ell\in G}(a_{i\ell} - a_{j\ell})b_{\ell k}\right)\,.$$

But

$$\sum_{\ell\in G}(a_{i\ell} - a_{j\ell})^+ = \sum_{\ell\in G}(a_{i\ell} - a_{j\ell})^- = \frac{1}{2}\sum_{\ell\in G}|a_{i\ell} - a_{j\ell}|,$$

and therefore

$$\begin{aligned}\sum_{k\in A}\sum_{\ell\in G}(a_{i\ell} - a_{j\ell})b_{\ell k} &= \sum_{\ell\in G}(a_{i\ell} - a_{j\ell})^+\sum_{k\in A}b_{\ell k} - \sum_{\ell\in G}(a_{i\ell} - a_{j\ell})^-\sum_{k\in A}b_{\ell k}\\ &\le \left(\frac{1}{2}\sum_{\ell\in G}|a_{i\ell} - a_{j\ell}|\right)\sup_{\ell\in G}\sum_{k\in A}b_{\ell k}\\ &\qquad - \left(\frac{1}{2}\sum_{\ell\in G}|a_{i\ell} - a_{j\ell}|\right)\inf_{\ell\in G}\sum_{k\in A}b_{\ell k}\\ &\le \left(\frac{1}{2}\sum_{\ell\in G}|a_{il} - a_{j\ell}|\right)\left(\sup_{\ell,\ell'\in G}\sum_{k\in A}(b_{\ell k} - b_{\ell' k})\right)\,.\end{aligned}$$

The announced inequality then follows from the identity

$$\sup_{\ell,\ell'\in G}\sup_{A\subseteq E}\sum_{k\in A}(b_{\ell k} - b_{\ell' k}) = \frac{1}{2}\sup_{\ell,\ell'\in G}\sum_{k\in E}|b_{\ell k} - b_{\ell' k}| = \delta(\mathbf{Q}_2)\,.$$

□

Theorem 15.2.5 *Let $\mathbf{P}$ be a stochastic matrix indexed by E, and let μ and ν be two probability distributions on E. Then*

$$d_V(\mu^T\mathbf{P}^n, \nu^T\mathbf{P}^n) \le d_V(\mu,\nu)\delta(\mathbf{P})^n. \tag{15.10}$$

Proof. The proof is by recurrence. Since

$$\begin{pmatrix}\mu^T\mathbf{P}^{n+1}\\ \nu^T\mathbf{P}^{n+1}\end{pmatrix} = \begin{pmatrix}\mu^T\mathbf{P}^{n}\\ \nu^T\mathbf{P}^{n}\end{pmatrix}\mathbf{P}$$

and (see Example 15.2.2)

$$d_V(\mu^T\mathbf{P}^n, \nu^T\mathbf{P}^n) = \delta\left(\begin{pmatrix}\mu^T\mathbf{P}^{n}\\ \nu^T\mathbf{P}^{n}\end{pmatrix}\right)\,.$$

Therefore, by Lemma 15.2.4,

$$d_V(\mu^T\mathbf{P}^{n+1}, \nu^T\mathbf{P}^{n+1}) \leq d_V(\mu^T\mathbf{P}^n, \nu^T\mathbf{P}^n)\delta(\mathbf{P})\,,$$

from which (15.10) follows by iteration. □

Corollary 15.2.6 *Let* $\mathbf{Q}_1$, $\mathbf{Q}_2$ *and* $\mathbf{P}$ *be stochastic matrices indexed by* $E \times E$. *Then*

$$|(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{P}| \leq |(\mathbf{Q}_1 - \mathbf{Q}_2)|\delta(\mathbf{P})\,.$$

Proof. Let μ_k and ν_k be the k-th row of $\mathbf{Q}_1$ and $\mathbf{Q}_2$ respectively. The inequality to be verified:

$$\sup_{k\in E} |\mu_k\mathbf{P} - \nu_k\mathbf{P}| \leq \sup_{k\in E} |\mu_k - \nu_k|\delta(\mathbf{P})\,,$$

is a direct consequence of Theorem 15.2.5. □

15.2.2 Ergodicity of Non-homogeneous Markov Chains

Let $\{X_n\}_{n\geq 0}$ be a non-homogeneous Markov chain with values in the countable set E, and define for all states $i, j \in E$ and all times $n \geq 0$

$$p_{n,i,j} := P(X_{n+1} = j | X_n = i).$$

The matrix

$$\mathbf{P}(n) := \{p_{n,i,j}\}_{i,j\in E}$$

is called the transition matrix at time n. Define for all $0 \leq m \leq k$

$$\mathbf{P}(m,k) := \mathbf{P}(m)\mathbf{P}(m+1)\cdots\mathbf{P}(k-1)\,.$$

It follows from the Bayes sequential rule that if the distribution of X_m is μ_m, the distribution of X_k is $\mu_m^T\mathbf{P}(m,k)$.

The Block Criterion

Definition 15.2.7 *The above non-homogeneous Markov chain is called weakly ergodic if for all* $m \geq 0$,

$$\lim_{k\uparrow\infty} \sup_{\mu,\nu} d_V\left(\mu^T\mathbf{P}(m,k), \nu^T\mathbf{P}(m,k)\right) = 0\,,$$

where the supremum is taken over all the probability distributions μ, ν *on* E.

Definition 15.2.8 *The chain is called strongly ergodic if there exists a probablity distribution* π *on* E *such that for all* $m \geq 0$,

$$\lim_{k\uparrow\infty} \sup_{\mu} d_V\left(\mu^T\mathbf{P}(m,k), \pi\right) = 0, \tag{15.11}$$

where the supremum is taken over all the probability distributions μ *on* E.

(Sometimes one says that the family of transition matrices $\{\mathbf{P}(n)\}_{n\geq 0}$ (rather than the chain) is weakly ergodic (resp., strongly ergodic).)

If the chain is homogeneous and ergodic, then it is strongly ergodic in the sense of Definition 15.2.8, by the theorem of convergence to steady state for ergodic HMC's.

Strong ergodicity implies weak ergodicity, since

$$d_V\left(\mu^T\mathbf{P}(m,k),\nu^T\mathbf{P}(m,k)\right)\leq d_V\left(\mu^T\mathbf{P}(m,k),\pi\right)+d_V\left(\nu^T\mathbf{P}(m,k),\pi\right).$$

However, there are weakly ergodic chains that are not strongly ergodic, as the following example shows.

EXAMPLE 15.2.9: WEAKLY, YET NOT STRONGLY ERGODIC. The state space $E=\{1,2\}$, $\mathbf{P}(0)=I$, the identity, and for $n\geq 1$,

$$\mathbf{P}(2n)=\begin{pmatrix}1/2n & 1-1/2n\\ 1/2n & 1-1/2n\end{pmatrix},\quad \mathbf{P}(2n+1)=\begin{pmatrix}1-1/(2n+1) & 1/(2n+1)\\ 1-1/(2n+1) & 1/(2n+1)\end{pmatrix}.$$

Elementary computations show that for any probability distribution μ on E,

$$\mu^T\mathbf{P}(m,2k+1)=\left(1-\frac{1}{2k+1},\frac{1}{2k+1}\right),\quad \mu^T\mathbf{P}(m,2k)=\left(\frac{1}{2k},1-\frac{1}{2k}\right),$$

and therefore, for all $k\geq m$,

$$\mu^T\mathbf{P}(m,k)-\nu^T\mathbf{P}(m,k)=0.$$

Thus, the chain is weakly ergodic. But it cannot be strongly ergodic, since $\mu^T\mathbf{P}(m,k)$ has, as $k\to\infty$, two limit vectors: $(1,0)$ and $(0,1)$.

As one might guess, weak ergodicity is in general not easy to check directly from the definition. Fortunately, there is a somewhat useable criterion in terms of Dobrushin's coefficient of ergodicity. It depends on the following lemma:

Lemma 15.2.10 *The chain is weakly ergodic if and only if for all $m\geq 0$,*

$$\lim_{k\uparrow\infty}\delta(\mathbf{P}(m,k))=0. \tag{15.12}$$

Proof. By Theorem 15.2.5 and observing that $d_V(\mu,\nu)\leq 1$,

$$d_V(\mu^T\mathbf{P}(m,k),\nu^T\mathbf{P}(m,k)\leq d_V(\mu,\nu)\delta(\mathbf{P}(m,k))\leq\delta(\mathbf{P}(m,k)).$$

Therefore (15.12) implies weak ergodicity. Conversely, it follows from the inequalities

$$\begin{aligned}\delta(\mathbf{P}(m,k)) &= \frac{1}{2}\sup_{i,j\in E}\sum_{\ell\in E}|p_{i\ell}(m,k)-p_{j\ell}(m,k)|\\ &\leq \frac{1}{2}\sup_{\mu,\nu}|\mu^T\mathbf{P}(m,k)-\nu^T\mathbf{P}(m,k)|\end{aligned}$$

that weak ergodicity implies (15.12). □

By Dobrushin's inequality, $\delta(\mathbf{P}(m,k)) \leq \prod_{r=m}^{k-1} \delta(\mathbf{P}(r))$, and therefore nullity of the infinite product $\prod_{r\geq 1} \delta(\mathbf{P}(r))$ is enough to guarantee weak ergodicity. However, in many applications, weak ergodicity occurs without the above infinite product diverging to zero. It turns out that the consideration of blocks gives a useful necessary and sufficient condition.

Theorem 15.2.11 *(Hajnal, 1958) The chain is weakly ergodic if and only if there exists a strictly increasing sequence of integers $\{n_s\}_{s\geq 0}$ such that*

$$\sum_{s=0}^{\infty}(1-\delta(\mathbf{P}(n_s,n_{s+1}))=\infty. \tag{15.13}$$

Proof. First observe that since $0 \leq \delta(\mathbf{P}(n_s,n_{s+1})) \leq 1$, (15.13) is equivalent to nullity of the infinite product $\prod_{s\geq 0}\delta(\mathbf{P}(n_s,n_{s+1}))$.

Denoting by i the first integer s such that $n_s \geq m$, and by j the last integer s such that $n_s \leq k-1$, Dobrushin's inequality gives

$$\begin{aligned}\delta(\mathbf{P}(m,k)) &\leq \delta\mathbf{P}(m,n_i))\left\{\prod_{s=i}^{j-1}\delta(\mathbf{P}(n_s,n_{s+1})\right\}\delta(\mathbf{P}(n_j,k)) \\ &\leq \prod_{s=i}^{j-1}\delta(\mathbf{P}(n_s,n_{s+1})).\end{aligned}$$

Since $j\to\infty$ as $k\to\infty$, we see that (15.13) implies weak ergodicity, by Lemma 15.2.10.

Conversely, if we assume weak ergodicity, then by Lemma 15.2.10, we can inductively construct for any $\gamma\in(0,1)$ a strictly increasing sequence of integers $\{n_s\}_{s\geq 0}$ by

$$n_0=1, n_{s+1}=\inf\{k>n_s;\ \delta(\mathbf{P}(n_s,k))\leq 1-\gamma\}.$$

For such sequences, the product $\prod_{s\geq 0}\delta(\mathbf{P}(n_s,n_{s+1}))$ is null, and this is equivalent to (15.13). □

It remains to decide when a weakly ergodic NHMC is strongly ergodic. No useful criterion of strong ergodicity is available, and we have to resort to sufficient conditions.

Theorem 15.2.12 *Suppose that the chain is weakly ergodic, and that for all $n\geq 0$, there exists a probability distribution $\pi(n)$ on E such that*

$$\pi^T(n)=\pi^T(n)\mathbf{P}(n)$$

and

$$\sum_{n=0}^{\infty}|\pi(n+1)-\pi(n)|<\infty\,. \tag{15.14}$$

The chain is then strongly ergodic.

Proof. In this proof we shall make use of the matrix norm

$$|A| = \sup_{i\in E} \sum_{j\in E} |a_{ij}|$$

defined for square matrices with real elements indexed by $E \times E$. Condition 15.14 implies the existence of a probability distribution π such that

$$\lim_{n\uparrow\infty} |\pi(n) - \pi| = 0\,. \qquad (\star)$$

Define Π_n (resp., Π) to be the matrix with all rows equal to $\pi(n)$ (resp., π). In particular, $|\Pi_n - \Pi| = |\pi(n) - \pi|$ and similarly $|\Pi_{n+1} - \Pi_n| = |\pi(n+1) - \pi(n)|$.

Also, for any probability distribution μ on E, $\mu^T\Pi = \pi$, and therefore (15.11) is equivalent to

$$\lim_{k\uparrow\infty} \sup_{\mu} \left|\mu^T \left(\mathbf{P}(m,k) - \Pi\right)\right| = 0,$$

which is in turn implied by

$$\lim_{k\uparrow\infty} |\mathbf{P}(m,k) - \Pi| = 0 \qquad (\star)\,.$$

We therefore proceed to the proof of $(\star)$, writing

$$\begin{aligned}\mathbf{P}(m,k) - \Pi =& \mathbf{P}(m,\ell)\mathbf{P}(\ell,k) - \Pi_{\ell+1}\mathbf{P}(\ell,k)\\ &+ \Pi_{\ell+1}\mathbf{P}(\ell,k) - \Pi_k + \Pi_k - \Pi.\end{aligned}$$

By the triangle inequality for matrix norms,

$$\begin{aligned}|\mathbf{P}(m,k) - \Pi| \le& |\mathbf{P}(m,\ell)\mathbf{P}(\ell,k) - \Pi_\ell\mathbf{P}(\ell,k)|\\ &+ |\Pi_\ell\mathbf{P}(\ell,k) - \Pi_{k-1}| + |\Pi_{k-1} - \Pi| = A + B + C\,.\end{aligned}$$

An upper bound for A is given by Corollary 15.2.6:

$$A \le |\mathbf{P}(m,\ell) - \Pi_\ell|\delta(\mathbf{P}(\ell,k)) \le 2\delta(\mathbf{P}(\ell,k))\,,$$

where the last inequality follows from the definition of the matrix norm used here.

In view of bounding B, we first observe that $\Pi_\ell\mathbf{P}(\ell) = \Pi_\ell$ and therefore

$$\Pi_\ell\mathbf{P}(\ell,k) = (\Pi_\ell - \Pi_{\ell+1})\mathbf{P}(\ell+1,k) + \Pi_{\ell+1}\mathbf{P}(\ell+1,k).$$

Iterating the process, we obtain

$$\Pi_\ell\mathbf{P}(\ell,k) = \sum_{j=\ell+1}^{k-1} (\Pi_{j-1} - \Pi_j)\mathbf{P}(j,k) + \Pi_{k-1},$$

and therefore

$$B \le \sum_{j=\ell+1}^{k-1} |\Pi_{j-1} - \Pi_j|\delta(\mathbf{P}(j,k)) \le \sum_{j=\ell+1}^{k-1} |\pi(j-1) - \pi(j)|\,,$$

where we have used the triangle inequality, Corollary 15.2.6, (15.14), and the observation $|\Pi_{j-1} - \Pi_j| = |\pi(j-1) - \pi(j)|$. As for matrix C, we have, using the last observation,

$$C = |\pi(k-1) - \pi|.$$

The rest of the proof is now clear: For a given $\epsilon > 0$, fix ℓ such that $B \leq \frac{\epsilon}{3}$ for all $k \geq \ell$ (use (15.14)), and take k large enough so that $A \leq \frac{\epsilon}{3}$ (use Dobrushin's inequality) and $C \leq \frac{\epsilon}{3}$ (use $(\star)$). □

It is *not* required that $\mathbf{P}(n)$ be an ergodic stochastic matrix, or that $\pi(n)$ be a unique stationary probability of $\mathbf{P}(n)$.

15.2.3 Bounded Variation Extensions

The question is: How useful is Theorem 15.2.12? It seems that in order to satisfy (15.14), one has to obtain a closed-form expression for $\pi(n)$, or at least sufficient information about $\pi(n)$. How much information? It turns out that very little is needed in practice. More precisely, a qualitative property of $\{\pi(n)\}_{n\geq 0}$ in terms of bounded variation extensions (to be defined) is sufficient to guarantee (15.14), and therefore strong ergodicity, if the chain is weakly ergodic.

We first recall a definition:

A function $f : (0, 1] \to \mathbb{R}$ is said to be of *bounded variation* (BV) if

$$\sup \left\{ \sum_{i=1}^{\infty} |f(x_i) - f(x_{i-1})|; 0 < x_i < \cdots < x_1 = 1 \text{ and } \lim_{i\to\infty} x_i = 0 \right\} < \infty.$$

Similarly, a vector function $\mu : (0, 1] \to \mathbb{R}^E$ is said to be of bounded variation if

$$\sup \left\{ \sum_{i=1}^{\infty} |\mu(x_i) - \mu(x_{i-1})|; 0 < x_i < \cdots < x_i = 1 \text{ and } \lim_{i\to\infty} x_i = 0 \right\} < \infty.$$

Definition 15.2.13 *The vector function $\bar{\pi} : (0, 1] \to \mathbb{R}^E$ is called a bounded variation extension of $\{\pi(n)\}_{n>0}$ if there exists a sequence $\{c_n\}_{n\geq 0}$ in $(0, 1]$ decreasing to 0 and such that $\bar{\pi}(c_n) = \pi(n)$ for all $n \geq 0$.*

Theorem 15.2.14 *(Anily and Federgruen, 1987) Suppose that $\{\mathbf{P}(n)\}_{n\geq 0}$ is weakly ergodic and that for all $n \geq 0$, there exists a probability vector $\pi(n)$ such that $\pi(n)\mathbf{P}(n) = \pi(n)$. If there exists a bounded variation extension $\bar{\pi}(c)$ of $\{\pi(n)\}_{n\geq 0}$, the chain is strongly ergodic.*

Proof. We have

$$\sum_{n\geq 0} |\pi(n+1) - \pi(n)| = \sum_{n\geq 0} |\bar{\pi}(c_{n+1}) - \bar{\pi}(c_n)| < \infty,$$

since $\bar{\pi}(c)$ is an extension of $\{\pi(n)\}_{n\geq 0}$ and of bounded variation, and the conclusion follows by Theorem 15.2.12. □

Let $\bar{\mathbf{P}}(c)$ be an extension of $\{\mathbf{P}(n)\}_{n\geq 0}$, that is, such that there exists a sequence $\{c_n\}_{n\geq 0}$ in $(0,1]$, decreasing to 0 as n goes to infinity and such that for all $n \geq 0$,

$$\bar{\mathbf{P}}(c_n) = \mathbf{P}(n)\,.$$

Suppose that for each $c \in (0,1]$, there exists a probability vector $\bar{\pi}(c)$ such that

$$\bar{\pi}(c)\bar{\mathbf{P}}(c) = \bar{\pi}(c)\,.$$

Is it enough for $\bar{\pi}(c)$ to be of bounded variation that $\bar{\mathbf{P}}(c)$ be of bounded variation, i.e., that

$$\sup\left\{\sum_{i=1}^{\infty} |\bar{\mathbf{P}}(x_{i+1}) - \bar{\mathbf{P}}(x_i)|; 0 < x_i < \cdots < x_1 = 1 \text{ and } \lim_{i\in\infty} x_i = 0\right\} < \infty?$$

A simple counterexample shows that this is not the case.

EXAMPLE 15.2.15: COUNTEREXAMPLE. Let

$$\begin{aligned} P(n) &= \begin{pmatrix} 1-e^{-n} & e^{-n} \\ e^{-n}\sin^2\left(\frac{n\pi}{2}\right) & 1-e^{-n}\sin^2\left(\frac{n\pi}{2}\right)\end{pmatrix}, \\ \bar{\mathbf{P}}(c) &= \begin{pmatrix} 1-e^{-1/c} & e^{-1/c} \\ e^{-1/c}\sin^2\left(\frac{\pi}{2c}\right) & 1-e^{-1/c}\sin^2\left(\frac{\pi}{2c}\right)\end{pmatrix}. \end{aligned}$$

Clearly, $\bar{\mathbf{P}}(c)$ is a bounded variation extension of $\{\mathbf{P}(n)\}_{n\geq 0}$. If the corresponding stationary probability $\bar{\pi}(c)$ were of bounded variation, then as shown in the proof of Theorem 15.2.14, $\sum_{n\geq 0} |\pi(n+1)-\pi(n)|$ would be finite. Computations give for the second coordinate of $\pi(n)$

$$\pi(n)_2 = \left(1+\sin^2\left(\frac{n\pi}{2}\right)\right)^{-1},$$

a quantity that oscillates between 1 and $\frac{1}{2}$. Therefore, $\sum_{n\geq 0} |\pi(n+1)-\pi(n)| = \infty$.

In order to give conditions on $\bar{\mathbf{P}}(c)$ ensuring that $\bar{\pi}(c)$ is of bounded variation, we shall first give a precise description of $\bar{\pi}(c)$ in terms of the entries of $\bar{\mathbf{P}}(c)$. This can be done in the case where E is finite, henceforth identified with $\{1,\ldots,N\}$ for simplicity. Indeed $\bar{\pi}(c)$ is a solution of the balance equations

$$\bar{\pi}(c)_i = \sum_{j=1}^{N} \bar{\mathbf{P}}(c)_{ji}\bar{\pi}(c)_j$$

for $1 \leq i \leq N-1$, together with the normalizing equation

$$\sum_{i=1}^{N} \bar{\pi}(c)_i = 1.$$

That is, in matrix form,

$$\bar{\pi}(c)\begin{pmatrix} 1-\bar{\mathbf{P}}(c) & \cdots & -\bar{\mathbf{P}}(c)_{N-1,1} & -\bar{\mathbf{P}}(c)_{N,1} \\ -\bar{\mathbf{P}}(c) & \cdots & -\bar{\mathbf{P}}(c)_{N-1,2} & -\bar{\mathbf{P}}(c)_{N,2} \\ \vdots & & \vdots & \vdots \\ -\bar{\mathbf{P}}(c)_{1,N-1} & \cdots & 1-\bar{\mathbf{P}}(c)_{N-1,N-1} & -\bar{\mathbf{P}}(c)_1 \\ 1 & & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Cramer's rule gives a solution in the form

$$\bar{\pi}(c)_i = \frac{A_i(c)}{B_i(c)},$$

where $A_i(c)$ and $B_i(c)$ are finite sums and differences of finite products of the entries of $\bar{\mathbf{P}}(c)$.

EXAMPLE 15.2.16: RATIONAL POLYNOMIAL-EXPONENTIAL MATRICES. This case covers most applications. The elements of $\bar{\mathbf{P}}(c)$ are ratios of functions of the type

$$\sum_{j=1}^{n} Q_j\left(\frac{1}{c}\right) e^{\lambda_j/c}, \qquad (\star)$$

where the Q_j are polynomial functions and the λ_j are real numbers. Then so are the elements of $\bar{\pi}(c)$, as well as their derivatives with respect to c. But ratios of terms of type $(\star)$ have for sufficiently small $c > 0$ a constant sign. Therefore a given element $\bar{\pi}(c)_i = \psi(c)$ is such that

(α) for some $c^* > 0, \psi : (0, c*] \to \mathbb{R}$ is monotone and bounded;

(β) $\psi : (0, 1] \to \mathbb{R}$ is continuously differentiable.

Properties (α) and (β) are sufficient to guarantee that $\psi : (0, 1] \to \mathbb{R}$ is of bounded variation.

We have spent some time explaining how the sufficient condition of strong ergodicity (15.14) can be checked. One may wonder whether this is really worthwhile, and whether a weaker and easier to verify condition is available. A natural candidate for a sufficient condition of strong ergodicity, *given weak ergodicity*, is

$$\lim_{n\uparrow\infty} |\pi(n) - \pi| = 0, \qquad (\dagger)$$

for some probability π. This is unfortunately not the case in general, as the following counterexample shows.

EXAMPLE 15.2.17: COUNTEREXAMPLE. Define for all $n \geq 1$,

$$\mathbf{P}(2n-1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \mathbf{P}(2n) = \begin{pmatrix} 0 & 1 \\ 1-\frac{1}{2n} & \frac{1}{2n} \end{pmatrix}.$$

The sequence $\{\mathbf{P}(n)\}_{n\geq 1}$ is weakly ergodic (Exercise 20.4.7). The corresponding stationary distributions are

$$\pi(2n-1) = \left(\frac{1}{2}, \frac{1}{2}\right) \ , \ \pi(2n+1) = \left(\frac{2n-1}{4n-1}, \frac{2n}{4n-1}\right),$$

and therefore (†) is satisfied with $\pi = (\frac{1}{2}, \frac{1}{2})$. On the other hand, if we define for all $k \geq 1$

$$R(k) = \mathbf{P}(2k)\mathbf{P}(2k+1) = \begin{pmatrix} 1 - \frac{1}{2k} & \frac{1}{2k} \\ 0 & 1 \end{pmatrix}$$

and

$$S(k) = \mathbf{P}(2k-1)\mathbf{P}(2k) = \begin{pmatrix} 1 & 0 \\ \frac{1}{2k} & 1 - \frac{1}{2k} \end{pmatrix},$$

then the sequences $\{R(k)\}_{k\geq 1}$ and $\{S(k)\}_{k\geq 1}$ are weakly ergodic (exercise), and their stationary distributions are constant, equal to $(1,0)$ and $(0,1)$, respectively. Therefore, by Theorem 15.2.12, they are strongly ergodic, and in particular,

$$\lim_{k\uparrow} \mathbf{P}(1)\mathbf{P}(2)\cdots\mathbf{P}(2k-1)\mathbf{P}(2k) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$\lim_{k\uparrow} \mathbf{P}(1)\left(\mathbf{P}(2)\mathbf{P}(3)\cdots\mathbf{P}(2k)\mathbf{P}(2k+1)\right) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Therefore, the sequence $\{\mathbf{P}(n)\}_{n\geq 1}$ is not strongly ergodic.

We shall quote without proof the following natural result[3].

Theorem 15.2.18 *Let $\{\mathbf{P}(n)\}_{n\geq 1}$ be a sequence of transition matrices each having at least one stationary distribution, and such that*

$$\lim_{n\uparrow\infty} |\mathbf{P}(n) - \mathbf{P}| = 0 \tag{15.15}$$

for some ergodic transition matrix $\mathbf{P}$. This sequence is strongly ergodic.

The requirement that $\mathbf{P}$ be ergodic will be found too stringent in the study of the convergence of simulated annealing algorithms, where typically the limit transition matrix is reducible.

Books for Further Information

For the algebraic aspects of Markov chains, see [Seneta, 1981]. The theory of discrete non-homogeneous Markov chains is seldom treated in textbooks, with the following exceptions: [Iosifescu, 1980], [Isaacson and Madsen, 1976], and [Seneta, 1981].

[3]Theorem V.4.5 of [Isaacson and Madsen, 1976].

15.3 Exercises

Exercise 15.3.1. Primitive transition matrices
Prove that a non-negative matrix is primitive if and only if it is irreducible and aperiodic.

Exercise 15.3.2. Rate of convergence
Give the convergence rate to equilibrium of the HMC with transition matrix

$$\mathbf{P} = \frac{1}{12}\begin{pmatrix} 0 & 6 & 6 \\ 4 & 3 & 5 \\ 8 & 3 & 1 \end{pmatrix}.$$

Exercise 15.3.3. Rate of convergence of a cyclic matrix
For the homogeneous Markov chain with state space $E = \{1, 2, 3\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 1-\beta & \beta \\ \gamma & 0 & 1-\gamma \end{pmatrix},$$

where $\alpha, \beta, \gamma \in (0, 1)$, compute $\lim_{n\uparrow\infty} \mathbf{P}^n$ and give the corresponding rate of convergence.

Exercise 15.3.4. Sibmating
Example 1.4.13 features a reproduction model among diploid organisms called random mating. We now consider *sibmating* (sister–brother mating), whereby two individuals are mated and two individuals from their offspring are chosen at random to be mated, and this incestuous process goes on through the generations.

We shall denote by X_n the genetic types of the mating pair at the n-th generation. Clearly, $\{X_n\}_{n\geq 0}$ is an HMC with six states representing the different pairs of genotypes $AA \times AA$, $aa \times aa$, $AA \times Aa$, $Aa \times Aa$, $Aa \times aa$, $AA \times aa$, denoted respectively 1, 2, 3, 4, 5, 6.

Find the quasi-stationary distribution in this case.

Exercise 15.3.5. Probability of absorption
Consider the chain with state space $E = \{1, 2, 3, 4, 5, 6, 7\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & & & & & \\ 0.8 & 0.2 & & & & & \\ & & 0 & 0.4 & 0.6 & & \\ & & 1 & 0 & 0 & & \\ & & 1 & 0 & 0 & & \\ 0.1 & 0 & 0.2 & 0.1 & 0.2 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0 & 0.1 & 0.2 & 0.4 \end{pmatrix}.$$

It has two recurrent classes $R_1 = \{1, 2\}$, $R_2 = \{3, 4, 5\}$ and one transient class $T = \{6, 7\}$. Compute the probability of absorption by class $\{3, 4, 5\}$ from transient state 6.

Exercise 15.3.6. SIBMATING
In the reproduction model called *sibmating* (sister–brother mating), two individuals are mated and two individuals from their offspring are chosen at random to be mated, and this incestuous process goes on through the subsequent generations.

Denote by X_n the genetic type of the mating pair at the n-th generation. Clearly, $\{X_n\}_{n\geq 0}$ is an HMC with six states representing the different pairs of genotypes $AA \times AA$, $aa \times aa$, $AA \times Aa$, $Aa \times Aa$, $Aa \times aa$, $AA \times aa$, denoted respectively 1, 2, 3, 4, 5, 6.

Identify the absorbing states and compute the absorption probability matrix.

Exercise 15.3.7. $v(i) = c(i) + \sum_{j\in E} p_{ij} v(j)$
Prove (17.16).

Exercise 15.3.8. TARGET TIME
Let π be the stationary distribution of an ergodic Markov chain with finite state space, and denote by T_i the return time to state i. Let S_Z be the time necessary to visit for the first time the random state Z chosen according to the distribution π, independently of the chain. Show that $E_i[S_Z]$ is independent of i, and give its expression in terms of the fundamental matrix.

Exercise 15.3.9. TRAVEL TIME IN A BIRTH-AND-DEATH PROCESS
Consider the birth-and-death process of Example 7.2.8 with $p = q = \frac{1}{2}$. Compute the travel time from a to $b > a$.

Exercise 15.3.10. QUASI-STATIONARY DISTRIBUTION
With the notation of subsection 6.3 on quasi-stationary distributions, show that for any transient state j,

$$\lim_{m\uparrow\infty, n\uparrow\infty} P(X_n = j | \nu > m + n) = \frac{u_1(j)v_1(j)}{\sum_{i\in T} u_1(i)v_1(i)}.$$

Exercise 15.3.11. DOBRUSHIN'S COEFFICIENT AND THE SLEM
Show that Dobrushin's coefficient is an upper bound of the SLEM.

Exercise 15.3.12. WEAKLY ERGODIC
Define for all $n \geq 1$,

$$\mathbf{P}(2n-1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \mathbf{P}(2n) = \begin{pmatrix} 0 & 1 \\ 1 - \frac{1}{2n} & \frac{1}{2n} \end{pmatrix}.$$

Prove that the sequence $\{\mathbf{P}(n)\}_{n\geq 1}$ is weakly ergodic.

Exercise 15.3.13. STRONGLY ERGODIC AND NOT STRONGLY ERGODIC
A. Prove that the sequence $\{\mathbf{P}(n)\}_{n\geq 0}$ defined by

$$\mathbf{P}(n) = \begin{pmatrix} \frac{1}{3}+\frac{1}{n} & \frac{2}{3}-\frac{1}{n} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

is strongly ergodic.

B. For $n \geq 0$, define

$$\mathbf{P}(2n-1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}, \ \mathbf{P}(2n) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Prove that the NHMC with transition matrices $\{\mathbf{P}(n)\}_{n\geq 0}$ is *not* strongly ergodic. (Hint: take $\mu = (0,1)$ as initial distribution, and compute the distributions of the chain at even and odd times.) Prove that this NHMC is weakly ergodic.

Exercise 15.3.14. ATTRACTION AND BINDING
Let $\mathbf{P}$ and $\mathbf{Q}$ be two transition matrices on the same state space E. Define the *attraction coefficient*

$$\alpha(\mathbf{P},\mathbf{Q}) = 1 - \frac{1}{2}\sup_{i,j\in E\ i\neq j}\sum_{k\in E}|p_{ik}-q_{jk}|$$

and the *binding coefficient*

$$\beta(\mathbf{P},\mathbf{Q}) = 1 - \frac{1}{2}\sup_{i\in E}\sum_{k\in E}|p_{ik}-q_{ik}|.$$

Construct a HMC $\{Z_n\}_{n\geq 0} := \{(X_n,Y_n)\}_{n\geq 0}$ where $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ are HMC's with the respective transition matrices $\mathbf{P}$ and $\mathbf{Q}$, as follows. If $X_n = i$, $Y_n = j$, construct X_{n+1} and Y_{n+1} in such a way that $P(X_{n+1}=k|X_n=i,Y_n=j) = p_{ik}$ and $P(Y_{n+1}=k|X_n=i,Y_n=j) = q_{jk}$, with maximal coupling. In particular, for $i \neq j$,

$$P(X_{n+1}=Y_{n+1}|X_n=i,Y_n=j) = 1 - d_V(p_{i\cdot},q_{j\cdot}) \geq \alpha(\mathbf{P},\mathbf{Q}),$$

and for $i = j$,

$$P(X_{n+1}=Y_{n+1}|X_n=i,Y_n=j) = 1 - d_V(p_{i\cdot},q_{j\cdot}) \geq \beta(\mathbf{P},\mathbf{Q}).$$

Thus, if the chains do not coincide at time n, they will at time $n+1$ with a probability at least $\alpha(\mathbf{P},\mathbf{Q})$, whence the appellation *attraction coefficient*. Similarly, if the chains coincide at time n, they will still coincide at time $n+1$ with a probability at least $\beta(\mathbf{P},\mathbf{Q})$, whence the appellation *binding coefficient*.

Prove that

$$\lim_{N\uparrow\infty}\frac{1}{N}\sum_{k=1}^{N}1_{\{X_k=Y_k\}} \geq \frac{\alpha(\mathbf{P},\mathbf{Q})}{1+\alpha(\mathbf{P},\mathbf{Q})-\beta(\mathbf{P},\mathbf{Q})}.$$

Chapter 16

The Coupling Method

16.1 Coupling Inequalities

16.1.1 Coupling and the Variation Distance

Coupling is a ubiquitous and versatile idea with many applications. In its simplest form, it consists in materializing two probability distributions of a single random element by two random elements with the said distributions, not independent, but correlated in a way that allows easy comparison of these distributions, in view for instance of determining if some event of interest is more probable under one distribution or the other. This technique will be applied to objects more sophisticated than random variables, such as random graphs (see the first example) or stochastic processes (see the proof of the fundamental theorem of the theory of Markov chains in section 16.2). The coupling method is often linked to the notion of variation distance, with applications to, for instance, the speed of convergence of the distribution of a stochastic process to its limit distribution, when the latter exists and is unique. Other more elaborate applications of the method will be seen, starting with Chen's approximation of a distribution on the integers by a Poisson distribution (section 16.3).

Coupling

Coupling two discrete probability distributions π' on E' and π'' on E'' consists in the construction of a probability distribution π on $E := E' \times E''$ such that the marginal distributions of π on E' and E'' are respectively π' and π'', that is,

$$\sum_{j \in E''} \pi(i,j) = \pi'(i) \text{ and } \sum_{i \in E'} \pi(i,j) = \pi''(j)\,.$$

Example 16.1.1: Coupled random graphs and monotone properties. Let $\mathbf{E_n}$ be the collection of all possible edges on the set of vertices $V = \{1, 2, \ldots, n\}$. Let be given a family of IID random variables $\{U_{\langle v,w\rangle}\}_{\langle v,w\rangle \in \mathbf{E_n}}$ uniformly distributed on $(0,1)$. A random graph $\mathcal{G}(n,p)$ may be generated as follows: admit $\langle v,w\rangle \in \mathbf{E_n}$ as an edge of $\mathcal{G}(n,p)$ if and only if $U_{\langle v,w\rangle} \leq p$.

A canonical coupling construction of the random graphs $\mathcal{G}(n,p_1), \mathcal{G}(n,p_2), \ldots, \mathcal{G}(n,p_k)$ on the set $V = \{1,2,\ldots,n\}$ of vertices is one for which an edge $\langle v,w\rangle \in \mathbf{E_n}$ is

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_16

admitted as an edge of $\mathcal{G}(n,p_i)$ if and only if $U_{\langle v,w\rangle} \le p_i$ $(1 \le i \le k)$. Clearly, if $p_i \le p_j$, then $\mathcal{G}(n,p_i) \subseteq \mathcal{G}(n,p_j)$.

Let $\mathcal{P}$ be some graph property. For instance, to be connected, to be a tree, to have no cliques of four vertices. If a graph G has this property, we write: $G \in \mathcal{P}$. A property $\mathcal{P}$ of graphs with the same set of vertices is called monotone increasing if whenever a graph G has this property, so does any graph $G' \supseteq G$. A monotone decreasing property is defined similarly, *mutatis mutandis*.

The canonical coupling of two random graphs $\mathcal{G}_{n,m_1}$ and $\mathcal{G}_{n,m_2}$ where $m_2 > m_1$ is the obvious one: Construct $\mathcal{G}_{n,m_1}$ and choose randomly in $\mathbf{E_n}\backslash\mathcal{E}_{n,m_1}$ the $m_2 - m_1$ edges to be added.

The following result is a direct consequence of the canonical coupling constructions:

Theorem 16.1.2 *Let $\mathcal{P}$ be a monotone increasing graph property. Then*

$$p' \ge p \Longrightarrow P(\mathcal{G}(n,p') \in \mathcal{P}) \ge P(\mathcal{G}(n,p) \in \mathcal{P})\,,$$

and

$$m_2 \ge m_1 \Longrightarrow P(\mathcal{G}_{n,m_2} \in \mathcal{P}) \ge P(\mathcal{G}_{n,m_1} \in \mathcal{P})\,.$$

Distance in Variation

Let E be a countable space. The distance in variation between two probability distributions α and β on E is the quantity

$$d_V(\alpha,\beta) := \frac{1}{2}\sum_{i\in E} |\alpha(i) - \beta(i)|. \tag{16.1}$$

That d_V is indeed a distance is clear.

Lemma 16.1.3 *Let α and β be two probability distributions on the same countable space E. Then*

$$d_V(\alpha,\beta) = \sup_{A\subseteq E}\{\alpha(A) - \beta(A)\} = \sup_{A\subseteq E}\{|\alpha(A) - \beta(A)|\}\,.$$

Proof. For the second equality observe that for each subset A there is a subset B such that $|\alpha(A) - \beta(A)| = \alpha(B) - \beta(B)$ (take $B = A$ or $\bar{A}$). For the first equality, write

$$\alpha(A) - \beta(A) = \sum_{i\in E} 1_A(i)\{\alpha(i) - \beta(i)\}$$

and observe that the right-hand side is maximal for $A = \{i \in E;\ \alpha(i) > \beta(i)\}$. Therefore, with $g(i) =:= \alpha(i) - \beta(i)$,

$$\sup_{A\subseteq E}\{\alpha(A) - \beta(A)\} = \sum_{i\in E} g^+(i) = \frac{1}{2}\sum_{i\in E} |g(i)|\ ,$$

where the equality $\sum_{i\in E} g(i) = 0$ was taken into account. □

The distance in variation between two random variables X and Y with values in E is the distance in variation between their probability distributions, and it is denoted (with a slight abuse of notation) by $d_V(X, Y)$. Therefore

$$d_V(X, Y) := \frac{1}{2}\sum_{i\in E} |P(X = i) - P(Y = i)| .$$

The distance in variation between a random variable X with values in E and a probability distribution α on E denoted (again with a slight abuse of notation) by $d_V(X, \alpha)$ is defined by

$$d_V(X, \alpha) := \frac{1}{2}\sum_{i\in E} |P(X = i) - \alpha(i)| .$$

Variation Distance and Hypothesis Testing

Suppose that we have to discriminate between two equiprobable hypotheses H_1 and H_2 concerning the distribution of a discrete random variable X taking values in E. Under H_1, the distribution is α, under H_2, it is β. For deciding which is the actual distribution of X, we use a partition (A, B) of E, deciding for α (resp., β) if X falls in A (resp., B). By doing so, we may have made the wrong guess. In fact, the probability of error is

$$\begin{aligned} P_E &= \frac{1}{2}P(X \in B \,|\, H_1) + \frac{1}{2}P(X \in A \,|\, H_2) \\ &= \frac{1}{2}\alpha(B) + \frac{1}{2}\beta(A) = \frac{1}{2}(1 - (\alpha(A) - \beta(A))) . \end{aligned}$$

The probability of error is minimal for a choice of A that maximizes $\alpha(A) - \beta(A)$. This occurs for the choice $A^* := \{i \in E \,|\, \alpha(i) \geq \beta(i)\}$ which gives for the minimal probability of error

$$P_E^* = \frac{1}{2}(1 - d_V(\alpha, \beta)) .$$

16.1.2 The First Coupling Inequality

For two probability distributions α and β on the countable set E, let $\mathcal{D}(\alpha, \beta)$ be the collection of pairs of random variables (X, Y) taking their values in $E \times E$, and with marginal distributions α and β, that is,

$$P(X = i) = \alpha(i),\ P(Y = i) = \beta(i) . \tag{16.2}$$

Theorem 16.1.4 *For any pair $(X, Y) \in \mathcal{D}(\alpha, \beta)$, we have the first coupling inequality*

$$d_V(\alpha, \beta) \leq P(X \neq Y), \tag{16.3}$$

and equality is attained by some pair $(X, Y) \in \mathcal{D}(\alpha, \beta)$, which is then said to realize maximal coincidence.

Proof. For arbitrary $A \subset E$,

$$P(X \neq Y) \geq P(X \in A, Y \in \bar{A}) = P(X \in A) - P(X \in A, Y \in A)$$
$$\geq P(X \in A) - P(Y \in A),$$

and therefore

$$P(X \neq Y) \geq \sup_{A \subset E} \{P(X \in A) - P(Y \in A)\} = d_V(\alpha, \beta).$$

We now construct $(X, Y) \in \mathcal{D}(\alpha, \beta)$ realizing equality. Let U, Z, V and W be independent random variables; U takes its values in $\{0, 1\}$, and Z, V, W take their values in E. The distributions of these random variables are given by

$$\begin{aligned} P(U = 1) &= 1 - d_V(\alpha, \beta), \\ P(Z = i) &= (\alpha(i) \wedge \beta(i))/\left(1 - d_V(\alpha, \beta)\right), \\ P(V = i) &= (\alpha(i) - \beta(i))^+/d_V(\alpha, \beta), \\ P(W = i) &= (\beta(i) - \alpha(i))^+/d_V(\alpha, \beta). \end{aligned}$$

Observe that $P(V = W) = 0$. Defining

$$(X, Y) = (Z, Z) \text{ if } U = 1$$
$$= (V, W) \text{ if } U = 0,$$

we have

$$\begin{aligned} P(X = i) &= P(U = 1, Z = i) + P(U = 0, V = i) \\ &= P(U = 1)P(Z = i) + P(U = 0)P(V = i) \\ &= \alpha(i) \wedge \beta(i) + (\alpha(i) - \beta(i))^+ = \alpha(i), \end{aligned}$$

and similarly, $P(Y = i) = \beta(i)$. Therefore, $(X, Y) \in \mathcal{D}(\alpha, \beta)$. Also, $P(X = Y) = P(U = 1) = 1 - d_V(\alpha, \beta)$, that is $P(X \neq Y) = d_V(\alpha, \beta)$. □

EXAMPLE 16.1.5: VARIATION DISTANCE OF TWO POISSON VARIABLES. Let λ and μ be two positive real numbers. Then

$$d_V(p_\lambda, p_\mu) \leq 1 - e^{-|\mu - \lambda|},$$

where p_α denotes the Poisson distribution with mean α. We prove this, assuming without loss of generality that $\mu > \lambda$. Let X be a Poisson variable with mean λ and let Z be a Poisson variable with mean $\mu - \lambda$ independent of X. In particular, $Y = X + Z$ is a Poisson variable with mean μ. Therefore $d_V(p_\lambda, p_\mu) = d_V(X, X+Z)$. Therefore, by Theorem 16.1.4,

$$d_V(p_\lambda, p_\mu) \leq P(X \neq X + Z) = P(Z > 0) = 1 - e^{-(\mu - \lambda)} = 1 - e^{-|\mu - \lambda|}.$$

EXAMPLE 16.1.6: POISSON LAW OF RARE EVENTS, TAKE 2. (Le Cam, 1960) Let $Y_1, \dots, Y_n$ be independent random variables taking their values in $\{0, 1\}$, with $P(Y_i = 1) = \pi_i$ $(1 \le i \le n)$. Let $X := \sum_{i=1}^n Y_i$ and $\lambda := \sum_{i=1}^n \pi_i$. Let p_λ be the Poisson distribution with mean λ. In order to bound the variation distance between the distribution q of X and p_λ, construct a coupling of the two distributions as follows. First, generate independent pairs $(Y_1, Y_1'), \dots, (Y_n, Y_n')$ such that

$$P(Y_i = j, Y_i' = k) = \begin{cases} 1 - \pi_i & \text{if } j = 0, k = 0, \\ e^{-\pi_i} \frac{\pi_i^k}{k!} & \text{if } j = 1, k \ge 1, \\ e^{-\pi_i} - (1 - \pi_i) & \text{if } j = 1, k = 0\,. \end{cases}$$

One verifies that for all $1 \le i \le n$, $P(Y_i = 1) = \pi_i$ and $Y_i' \sim \text{Poi}(\pi_i)$. In particular, $X' := \sum_{i=1}^n Y_i'$ is a Poisson variable with mean λ. Now

$$\begin{aligned} P(X \ne X') &= P\left(\sum_{i=1}^n Y_i \ne \sum_{i=1}^n Y_i'\right) \\ &\le P\left(Y_i \ne Y_i' \text{ for some } i\right) \le \sum_{i=1}^n P\left(Y_i \ne Y_i'\right)\,. \end{aligned}$$

But

$$\begin{aligned} P\left(Y_i \ne Y_i'\right) &= e^{-\pi_i} - (1 - \pi_i) + \sum_{k \ge 2} e^{-\pi_i} \frac{\pi_i^k}{k!} \\ &= \pi_i \left(1 - e^{-\pi_i}\right) \le \pi_i^2\,. \end{aligned}$$

Therefore $P(X \ne X') \le \sum_{i=1}^n \pi_i^2$ and by the coupling inequality

$$d_V(q, p_\lambda) \le \sum_{i=1}^n \pi_i^2\,.$$

Remark 16.1.7 Observe that $\sum_{i=1}^n \pi_i^2 = \lambda - \text{Var}(X)$ and that if X is a Poisson variable with mean λ, then $\text{Var}(X) = \lambda$, and therefore in this case, the bound is the tightest possible.

Remark 16.1.8 With $\pi_i = p := \frac{\lambda}{n}$, we have

$$d_V(q, p_\lambda) \le \frac{\lambda^2}{n}\,.$$

In other terms the binomial distribution of size n and mean λ differs in variation from a Poisson variable with the same mean of less than $\frac{\lambda^2}{n}$. Le Cam's inequality is therefore a refinement of the elementary Poisson law of rare events since it gives an exploitable estimate for finite n.

16.1.3 The Second Coupling Inequality

Definition 16.1.9 *(A) A sequence $\{\alpha_n\}_{n\geq 0}$ of probability distributions on E is said to converge in variation to the probability distribution β on E if*

$$\lim_{n\uparrow\infty} d_V(\alpha_n, \beta) = 0\,.$$

(B) An E-valued random sequence $\{X_n\}_{n\geq 0}$ such that for some probability distribution π on E,

$$\lim_{n\uparrow\infty} d_V(X_n, \pi) = 0\,, \tag{16.4}$$

is said to converge in variation to π.

EXAMPLE 16.1.10: CONVERGENCE IN VARIATION OF POISSON VARIABLES. Let p_λ denote the Poisson distribution with mean λ. If $\{\lambda_n\}_{n\geq 1}$ is a sequence of positive real numbers converging to $\lambda > 0$,

$$\lim_{n\uparrow\infty} d_V(p_\lambda, p_{\lambda_n}) = 0\,.$$

But from Example 16.1.5, $d_V(p_\lambda, p_{\lambda_n}) \leq 1 - e^{-|\lambda_n - \lambda|}$.

Definition 16.1.11 *Two stochastic processes $\{X'_n\}_{n\geq 0}$ and $\{X''_n\}_{n\geq 0}$ taking their values in the same state space E are said to couple if there exists an almost surely finite random time τ such that*

$$n \geq \tau \Rightarrow X'_n = X''_n. \tag{16.5}$$

The random variable τ is called a coupling time of the two processes.

Theorem 16.1.12 *For any coupling time τ of $\{X'_n\}_{n\geq 0}$ and $\{X''_n\}_{n\geq 0}$, we have the second coupling time inequality*

$$d_V(X'_n, X''_n) \leq P(\tau > n)\,. \tag{16.6}$$

Proof. For all $A \subseteq E$,

$$\begin{aligned} P(X'_n \in A) - P(X''_n \in A) &= P(X'_n \in A,\ \tau \leq n) + P(X'_n \in A,\ \tau > n) \\ &- P(X''_n \in A,\ \tau \leq n) - P(X''_n \in A, \tau > n) \\ &= P(X'_n \in A,\ \tau > n) - P(X''_n \in A,\ \tau > n) \\ &\leq P(X'_n \in A,\ \tau > n) \leq P(\tau > n). \end{aligned}$$

Inequality (16.6) then follows from Lemma 16.1.3. □

Theorem 16.1.12 will be exploited in the proof of the limit theorem for Markov chains (Theorem 15.1.1).

16.2 Limit Distribution via Coupling

16.2.1 Doeblin's Idea

The original idea is that of (Doeblin, 1937); see also (Pitman, 1976) and (Griffeath, 1978).

Observe that Definition 16.1.9 concerns only the marginal distributions of the stochastic process, not the stochastic process itself. Therefore, if there exists another stochastic process $\{X'_n\}_{n\geq 0}$ such that $X_n \overset{\mathcal{D}}{\sim} X'_n$ for all $n \geq 0$, and if there exists a third one $\{X''_n\}_{n\geq 0}$ such that $X''_n \overset{\mathcal{D}}{\sim} \pi$ for all $n \geq 0$, then (16.4) follows from

$$\lim_{n\uparrow\infty} d_V(X'_n, X''_n) = 0. \tag{16.7}$$

This trivial observation is useful because of the resulting freedom in the choice of $\{X'_n\}$ and $\{X''_n\}$. An interesting situation occurs when there exists a finite random time τ such that $X'_n = X''_n$ for all $n \geq \tau$.

Proof. We prove that, for all probability distributions μ and ν on E,

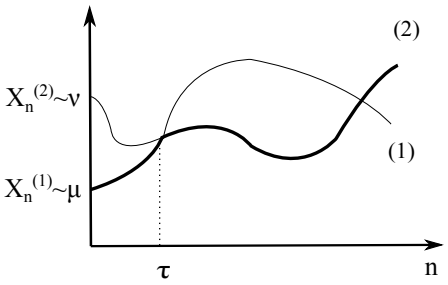
$$\lim_{n\uparrow\infty} d_V(\mu^T\mathbf{P}^n, \nu^T\mathbf{P}^n) = 0.$$

The announced results correspond to the particular case where ν is the stationary distribution π, and particularizing further, $\mu = \delta_j$. From the discussion preceding Definition 16.1.11, it suffices to construct two coupling chains with initial distributions μ and ν, respectively. This is done in the next theorem. □

Theorem 16.2.1 *Let $\{X_n^{(1)}\}_{n\geq 0}$ and $\{X_n^{(2)}\}_{n\geq 0}$ be two independent ergodic* HMC*s with the same transition matrix $\mathbf{P}$ and initial distributions μ and ν, respectively. Let $\tau = \inf\{n \geq 0;\ X_n^{(1)} = X_n^{(2)}\}$, with $\tau = \infty$ if the chains never intersect. Then τ is, in fact, almost surely finite. Moreover, the process $\{X'_n\}_{n\geq 0}$ defined by*

$$X'_n = \begin{cases} X_n^{(1)} & \text{if } n \leq \tau, \\ X_n^{(2)} & \text{if } n \geq \tau \end{cases} \tag{16.8}$$

is an HMC *with transition matrix $\mathbf{P}$ (see the figure below).*



Proof. STEP 1. Consider the product HMC $\{Z_n\}_{n\geq 0}$ defined by $Z_n = (X_n^{(1)}, X_n^{(2)})$. It takes values in $E \times E$, and the probability of transition from (i,k) to (j,ℓ) in n steps is $p_{ij}(n)p_{k\ell}(n)$. We first show that this chain is irreducible. The probability of transition from (i,k) to (j,ℓ) in n steps is $p_{ij}(n)p_{k\ell}(n)$. Since $\mathbf{P}$ is irreducible and *aperiodic*, by Theorem 6.1.21, there exists m such that for all pairs (i,j) and (k,ℓ), $n \geq m$ implies $p_{ij}(n)p_{k\ell}(n) > 0$. This implies irreducibility. (Note the essential role of aperiodicity. A simple counterexample is that of the symmetric random walk on $\mathbb{Z}$, which is irreducible but of period 2. The product of two independent such HMC's is the symmetric random walk on $\mathbb{Z}^2$ which has two communication classes.)

STEP 2. Next we show that the two independent chains meet in finite time. Clearly, the distribution $\tilde{\sigma}$ defined by $\tilde{\sigma}(i,j) := \pi(i)\pi(j)$ is a stationary distribution for the product chain, where π is the stationary distribution of $\mathbf{P}$. Therefore, by the stationary distribution criterion, the product chain is positive recurrent. In particular, it reaches the diagonal of E^2 in finite time, and consequently, $P(\tau < \infty) = 1$.

It remains to show that $\{X'_n\}_{n\geq 0}$ given by (16.8) is an HMC with transition matrix $\mathbf{P}$. For this we use the following lemma.

Lemma 16.2.2 *Let $X_0^1, X_0^2, Z_n^1, Z_n^2$ $(n \geq 1)$ be independent random variables, and suppose moreover that Z_n^1, Z_n^2 $(n \geq 1)$ are identically distributed. Let τ be a non-negative integer-valued random variable such that for all $m \in \mathbb{N}$, the event $\{\tau = m\}$ is expressible in terms of $X_0^1, X_0^2, Z_n^1, Z_n^2\, (n \leq m)$. Define the sequence $\{Z_n\}_{n\geq 1}$ by*

$$\begin{aligned} Z_n &= Z_n^1 \text{ if } n \leq \tau\,, \\ &= Z_n^2 \text{ if } n > \tau\,. \end{aligned}$$

Then, $\{Z_n\}_{n\geq 1}$ has the same distribution as $\{Z_n^1\}_{n\geq 1}$ and is independent of X_0^1, X_0^2.

Proof. For any sets C_1, C_2, A_1, …, A_k in the appropriate spaces,

$$\begin{aligned} &P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell \in A_\ell, 1 \leq \ell \leq k) \\ &= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell \in A_\ell, 1 \leq \ell \leq k, \tau = m) \\ &\quad + P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1 \in A_1, \ldots, Z_k \in A_k, \tau > k) \\ &= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq m, \tau = m, Z_r^2 \in A_r, m+1 \leq r \leq k) \\ &\quad + P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \leq \ell \leq k, \tau > k)\,. \end{aligned}$$

Since the event $\{\tau = m\}$ is independent of $Z_{m+1}^2 \in A_{m+1}, \ldots, Z_k^2 \in A_k$ $(k \geq m)$, this is equal to

$$\sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \le \ell \le m, \tau = m) P(Z_r^2 \in A_r, m+1 \le r \le k)$$
$$+ P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \le \ell \le k, \tau > k)$$
$$= \sum_{m=0}^{k} P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \le \ell \le m, \tau = m, Z_r^1 \in A_r, m+1 \le r \le k)$$
$$+ P(X_0^1 \in C_1, X_0^2 \in C_2, Z_\ell^1 \in A_\ell, 1 \le \ell \le k, \tau > k)$$
$$= P(X_0^1 \in C_1, X_0^2 \in C_2, Z_1^1 \in A_1, \ldots, Z_k^1 \in A_k)\,.$$

□

STEP 3. We now complete the proof. The statement of the theorem concerns only the distributions of $\{X_n^1\}_{n\ge 0}$ and $\{X_n^2\}_{n\ge 0}$, and therefore we can assume a representation

$$X_{n+1}^\ell = f(X_n^\ell, Z_{n+1}^\ell) \quad (\ell = 1, 2)\,,$$

where $X_0^1, X_0^2, Z_n^1, Z_n^2$ $(n \ge 1)$ satisfy the conditions stated in Lemma 16.2.2. The random time τ satisfies the condition of Lemma 16.2.2. Defining $\{Z_n\}_{n\ge 1}$ in the same manner as in this lemma, we therefore have

$$X_{n+1} = f(X_n, Z_{n+1})\,,$$

which proves the announced result. □

16.2.2 The Null Recurrent Case

Theorem 15.1.1 concerns the positive recurrent case. The proof of the null recurrent version requires more care:

Theorem 16.2.3 *(Orey, 1971) Let* $\mathbf{P}$ *be an irreducible null recurrent transition matrix on* E*. Then for all* $i, j \in E$*,*

$$\lim_{n\uparrow\infty} p_{ij}(n) = 0\,. \tag{16.9}$$

Proof. The periodic case follows from the aperiodic case by considering the restriction of $\mathbf{P}^d$ to C_0, an arbitrary cyclic class, and observing that this restriction is also null recurrent. Therefore, $\mathbf{P}$ will be assumed aperiodic.

In this case, we have seen that the product HMC $\{Z_n\}_{n\ge 0} = \{X_n^{(1)}, X_n^{(2)})\}_{n\ge 0}$ is irreducible and aperiodic. However, it cannot be argued that it is recurrent, even if each of its components is recurrent. One must therefore separate the two possible cases.

First, suppose the product chain transient. Its n-step transition probability from (i, i) to (j, j) is $[p_{ij}(n)]^2$, and it tends to 0 as $n \to \infty$. The result is therefore proved in this particular case.

Suppose now that the product chain is recurrent. The coupling argument used in the aperiodic case applies and yields

$$\lim_{n\uparrow\infty} |\mu^T \mathbf{P}^n - \nu^T \mathbf{P}^n| = 0 \tag{16.10}$$

for arbitrary initial distributions μ and ν. Suppose now that for some $i, j \in E$, (16.9) is not true. One can then find a sequence $\{n_k\}_{k\geq 0}$ of integers strictly increasing to ∞ such that

$$\lim_{k\uparrow\infty} p_{ij}(n_k) = \alpha > 0\,.$$

For fixed $i \in E$ chosen as above, the sequence $(\{p_{is}(n_k), s \in E\})_{k\geq 0}$ of vectors of $[0,1]^E$ is compact in the topology of pointwise convergence. Therefore (see Theorem 1.10 of the Appendix for an elementary proof), there exists a subsequence $\{m_\ell\}_{\ell\geq 0}$ of integers strictly increasing to ∞ and a vector $\{x_s, s \in E\} \in [0,1]^E$ such that for all $s \in E$,

$$\lim_{\ell\uparrow\infty} p_{is}(m_\ell) = x_s\,.$$

Now, $x_j = \alpha > 0$, and therefore $\{x_s, s \in E\}$ is nontrivial. Since $\sum_{s\in E} p_{is}(m_\ell) = 1$, it follows from Fatou's lemma that $\sum_{s\in E} x_s \leq 1$. Apply Fatou's lemma to the right-hand side of

$$p_{is}(m_\ell + 1) = \sum_{k\in E} p_{ik}(m_\ell) p_{ks}$$

to obtain

$$x_s \geq \sum_{k\in E} x_k p_{ks}\,.$$

Summing with respect to s:

$$\sum_{s\in E} x_s \geq \sum_{s\in E}\sum_{k\in E} x_k p_{ks} = \sum_{k\in E}\left(x_k \sum_{s\in E} x_k p_{ks}\right) = \sum_{k\in E} x_k\,.$$

Therefore the inequality can only be an equality. In other words, $\{x_s, s \in E\}$ is an invariant measure of $\mathbf{P}$. It has finite mass, which implies that $\mathbf{P}$ is positive recurrent, a contradiction. Therefore, (16.9) cannot be contradicted. □

16.3 Poisson Approximation

16.3.1 Chen's Variation Distance Bound

Let $Y_i \sim Bern(\pi_i)$ $(1 \leq i \leq n)$ be Bernoulli variables, and define $X := \sum_{i=1}^n Y_i$, a random variable with mean $\lambda := \sum_{i=1}^n \pi_i$. This is the situation considered in Example 16.1.6, except that the independence assumption for the variables Y_i is now replaced by the following one:

Assumption H: There exist random variables U_k and V_k $(1 \leq k \leq n)$ defined on the same probability space and such that

(i) U_k has the distribution of $X = \sum_{i=1}^n Y_i$, and

(ii) $1 + V_k$ has the distribution of X conditioned by $Y_k = 1$.

Theorem 16.3.1 *(Chen, 1975) Under Assumption H,*

$$d_V(X, p_\lambda) \leq \left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{k=1}^{n} \pi_k E\left[|U_k - V_k|\right] . \tag{16.11}$$

Since $\frac{1-e^{-\lambda}}{\lambda} \leq 1 \wedge \lambda^{-1}$, the above bound sometimes appears in the form

$$d_V(X, p_\lambda) \leq \left(1 \wedge \lambda^{-1}\right) \sum_{k=1}^{n} \pi_k E\left[|U_k - V_k|\right] .$$

The following corollary improves the result of Example 16.1.6 when λ is large.

EXAMPLE 16.3.2: RECOVERING LE CAM'S THEOREM. Under the conditions prevailing in Example 16.1.6 (including the independence of the Y_i's), we have that

$$d_V(X, p_\lambda) \leq \left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{i=1}^{n} \pi_i^2 . \tag{16.12}$$

Proof. Due to the independence hypothesis, Assumption H is satisfied with $U_k = X$ and $V_k = \sum_{i \neq k} Y_i = X - Y_k$. Inequality (16.16) then follows from (16.11) since $E\left[|U_k - V_k|\right] = E\left[Y_k\right] = \pi_k$. □

Corollary 16.3.3 *Suppose that there exists for each k $(1 \leq k \leq n)$ a collection of random variables Z_{kj} $(1 \leq j \leq n, j \neq k)$ whose distribution is the same as that of Y_j $(1 \leq j \leq n, j \neq k)$ conditioned by $Y_k = 1$. Then*

$$d_V(X, p_\lambda) \leq \left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{k=1}^{n} \pi_k \left(\pi_k + \sum_{1 \leq j \leq n,\, j \neq k} E\left[|Y_j - Z_{kj}|\right]\right) . \tag{16.13}$$

Proof. The requirements of Theorem 16.3.1 are met with $V_k = \sum_{1 \leq j \leq n,\, j \neq k} Z_{kj}$. In this case, we have

$$\begin{aligned} E\left[|U_k - V_k|\right] &= E\left[\left|\sum_{1 \leq j \leq n} Y_k - \sum_{1 \leq j \leq n,\, j \neq k} Z_{kj}\right|\right] \\ &\leq E\left[Y_k\right] + \sum_{1 \leq j \leq n,\, j \neq k} E\left[|Y_j - Z_{kj}|\right] = \pi_k + \sum_{1 \leq j \leq n,\, j \neq k} E\left[|Y_j - Z_{kj}|\right] . \end{aligned}$$

□

EXAMPLE 16.3.4: ISOLATED NODES IN ERDÖS–RÉNYI RANDOM GRAPHS. Consider the random graph $\mathcal{G}(n, p_n)$ with

$$np_n = \log n + c$$

for some $c > 0$. Let X be the number of isolated nodes in this graph. This random variable converges in variation as $n \uparrow \infty$ to a Poisson variable with mean e^{-c}.

Proof. Let $V := \{1, 2, \ldots, n\}$ be the set of vertices. The random graph $\mathcal{G}(n, p)$ can be constructed from an IID family of $\{0,1\}$-valued random variables $\{\xi_{ij}\}_{1 \le i,j \le n}$ with $P(\xi_{ij} = 1) = p := p_n$, where $\xi_{ij} = 1$ indicates that vertices i and j are connected. The indicator Y_j of the event that vertex j is isolated can then be represented as

$$Y_j = \prod_{u;\, u \neq j} (1 - \xi_{ju}).$$

The number of isolated nodes is $X = \sum_{i=1}^{n} Y_i$. Let

$$Z_{kj} := \prod_{u;\, u \notin \{j,k\}} (1 - \xi_{ju}) \,.$$

Conditioning on $Y_k = 1$ is equivalent to conditioning on the event that $\xi_{ku} = 0$ for all $u \neq k$. By independence of the edge variables, conditioning on this event does not change the distribution of the edge variables other than those involved in the said event. In particular, the distribution of $(Z_{kj}, 1 \le j \le n, j \neq k)$ is the same as that of $(Y_j, 1 \le j \le n, j \neq k)$ conditioned by $Y_k = 1$. We are therefore in the framework of Theorem 16.3.3, which says

$$d_V(X, p_\lambda) \le \left(\frac{1 - e^{-\lambda}}{\lambda}\right) \sum_{k=1}^{n} \pi_k \left(\pi_k + \sum_{1 \le i \le n,\, i \neq k} E\left[|Y_j - Z_{kj}|\right]\right) .$$

We now identify the quantities involved in the previous inequality. First $\pi_k = (1-p)^{n-1} := \pi$ and $\lambda = E[X] = n\pi$. Observing that $|Y_j - Z_{kj}| = \xi_{kj} \prod_{u;\, u \notin \{k,j\}} (1 - \xi_{uj})$, we have that $E\left[|Y_j - Z_{kj}|\right] = p(1-p)^{n-2}$. Therefore

$$\begin{aligned} d_V(X, p_\lambda) &\le \left(\frac{1 - e^{-\lambda}}{\lambda}\right) n\pi \left[\pi + (n-1)p(1-p)^{n-2}\right] \\ &\le \left(\frac{1 - e^{-\lambda}}{\lambda}\right) \lambda \left[\pi + (n-1)p(1-p)^{n-2}\right] \\ &= \left(1 - e^{-\lambda}\right) \left[\pi + \frac{(n-1)p\pi}{1-p}\right], \end{aligned}$$

and finally

$$d_V(X, p_\lambda) \le \left(\pi + \lambda \frac{p}{1-p}\right). \qquad (\star)$$

With $p = p_n$,

$$\lambda = n\pi = n\left(1 - \frac{\log n + c}{n}\right)^{n-1} \to e^{-c},$$

so that the upper bound in $(\star)$ is of the order of $\frac{\log n}{n}$, and therefore $d_V(X, p_\lambda) \to 0$. By the triangle inequality

$$d_V(X, p_{e^{-c}}) \leq d_V(X, p_\lambda) + d_V(p_{e^{-c}}, p_\lambda)\,.$$

But $d_V(p_{e^{-c}}, p_\lambda) \to 0$ (Example 16.1.5) and the result then follows. □

Example 16.3.4 showed how Theorem 16.3.1 allows one to treat cases where the independence assumption is relaxed. It can also be viewed as an application of the following general result:

Theorem 16.3.5 *Let be given for each $i \in \{1, 2, \ldots, n\}$ a set $\mathcal{N}_i \subseteq \{1, 2, \ldots, n\} \backslash \{i\}$ such that Y_i is independent of $(Y_j\,;\, j \notin \mathcal{N}_i \cup \{i\})$. The Y_i's are then said to be dissociated relatively to the neighbourhoods $\mathcal{N}_i \in \{1, 2, \ldots, n\}$. Then*

$$d_V(X, p_\lambda) \leq \left(\frac{1 - e^{-\lambda}}{\lambda}\right) \sum_{i=1}^{n} \left(\pi_i^2 + \sum_{j \in \mathcal{N}_i} (\pi_i \pi_j + E\,[Y_i Y_j])\right)\,. \tag{16.14}$$

Proof. Apply Theorem 16.3.1 with $U_k = X$ and

$$V_k = \sum_{j \notin \mathcal{N}_k \cup \{k\}} Y_j + \sum_{j \in \mathcal{N}_k} Y_j^{(k)}\,,$$

where

$$Y_j^{(k)} \sim X_j \mid X_k = 1\,.$$

The announced result follows from Theorem 16.3.1 and the bound

$$\begin{aligned} E\,[|U_k - V_k|] &= E\left[|Y_k + \sum_{j \in \mathcal{N}_k} (Y_j - Y_j^{(k)})|\right] \\ &\leq E\,[|Y_k|] + \sum_{j \in \mathcal{N}_k} \left(E\,[|Y_j|] + E\left[|Y_j^{(k)}|\right]\right) \\ &= \pi_k + \sum_{j \in \mathcal{N}_k} (\pi_j + E\,[Y_j \mid Y_k = 1]) \\ &= \pi_k + \sum_{j \in \mathcal{N}_k} \left(\pi_j + \frac{E\,[Y_j Y_k]}{\pi_k}\right)\,. \end{aligned}$$

□

Example 16.3.6: Wedges in the Erdös–Rényi random graph. Consider the random graph $\mathcal{G}(n, p)$. A wedge of a graph $G = (V, \mathcal{E})$ is a pair $(u, \{v, w\})$ where the vertices u, v, w are distinct, $u \sim v$, $u \sim w$ ($u \sim v$ means that $\langle u, v\rangle$ is an edge of G). The order of v and w is irrelevant, that is $(u, \{v, w\})$ and $(u, \{w, v\})$ is the same wedge, henceforth denoted by $u.vw$. Let X be the number of wedges in $\mathcal{G}(n, p)$. Since there are $n\binom{n-1}{2}$ pairs $(u, \{v, w\})$ where the vertices u, v, w are distinct, and since the probability that they form a wedge is p^2, we have that $\lambda := E[X] = n\binom{n-1}{2}p^2$.

Let Y_{vw} (resp., $Y_{u.vw}$) be the indicator variable that expresses that $\langle v, w\rangle$ is an edge (resp., $u.vw$ is a wedge) of $\mathcal{G}(n,p)$. In particular

$$Y_{u.vw} = Y_{uv}Y_{uw} \sim \text{Bern}(\pi_{u.vw})\,,$$

where $\pi_{u.vw} = p^2$. We have that

$$X = \sum_{u=1}^{n} \sum_{v<w\,;\,u\notin\{v,w\}} Y_{u.vw}\,.$$

Define

$$\mathcal{N}_{u.vw} := \{(a.bc)\,;\, |\{\langle u,v\rangle, \langle u,w\rangle\} \cap \{\langle a,b\rangle, \langle a,c\rangle\}| = 1\}\ .$$

The variables $Y_{u.vw}$ are dissociated with respect to the neighbourhoods $\mathcal{N}_{u.vw}$. Moreover,

$$|\mathcal{N}_{u.vw}| = 2(n-3) + 2(n-2) = 2(n-5)$$

(the first term counts the wedges in $\mathcal{N}_{u.vw}$ for which the central vertex is u, whereas the second term counts the wedges in $\mathcal{N}_{u.vw}$ with the leg u, v or u, w). Application of Formula (16.14) yields

$$\begin{aligned} d_V(X, p_\lambda) &\le \left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{u.vw} \left(\pi_{u.vw}^2 + \sum_{a.bc\in\mathcal{N}_{u.vw}} (\pi_{u.vw}\pi_{a.bc} + E\left[Y_{u.vw}Y_{a.bc}\right])\right) \\ &= \min(1,\lambda^{-1}) n \binom{n-1}{2} (p^4 + 2(2n-5)(p^4+p^3)) \\ &\le \min(\lambda,1)(p^2+8np) \le \min(\frac{1}{2}n^3p^2, 1)(p+8np)\,. \end{aligned}$$

In particular, if $p_n = \alpha^{\frac{1}{2}}\frac{1}{n\sqrt{n}}$, $E[X] \sim \alpha$, $np \to 0$ while $n \uparrow +\infty$, and therefore

$$d_V(X, p_\alpha) \to 0\,.$$

In other words, as $n \uparrow +\infty$, the number of wedges converge in variation to a Poisson variable of mean α.

16.3.2 Proof of Chen's Bound

The proof of Theorem 16.3.1 will given after a few preliminaries.

Denote by $E_\lambda\left[h(Z)\right]$ the expectation of $h(Z)$ when Z is a Poisson variable with mean λ. The equation

$$h(i) - E_\lambda\left[h(Z)\right] = \lambda f(i+1) - if(i)\,, \tag{16.15}$$

where the unknown is the function $f : \mathbb{N} \to \mathbb{R}$ and the *data* is the function $h : \mathbb{N} \to \mathbb{R}_+$, is called Chen's equation. It has up to an arbitrary value of $f(0)$ (which we from now on take to be 0) a unique solution obtained by recurrence from (16.15) itself. In particular, if f_1 and f_2 are solutions corresponding to h_1 and h_2 respectively, then $f_1 + f_2$ is the solution corresponding to $h_1 + h_2$.

Proof. (of Theorem 16.3.1.) Denoting by f_A the solution of Chen's equation corresponding to the data $h = 1_A$, where A is any subset of $\mathbb{N}$,

$$1_A(i) - E_\lambda\left[1_A(Z)\right] = \lambda f_A(i+1) - i f_A(i)$$

and consequently, for any integer-valued random variable X,

$$P(X \in A) - p_\lambda(A) = E\left[\lambda f_A(X+1)\right] - E\left[X f_A(X)\right] ,$$

where $p_\lambda(A) := P_\lambda(Z \in A)$. From this it follows that

$$d_V(X, p_\lambda) \leq \sup_{A \subseteq \mathbb{N}} \left\{E\left[\lambda f_A(X+1)\right] - E\left[X f_A(X)\right]\right\} .$$

With $f = f_A$,

$$\begin{aligned} E\left[\lambda f(X+1)\right] - E\left[X f(X)\right] &= \sum_{k=1}^{n} \left(\pi_k E\left[f(X+1)\right] - E\left[Y_k f(X)\right]\right) \\ &= \sum_{k=1}^{n} \pi_k \left(E\left[f(X+1)\right] - E\left[f(X) \,|\, Y_k = 1\right]\right) \\ &= \sum_{k=1}^{n} \pi_k E\left[f(U_k+1) - f(V_k+1)\right] . \end{aligned}$$

Therefore, if

$$|f_A(m) - f_A(\ell)| \leq \frac{1 - e^{-\lambda}}{\lambda} |m - \ell| , \qquad (\star)$$

for all integers m, ℓ and all $A \subseteq \mathbb{N}$, the bound (16.11) is proved.

For the proof of $(\star)$, first obtain the following expression of f_A with $f_A(0) = 0$: for $i \in \mathbb{N}$,

$$f_A(i+1) = \frac{p_\lambda(A \cap N_i) - p_\lambda(A) p_\lambda(N_i)}{\lambda p_\lambda(i)}$$

where $N_i := \{0, 1, \ldots, i\}$. This equality follows from the following calculations:

$$\begin{aligned} f_A(i+1) &= \lambda^{-1}\left(1_A(i) + i f_A(i) - p_\lambda(A)\right) \\ &= \lambda^{-1} 1_A(i) + \lambda^{-2} 1_A(i-1) - p_\lambda(A)\left(\lambda^{-1} + i\lambda^{-2}\right) + \lambda^{-2} i(i-1) f_A(i-1) \\ &= \cdots \\ &= \lambda^{-1} 1_A(i) + i\lambda^{-2} 1_A(i-1) + \cdots + i!\lambda^{-(i+1)} 1_A(0) \\ &\qquad - \left(\lambda^{-1} + i\lambda^{-2} + \cdots + i!\lambda^{-(i+1)}\right) p_\lambda(A) \\ &= i!\lambda^{-(i+1)} e^{\lambda} \left(p_\lambda(A \cap N_i) - p_\lambda(A) p_\lambda(N_i)\right) . \end{aligned}$$

To obtain the bound in $(\star)$, we first observe that, by linearity of Chen's equation,

$$f_A(i+1) - f_A(i) = \sum_{j \in A} \left(f_{\{j\}}(i+1) - f_{\{j\}}(i)\right) . \qquad (\dagger)$$

Now, for $j \neq i$, $f_{\{j\}}(i+1) - f_{\{j\}}(i) \leq 0$, as we now check. For $j > i$,

$$f_{\{j\}}(i+1) = -\frac{p_\lambda(j)p_\lambda(N_i)}{\lambda p_\lambda(i)} = -p_\lambda(j)\sum_{k=0}^{i}\binom{i}{k}\frac{k!}{\lambda^{k+1}}\,,$$

a quantity decreasing in i. For $j < i$,

$$f_{\{j\}}(i+1) = \frac{p_\lambda(j)\,(1-p_\lambda(N_i))}{\lambda p_\lambda(i)} = p_\lambda(j)\sum_{k=0}^{\infty}\frac{1}{\binom{k+i+1}{k+1}}\frac{\lambda^k}{(k+1)!}\,,$$

which is also decreasing in i. Therefore, by (†) and $p_\lambda(i) = e^{-\lambda}\frac{\lambda^i}{i!}$,

$$\begin{aligned} f_A(i+1) - f_A(i) &\le f_{\{i\}}(i+1) - f_{\{i\}}(i) \\ &= \frac{p_\lambda(i)\,(1-p_\lambda(N_i))}{\lambda p_\lambda(i)} + \frac{p_\lambda(i)p_\lambda(N_{i-1})}{\lambda p_\lambda(i-1)} \\ &= \frac{1}{\lambda}\left(1 - p_\lambda(N_i) + \frac{\lambda}{i}p_\lambda(N_{i-1})\right) \\ &\le \frac{1}{\lambda}\left(1 - p_\lambda(N_i) + p_\lambda(N_i\backslash\{0\})\right) \\ &= \frac{1-p_\lambda(0)}{\lambda} = \frac{1-e^{-\lambda}}{\lambda}. \end{aligned}$$

Therefore

$$f_A(i+1) - f_A(i) \le \frac{1-e^{-\lambda}}{\lambda}\,.$$

Replacing A by its complement and observing that $f_A = -f_{\bar{A}}$, we have that

$$-(f_A(i+1) - f_A(i)) = f_{\bar{A}}(i+1) - f_{\bar{A}}(i) \le \frac{1-e^{-\lambda}}{\lambda}$$

and therefore

$$|f_A(i+1) - f_A(i)| \le \frac{1-e^{-\lambda}}{\lambda}\,,$$

from which ($\star$) follows. □

Books for Further Information

The next two references require familiarity with advanced probability theory. [Lindvall, 1992] is a concise presentation of coupling with an abundance of fine examples. [Barbour, Holst, Janson and Spencer, 1992] is the fundamental reference on Poisson approximation, for random variables of course, but also for point processes.

16.4 Exercises

Exercise 16.4.1. MAXIMAL COINCIDENCE OF BIASED COINS
Find a pair of $\{0,1\}$-valued random variables with prescribed marginals

$$P(X=1) = a\,,\ P(Y=1) = b,$$

where $a, b \in (0,1)$, and such that $P(X = Y)$ is maximal.

Exercise 16.4.2. PROPERTIES OF THE VARIATION DISTANCE
1. Variation distance and image distributions. Let α and β be two probability distributions on the countable space E, and let $f : E \mapsto F$ where F is another countable space. Define the probability distribution αf^{-1} on F by $\alpha f^{-1}(B) = \alpha(f^{-1}(B))$, and define likewise βf^{-1}. Prove that

$$d_V(\alpha, \beta) \geq d_V\left(\alpha f^{-1}, \beta f^{-1}\right) .$$

2. Convexity. Let α_k and β_k be probability distributions on the countable space E $(1 \leq k \leq m)$. Show that if $\lambda_k \in [0,1]$ and $\sum_{k=1}^m \lambda_k = 1$, then

$$d_V\left(\sum_{k=1}^m \lambda_k \alpha_k, \sum_{k=1}^m \lambda_k \beta_k\right) \leq \sum_{i=k}^m \lambda_k d_V(\alpha_k, \beta_k) .$$

3. Prove the above properties using the interpretation of variation distance in terms of hypothesis testing (Subsection 16.1.1).

Exercise 16.4.3.
Let α and β be two probability distributions on the countable space E. Show that

$$d_V(\alpha, \beta) = \frac{1}{2} \sup_{|f| \leq 1} \left(\sum_i f(i)\alpha(i) - \sum_i f(i)\beta(i) \right)$$

where $|f| := \sup_{i \in E} |f(i)|$.

Exercise 16.4.4. CONVERGENCE SPEED VIA COUPLING
Suppose that the coupling time τ in Theorem 15.1.1 satisfies $E[\psi(\tau)] < \infty$ for some non-decreasing function $\psi : \mathbb{N} \to \mathbb{R}_+$ such that $\lim_{n \uparrow \infty} \psi(n) = \infty$. Show that for any initial distributions μ and ν

$$|\mu^T \mathbf{P}^n - \nu^T \mathbf{P}^n| = o\left(\frac{1}{\psi(n)}\right) .$$

Exercise 16.4.5.
Let $\{Z_n\}_{n \geq 1}$ be an IID sequence of IID $\{0,1\}$-valued random variables, $P(Z_n = 1) = p \in (0,1)$. Show that for all $k \geq 1$,

$$\lim_{n \uparrow \infty} P(Z_1 + Z_2 + \cdots Z_n \text{ is divisible by } k) = 1 .$$

Exercise 16.4.6.
Let $\mathbf{P}$ be an ergodic transition matrix on the *finite* state space E. Prove that for any initial distributions μ and ν, one can construct two HMC's $\{X_n\}_{n \geq 0}$ and $\{Y_n\}_{n \geq 0}$ on E with the same transition matrix $\mathbf{P}$, and the respective initial distributions μ

and ν, in such a way that they couple at a finite time τ such that $E[e^{\alpha\tau}] < \infty$ for some $\alpha > 0$.

Exercise 16.4.7. The lazy walk on the circle
Consider N points on the circle forming the state space $E := \{0, 1, \ldots, N-1\}$. Two points i, j are said to be neighbours if $j = i \pm 1$ modulo n. Consider the Markov chain $\{(X_n, Y_n)\}_{n\geq 0}$ with state space $E \times E$ and representing two particles moving on E as follows. At each time n choose X_n or Y_n with probability $\frac{1}{2}$ and move the corresponding particle to the left or to the right, equiprobably while the other particle remains still. The initial positions of the particles are a and b respectively. Compute the average time it takes until the two particles collide (the average coupling time of two lazy random walks).

Exercise 16.4.8. Coupling time for the 2-state hmc
Find the distribution of the first meeting time of two independent hmc with state space $E = \{1, 2\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where $\alpha, \beta \in (0, 1)$, when their initial states are different.

Exercise 16.4.9. Chen's characterization of the Poisson distribution
Show that Z is a Poisson variable with mean λ if and only if,

$$E\left[\lambda f(Z+1)\right] - E\left[Z f(Z)\right] = 0$$

whenever the expectation is well defined.

Exercise 16.4.10. Alternative expressions of the variation distance
Let α and β be two probability distributions on the countable space E. Show that

$$d_V(\alpha, \beta) = 1 - \sum_{i\in E} \alpha(i) \wedge \beta(i) = \sum_{i\in E} (\alpha(i) - \beta(i))^+ = \sum_{i\in E} (\beta(i) - \alpha(i))^+ .$$

Exercise 16.4.11. Birthdays
Consider an assembly of 73 persons, each one having 10 friends in this assembly. The birthday dates of these persons are supposed independent and uniformly distributed over the year (take 365 days). Apply Theorem 16.3.5 to the proof that if X is the number of persons sharing a birthday, then

$$d_V(X, \mathrm{Poi}(1)) \leq \frac{37}{365}.$$

Exercise 16.4.12.
The framework is that of Theorem 16.3.1 with the additional assumption that $U_k \geq V_k$ $(1 \leq k \leq n)$. Prove that

$$d_V(X, p_\lambda) \leq \left(\frac{1 - e^{-\lambda}}{\lambda}\right)(\lambda - \text{ Var }(X)). \tag{16.16}$$

Exercise 16.4.13. Positively associated variables
Let $Y_i \sim Bern(\pi_i)$ $(1 \leq i \leq n)$ be Bernoulli variables, and define $X := \sum_{i=1}^n Y_i$, a random variable with mean $\lambda := \sum_{i=1}^n \pi_i$. Suppose there exists for each k a family of random variables $Y_i^{(k)}$ $(1 \leq i \leq n, i \neq k)$ such that

$$(Y_i^{(k)} : i \neq k) \sim (Y_i : i \neq k) \mid Y_k = 1$$

and

$$Y_i^{(k)} \geq Y_i \, (i \neq k).$$

Prove that

$$d_V(X, p_\lambda) \leq \min(1, \lambda^{-1}) \left(\text{Var}(X) - \lambda + 2 \sum_i \pi_i^2\right).$$

Exercise 16.4.14. Adjacent failures on a circle
(For this exercise, use the result of Exercise 16.4.13.) Consider n points regularly arranged on a circle. Let Z_i be a Bernoulli variable attached to point i, $P(Z_i = 1) = p$, and suppose that the sequence $\{Z_n\}_{1 \leq i \leq n}$ iid. Let X_i be the indicator of the event $Z_i = Z_{i+1}$ (here, $+$ is the addition modulo n). Let $W := \sum_{i=1}^n X_i$. Show that

$$d_V(W, \text{Poi}(np^2)) \leq \min\{np^2, 1\} p(2 - p).$$

Exercise 16.4.15. Negatively associated variables
Let $Y_i \sim Bern(\pi_i)$ $(1 \leq i \leq n)$ be Bernoulli variables, and define $X := \sum_{i=1}^n Y_i$, a random variable with mean $\lambda := \sum_{i=1}^n \pi_i$. Suppose there exists for each k a family of random variables $Y_i^{(k)}$ $(1 \leq i \leq n, i \neq k)$ such that

$$(Y_i^{(k)}; i \neq k) \sim (Y_i; i \neq k) \mid Y_k = 1$$

and

$$Y_i^{(k)} \leq Y_i \qquad (i \neq k).$$

Prove that

$$d_V(X, p_\lambda) \leq \min(1, \lambda^{-1}) (\lambda - \text{Var}(X)).$$

Chapter 17

Martingale Methods

17.1 Martingales

17.1.1 Definition and Examples

In casino parlance, a martingale is a strategy that beats the bank. For a probabilist, it is a sequence of random variables that has no tendency to increase or decrease (the precise definition will be given soon). This notion has to do with casino games but it is also useful outside Las Vegas. In fact, it is one of the foundations of modern probability theory.

Let $\{X_n\}_{n\geq 0}$ be a sequence of discrete real-valued random variables. Such a sequence is also called a (discrete-time) (real-valued) discrete stochastic process.

Definition 17.1.1 *A real-valued stochastic process $\{Y_n\}_{n\geq 0}$ such that for each $n \geq 0$*

(i) Y_n is a function of n and $X_0^n := (X_0, \ldots, X_n)$, and

(ii) $E[|Y_n|] < \infty$ or $Y_n \geq 0$,

is called a martingale (resp., submartingale, supermartingale) with respect to $\{X_n\}_{n\geq 0}$ if, moreover,

$$E[Y_{n+1} \,|\, X_0^n] = Y_n \,(resp.,\ \geq Y_n, \leq Y_n). \tag{17.1}$$

For short, one may say "X_0^n-martingale" for "martingale with respect to $\{X_n\}_{n\geq 0}$", with similar abbreviations for supermartingales and submartingales.

Observe that a martingale is a submartingale *and* a supermartingale.

Remark 17.1.2 This book is concerned with discrete random variables, and therefore the precision "discrete" will be omitted in the definitions and results that apply *verbatim* to the general case.

EXAMPLE 17.1.3: SUMS OF IID RANDOM VARIABLES. Let $\{X_n\}_{n\geq 0}$ be a sequence of IID random variables with mean 0. The stochastic process

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_17

$$Y_n := X_0 + X_1 + \cdots + X_n \quad (n \geq 0)$$

is an X_0^n-martingale. In fact, for all $n \geq 1$,

$$E[Y_{n+1} \,|\, X_0^n] = E[Y_n \,|\, X_0^n] + E[X_{n+1} \,|\, X_0^n] = Y_n + E[X_{n+1}] = Y_n \,,$$

where the second equality is due to the facts that Y_n is a function of X_0^n (Theorem 2.3.7) and that X_0^n and X_{n+1} are independent (Theorem 2.3.8).

EXAMPLE 17.1.4: PRODUCTS OF IID RANDOM VARIABLES. Let $\{X_n\}_{n\geq 0}$ be a sequence of integrable IID random variables with mean 1. The stochastic process

$$Y_n := \prod_{k=0}^{n} X_k \quad (n \geq 0)$$

is an X_0^n-martingale. In fact,

$$E[Y_{n+1} \,|\, X_0^n] = E\left[X_{n+1}\left(\prod_{k=0}^{n} X_k\right) \,|\, X_0^n\right] = E[X_{n+1} \,|\, X_0^n] \prod_{k=0}^{n} X_k$$
$$= E[X_{n+1}] \prod_{k=1}^{n} X_k = 1 \times Y_n = Y_n \,,$$

where the second equality is due to the fact that $\prod_{k=0}^{n} X_k$ is a function of X_0^n (Theorem 2.3.7) and the third to the fact that X_0^n and X_{n+1} are independent (Theorem 2.3.8).

EXAMPLE 17.1.5: EMPTY BINS, TAKE 1. There are m balls to be placed in N bins. Each ball is assigned to a bin randomly and independently of the others. The mean of the number Z of empty bins is $\mu := E[Z] = N\,(1 - 1/N)^m$. From the coupon collector's point of view: there are N coupons in the complete collection, and Z is the number of missing coupons when m chocolate tablets have been bought. The stochastic process

$$M_n := Y_n \left(1 - \frac{1}{N}\right)^{m-n} \qquad (0 \leq n \leq m) \,,$$

where Y_n is the number of empty bins at time n (that is, immediately after the n-th ball has been placed) is a Y_0^n-martingale. (Note also that $M_m = Z$ and that $M_0 = N\,(1 - 1/N)^m = E[Z] = \mu$.)

Proof. Given Y_0^{n-1}, with probability $1 - \frac{Y_{n-1}}{N}$ the n-th ball falls into a currently non-empty bin (and then $Y_n = Y_{n-1}$) and with probability $\frac{Y_{n-1}}{N}$ into a currently empty bin (and then $Y_n = Y_{n-1} - 1$). Therefore

$$E\left[Y_n \,|\, Y_0^{n-1}\right] = \left(1 - \frac{Y_{n-1}}{N}\right) Y_{n-1} + \frac{Y_{n-1}}{N}(Y_{n-1} - 1) = Y_{n-1}\left(1 - \frac{1}{N}\right) \,,$$

and consequently

$$E\left[M_n \,|\, Y_0^{n-1}\right] = E\left[Y_n \,|\, Y_0^{n-1}\right]\left(1-\frac{1}{N}\right)^{m-n} = Y_{n-1}\left(1-\frac{1}{N}\right)^{m-n+1} = M_{n-1}\,.$$

□

EXAMPLE 17.1.6: GAMBLING, TAKE 1. Consider the stochastic process $\{Y_n\}_{n\geq 0}$ with values in $\mathbb{N}$ defined by $Y_0 = a \in \mathbb{N}$ and, for $n \geq 0$,

$$Y_{n+1} = Y_n + X_{n+1}\, b_{n+1}(Y_0^n),$$

where $\{X_n\}_{n\geq 1}$ is an IID sequence of random variables taking the values $+1$ or -1 equiprobably, and where the family of functions $b_n : \mathbb{N} \to \mathbb{N}$, $n \geq 1$, is a given betting strategy, that is, $b_{n+1}(Y_0^n)$ is the stake at time $n+1$ of a gambler given the observed history $Y_0^n := (Y_0, \ldots, Y_n)$ of his fortune up to time n. The initial conditions are $X_0 = Y_0 = a$. Admissible bets must guarantee that the fortune Y_n remains non-negative at all times n, that is, $b_{n+1}(y_0^n) \leq y_n$, and the game ends as soon as the gambler is ruined. The process so defined is an X_0^n-martingale. Indeed, Y_n is a function of X_0^n (observe that Y_0^n is a function of X_0^n) and

$$\begin{aligned} E\left[Y_{n+1} \,|\, X_0^n\right] &= E\left[Y_n \,|\, X_0^n\right] + E\left[X_{n+1} b_{n+1}(Y_0^n) \,|\, X_0^n\right] \\ &= Y_n + E\left[X_{n+1} \,|\, X_0^n\right] b_{n+1}(Y_0^n) = Y_n\,, \end{aligned}$$

where the second equality uses Theorem 2.3.7 (again: Y_0^n is a function of X_0^n) and the assumption that X_{n+1} is independent of X_0^n and of mean 0 (Theorem 2.3.8).

17.1.2 Martingale Transforms

Let $\{X_n\}_{n\geq 0}$ be some sequence of random variables with values in the denumerable set $\mathcal{X}$. The sequence of complex-valued random variables $\{H_n\}_{n\geq 1}$ is called an X_0^n-predictable process if for all $n \geq 1$,

$$H_n = g_n(X_0^{n-1})$$

for some function $g_n : \mathcal{X}^n \to \mathbb{C}$. Let $\{Y_n\}_{n\geq 0}$ be another sequence of complex random variables. The sequence $\{(H \circ Y)_n\}_{n\geq 1}$ defined by

$$(H \circ Y)_n := \sum_{k=1}^{n} H_k(Y_k - Y_{k-1}) \quad (n \geq 1)$$

is called the transform of $\{Y_n\}_{n\geq 0}$ by $\{H_n\}_{n\geq 1}$.

Theorem 17.1.7 *(a) Let $\{Y_n\}_{n\geq 0}$ be a X_0^n-submartingale (resp., martingale) and let $\{H_n\}_{n\geq 1}$ be a bounded non-negative X_0^n–predictable process. Then $\{(H\circ Y)_n\}_{n\geq 1}$ is a X_0^n-submartingale (resp., martingale).*

(b) If $\{Y_n\}_{n\geq 0}$ is a X_0^n-martingale and if $\{H_n\}_{n\geq 1}$ is bounded and X_0^n–predictable, then $\{(H \circ Y)_n\}_{n\geq 1}$ is a X_0^n-martingale.

The proof is left as an exercise (Exercise 17.1.7).

Theorem 17.1.7 has the *stopped martingale* theorem for corollary:

Corollary 17.1.8 *Let $\{Y_n\}_{n\geq 0}$ be a X_0^n–submartingale (resp., martingale), and let τ be a X_0^n-stopping time. Then $\{Y_{n\wedge\tau}\}_{n\geq 0}$ is a X_0^n–submartingale (resp., martingale). In particular,*

$$E[Y_{n\wedge\tau}] \geq E[Y_0] \quad (resp., \ = E[Y_0]) \quad (n \geq 0)\,. \tag{17.2}$$

Proof. Let $H_n := 1_{\{n\leq\tau\}}$. The stochastic process $\{H_n\}_{n\geq 1}$ is X_0^n–predictable since $\{H_n = 0\} = \{\tau \leq n-1\}$ is of the form $g(X_0^{n-1})$. We have

$$\begin{aligned} Y_{n\wedge\tau} &= Y_0 + \sum_{k=1}^{n\wedge\tau} (Y_k - Y_{k-1}) \\ &= Y_0 + \sum_{k=1}^{n} 1_{\{k\leq\tau\}} (Y_k - Y_{k-1}) \end{aligned}$$

The result then follows by Theorem 17.1.7. □

17.1.3 Harmonic Functions of Markov Chains

Let $\{X_n\}_{n\geq 0}$ be an HMC on the countable space E with transition matrix $\mathbf{P}$. A function $h : E \to \mathbb{R}$, represented as a column vector of the dimension of E, is called harmonic (resp., subharmonic, superharmonic) iff

$$\mathbf{P}h = h \,(\text{resp.}, \ \geq h, \leq h)\,, \tag{17.3}$$

that is, in developed form, for all $i \in E$,

$$\sum_{j\in E} p_{ij}h(j) = h(i) \,(\text{resp.}, \ \geq h(i), \leq h(i)).$$

Superharmonic functions are also called excessive functions.

Equation (17.3) is equivalent to

$$E[h(X_{n+1}) \mid X_n = i] = h(i) \,(\text{resp.}, \ \geq h(i), \leq h(i))\,,$$

for all $i \in E$. In view of the Markov property, the left-hand side of the above equality is also equal to

$$E[h(X_{n+1}) \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0],$$

and therefore (17.3) is equivalent to

$$E[h(X_{n+1} \mid X_0^n] = h(X_n) \,(\text{resp.}, \ \leq h(X_n), \geq h(X_n)). \tag{17.4}$$

Therefore, if either $E[|h(X_n)|] < \infty$ for all $n \geq 0$, or $h \geq 0$, the process $\{h(X_n)\}_{n\geq 0}$ is a martingale (resp., submartingale, supermartingale) with respect to $\{X_n\}_{n\geq 0}$.

17.2 Hoeffding's Inequality

17.2.1 The Basic Inequality

Theorem 17.2.1 *(Hoeffding, 1963) Let $\{M_n\}_{n\geq 0}$ be a real X_0^n-martingale such that, for some sequence $c_1, c_2, \ldots$ of real numbers,*

$$P(|M_n - M_{n-1}| \leq c_n) = 1 \quad (n \geq 1)\,.$$

Then, for all $x \geq 0$,

$$P(|M_n - M_0| \geq x) \leq 2\exp\left(-\frac{1}{2}x^2 \Big/ \sum_{i=1}^{n} c_i^2\right). \tag{17.5}$$

Proof. For $|z| \leq 1$, $\lambda := \frac{1}{2}(1-z) \in [0,1]$, and for any $a \in \mathbb{R}$,

$$az = \lambda(-a) + (1-\lambda)a\,.$$

Therefore, by convexity of the function $z \mapsto e^{az}$,

$$e^{az} \leq \frac{1}{2}(1-z)e^{-a} + \frac{1}{2}(1+z)e^{+a}.$$

In particular, if Z is a centered random variable such that $P(|Z| \leq 1) = 1$,

$$E[e^{aZ}] \leq \frac{1}{2}(1-E[Z])e^{-a} + \frac{1}{2}(1+E[Z])e^{+a} = \frac{1}{2}e^{-a} + \frac{1}{2}e^{+a} \leq e^{\frac{1}{2}a^2}$$

(see Example 3.2.5 for the last inequality). With $Z := (M_n - M_{n-1})/c_n$ and by similar arguments, for all $a \in \mathbb{R}$,

$$\begin{aligned}
&E\left[e^{a\left(\frac{M_n - M_{n-1}}{c_n}\right)}\,\middle|\, X_0^{n-1}\right] \\
&\leq \tfrac{1}{2}\left(1 - E\left[\tfrac{M_n - M_{n-1}}{c_n}\,\middle|\, X_0^{n-1}\right]\right)e^{-a} + \tfrac{1}{2}\left(1 + E\left[\tfrac{M_n - M_{n-1}}{c_n}\,\middle|\, X_0^{n-1}\right]\right)e^{+a} \\
&= \tfrac{1}{2}e^{-a} + \tfrac{1}{2}e^{+a} \leq e^{\frac{1}{2}a^2}\,,
\end{aligned}$$

because $E\left[M_n - M_{n-1} \,|\, X_0^{n-1}\right] = 0$ by definition of a martingale. Replacing a by $c_n a$ in the last chain of inequalities gives

$$E\left[e^{a(M_n - M_{n-1})}\big|X_0^{n-1}\right] \leq e^{\frac{1}{2}a^2 c_n^2}.$$

Therefore,

$$\begin{aligned}
E\left[e^{a(M_n - M_0)}\right] &= E\left[e^{a(M_{n-1} - M_0)}e^{a(M_n - M_{n-1})}\right] \\
&= E\left[e^{a(M_{n-1} - M_0)}E\left[e^{a(M_n - M_{n-1})}\big|X_0^{n-1}\right]\right] \\
&\leq E\left[e^{a(M_{n-1} - M_0)}\right] \times e^{\frac{1}{2}a^2 c_n^2}\,,
\end{aligned}$$

and by iteration

$$E\left[e^{a(M_n-M_0)}\right] \le e^{\frac{1}{2}a^2\sum_{i=1}^n c_i^2}.$$

In particular, by Markov's inequality, with $a > 0$,

$$P(M_n - M_0 \ge x) \le e^{-ax} E\left[e^{a(M_n-M_0)}\right] \le e^{-ax+\frac{1}{2}a^2\sum_{i=1}^n c_i^2}.$$

Minimization of the right-hand side with respect to a gives

$$P(M_n - M_0 \ge x) \le e^{-\frac{1}{2}x^2 \Big/ \sum_{i=1}^n c_i^2}. \tag{17.6}$$

By the same argument with $M_0 - M_n$ instead of $M_n - M_0$,

$$P(-(M_n - M_0) \ge x) \le e^{-\frac{1}{2}x^2 \Big/ \sum_{i=1}^n c_i^2}.$$

The announced bound then follows from these two bounds since for any real random variable X, all $x \in \mathbb{R}_+$, $P(|X| \ge x) = P(X \ge x) + P(X \le -x)$. □

EXAMPLE 17.2.2: EMPTY BINS, TAKE 2. Refer to Example 17.1.5. We shall derive the following inequality concerning the number Z of empty bins:

$$P\left(|Z - \mu| \ge \lambda\right) \le 2\exp\left\{-\frac{\lambda^2(N-\frac{1}{2})}{N^2-\mu^2}\right\}. \tag{17.7}$$

For this recall from Example 17.1.5 that

$$M_n := Y_n\left(1-\frac{1}{N}\right)^{m-n},$$

where Y_n is the number of empty bins at time n (that is, immediately after the n-th ball has been placed) is a Y_0^n-martingale. Also, $M_m = Z$ and $M_0 = N\left(1-\frac{1}{N}\right)^m = E[Z] = \mu$. We have that

$$M_n - M_{n-1} = \left(Y_n - Y_{n-1}\left(1-\frac{1}{N}\right)\right)\left(1-\frac{1}{N}\right)^{m-n}$$

and since $Y_n \le Y_{n-1}$ and $Y_{n-1} \le N$,

$$Y_n - Y_{n-1}\left(1-\frac{1}{N}\right) \le Y_{n-1}\left(1-1+\frac{1}{N}\right) = Y_{n-1}\left(\frac{1}{N}\right) \le +1.$$

Also

$$Y_n - Y_{n-1}\left(1-\frac{1}{N}\right) \ge Y_n - Y_{n-1} \ge -1\,.$$

Therefore,

$$|M_n - M_{n-1}| \le c_n := \left(1-\frac{1}{N}\right)^{m-n}.$$

The result follows from Hoeffding's inequality applied to $M_m - M_0 = Z - \mu$ and the identity

$$\sum_{n=1}^m c_n^2 = \frac{1-\beta^{2m}}{1-\beta^2},$$

where $\beta = \frac{N-1}{N}$. But since $\mu = N\beta^m$, the latter quantity is equal to $\frac{N^2-\mu^2}{2N-1}$.

17.2.2 The Lipschitz Condition

The following is a general framework of application of Hoeffding's inequality that will be applied to the Erdös–Rényi random graphs.

Let $\mathcal{X}$ be a finite set and let N be a positive integer. Let $f : \mathcal{X}^N \to \mathbb{R}$ be a given function. Remember the notation $x = (x_1, \dots, x_N)$ and $x_1^k = (x_1, \dots, x_k)$. In particular, $x = x_1^N$. For $x \in \mathcal{X}^N$, $z \in \mathcal{X}$ and $1 \le k \le N$, define

$$f_k(x, z) := f(x_1, \dots, x_{k-1}, z, x_{k+1}, \dots, x_N)\,.$$

The function f is said to satisfy a Lipschitz condition with bound c if for all $x \in \mathcal{X}^N$, $z \in \mathcal{X}$ and $1 \le k \le N$,

$$|f_k(x, z) - f(x)| \le c\,.$$

In other words, changing a single coordinate entails a change not bigger than c in absolute value. Let $X_1, X_2, \dots, X_N$ be independent random variables with values in $\mathcal{X}$. Define the martingale $\{M_n\}_{n\ge 0}$ by $M_0 := E\,[f(X)]$, and for $n \ge 1$,

$$M_n := E\,[f(X)\,|\,X_1^n] := E\,[f(X_1, \dots, X_N)\,|\,X_1^n]\,.$$

By the independence assumption and Theorem 2.3.9, with obvious notations,

$$E\,[f(X)\,|\,X_1^n] = \sum_{x_{n+1}^N} f(X_1^{n-1}, X_n, x_{n+1}^N) P(X_{n+1}^N = x_{n+1}^N)$$

and

$$E\left[f(X)\,|\,X_1^{n-1}\right] = \sum_{x_{n+1}^N} \sum_{x_n} f(X_1^{n-1}, x_n, x_{n+1}^N) P(X_n = x_n) P(X_{n+1}^N = x_{n+1}^N)\,.$$

Therefore

$$\begin{aligned}
&|M_n - M_{n-1}| \le \\
&\sum_{x_{n+1}^N} \sum_{x_n} |f(X_1^{n-1}, x_n, x_{n+1}^N) - f(X_1^{n-1}, X_n, x_{n+1}^N)| P(X_n = x_n) P(X_{n+1}^N = x_{n+1}^N) \\
&\qquad\qquad \le c \sum_{x_{n+1}^N} \sum_{x_n} P(X_n = x_n) P(X_{n+1}^N = x_{n+1}^N) = c\,.
\end{aligned}$$

EXAMPLE 17.2.3: EXPOSURE MARTINGALES IN RANDOM GRAPHS. The random graph $\mathcal{G}(n, p)$ may be generated as follows. Enumerate the $N = \binom{n}{2}$ edges of the complete graph on $V = V_n := \{1, 2, \dots, n\}$ from $i = 1$ to $i = N$. Generate a random vector $X = (X_1, \dots, X_N)$ with independent and identically distributed variables with values in $\{0, 1\}$ and common distribution $P(X_i = 1) = p$. Then include edge i in $\mathcal{G}(n, p)$ if and only if $X_i = 1$. Any functional of $\mathcal{G}(n, p)$ can always be written as $f(X)$. The edge exposure martingale corresponding to this functional is the X_0^k-martingale defined by $M_0 = E\,[f(X)]$ and for $k \ge 1$,

$$M_k := E\left[f(X)\,|\,X_1^k\right].$$

Since the X_k's are independent, the general method just presented can be applied provided the Lipschitz condition is satisfied.

Another type of martingale related to a $\mathcal{G}(n,p)$ graph is useful. For $1 \leq i \leq n$, define the graph G_i to be the restriction of $\mathcal{G}(n,p)$ to V_i. A functional of $\mathcal{G}(n,p)$ can always be written as $f(G)$, where $G := (G_1, \ldots, G_n)$. The vertex exposure martingale corresponding to this functional is the G_1^i-martingale defined by $M_0 = E[f(G)]$ and for $i \geq 1$,

$$M_i := E\left[f(G) \,|\, G_1^i\right] .$$

Example 17.2.4: Chromatic number of a random graph. (Shamir and Spencer, 1987) The chromatic number of a graph G is the minimal number of colours needed to colour the vertices in such a way that no adjacent vertices receive the same colour. Call $f(G)$ the chromatic number of $\mathcal{G}(n,p)$. Since the difference between $f(G_0^{i-1}, G_i, g_{i+1}^n)$ and $f(G_0^{i-1}, g_i, g_{i+1}^n)$ is at most one for all g_i, g_{i+1}^n, one may attempt to apply Hoeffding's bound in the form (17.6) to obtain

$$P\left(f(G) - E[f(G)] \geq \lambda\sqrt{n}\right) \leq e^{-2\lambda^2} .$$

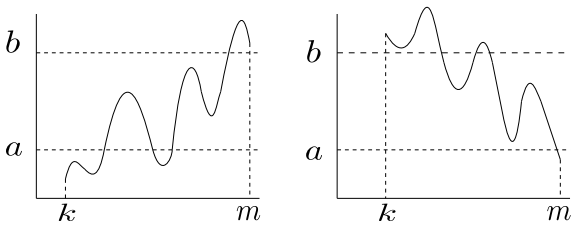
However, the G_i's are not independent. Nevertheless, the general method can be applied modulo a slight change of point of view. Let X_1 be an arbitrary constant, and for $2 \leq i \leq n$, let $X_i = \{X_{\langle i,j\rangle},\ 1 \leq j \leq i-1\}$ (recall the definition of $X_{\langle u,v\rangle}$ in Definition 2.1.54). The passage from subgraph G_{i-1} to subgraph G_i is represented by the "difference" X_i between these two subgraphs. Then $f(G)$ can be rewritten as $h(X) = h(X_1, \ldots, X_n)$ and the general method applies since the X_i's are independent.

17.3 The Two Pillars of Martingale Theory

17.3.1 The Martingale Convergence Theorem

The martingale convergence theorem is in fact the submartingale convergence theorem (but of course a martingale is a special case of submartingale). It is the probabilistic counterpart of the convergence of a bounded non-decreasing sequence of real numbers to a finite limit. Its proof rests on the upcrossing inequality.

An upcrossing of the interval $[a,b]$ by a real random sequence $\{X_n\}_{n\geq 0}$ is said to occur if for some k, m $(0 \leq k \leq m)$, $X_k \leq a$, $X_m \geq b$ and $X_\ell < b$ for all ℓ $(k < \ell < m)$. A downcrossing is defined in a similar way.



Theorem 17.3.1 *Let $\{S_n\}_{n\geq 0}$ be an X_0^n-submartingale. Let $a, b \in \mathbb{R}$, $a < b$, and let ν_n be the number of upcrossings of $[a, b]$ before ($\leq$) time n. Then*

$$(b-a)E[\nu_n] \leq E[(S_n - a)^+]. \tag{17.8}$$

Proof. Since ν_n is the number of upcrossings of the interval $[0, b-a]$ by the submartingale $\{(S_n - a)^+\}_{n\geq 1}$, we may suppose without loss of generality that $S_n \geq 0$ and take $a = 0$. We then just need to prove that

$$bE[\nu_n] \leq E[S_n - S_0]\,. \tag{17.9}$$

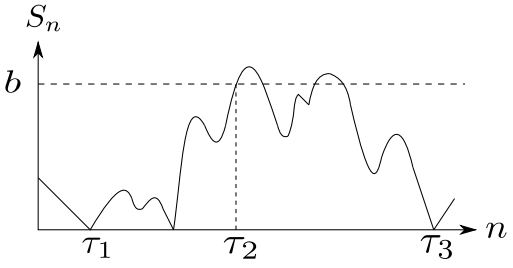
Define a sequence of X_0^n-stopping times as follows: $\tau_0 = 0$,

$$\begin{aligned}\tau_1 &= \inf\{n > \tau_0\,;\, S_n = 0\}\\ \tau_2 &= \inf\{n > \tau_1\,;\, S_n \geq b\}\end{aligned}$$

and more generally,

$$\begin{aligned}\tau_{2k+1} &= \inf\{n > \tau_{2k}\,;\, S_n = 0\}\\ \tau_{2k+2} &= \inf\{n > \tau_{2k+1}\,;\, S_n \geq b\}\,,\end{aligned}$$

with the usual convention $\inf \varnothing = \infty$.



For $i \geq 1$, let

$$\begin{aligned}\phi_i &= 1 \text{ if } \tau_m < i \leq \tau_{m+1} \text{ for some odd } m\\ &= 0 \text{ if } \tau_m < i \leq \tau_{m+1} \text{ for even odd } m.\end{aligned}$$

Observe that $\{\phi_i = 1\} = \cup_{odd\, m}\,(\{\tau_m < i\}\backslash\{\tau_{m+1} < i\})$ is a set defined in terms of X_0^{i-1} and that

$$b\nu_n \le \sum_{i=1}^{n} \phi_i(S_i - S_{i-1}).$$

Therefore

$$\begin{aligned} bE[\nu_n] &\le E\left[\sum_{i=1}^{n} \phi_i(S_i - S_{i-1})\right] = \sum_{i=1}^{n} E[\phi_i(S_i - S_{i-1})] \\ &= \sum_{i=1}^{n} E\left[\phi_i E[(S_i - S_{i-1})|X_0^{i-1}]\right] = \sum_{i=1}^{n} E\left[\phi_i(E[S_i|X_0^{i-1}] - S_{i-1})\right] \\ &\le \sum_{i=1}^{n} E\left[(E[S_i|X_0^{i-1}] - S_{i-1})\right] \le \sum_{i=1}^{n}(E[S_i] - E[S_{i-1}]) = E[S_n - S_0]\,. \end{aligned}$$

□

Theorem 17.3.2 *Let $\{S_n\}_{n\ge 0}$ be an X_0^n-submartingale, L_1-bounded, that is such that*

$$\sup_{n\ge 0} E[|S_n|] < \infty. \tag{17.10}$$

Then S_n converges P-a.s. to an integrable random variable S_∞.

Proof. Let ν_n be the number of upcrossings of an interval $[a, b]$ prior ($\le$) to time n, and let $\nu_\infty = \lim_{n\uparrow\infty} \nu_n$. By the upcrossing inequality (17.8),

$$\begin{aligned} (b-a)E[\nu_n] &\le E[(S_n - a)^+] \le E[S_n^+] + |a| \\ &\le \sup_{k\ge 0} E[S_k^+] + |a| \le \sup_{k\ge 0} E[|S_k|] + |a|\,. \end{aligned}$$

Therefore, letting $n \uparrow \infty$,

$$(b-a)E[\nu_\infty] \le \sup_{k\ge 0} E[[S_k|] + |a| < \infty.$$

In particular, $\nu_\infty < \infty$, P-a.s. Therefore, for all $a, b \in \mathbb{R}$, $a < b$,

$$P(\liminf_{n\uparrow\infty} S_n < a < b < \limsup_{n\uparrow\infty} S_n) = 0.$$

Since, denoting by $\mathbb{Q}$ the set of rationals numbers,

$$\{\liminf_{n\uparrow\infty} S_n < \limsup_{n\uparrow\infty} S_n\} = \bigcup_{a,b\in\mathbb{Q};\, a<b} \{\liminf_{n\uparrow\infty} S_n < a < b < \limsup_{n\uparrow\infty} S_n\},$$

we have

$$P(\liminf_{n\uparrow\infty} S_n < \limsup_{n\uparrow\infty} S_n) = 0,$$

which implies that $\{S_n\}_{n\ge 0}$ converges P-a.s. By Fatou's lemma,

$$E[\lim_{n\uparrow\infty} S_n] \le \liminf_{n\uparrow\infty} E|S_n| \le \sup_{n\ge 0} E|S_n| < \infty.$$

□

Remark 17.3.3 The statement of Theorem 17.3.2 does not make any restriction concerning the range of the random variables concerned, which in this book are discrete. In fact, the result remains true in the general case, with exactly the same proof. Note that the proof given above in this book only features exclusively discrete random variables if the martingale $\{S_n\}_{n\geq 0}$ takes discrete values.

Corollary 17.3.4 *(a) Any non-positive submartingale converges to an integrable random variable.*
(b) Any non-negative supermartingale converges to an integrable random variable.

Proof. (b) follows from (a) by changing signs. For (a), we have

$$E[|S_n|] = -E[S_n] \leq -E[S_0] = E[|S_0|] < \infty.$$

Therefore condition (17.10) is satisfied and the conclusion follows from Theorem 17.3.2. □

EXAMPLE 17.3.5: GAMBLING, TAKE 2. Consider the situation in Example 17.1.6, assuming that the initial fortune a is a positive integer, and that the bets are also positive integers. Therefore the process $\{Y_n\}_{n\geq 0}$ is a non-negative $\mathcal{F}_n^X$-martingale, and by the martingale convergence theorem, it almost surely has a finite limit. Since the bets are assumed positive integers when the fortune of the player is positive, this limit cannot be other than 0. Since Y_n is a non-negative integer for all $n \geq 0$, this can happen only if the fortune of the gambler becomes null in finite time.

EXAMPLE 17.3.6: BRANCHING PROCESS VIA MARTINGALES. Consider the branching process $\{X_n\}_{n\geq 0}$ of Section 10.1. The stochastic process

$$Y_n = \frac{X_n}{m^n},$$

where m is the average number of sons of a given individual, is a martingale with respect to $\{X_n\}_{n\geq 0}$. Indeed, each of the X_n members of the n-th generation gives on average m sons, and they do this independently. Therefore, $E[X_{n+1}|X_n] = mX_n$, and

$$E\left[\frac{X_{n+1}}{m^{n+1}}|X_0^n\right] = E\left[\frac{X_{n+1}}{m^{n+1}}|X_n\right] = \frac{X_n}{m^n}.$$

By the martingale convergence theorem, almost surely

$$\lim_{n\uparrow\infty} \frac{X_n}{m^n} = Y < \infty.$$

In particular, if $m < 1$, then $\lim_{n\uparrow\infty} X_n = 0$ almost surely. Since X_n takes integer values, this implies that the branching process eventually becomes extinct.

If $m = 1$, then $\lim_{n\uparrow\infty} X_n = X_\infty < \infty$, and it is easily argued that this limit must be 0. Therefore, in this case as well the process eventually becomes extinct.

For the case $m > 1$, we consider the unique solution in $(0, 1)$ of $x = g(x)$ (Theorem 2.2.8). Suppose we can show that $Z_n = x^{X_n}$ is a martingale. Then, by the martingale convergence theorem, it converges to a finite limit, and therefore X_n has a limit X_∞, which, however, can be infinite. One can easily argue that this limit cannot be other than 0 (extinction) or ∞ (nonextinction). Since $\{Z_n\}_{n\geq 0}$ is a martingale, $x = E[Z_0] = E[Z_n]$, and therefore, by dominated convergence, $x = E[Z_\infty] = E[x^{X_\infty}] = P(X_\infty = 0)$. Therefore, x is the probability of extinction.

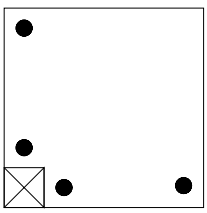
It remains to show that $\{Z_n\}_{n\geq 0}$ is a martingale. We have

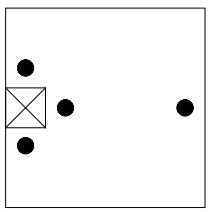
$$E[x^{X_{n+1}}|X_n = i] = x^i.$$

This is obvious if $i = 0$, and if $i > 0$, X_{n+1} is the sum of i independent random variables with the same generating function g. Therefore, $E[x^{X_{n+1}}|X_n = i] = g(x)^i = x^i$. From this last result and the Markov property,

$$E[x^{X_{n+1}}|X_0^n] = E[x^{X_{n+1}}|X_n] = x^{X_n}.$$

EXAMPLE 17.3.7: A CELLULAR AUTOMATON. Consider a chessboard of size $N \times N$, on which are placed stones, exactly one on each square. Each stone has one among k possible colours. The state X_n of the process at time n is the $N \times N$ matrix with elements in $\{1, \ldots, k\}$ describing the chessboard and the colour of the stone in each square. The transition from X_n to X_{n+1} is as follows. Select one case of the chessboard at random, and change the color of the stone there, the new colour being the colour of a stone chosen at random among the four neighboring stones. To avoid boundary effects, we shall suppose that the chessboard is a bi-torus in the sense of the figure below, where the black dots represent the neighbours of the crossed case.





We shall see that in finite time the chessboard becomes monochromatic and prove, using a martingale argument, that the probability of being absorbed in the, say red, monochromatic state is equal to the initial proportion of red states.

Denote by Y_n the proportion of red stones at stage n. The process $\{Y_n\}_{n\geq 0}$ is a martingale with respect to $\{X_n\}_{n\geq 0}$. Indeed, Y_n is a function of X_n and is

integrable, since it is bounded by 1. Also, $E[Y_{n+1}|X_0^n] = Y_n$, as the following exchange argument shows. Let α_{n+1} be the box selected at time $n+1$ and let β_{n+1} be the selected neighbor of α_{n+1}. Then, for any pair (α, β) of such boxes, $P(\alpha_{n+1} = \alpha, \beta_{n+1} = \beta | X_0^n) = P(\alpha_{n+1} = \beta, \beta_{n+1} = \alpha | X_0^n) = \frac{1}{8N^2}$. Clearly, if the result $\alpha_{n+1} = \alpha, \beta_{n+1} = \beta$ changes Y_n to $Y_{n+1} = Y_n + \Delta Y_{n+1}$, the result $\alpha_{n+1} = \beta$, $\beta_{n+1} = \alpha$ changes Y_n to $Y_{n+1} = Y_n - \Delta Y_{n+1}$. Since these two situations are equiprobable, the martingale property follows.

By the martingale convergence theorem, $\lim_{n\uparrow\infty} Y_n = Y$ exists, and by dominated convergence $E[Y] = \lim_{n\uparrow\infty} E[Y_n]$. Therefore, since $E[Y_n] = E[Y_0]$, we have $E[Y] = E[Y_0] = y_0$, where y_0 is the initial proportion of red stones. Because $|\Delta Y_n| = 0$ or $\frac{1}{N^2}$ for all n, $\{Y_n\}_{n\geq 0}$ can converge only if it remains constant after some (random) time, and this constant is either 0 or 1. Since the limit 1 corresponds to absorption by the "all-red" state, we see that the probability of being absorbed by the "all-red" state is equal to the initial proportion of red stones.

A Transience Criterion

The following simple application of the martingale convergence theorem generalizes Example 6.1.18:

Theorem 17.3.8 *An irreducible recurrent* HMC *has no non-negative superharmonic or bounded subharmonic functions besides the constant functions.*

Proof. If h is non-negative superharmonic (resp., bounded subharmonic), then the stochastic sequence $\{h(X_n)\}_{n\geq 0}$ is a non-negative supermartingale (resp., bounded submartingale) and therefore, by the martingale convergence theorem (Corollary 17.3.4), it converges to a finite limit Y. Since the chain visits any state $i \in E$ infinitely often, one must have $Y = h(i)$ almost surely for all $i \in E$. In particular, h is a constant. □

Theorem 17.3.9 *Let the transition matrix* $\mathbf{P}$ *on the discrete state space* E *be irreducible, and suppose that there exists a non-negative function* $h : E \to \mathbb{R}$ *such that*

$$\sum_{k\in E} p_{ik} h(k) \leq h(i), \text{ for all } i \notin F, \tag{17.11}$$

for some finite subset $F \subset E$. *Then the corresponding* HMC *is recurrent.*

Remark 17.3.10 The conditions of the above result are also necessary (we shall not prove this here), and this is why it is called a criterion.

Proof. Let $\tau = \tau(F)$ be the return time to F, and define $Y_n = h(X_n)1_{\{n<\tau\}}$. Equality (17.11) is just $E[h(X_{n+1}) \mid X_n = i] \leq h(i)$ for all $i \notin F$. For $i \notin F$, we have, using the basic rules for conditional expectation (Theorems 2.3.8, 2.3.7 and 2.3.6)

$$\begin{aligned} E_i[Y_{n+1} \mid X_0^n] &= E_i[Y_{n+1}1_{\{n<\tau\}} \mid X_0^n] + E_i(Y_{n+1}1_{\{n\geq\tau\}} \mid X_0^n] \\ &= E_i[Y_{n+1}1_{\{n<\tau\}} \mid X_0^n] \leq E_i[h(X_{n+1})1_{\{n<\tau\}} \mid X_0^n] \\ &= 1_{\{n<\tau\}}E_i[h(X_{n+1}) \mid X_0^n] = 1_{\{n<\tau\}}E_i[h(X_{n+1}) \mid X_n] \\ &\leq 1_{\{n<\tau\}}h(X_n), \end{aligned}$$

where the third *equality* comes from the fact that $1_{\{n<\tau\}}$ is a function of X_0^n, the fourth *equality* is the Markov property, and the last *inequality* is true because P_i-a.s., $X_n \notin F$ on $n < \tau$. Therefore, P_i-a.s., for $i \notin F$, P_i-a.s,

$$E_i[Y_{n+1}|X_0^n] \leq Y_n,$$

that is, $\{Y_n\}_{n\geq 0}$ is, under P_i, a non-negative supermartingale with respect to $\{X_n\}_{n\geq 0}$. By the martingale convergence theorem, $\lim_{n\uparrow\infty} Y_n = Y_\infty$ exists and is finite, P_i-a.s.

Suppose, in view of contradiction, that the chain is transient. It must then visit any finite subset of the state space only a finite number of times. In particular, for arbitrary K, we can have $h(X_n) < K$ only for a finite (random) number of indices n. This implies that $\lim_{n\to\infty} h(X_n) = +\infty, P_j$-a.s. (for any $j \in E$). For this to be compatible with the fact that $\{1_{\{n<\tau\}}h(X_n)\}$ has P_i-a.s. a finite limit for $i \notin F$, we must have $P_i(\tau < \infty) = 1$.

In summary, $P_i(\tau < \infty) = 1$ for all $i \notin F$. Since F is finite, some state in F must be recurrent, hence the announced contradiction. □

17.3.2 Optional Sampling

Recall the definition of a stopping time.

Definition 17.3.11 *Let $\{X_n\}_{n\geq 0}$ be some sequence of random variables taking their values in the discrete space $\mathcal{X}$. A random variable T taking integer values and possibly the value ∞ is called an X_0^n-stopping time if for all integers m the event $\{T = m\}$ is expressible in terms of X_0^n, that is, more precisely, there exists a function $g_m : \mathcal{X}^{m+1} \to \{0, 1\}$ such that*

$$1_{\{T=m\}} = g_m(X_0^{m+1}).$$

The following result is a weak form of Doob's optional sampling theorem.

Theorem 17.3.12 *Let $\{M_n\}_{n\geq 0}$ be an X_0^n-martingale, and let T be an X_0^n-stopping time. Suppose that at least one of the following conditions holds:*

(α) $P(T \leq n_0) = 1$ for some $n_0 \geq 0$.

(β) $P(T < \infty) = 1$ and $|M_n| \leq K < \infty$ when $n \leq T$.

Then

$$E[M_T] = E[M_0]. \tag{17.12}$$

Proof. (α) Write

$$M_T - M_0 = \sum_{k=0}^{n_0-1} (M_{k+1} - M_k) 1_{\{k<T\}}.$$

Since T is a stopping time of $\{X_n\}_{n\geq 0}$,

$$1_{\{k<T\}} = \varphi(X_0^k)$$

for some function φ, and therefore, using the basic rules of conditioning (Theorems 2.3.6 and 2.3.10)

$$\begin{aligned} E[(M_{k+1} - M_k)1_{\{k<T\}}] &= E[(M_{k+1} - M_k)\varphi(X_0^k)] \\ &= E\left[E[(M_{k+1} - M_k)\varphi(X_0^k)|X_0^k)]\right] \\ &= E\left[E[(M_{k+1} - M_k)|X_0^k)]\varphi(X_0^k)\right] = 0. \end{aligned}$$

Therefore,

$$E[M_T - M_0] = \sum_{k=0}^{n_0-1} E[(M_{k+1} - M_k)1_{\{k<T\}}] = 0.$$

(β) Apply the result of (α) to the finite stopping time $T \wedge n_0$ to obtain

$$E[M_{T\wedge n_0}] = E[M_0].$$

Therefore,

$$\begin{aligned} |E[M_T] - E[M_0]| &= |E[M_T] - E[M_{T\wedge n_0}]| \\ &\leq E[|M_T - M_{T\wedge n_0}|] \\ &= E[\sum_{k=n_0+1}^{\infty} |M_k - M_{k\wedge n_0}|\, 1_{\{k=T\}}] \\ &\leq E[\sum_{k=n_0+1}^{\infty} 2K 1_{\{k=T\}}] = 2KP(T > n_0). \end{aligned}$$

Since T is finite, $\lim_{n_0\uparrow\infty} P(T > n_0) = 0$, and therefore $E[M_T] = E[M_0]$. □

EXAMPLE 17.3.13: THE GAMBLER'S RUIN, TAKE 3. Consider the symmetric random walk $\{X_n\}_{n\geq 0}$ on $\mathbb{Z}$ with initial state 0. It is an X_0^n-martingale. Let T be the first time n for which $X_n = -a$ or $+b$, where $a, b > 0$. This is an X_0^n-stopping time, and moreover $T < \infty$. We can apply Theorem 17.3.12 (optional sampling), part (β), with $K = \sup(a, b)$, to obtain $0 = E[X_0] = E[X_T]$. With $v := P(-a \text{ is hit before } b)$, we have $E[X_T] = -av + b(1 - v)$, and therefore

$$P(-a \text{ is hit before } b) = \frac{b}{a+b}.$$

EXAMPLE 17.3.14: A COUNTEREXAMPLE. Consider the symmetric random walk of the previous example, but now define T to be the hitting time of $b > 0$. We know that $T < \infty$, since the symmetric walk on $\mathbb{Z}$ is recurrent. If the optional sampling theorem applied, we would have

$$0 = E[X_0] = E[X_T] = b,$$

an obvious contradiction. The optional sampling theorem (Theorem 17.3.12) does not apply because neither condition (α) nor (β) thereof is satisfied.

The Maximum Principle

The general appoach to the absorption problem given below is in terms of harmonic functions. However, the actual implementation of this approach requires one to find explicit forms of harmonic functions satisfying some boundary conditions, which is not always too easy. In contrast, the purely algebraic method can always be implemented in the finite state space case (at the cost of matrix computations).

Let $\{X_n\}_{n\geq 0}$ be an HMC with countable state space E and transition matrix $\mathbf{P}$. Let D be an arbitrary subset of E, called the domain, and let $\overline{D} := E\backslash D$. Let $c : D \to \mathbb{R}$ and $\varphi : \overline{D} \to \mathbb{R}$ be non-negative functions called the unit time gain function and the final gain function, respectively. Let τ be the hitting time of $\overline{D}$.

For each state $i \in E$, define

$$v(i) = E_i \left[\sum_{0\leq k<\tau} c(X_k) + \varphi(X_\tau) 1_{\{\tau<\infty\}} \right]. \tag{17.13}$$

The function $v : E \to \overline{\mathbb{R}}$ so defined is non-negative and possibly infinite. Note that τ is not required to be finite, and that $\overline{D}$ may be empty.

In the context of control theory, v is called the average reward function, since $v(i)$ is the average cost incurred when starting from state i, from the initial time $n = 0$ to the final time $n = \tau$, $c(X_n)$ being the running gain at time n and $\varphi(X_\tau)$ the final reward.

Theorem 17.3.15 *The function $v : E \to \overline{\mathbb{R}}_+$ defined by (17.13) satisfies the following properties:*

(i) it is non-negative and satisfies

$$v = \begin{cases} \mathbf{P}v + c & \text{on } D, \\ \varphi & \text{on } \overline{D}, \end{cases} \tag{17.14}$$

(ii) it is majored by any non-negative function $u : E \to \overline{\mathbb{R}}$ such that

$$u \geq \begin{cases} \mathbf{P}u + c & \text{on } D, \\ \varphi & \text{on } \overline{D}, \end{cases} \tag{17.15}$$

(iii) and moreover, if for all $i \in E$, $P_i(\tau < \infty) = 1$, then (17.14) has at most one non-negative bounded solution.

Proof. (i) Properties $v \geq 0$ and $v = \varphi$ on $\overline{D}$ are satisfied by definition. For $i \in D$, first-step analysis gives (Exercise 15.3.7)

$$v(i) = c(i) + \sum_{j \in E} p_{ij} v(j)\,. \tag{17.16}$$

(ii) Define for $n \geq 0$ the non-negative function $v_n : E \to \mathbb{R}$ by

$$v_n(i) = E_i \left[\sum_{k=0}^{n-1} c(X_k) 1_{\{k<\tau\}} + \varphi(X_\tau) 1_{\{\tau<n\}} \right]. \tag{17.17}$$

Observe that $v_0 \equiv 0$ and, by monotone convergence, $\lim_{n\uparrow\infty} \uparrow v_n = v$. Also, with a proof similar to that of (i),

$$v_{n+1} = \begin{cases} \mathbf{P}v_n + c & \text{on } D, \\ \varphi & \text{on } \overline{D}. \end{cases} \tag{17.18}$$

With u as in (17.15), we have $u \geq v_0$. We show by induction that $u \geq v_n$. This is true for $n = 0$. Suppose it is true for some n. We have $u \geq \mathbf{P}u + c \geq \mathbf{P}v_n + c = v_{n+1}$ on D, and $u \geq \varphi = v_{n+1}$ on $\overline{D}$. Therefore, $u \geq v_{n+1}$. Since $u \geq v_n$ for all $n \geq 0$, $u \geq \lim_{n\to\infty} v_n = v$.

(iii) Suppose that u satisfies

$$u = \begin{cases} \mathbf{P}u + c & \text{on } D, \\ \varphi & \text{on } \overline{D}\,. \end{cases}$$

Suppose in addition that it is bounded (note that this implies that c and φ are bounded) and non-negative. Then by Exercise 17.4.9,

$$M_n = u(X_n) - u(X_0) - \sum_{k=0}^{n-1} (\mathbf{P} - I) u(X_k) \tag{17.19}$$

is an X_0^n-martingale. By the optional sampling theorem (Theorem 17.3.12), for all integers K, $E_i[M_{\tau\wedge K}] = E_i[M_0] = 0$, and therefore, observing that $(I - \mathbf{P})\, u = c$ on D,

$$u(i) = E_i[u(X_{\tau\wedge K})] - E_i \left[\sum_{k=0}^{\tau\wedge K-1} (\mathbf{P} - I) u(X_k) \right] = E_i \left[u(X_{\tau\wedge K}) + \sum_{k=0}^{\tau\wedge K-1} c(X_k) \right].$$

Since $P_i(\tau < \infty) = 1$, $\lim_{K\uparrow\infty} E_i[u(X_{\tau\wedge K}] = E_i[u(X_\tau)]$ by dominated convergence. But $u(X_\tau) = \varphi(X_\tau)$ because $u = \varphi$ on $\overline{D}$. Therefore $\lim_{K\uparrow\infty} E_i[u(X_{\tau\wedge K}] = E_i[\varphi(X_\tau)]$. Also, $\lim_{K\uparrow\infty} E_i[\sum_{k=0}^{\tau\wedge K-1} c(X_k)] = E_i[\sum_{k=0}^{\tau-1} c(X_k)]$ by monotone convergence. Finally,

$$u(i) = E_i \left[\sum_{k=0}^{\tau-1} c(X_k) + \varphi(X_\tau) \right] = v(i).$$

□

Theorem 17.3.15 can be rephrased as follows. The function v given by (17.13) is a minorant of all non-negative solutions of (17.15), and for $u = v$, the inequalities in (17.15) become equalities. Moreover, if v is bounded and $P_i(\tau < \infty) = 1$ for all $i \in E$, then v is the *unique* bounded solution of (17.14).

Definition 17.3.16 *If $\mathbf{P}h = h$ on $A \subseteq E$, we say that h is harnonic on A.*

Corollary 17.3.17 *Let $\varphi : E \to \mathbb{R}$ be a bounded non-negative function, and let τ_B be the hitting time of $B \subset E$. Then, if $P_i(\tau_B < \infty) = 1$ for all $i \in E$,*

$$v(i) := E_i \left[\varphi(X_{\tau_B}) \right]$$

defines the unique bounded non-negative function $v : E \to \mathbb{R}$ that is harmonic on $\overline{B}$ and equal to φ on B.

EXAMPLE 17.3.18: APPLICATION TO THE ABSORPTION PROBLEM. Suppose that the transient set T is finite and that the recurrent classes R_1, R_2, ... are singletons, and therefore absorbing states, denoted by r_1, r_2, ... (As shown before, the general case can always be reduced to this one as far as absorption probabilities are concerned.) In Corollary 17.3.17, take for B the set of absorbing states, and therefore $\overline{B} = T$. Let $\varphi = 1_{\{r_1\}}$. As T is assumed finite, the time to absorption in one of the absorbing states is finite. The quantity $v(i)$ is just the probability of absorption in r_1. Therefore v is in this case the unique bounded non-negative function $v : E \to \mathbb{R}$ that is harmonic on T and equal to $\varphi = 1_{\{r_1\}}$ on $R := \{r_1, r_2, \ldots\}$.

Suppose that we want to compute the average time to absorption $E_i[\tau_R]$, $i \in T$. For this, we take in Theorem 17.3.15 $D = R$, $\tau = \tau_R$, $c(i) \equiv 1$, $\varphi \equiv 0$. Then v defined by $v(i) := E_i[\tau_R]$ is the unique bounded non-negative function such that $v = \mathbf{P}v + 1$ on T and $= 0$ on R.

EXAMPLE 17.3.19: APPLICATION TO OPTIMAL CONTROL. Consider a stochastic process $\{X_n\}_{n \geq 0}$ with values in E, that is controlled in the following way. Let $\{\mathbf{P}(a); a \in A\}$, where A is some set, the set of *actions*, be a family of transition matrices on E, with the interpretation that, if at time n the controlled process is in state i, and if the controller takes action a, then at time $n+1$ the state will be j with probability $p_{ij}(a)$. A *control strategy* u is a (measurable) function $u : E \to A$ which prescribes to take action $u(i)$ when the process is in state i. Therefore, under the strategy u, the controlled process is an HMC with transition matrix $\mathbf{P}^u$, where

$$p_{ij}^u = p_{ij}(u(i)).$$

There is a cost $V^u(i)$ associated with each strategy u and each initial state i, of the form

$$V^u(i) = E_i^u[\sum_{0\le k<T} c^u(X_k) + \varphi^u(X_T)1_{\{T<\infty\}}],$$

where c^u, φ^u and T are as in Theorem 17.3.15, with D fixed, and moreover, $c^u(i) = c(i, u(i))$ and $\varphi^u(i) = \varphi(i, u(i))$, for appropriate functions c and φ. The problem of *optimal control* is that of finding, if it exists, an *optimal strategy* u^*, such that

$$V^{u^*}(i) \ge V^u(i),$$

for all states i and all strategies u.

We have the following result. Suppose that there exists a function $V : E \to \mathbb{R}$ such that

$$V(i) = \sup_{a\in A}\left\{\sum_{j\in E} p_{ij}(a)V(j) + c(i,a)\right\} \text{ for all } i \in D,$$

and

$$V(i) = \sup_{a\in A} \varphi(i,a) \text{ for all } i \in \partial D,$$

and that the suprema above are attained for $a = u^*(i)$, for some (measurable) function $u^* : E \to A$. Then, u^* is an optimal control and $V = V^{u^*}$.

Proof. Since for all controls u,

$$V \ge \mathbf{P}^u V + c^u \text{ on } D,$$

and

$$V \ge \varphi^u \text{ on } \partial D,$$

it follows from Theorem 17.3.15 that

$$V \ge V^u$$

for all controls u. Also, $V = V^{u^*}$ and therefore u^* is an optimal control. □

17.4 Exercises

Exercise 17.4.1. POLYA'S URN
An urn initially contains b black balls and w white balls. At each step of the sequential replacement procedure, a ball is drawn at random and replaced by c balls of the same colour. Let B_n and W_n be the number of black and white balls respectively in the urn at the n-th step, so that

$$Y_n = \frac{B_n}{B_n + W_n}$$

is the fraction of black balls at the n-th step. Let $X_n := (B_n, W_n)$. Prove that the stochastic process $\{Y_n\}_{n\ge0}$ is a martingale with respect to $\{X_n\}_{n\ge0}$. Is there a limit as $n \to \infty$ for X_n?

Exercise 17.4.2. MARTINGALE TRANSFORMS
Let $\{X_n\}_{n\geq 0}$ be some sequence of random variables with values in the denumerable set $\mathcal{X}$. The sequence of complex-valued random variables $\mathbf{H} := \{H_n\}_{n\geq 1}$ is called an X_0^n-predictable process if for all $n \geq 1$,

$$H_n = g_n(X_0^{n-1})$$

for some function $g_n : \mathcal{X}^n \to \mathbb{C}$. Let $\mathbf{Y} := \{Y_n\}_{n\geq 0}$ be another sequence of complex random variables. The sequence $\mathbf{H} \circ \mathbf{Y} := \{(H \circ Y)_n\}_{n\geq 1}$ defined by

$$(H \circ Y)_n := \sum_{k=1}^{n} H_k(Y_k - Y_{k-1}), \quad n \geq 1$$

is called the transform of $\mathbf{Y}$ by $\mathbf{H}$. Prove the following:

(a) Let $\mathbf{Y}$ be an X_0^n-submartingale (resp., martingale) and let $\mathbf{H}$ be a bounded non-negative X_0^n-predictable process. Then $\mathbf{H} \circ \mathbf{Y}$ is an X_0^n-submartingale (resp., martingale).

(b) If $\mathbf{Y}$ is an X_0^n-martingale and if $\mathbf{H}$ is bounded and X_0^n-predictable, then $\mathbf{H} \circ \mathbf{Y}$ is an X_0^n-martingale.

Exercise 17.4.3. LIKELIHOOD RATIO MARTINGALE
Let $\{X_n\}_{n\geq 0}$ be a sequence of discrete random variables with values in E. Let for $n \geq 0$, $x_0, x_1, \ldots, x_n \in E$

$$p_n(x_0, x_1, \ldots, x_n) := P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n)\,.$$

Show that $p_n(X_0, X_1, \ldots, X_n) > 0$ almost surely. For each $n \geq 0$, let q_n be a function from E^{n+1} to $[0, 1]$ such that

$$\sum_{x_0, x_1, \ldots, x_n \in E} q_n(x_0, x_1, \ldots, x_n) = 1\,.$$

Show that the sequence $\{M_n\}_{n\geq 0}$ defined by

$$M_n := \frac{q_n(X_0, X_1, \ldots, X_n)}{p_n(X_0, X_1, \ldots, X_n)}$$

is a martingale.

Exercise 17.4.4.
Let $\{X_n\}_{n\geq 1}$ be an IID sequence of random variables with values in $\{-1, +1\}$, and such that $P(X_n = -1) = P(X_n = -1) = \frac{1}{2}$. Let $S_n := X_0 + X_1 + \cdots + X_n$ $(n \geq 0)$. Show that $\{S_n^2 - n\}_{n\geq 0}$ is an X_0^n-martingale.

Exercise 17.4.5. MARTINGALE WITH RESPECT TO A POINT PROCESS
Let $\{X_n\}_{n\geq 0}$ be a sequence of $\{0, 1\}$-valued random variables such that for all $n \geq 0$,

$$P(X_{n+1} = 1 \,|\, X_0^n) = \alpha_n(X_0^n)\,,$$

where α_n is a function from $\{0,1\}^{n+1}$ into $[0,1]$. Let $\{M_n\}_{n\geq 0}$ be a real-valued integrable X_0^n-martingale, necessarily (by definition) of the form $M_n = f_n(X_0^n)$ where f_n is a function from $\{0,1\}^{n+1}$ into $[0,1]$. Show that it can always be represented as

$$M_n = M_0 + \sum_{i=1}^{n} \varphi_{i-1}(X_0^{i-1})(X_i - \alpha_{i-1}(X_0^{i-1}))\,,$$

where for some functions $\varphi_{i-1} : \{0,1\}^i \to \mathbb{R}$ $(i \geq 1)$ such that for all $n \geq 1$,

$$\sum_{i=1}^{n} |\varphi_{i-1}(X_0^{i-1})|\alpha_{i-1}(X_0^{i-1}) < \infty\,.$$

Exercise 17.4.6. Convex functions of martingales
Let I be an interval of $\mathbb{R}$ of arbitrary nature with non-empty interior and let $\phi : I \to \mathbb{R}$ be a convex function.

A. Let $Y = \{Y_n\}_{n\geq 0}$ be an X_0^n-martingale such that $P(Y_n \in I) = 1$ for all $n \geq 0$. Assume that $E\,[|\phi(Y_n)|] < \infty$ for all $n \geq 0$. Show that the process $\{\phi(Y_n)\}_{n\geq 0}$ is an X_0^n-submartingale.

B. Assume in addition that ϕ is non-decreasing and suppose this time that Y is an X_0^n-submartingale. Show that the process $\{\phi(Y_n)\}_{n\geq 0}$ is an X_0^n-submartingale.

C. Let $Y = \{Y_n\}_{n\geq 0}$ be an X_0^n-martingale and let $p \geq 1$. Prove that $\{|Y_n|^p\}_{n\geq 0}$ and $\{Y_n^+\}_{n\geq 0}$ are X_0^n-submartingales.

Exercise 17.4.7. Hit probability
Let X be an hmc with state space E, and let B be a closed subset of states, that is,

$$i \in B \Rightarrow \sum_{j\in B} p_{ij} = 1.$$

Let T be the hitting time of B, and define for $i \in E$,

$$h(i) = P_i(T < \infty).$$

Show that $\{h(X_n)\}_{n\geq 0}$ is a martingale with respect to $\{X_n\}_{n\geq 0}$.

Exercise 17.4.8. The Doob transform
Let $\mathcal{P}$ be the transition matrix of an hmc $\{X_n\}_{n\geq 0}$ with state space E, and let $B \subset E$, that is either empty or with all states absorbing ($p_{ii} = 1$ for all $i \in B$). Let h be a positive harmonic function on $E\backslash B$. Define for all $i, j \in E$

$$\tilde{p}_{ij} := \frac{p_{ij}h(j)}{h(i)}\,.$$

(i) Show that $\tilde{\mathbf{P}} := \{\tilde{p}_{ij}\}_{i,j\in E}$ is a transition matrix. (It is called the Doob h-transform of $\mathbf{P}$.)

(ii) Take $B := \{a, b\}$. Let τ_a and τ_b be the hitting times of a and b respectively. Show that

$$h(i) := P_i(X_{\tau_a \wedge \tau_b} = b) \qquad (\star)$$

defines a harmonic function on $E \backslash B$.

(iii) Suppose that h as defined in $(\star)$ is positive. Let $\tilde{\mathbf{P}}$ be the transition matrix defined in (i). Show that

$$\tilde{p}_{ij} = P_i(X_1 = j \,|\, \tau_b < \tau_a)\,.$$

Exercise 17.4.9. The Lévy martingale
(i) Let $\{X_n\}_{n\geq 0}$ be an HMC with transition matrix $\mathbf{P}$ and state space E, and let $f : E \to \mathbb{R}$ be a bounded function. Show that the process

$$M_n^f = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (\mathbf{P} - I) f(X_k)$$

is a martingale with respect to $\{X_n\}_{n\geq 0}$.

(ii) Let $\{X_n\}_{n\geq 0}$ be a stochastic process with values in E. Let $\mathbf{P}$ be some transition matrix on E. Prove that if for all bounded $f : E \to \mathbb{R}$, $\left\{M_n^f\right\}_{n\geq 0}$ is a martingale with respect to $\{X_n\}_{n\geq 0}$, then $\{X_n\}_{n\geq 0}$ is a HMC with transition matrix $\mathbf{P}$.

Exercise 17.4.10. The unlimited gambler
Consider the gambling situation of Example 17.1.6 when the stakes are bounded, say by M, and when the initial fortune of the gambler is a. But we suppose that the gambler can borrow whatever amount he needs, so that his "fortune" Y_n at any time n can take arbitrary values. Prove that

$$P(|Y_n - a| \geq \lambda \leq 2 \exp\left(-\frac{\lambda^2}{2nM^2}\right)\,.$$

Exercise 17.4.11. Fair coin tosses
Consider a Bernoulli sequence of parameter $\frac{1}{2}$ representing a fair game of HEADS and TAILS. Let X be the number of HEADS after n tosses. Use Hoeffding's inequality to prove that

$$P(|X - E[X]| \geq \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{n}\right)\,.$$

Exercise 17.4.12. Empty bins
Consider the usual "balls and bins" setting with n bins and m balls (the multinomial distribution). Let X be the number of empty bins. Prove that

$$P(|X - E[X]| \geq \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{m}\right)\,.$$

Exercise 17.4.13. PATTERN MATCHING
Let $f(x)$ to be the number of occurences of the fixed pattern $b = (b_1, \dots, b_k)$ $(k \le N)$ in a sequence $x = (x_1, \dots, x_N)$ of elements of a finite set $\mathcal{X}$, that is

$$f(x) = \sum_{i=1}^{N-k+1} 1_{\{x_i = b_1, \dots, x_{i+k-1} = b_k\}} .$$

The mean number of matches in an IID sequence $X := (X_1, \dots, X_N)$ with uniform distribution on $\mathcal{X}$ is therefore

$$E[f(X)] = \sum_{i=1}^{N-k+1} E\left[1_{\{X_i = b_1, \dots, X_{i+k-1} = b_k\}}\right] = \sum_{i=1}^{N-k+1} \left(\frac{1}{|\mathcal{X}|}\right)^k$$

that is

$$E[f(X)] = (N-k+1)\left(\frac{1}{|\mathcal{X}|}\right)^k .$$

Prove that

$$P(|f(X) - E[f(X)]| \ge \lambda) \le 2e^{-\frac{1}{2}\frac{\lambda^2}{Nk^2}} .$$

Exercise 17.4.14. AN EXTENSION OF HOEFFDING'S INEQUALITY
Let M be a real X_0^n-martingale such that, for some sequence $d_1, d_2, \dots$ of real numbers,

$$P(B_n \le M_n - M_{n-1} \le B_n + d_n) = 1, \quad n \ge 1 ,$$

where for each $n \ge 1$, B_n is a function of X_0^{n-1}. Prove that, for all $x \ge 0$,

$$P(|M_n - M_0| \ge x) \le 2\exp\left(-2x^2 \Big/ \sum_{i=1}^{n} d_i^2\right) .$$

Exercise 17.4.15. RUINED AGAIN!
Show that the function $h(i) = \left(\frac{q}{p}\right)^i$ is harmonic for the nonsymmetric random walk on $\mathbb{Z}$ (with $p_{i,i+1} = p, p_{i,i-1} = q = 1-p$, $p \ne \frac{1}{2}$), where $p \in (0,1), p \ne \frac{1}{2}$. Apply the optional sampling theorem to obtain the ruin probability in the ruin problem of Example 6.1.3.

Exercise 17.4.16. MEAN HITTING TIME VIA MARTINGALES
Let X be a symmetric random walk on $\mathbb{Z}$. Show that X and $\{X_n^2 - n\}_{n\ge 0}$ are martingales with respect to $\{X_n\}_{n\ge 0}$. Deduce from this the mean of T of the hitting time of $-a, b$, where a and b are positive integers.

Exercise 17.4.17. ABSORPTION PROBABILITY
Consider the homogeneous Markov chain $\{X_n\}_{n\ge 1}$ with state space $E = \{0, 1, \dots, m\}$ and transition probabilities

$$p_{ij} = \binom{m}{j}\left(\frac{i}{m}\right)^j \left(1 - \frac{i}{m}\right)^{m-j} .$$

In particular, 0 and m are absorbing states.

(a) Show that $\{X_n\}_{n\geq 1}$ is a martingale.

(b) Compute the probability of absorption by state 0.

Exercise 17.4.18. THE BALLOT PROBLEM
In the ballot problem, let X_k be the number of votes in advance (can be negative) for candidate I after disclosure of the k-th bulletin, and define for $0 \leq k \leq n-1$, where $n := a + b$,

$$M_k := \frac{X_{n-k}}{n-k}\,.$$

(i) Prove that the sequence $M_0, M_1, \ldots, M_{n-1}$ forms an M_0^k-martingale.

(ii) Let A be the event that candidate I leads all the way to victory. Prove that

$$P(A) = \frac{a-b}{a+b}\,.$$

(Hint: consider the time τ at which $X_k = 0$ if such k exists, or $n-1$ otherwise.)

Chapter 18

Discrete Renewal Theory

18.1 Renewal processes

18.1.1 The Renewal Equation

In the analytic approach to Markov chains, the proof of convergence to steady state of an ergodic HMC is a consequence of a result on power series called the renewal theorem by the probabilists. This result forms the matter of this section. However, the renewal theorem will not be used as the essential step towards the convergence theorem, but on the contrary, it will be obtained as a corollary of the latter.

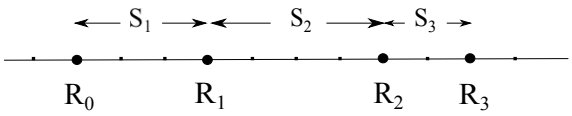
We start with the basic definitions. Let $\{S_n\}_{n\geq 1}$ be an IID sequence of random variables with values in $\{1, 2, \ldots, +\infty\}$ and with the probability distribution

$$P(S_1 = k) = f_k\,. \tag{18.1}$$

Define for $n \geq 0$,

$$R_{n+1} = R_n + S_{n+1}, \tag{18.2}$$

where R_0 is an arbitrary random variable with values in $\mathbb{N}$ (in particular, $R_0 < \infty$). The sequence $\{R_n\}_{n\geq 0}$ is called a delayed (by R_0) renewal sequence with the renewal distribution $\{f_k\}_{k\geq 1}$. If $R_0 \equiv 0$, one speaks of an undelayed renewal sequence, or, more simply, of a renewal sequence. If $P(S_1 = \infty) = 0$, the renewal sequence (delayed or not) is called a proper renewal sequence, and $\{f_k\}_{k\geq 1}$ is called a proper renewal distribution. Otherwise, one speaks of a defective renewal sequence and of a defective renewal distribution.



The quantity

$$\alpha := P(S_1 = \infty)$$

is the defect of the renewal distribution. The random time R_k is the kth renewal time, and the sequence $\{S_n\}_{n\geq 1}$ is the inter-renewal sequence.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_18

With the renewal distribution $\{f_k\}_{k\geq 1}$ is associated the *renewal equation*

$$u_n = v_n + \sum_{k=1}^{n} f_k u_{n-k} \tag{18.3}$$

(for $n = 0$, this reduces to $u_0 = v_0$). The sequence $\{u_n\}_{n\geq 0}$ is the *unknown sequence*, and $\{v_n\}_{n\geq 0}$ is the *data*, a sequence of real numbers such that

$$\sum_{k=0}^{\infty} |v_k| < \infty\,. \tag{18.4}$$

Since u_n can be computed recursively as a function of $u_0, \ldots, u_{n-1}, v_0, \ldots, v_n$, a solution of the renewal equation always exists and is unique.

EXAMPLE 18.1.1: LIFETIME OF A DEFECTIVE RENEWAL SEQUENCE. Define the lifetime L of a defective renewal sequence by

$$L := \inf\{R_k; k \geq 0, S_{k+1} = \infty\}\,.$$

It is the last renewal time at finite distance. We shall see that $u_n = P(L > n)$ satisfies a renewal equation. For this, write

$$1_{\{L>n\}} = 1_{\{L>n\}} 1_{\{S_1>n\}} + 1_{\{L>n\}} 1_{\{S_1\leq n\}}\,.$$

Observe that $\{L > n, S_1 > n\} = \{n < S_1 < \infty\}$. Also, denoting by $\hat{L}$ the lifetime associated with the renewal sequence $\{R_{n+1} - R_1\}_{n\geq 0}$, we have the set identity $\{L > n, S_1 \leq n\} = \{\hat{L} > n - S_1, S_1 \leq n\}$. Therefore,

$$P(L > n) = P(n < S_1 < \infty) + P(\hat{L} > n - S_1, S_1 \leq n)\,.$$

Now, L and $\hat{L}$ have the same distribution, and $\hat{L}$ is independent of S_1. Therefore,

$$P(\hat{L} > n-S_1, S_1 \leq n) = \sum_{k=1}^{n} P(\hat{L} > n-k)P(S_1 = k) = \sum_{k=1}^{n} P(L > n-k)P(S_1 = k)\,.$$

This shows that u_n satisfies the renewal equation with data $v_n = P(n < S_1 < \infty)$.

Definition 18.1.2 *Define the Dirac sequence* $\{\delta_n\}_{n\geq 0}$ *by* $\delta_0 = 1, \delta_n = 0$ *for* $n \geq 1$. *When the data is the Dirac sequence, the renewal equation is called the basic renewal equation, and its solution the fundamental solution.*

The fundamental solution will be denoted by $\{h_n\}_{n\geq 0}$, and therefore $h_0 = 1$, and for $n \geq 1$,

$$h_n = \sum_{k=1}^{n} f_k h_{n-k}\,. \tag{18.5}$$

The fundamental solution has a very simple interpretation. Indeed, h_n is the probability that n is a renewal time (we then say, for short, "n is renewal"). It suffices

to show that $u_n = P(n \text{ is renewal})$ is the unique solution of the basic renewal equation. Clearly, $u_0 = 1$. Also,

$$\begin{aligned}
P(n \text{ is renewal}) &= \sum_{k=0}^{n-1} P(n \text{ is renewal, last renewal strictly before } n \text{ is } k) \\
&= \sum_{i=0}^{\infty}\sum_{k=0}^{n-1} P(S_{i+1} = n-k,\ k = R_i) \\
&= \sum_{i=0}^{\infty}\sum_{k=0}^{n-1} P(S_{i+1} = n-k)P(k = R_i) \\
&= \sum_{k=0}^{n-1} P(S_1 = n-k)\left(\sum_{i=0}^{\infty} P(k = R_i)\right) \\
&= \sum_{k=0}^{n-1} P(S_1 = n-k)P(k \text{ is renewal }) \\
&= \sum_{k=0}^{n-1} u_k f_{n-k} = \sum_{k=1}^{n} f_k u_{n-k}\,.
\end{aligned}$$

Therefore,

$$h_k = P(k \text{ is a renewal time}). \tag{18.6}$$

In particular, if ν_n is the number of renewal times R_k in the interval $[0, n]$, then

$$\nu_n = \sum_{k=0}^{n} h_k\,. \tag{18.7}$$

We now introduce a definition and a convenient notation. The *convolution* of two real sequences $\{x_n\}_{n\geq 0}$ and $\{y_n\}_{n\geq 0}$ is the real sequence $\{z_n\}_{n\geq 0}$ defined by

$$z_n = \sum_{k=0}^{n} x_k y_{n-k}\,.$$

This is written for short as $z = x * y$.

Theorem 18.1.3 *The renewal equation (18.3) has a unique solution*

$$u = h * v\,. \tag{18.8}$$

Proof. Existence and uniqueness have already been observed. To check that the announced solution is correct, write the renewal equation as $u = v + f * u$ (with $f_0 = 0$) and the fundamental equation as $h = \delta + f * h$. Inserting (18.8) into the renewal equation gives $h * v = v + f * (h * v)$ which is indeed true, since the right-hand side is $v + (f * h) * v = v + (h - \delta) * v = v + h * v - \delta * v$, that is, $h * v$, because $\delta * v = v$. □

EXAMPLE 18.1.4: GEOMETRIC INTER-RENEWAL TIMES. When the distribution of the typical inter-renewal time is geometric, i.e., for $k \geq 1$,

$$P(S_1 = k) = p(1-p)^{k-1},$$

the fundamental solution is given by $h_0 = 1$, and

$$h_n = p,$$

for $n \geq 1$, as can be readily checked. The solution of the general renewal equation is then

$$u_n = v_n + p(v_0 + \cdots + v_{n-1}).$$

One observes in this particular case that since $\lim_{n\uparrow\infty} v_n = 0$ in view of assumption (18.4),

$$\lim_{n\uparrow\infty} u_n = p\sum_{k=0}^{\infty} v_k = \frac{\sum_{k\geq 0} v_k}{\sum_{k\geq 1} k f_k}.$$

This result will be generalized by the renewal theorem.

18.1.2 Renewal Theorem

The renewal distribution $\{f_k\}_{k\geq 1}$ is called lattice (resp., non-lattice) if $d := g.c.d.\{k \ ; \ k \geq 1, f_k > 0\} > 1$ (resp., $= 1$); the integer d is called the span of the renewal distribution.

Theorem 18.1.5 *Let $\{f_k\}_{k\geq 1}$ be a non-lattice and proper renewal distribution. For the unique solution of the renewal equation with data satisfying assumption (18.4),*

$$\lim_{n\uparrow\infty} u_n = \frac{\sum_{k\geq 0} v_k}{\sum_{k\geq 1} k f_k}, \tag{18.9}$$

where the ratio on the right-hand side is 0 if $\sum_{k\geq 1} k f_k = \infty$.

Proof. A. Assume the result true for the fundamental solution, that is,

$$\lim_{n\uparrow\infty} h_n = \frac{1}{\sum_{k\geq 1} k f_k} := h_\infty. \tag{18.10}$$

From expression (18.8) of the solution in terms of the fundamental solution, we obtain

$$\sum_{k=0}^{n} (h_{n-k} - h_\infty) v_k = u_n - h_\infty \sum_{k=0}^{n} v_k.$$

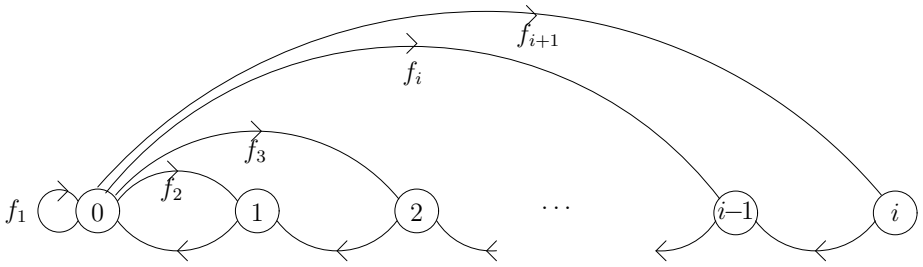
The result follows if we can prove that the left-hand side of the above equality converges to 0 as $n \to \infty$. Indeed, with $g(n,k) = (h_{n-k} - h_\infty) v_k 1_{\{k\leq n\}}$, we have for fixed k, $\lim_{n\uparrow\infty} g(n,k) = 0$, and $|g(n,k)| \leq |v_k|$, where $\sum_{k\geq 0} |v_k| < \infty$. Therefore, by dominated convergence for series, $\lim_{n\uparrow\infty} \sum_{k\geq 0} g(n,k) = \sum_{k\geq 0} \lim_{n\uparrow\infty} g(n,k) = 0$.

B. It remains to prove (18.10). For this, introduce a Markov chain with state space $E = \mathbb{N}$ if the support of $\{f_k\}_{k\geq 1}$ is unbounded, and state space $E = \{0, \ldots, M-1\}$ if $M < \infty$ is the largest value of S_1. We suppose for definiteness that $E = \mathbb{N}$. The nonzero entries of the transition matrix are

$$\begin{aligned} p_{i,i-1} &= 1,\ i \geq 1\,, \\ p_{0i} &= f_{i+1},\ i \geq 0\,. \end{aligned}$$

The corresponding transition graph is shown in the figure below. Note that this is the transition graph of the forward recurrence time HMC $\{X_n\}_{n\geq 0}$ defined by

$$X_n = \inf\{R_k\,;\ R_k \geq n\} - n\,.$$



Transition graph of the forward recurrence chain

This chain is clearly irreducible. The distribution of the return time to state 0 is

$$P_0(T_0 = n) = f_n\,.$$

Event $\{T_0 = n\}$ implies event $\{X_n = 0\}$, and therefore $P_0(X_n = 0) \geq P_0(T_0 = n)$. Consequently, the set $A = \{n \geq 1; p_{00}(n) > 0\}$ contains the set $B = \{n \geq 1; f_n > 0\}$, and therefore the g.c.d. of A is smaller than or equal to the g.c.d. of B. Therefore, the g.c.d. of A equals 1, that is, the chain is aperiodic.

Since the renewal distribution is assumed proper, we have $P_0(T_0 < \infty) = \sum_{n\geq 1} f_n = 1$, and therefore the chain is recurrent. If $E_0[T_0] < \infty$, it is ergodic, and then

$$\lim_{n\uparrow\infty} p_{00}(n) = \pi_0 = \frac{1}{E_0[T_0]}.$$

If the chain is not ergodic but only null recurrent, then $\lim_{n\uparrow\infty} p_{00}(n) = 0$ by Orey's theorem. In both cases, since $E_0[T_0] = \sum_{k\geq 1} k f_k$,

$$\lim_{n\uparrow\infty} p_{00}(n) = \frac{1}{\sum_{k\geq 1} k f_k}.$$

The proof of (18.9) is complete because $p_{00}(n) = P(n \text{ is renewal}) = h_n$. □

Corollary 18.1.6 *Under the same conditions as in Theorem 18.1.5, except that the span d of the renewal distribution is now strictly greater than 1, the solution of the renewal equation (18.3) with data satisfying (18.4) satisfies, for all $r \in [0, d-1]$,*

$$\lim_{N\uparrow\infty} u_{r+Nd} = d\frac{\sum_{k\geq 0} v_{r+kd}}{\sum_{k\geq 1} k f_k}\,. \tag{18.11}$$

Proof. Observe that when $\{f_k\}_{k\geq 1}$ is proper and lattice with span d, the distribution $\{f_{Nd}\}_{N\geq 1}$ is proper and non-lattice. On the other hand, the renewal equation (18.3) splits into d renewal equations. The rth one ($r \in [0, d-1]$) is

$$u_{r+Nd} = v_{r+Nd} + \sum_{\ell=1}^{N} f_{\ell d} u_{r+Nd-\ell d}\,,$$

where N is the time variable. The renewal theorem can be applied to each one, and we obtain (18.11) after observing that

$$\sum_{N=1}^{\infty} N f_{Nd} = \frac{1}{d}\sum_{N=1}^{\infty} N d f_{Nd} = \frac{1}{d}\sum_{k=1}^{\infty} k f_k\,.$$

□

18.1.3 Defective Renewal Theorem

Theorem 18.1.7 *Suppose that the renewal distribution is defective, and that the data sequence of the renewal equation is nonnegative and satisfies (instead of (18.4))*

$$\lim_{n\uparrow\infty} v_n = v_\infty < \infty. \tag{18.12}$$

The solution of the renewal equation then satisfies

$$\lim_{n\uparrow\infty} u_n = \frac{v_\infty}{\alpha}, \tag{18.13}$$

where $\alpha = P(S_1 = \infty)$ is the defect of the renewal distribution.

Proof. The forward recurrence HMC in the proof of Theorem 18.1.5 now has $E = \mathbb{N} \cup \{+\infty\}$ for state space. All states besides $+\infty$ are transient. In particular, the average number of visits to 0 is finite:

$$\nu_\infty = \sum_{k=0}^{\infty} h_k < \infty.$$

From the expression of the solution

$$u_n = \sum_{k=0}^{n} h_k v_{n-k}\,,$$

we therefore obtain by the dominated convergence for series (see Theorem 1.6 of the Appendix),

$$\lim_{n\uparrow\infty} u_n = (\sum_{k=0}^{\infty} h_k)v_\infty = \nu_\infty v_\infty.$$

Now, the probability of n visits to 0 is $(1-\alpha)^{n-1}\alpha$, and therefore the average number of visits to 0 is $\nu_\infty = \frac{1}{\alpha}$. □

Theorem 18.1.8 *Suppose that the renewal distribution is non-lattice and defective, that there exists $\gamma > 1$ such that*

$$\sum_{k=0}^{\infty} \gamma^n f_n = 1, \tag{18.14}$$

and that the data sequence satisfies

$$\sum_{k=0}^{\infty} \gamma^n |v_n| < \infty. \tag{18.15}$$

The solution of the renewal equation then satisfies

$$\lim_{n\uparrow\infty} \gamma^n u_n = \frac{\sum_{k=0}^{\infty} \gamma^k v_k}{\sum_{k=0}^{\infty} k\gamma^k f_k}. \tag{18.16}$$

Proof. Observe that if we define $\tilde{f}_n = \gamma^n f_n$, $\tilde{v}_n = \gamma^n v_n$, $\tilde{u}_n = \gamma^n u_n$, then

$$\tilde{u}_n = \tilde{v}_n + \sum_{k=1}^{n} \tilde{f}_k \tilde{u}_{n-k}.$$

This renewal equation is non-lattice and proper, and therefore the announced result follows from the renewal theorem. □

A consequence of (18.14) is the exponential decay of the renewal distribution. This shows in particular that (18.14) is an assumption that is *not always satisfied*.

EXAMPLE 18.1.9: CONVERGENCE RATE IN THE DEFECTIVE CASE. The situation is that of Theorem 18.1.10, where in addition the renewal distribution is assumed non-lattice, and moreover, (18.14) is true for some $\gamma > 1$. We seek to understand how u_n tends to u_∞. For this we define $\hat{u}_n = u_n - u_\infty$. Rewriting the renewal equation for u_n as

$$u_n - u_\infty = v_n - u_\infty + \sum_{k=1}^{n} f_k(u_{n-k} - u_\infty) + u_\infty \sum_{k=1}^{n} f_k,$$

we see that $\hat{u}_n$ satisfies the renewal equation with data

$$\hat{v}_n = v_n - u_\infty P(S_1 > n).$$

We can therefore apply Theorem 3.5 to obtain, after rearrangement,

$$\lim_{n\uparrow\infty} \gamma^n (u_n - \frac{v_\infty}{\alpha}) = \frac{1}{\gamma} \left\{ \frac{\sum_{k=0}^{\infty} \gamma^k v_k}{\sum_{k=0}^{\infty} \gamma^k P(S_1 > k)} - \frac{v_\infty}{P(S_1 = \infty)} \right\}.$$

An *excessive* renewal equation is one for which $\sum_{k=1}^{\infty} f_n > 1$. Theorem 18.1.10 then has an obvious counterpart. Note that in the excessive case (18.14) *always* has a solution γ, and of course it is in $(0, 1)$.

EXAMPLE 18.1.10: THE LOTKA–VOLTERRA MODEL. At each time $n \in \mathbb{Z}$, an average number u_n of daughters is born. Each of them gives birth independently of the other women. The average number of daughters of any given woman in the kth year of her life, $k \geq 1$, is f_k. At time 0 the population has $\alpha(i)$ women of age i. Expressing that u_n is the sum of v_n, the average number of daughters born at time n from mothers born at or before time 0, and of r_n, the average number of daughters born at time n from mothers born strictly after time 0 and up to time n, we obtain the renewal equation with data sequence

$$v_n = \sum_{i=0}^{\infty} \alpha(i) f_{n+i}.$$

In this context, the renewal equation is known as the *Lotka–Volterra equation*. Denote by

$$\rho = \sum_{k=1}^{\infty} f_n$$

the average number of daughters of any given woman, and assume that this number is positive and finite. Assume also that it is different from 1. Assume that γ defined by (18.14) exists and that the renewal distribution is non-lattice. Denoting by C the right-hand side of (18.16),

$$\lim_{n\uparrow\infty} \gamma^n u_n = C.$$

Note that $\gamma < 1$ if $\rho > 1$, and $\gamma > 1$ if $\rho < 1$. The first case corresponds to exponential explosion, whereas the second case is that of exponential extinction.

18.1.4 Renewal Reward Theorem

Theorem 18.1.11 *Let $\{S_n\}_{n\geq 1}$ be an* IID *sequence of positive random variables such that $E[S_1] < \infty$, and let R_0 be a finite non-negative random variable independent of this sequence. Define for all $n \geq 0$, $R_{n+1} = R_n + S_{n+1}$ and for $n \geq 0$, $N(n) = \sum_{k=1}^{\infty} 1_{\{R_k \leq n\}}$. Now let $\{Y_n\}_{n\geq 1}$ be an* IID *sequence of random variables such that $E[|Y_1|] < \infty$. Then*

$$\lim_{n\uparrow\infty} \frac{N(n)}{n} = \frac{1}{E[S_1]}$$

and

$$\lim_{n\uparrow\infty} \frac{\sum_{k=1}^{N(n)} Y_k}{n} = \frac{E[Y_1]}{E[S_1]}\,.$$

Proof. Since $R_{N(n)} \leq n < R_{N(n)+1}$, we have

$$\frac{N(n)}{R_{N(n)+1}} < \frac{N(n)}{n} \leq \frac{N(n)}{R_{N(n)}}.$$

But the right-most term is the inverse of

$$\frac{R_{N(n)}}{N(n)} = \frac{R_0 + \sum_{k=1}^{N(n)} S_k}{N(n)}.$$

By the strong law of large numbers and the fact that $\lim_{n\uparrow\infty} N(n) = \infty$ (the S_n's are finite), this quantity tends to $E[S_1]$, and similarly, $\frac{R_{N(n)+1}}{N(n)} = \frac{R_{N(n)+1}}{N(n)+1}\frac{N(n)+1}{N(n)}$ tends to $E[S_1]$ as $n \to \infty$. The proof of the second formula follows from the strong law of large numbers and the first formula, since

$$\frac{\sum_{k=1}^{N(n)} Y_k}{n} = \frac{\sum_{k=1}^{N(n)} Y_k}{N(n)} \cdot \frac{N(n)}{n}.$$

□

18.2 Regenerative Processes

18.2.1 Basic Definitions and Examples

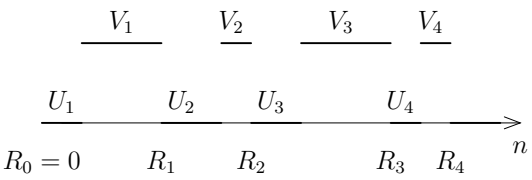
In the introductory lines of the previous section, we mentioned that the renewal theorem can be proven independently of the basic results of convergence to steady state, and that it could even be used to prove such convergence results. Therefore, it seems that the probabilistic approach to convergence gives a negligible status to the renewal theorem, which enjoys a central position in the analytic approach. However, the renewal theorem remains indispensable in the proof of convergence to equilibrium of stochastic processes of a more general nature than homogeneous Markov chains, namely regenerative processes. The common feature that such processes share with the homogeneous recurrent Markov chains is the existence of regenerative cycles.

Definition 18.2.1 *Let $\{Z_n\}_{n\geq 0}$ be a stochastic process with values in an arbitrary state space E and let $\{R_n\}_{n\geq 0}$ be a delayed renewal sequence. The process $\{Z_n\}_{n\geq 0}$ is said to be regenerative with respect to the renewal sequence $\{R_n\}_{n\geq 0}$ if for all $k \geq 0$, $\{Z_{n+R_k}\}_{n\geq 0}$ is independent of $R_0, S_1, \ldots, S_k$ and has the same distribution as $\{Z_{n+R_0}\}_{n\geq 0}$.*

Note that the definition does not require that $\{Z_{n+R_k}\}_{n\geq 0}$ be independent of $\{Z_n, n \in [0, R_k - 1]\}$, although in many examples this is satisfied. The freedom resulting from the relaxed conditions can be very useful.

EXAMPLE 18.2.2: RECURRENT MARKOV CHAINS. Let $\{X_n\}_{n\geq 0}$ be an irreducible recurrent HMC, with arbitrary initial distribution. Let $\{R_n\}_{n\geq 0}$ be the successive *hitting* times of state 0. The regenerative cycle theorem (Theorem 7.4 of Chapter 2) tells us that $\{X_n\}_{n\geq 0}$ is regenerative with respect to $\{R_n\}_{n\geq 0}$.

EXAMPLE 18.2.3: RELIABILITY. Let $\{U_n\}_{n\geq 1}$ and $\{V_n\}_{n\geq 1}$ be two independent IID sequences of positive integer-valued random variables. Define the sequence $\{S_n\}_{n\geq 1}$ by $S_n = U_n + V_n$, and let $\{R_n\}_{n\geq 0}$ be the associated nondelayed renewal sequence ($R_0 \equiv 0$). Define a $\{0, 1\}$-valued process $\{Z_n\}_{n\geq 0}$ as in the figure below. Clearly, $\{Z_n\}_{n\geq 0}$ is a regenerative process with respect to $\{R_n\}_{n\geq 0}$.



A sample path of the reliability process

Regenerative processes generate renewal equations and are the main motivation for the study of such equations. For instance, if $f : E \to \mathbb{R}$ is a nonnegative function, and if $\{Z_n\}_{n\geq 0}$ is an E-valued process regenerative with respect to the nondelayed renewal sequence $\{R_n\}_{n\geq 0}$, then the sequence $\{u_n\}_{n\geq 0}$, where $u_n = E[f(Z_n)]$, satisfies a renewal equation. Indeed,

$$E[f(Z_n)] = E[f(Z_n)1_{\{n<S_1\}}] + E[f(Z_n)1_{\{n\geq S_1\}}],$$

and, setting $\tilde{Z}_n = Z_{n+S_1}$, we have

$$\begin{aligned}
E[f(Z_n)1_{\{n\geq S_1\}}] &= E[f(\widetilde{Z}_{n-S_1})1_{\{n\geq S_1\}}] \\
&= \sum_{k=1}^{\infty} E[f(\widetilde{Z}_{n-S_1})1_{\{n\geq S_1\}}1_{\{S_1=k\}}] \\
&= \sum_{k=1}^{n} E[f(\widetilde{Z}_{n-k})1_{\{S_1=k\}}] \\
&= \sum_{k=1}^{n} E[f(\widetilde{Z}_{n-k})]P(S_1 = k) \\
&= \sum_{k=1}^{n} E[f(Z_{n-k})]P(S_1 = k),
\end{aligned}$$

where the independence of S_1 and $\{\widetilde{Z}_n\}_{n\geq 0}$, as well as the assumption of equidistribution of $\{\widetilde{Z}_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$, have been taken into account. Therefore,

$$E[f(Z_n)] = E[f(Z_n)1_{\{n<S_1\}}] + \sum_{k=1}^{n} E[f(Z_{n-k})]P(S_1 = k),$$

which is precisely the renewal equation with data

$$v_n = E[f(Z_n)1_{\{n<S_1\}}].$$

18.2.2 The Regenerative Theorem

Observe that

$$\sum_{n=0}^{\infty} |v_n| = \sum_{n=0}^{\infty} |E[f(Z_n)1_{\{n<S_1\}}]| \leq E[\sum_{n=0}^{\infty} |f(Z_n)|1_{\{n<S_1\}}] = E[\sum_{n=0}^{S_1-1} |f(Z_n)|].$$

Therefore, by the renewal theorem, we have the following:

Theorem 18.2.4 *Let $\{Z_n\}_{n\geq 0}$ be a nondelayed ($R_0 = 0$) regenerative process and let $f : E \to \mathbb{R}$ be such that*

$$E\left[\sum_{n=0}^{S_1-1} |f(Z_n)|\right] < \infty. \tag{18.17}$$

If the distribution of S_1 is proper and non-lattice, then

$$\lim_{n\uparrow\infty} E[f(Z_n)] = \frac{E\left[\sum_{n=0}^{S_1-1} f(Z_n)\right]}{E[S_1]}. \tag{18.18}$$

EXAMPLE 18.2.5: RELIABILITY FORMULA. This is a continuation of Example 18.2.3. We assume that $S_1 = U_1 + V_1$ is proper and non-lattice. Applying the regenerative theorem with $f(z) = 1_{\{0\}}(z)$, and assuming $E[U_1] < \infty$, we find that

$$\lim_{n\uparrow\infty} P(Z_n = 0) = \frac{E[U_1]}{E[U_1] + E[V_1]},$$

since $E[f(Z_n)] = E[1_{\{0\}}(Z_n)] = P(Z_n = 0)$, and $\sum_{n=0}^{S_1-1} 1_{\{0\}}(Z_n) = U_1$.

EXAMPLE 18.2.6: THE BUS PARADOX. Consider the renewal sequence with $R_0 = 0$. Define for each $n \geq 0$ the backward recurrence time B_n and the forward recurrence time F_n by

$$B_n = n - L_n, \; F_n = N_n - n,$$

where

$$L_n = \sup\{R_k;\ k \geq 0,\ R_k \leq n\}$$

and

$$N_n = \inf\{R_k; k > 0,\ R_k > n\}.$$

In particular, if $n = R_m$ for some m, then $B_n = 0$ and $F_n = R_{m+1} - R_m = S_{m+1}$. Observe that $F_n \geq 1$ for all $n \geq 0$. Also, if $n \in [R_m, R_{m+1})$, then $B_n + F_n = S_{m+1}$. The regenerative theorem with $Z_n = (B_n, F_n)$ and $f(Z_n) = 1_{\{B_n=i\}}1_{\{F_n=j\}}$ gives, provided that the distribution of S_1 is proper and non-lattice,

$$\lim_{n\uparrow\infty} P(B_n = i, F_n = j) = \frac{P(S_1 = i + j)}{E[S_1]}\,, \qquad (\star)$$

Indeed the sum $\sum_{n=0}^{S_1-1} 1_{\{B_n=i,F_n=j\}}$ has at most one non-zero term, in which case it is equal to 1. For this term, say corresponding to the index $n = n_0$, $B_{n_0} + F_{n_0} = S_1 = i + j$. Therefore the sum is equal to $1_{\{S_1=i+j\}}$.

Summing $(\star)$ from $j = 1$ to ∞, and recalling that $F_n \geq 1$, one obtains

$$\lim_{n\uparrow\infty} P(B_n = i) = \frac{P(S_1 > i)}{E[S_1]}. \qquad (18.19)$$

Similarly, for the forward recurrence time,

$$\lim_{n\uparrow\infty} P(F_n = j) = \frac{P(S_1 \geq j)}{E[S_1]}.$$

The roles of B_n and F_n are not symmetric. To restore symmetry, one must consider B_n and $F'_n = F_n - 1$ (recall that $F_n \geq 1$). Then

$$\lim_{n\uparrow\infty} P(F'_n = j) = \frac{P(S_1 > j)}{E[S_1]}.$$

Since $B_n + F_n = S_m$ for some (random) m determined by the condition $n \in [R_m, R_{m+1})$, one might expect that $P(B_n + F_n = k) = P(S_1 = k)$. But this is in general false, and constitutes the apparent *paradox of recurrence times*, also called the *bus paradox* (Exercise 18.3.7). It is true that $P(B_n + F_n = k) = P(S_m = k)$, but m is random, and therefore there is no reason why S_m should have the same distribution as S_1. As a matter of fact,

$$\begin{aligned}\lim_{n\uparrow\infty} P(B_n + F_n = k) &= \lim_{n\uparrow\infty} \sum_{\substack{i,j\\ i+j=k}} P(B_n = i, F_n = j)\\ &= \sum_{\substack{i,j\\ i+j=k}} \frac{P(S_1 = i + j)}{E[S_1]} = \frac{kP(S_1 = k)}{E[S_1]}.\end{aligned}$$

Theorem 18.2.7 *Let $\{Z_n\}_{n\geq 0}$ be a possibly delayed regenerative process (recall, however, that $R_0 < \infty$). Let $f : E \to \mathbb{R}$ be such that*

$$\lim_{n\uparrow\infty} E[f(Z_n)1_{\{n<R_0\}}] = 0$$

and

$$E\left[\sum_{k=R_0}^{R_1-1} |f(Z_k)|\right] < \infty\,.$$

Then, if the renewal distribution is proper and non-lattice,

$$\lim_{n\uparrow\infty} E[f(Z_n)] = \frac{E\left[\sum_{k=R_0}^{R_1-1} f(Z_k)\right]}{E[S_1]}\,. \tag{18.20}$$

Proof. It suffices to show that the limit of $E[f(Z_n)1_{n\geq R_0}]$ equals the right-hand side of (18.20). Introduce $\{\widetilde{Z}_n\}_{n\geq 0} = \{Z_{n+R_0}\}_{n\geq 0}$, and observe that this process is an undelayed regenerative process with respect to $\{R_n - R_0\}_{n\geq 1}$ that is proper and non-lattice. We have

$$E[f(Z_n)1_{\{n\geq R_0\}}] = E[f(\widetilde{Z}_{n-R_0})1_{\{n\geq R_0\}}] = \sum_{k=0}^{n} E[f(\widetilde{Z}_{n-k})P(R_0 = k).$$

By the non-delayed version of the regenerative theorem, we have that

$$\lim_{n\uparrow\infty} E[f(\widetilde{Z}_n)] = \frac{E\left[\sum_{k=R_0}^{R_1-1} f(Z_k)\right]}{E[R_1 - R_0]},$$

and therefore, by dominated convergence for series (see Theorem 1.6 of the Appendix),

$$\lim_{n\uparrow\infty} \sum_{k=0}^{n} E[f(\widetilde{Z}_{n-k})]P(R_0 = k) = \frac{E\left[\sum_{k=R_0}^{R_1-1} f(Z_k)\right]}{E[R_1 - R_0]}\,.$$

□

A useful case where the conditions are satisfied is when f is bounded (use dominated convergence).

Books for Further Information

Discrete regenerative theory is treated in a few classic texts on probability, and is the theme of the more specialized and mathematical monograph [Kingman, 1972].

18.3 Exercises

Exercise 18.3.1. A PEDESTRIAN

At a crosswalk, cars pass on a single lane at times $R_0 = 0, R_1, R_2, \ldots$, where $\{R_n\}_{n\geq 0}$ is a proper renewal sequence. A pedestrian arriving at time 0 crosses the lane as soon as he sees a time interval $x > 0$ between two consecutive cars. How long must he wait, on average?

Exercise 18.3.2. LIFETIME
Let L be the lifetime of a defective renewal sequence. Show that $\lim_{n\uparrow\infty} P(L > n) = 0$, and give the rate of convergence. Treat in detail the case where the inter-renewal sequence is geometric.

Exercise 18.3.3. THE BASIC RENEWAL THEOREM
Let $\nu((a, b])$ be the average number of renewal epochs in the integer interval $(a, b]$ of a proper non-lattice renewal sequence. What is the limit as $n \uparrow \infty$ of $\nu((n + a, n + b])$?

Exercise 18.3.4. RENEWAL THEOREM WITH MULTIPLE EVENTS
Suppose that the typical inter-renewal time S_1 of a renewal sequence is proper but that $P(S_1 = 0) = f_0 > 0$. Otherwise, suppose that $\text{GCD}\{n \geq 1; f_n > 0\} = 1$. Show that the solution of the *extended* renewal equation

$$u_0 = v_0,\ u_n = v_n + \sum_{k=0}^{n} f_k u_{n-k},\ n \geq 1$$

(notice the additional term in the sum, corresponding to $k = 0$) satisfies, under the summability condition (18.4),

$$\lim_{n\uparrow\infty} u_n = \frac{\sum_{k\geq 0} v_k}{\sum_{k=1} k f_k}.$$

Exercise 18.3.5. ASYMPTOTICS OF THE LOTKA–VOLTERRA MODEL
In the population model of Example 18.1.10 what is, in the critical case $\rho = 1$, the average number of daughters born at a given large time, when $\alpha(i) = 0$ for all $i > 0$, and $\alpha(0) = 1$? Suppose now that $f_k = e^{-\beta}\frac{\theta^{k-1}}{(k-1)!}$ for $k \geq 1$. Discuss the asymptotic behavior of u_n, the average number of daughters born at time n, in terms of the positive parameters β and θ (use the same initial conditions as in Example 18.1.10).

Exercise 18.3.6. MAINTENANCE
A given machine can be in either one of three states: G (*good*), M (*in maintenance*), or R (*in repair*). Its successive periods where it is in state G (resp., M, R) form an independent and identically distributed sequence $\{S_n\}_{n\geq 0}$ (resp., $\{U_n\}_{n\geq 0}$, $\{V_n\}_{n\geq 0}$) with finite mean. All these sequences are assumed mutually independent. The maintenance policy uses a number $T > 0$. If the machine has age T and has not failed, it goes to state M. If it fails before it has reached age T, it enters state R. From states M and R, the next state is G. Find the steady state probability that the machine is operational. (Note that "good" does not mean "operational." The machine can be"good" but, due to the operations policy, in maintenance, and therefore not operational. However, after a period of maintenance or of repair, we consider that the machine starts anew, and enters a G period.)

Exercise 18.3.7. THE BUS PARADOX
In Example 18.2.6, when the typical inter-renewal time is geometric, compute $\lim_{n\uparrow\infty} P(F_n+B_n = k)$, $\lim_{n\uparrow\infty} P(F_n = k)$, and $\lim_{n\uparrow\infty} P(F_n+B_n = k)$. In Example 18.2.6, under what circumstances do we have $\lim_{n\uparrow\infty} P(F_n + B_n = k) = P(S_1 = k)$ for all $k \geq 1$?

Chapter 19

Monte Carlo

19.1 Approximate Sampling

19.1.1 Basic Principle and Algorithms

Recall the method of the inverse in order to generate a discrete random variable Z with distribution $P(Z = a_i) = p_i \quad (0 \leq i \leq K)$. A crude algorithm based on this method would perform successively the tests $U \leq p_0?$, $U \leq p_0 + p_1?$, ..., until the answer is positive. Although very simple in principle, the inverse method has the following drawbacks when the size r of the state space E is large.

(a) Problems arise that are due to the small size of the intervals partitioning $[0, 1]$ and to the cost of precision in computing.

(b) In random field simulation, another, maybe more important, reason is the necessity to enumerate the configurations, which implies coding and decoding of a mapping from the integers to the usually very large configuration space.

(c) Another situation is that in which the probability density π is known only up to a normalizing factor, that is, $\pi(i) = K\tilde{\pi}(i)$, and when the sum $\sum_{i \in E} \pi(i) = K^{-1}$ that gives the normalizing factor is prohibitively difficult to compute. In physics, this is a frequent case.

The quest for a random generator without these ailments is at the origin of the Monte Carlo Markov chain (MCMC) sampling methodology.

The basic principle is the following. One constructs an irreducible aperiodic HMC $\{X_n\}_{n \geq 0}$ with state space E and stationary distribution π. Since the state space is finite, the chain is ergodic, and therefore, by Theorem 15.1.1, for any initial distribution μ and all $i \in E$,

$$\lim_{n \to \infty} P_\mu(X_n = i) = \pi(i)\,. \tag{19.1}$$

Therefore, for large n, X_n has a distribution close to π.

The first task is that of designing the MCMC algorithm. One must find an ergodic transition matrix $\mathbf{P}$ on E, with stationary distribution π. In the Monte Carlo context, the transition mechanism of the chain is called a *sampling algorithm*,

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_19

and the asymptotic distribution π is called the *target distribution*, or *sampled distribution*.

There are infinitely many transition matrices with a given target distribution, and among them there are infinitely many that correspond to a reversible chain, that is, such that

$$\pi(i)p_{ij} = \pi(j)p_{ji}.$$

We seek solutions of the form

$$p_{ij} = q_{ij}\alpha_{ij} \tag{19.2}$$

for $j \neq i$, where $Q = \{q_{ij}\}_{i,j\in E}$ is an arbitrary irreducible transition matrix on E, called the *candidate-generator* matrix. When the present state is i, the next *tentative* state j is chosen with probability q_{ij}. When $j \neq i$, this new state is accepted with probability α_{ij}. Otherwise, the next state is the same state i. Hence, the resulting probability of moving from i to j when $i \neq j$ is given by (19.2). It remains to select the *acceptance* probabilities α_{ij}.

EXAMPLE 19.1.1: METROPOLIS, TAKE 1. In this algorithm (Metropolis et al., 1953) , $\alpha_{ij} = \min(1, (\pi(j)q_{ji})/\pi(i)q_{ij}))$. In Physics, one often finds distributions of the form

$$\pi(i) = \frac{e^{-U(i)}}{Z}, \tag{19.3}$$

where $U : E \to \mathbb{R}$ is the "energy function" and Z is the "partition function", the normalizing constant ensuring that π is indeed a probability vector. The acceptance probability of the transition from i to j is then, assuming the candidate-generating matrix to be *symmetric*,

$$\alpha_{ij} = \min\left(1, e^{-(U(j)-U(i))}\right).$$

EXAMPLE 19.1.2: BARKER'S ALGORITHM. (Barker, 1965) This algorithm, corresponds to the choice $\alpha_{ij} = (\pi(j)q_{ji})/(\pi(j)q_{ji} + \pi(i)q_{ij})$. When the distribution π is of the form (19.3), the acceptance probability of the transition from i to j is, with a *symmetric* candidate-generating matrix,

$$\alpha_{ij} = \frac{e^{-U(i)}}{e^{-U(i)} + e^{-U(j)}}.$$

This corresponds to the basic principle of statistical thermodynamics: when there are two states 1 and 2 with energies E_1 and E_2, Nature chooses 1 with probability $\frac{e^{-E_1}}{e^{-E_1}+e^{-E_2}}$.

The interest of the above algorithms resides in the fact that their implementation requires the knowledge of the target distribution π only up to a normalizing constant, since it depends only on the ratios $\pi(j)/\pi(i)$ (this in particular avoids

the need to compute the normalizing constant Z in (19.3), which is often inaccessible to exact computation). The latter statement is true only as long as the candidate-generating matrix Q is known.

EXAMPLE 19.1.3: THE GIBBS ALGORITHM. Consider a multivariate probability distribution

$$\pi(x(1), \ldots, x(N))$$

on a set $E = \Lambda^N$, where Λ is countable. The basic step of the Gibbs sampler for π consists in selecting a coordinate index i $(1 \leq i \leq N)$ at random, and choosing the new value $y(i)$ of the corresponding coordinate, given the present values $x(1), \ldots, x(i-1), x(i+1), \ldots, x(N)$ of the other coordinates, with probability

$$\pi(y(i) \mid x(1), \ldots, x(i-1), x(i+1), \ldots, x(N)).$$

One checks as above that π is the stationary distribution of the corresponding chain.

19.1.2 Sampling Random Fields

Let $X \in \Lambda^V$ be a random field on the finite set of vertices V, finite phase space Λ and probability distribution π. In the following examples, we apply the above general method for sampling π by constructiong an ergodic Markov chain $\{X_n\}_{n \geq 0}$ with state space $E = \Lambda^V$ with stationary distribution π.

EXAMPLE 19.1.4: GIBBS SAMPLER. The Gibbs sampler uses a strictly positive probability distribution $(q_v, v \in V)$ on V, and the transition from $X_n = x$ to $X_{n+1} = y$ is made according to the following rule. The new state y is obtained from the old state x by changing (or not) the value of the phase at *one site only*. The site v whose phase is to be modified (or not) at time n is chosen independently of the past with probability q_v. When site v has been selected, the current configuration x is changed into y as follows: $y(V \backslash v) = x(V \backslash v)$, and the new phase $y(v)$ at site v is selected with probability $\pi(y(v) \mid x(V \backslash v))$. Thus, configuration x is changed into $y = (y(v), x(V \backslash v))$ with probability $q_v \pi(y(v) \mid x(V \backslash v)$, according to the local specification at site v. This gives for the nonzero entries of the transition matrix

$$P(X_{n+1} = y \mid X_n = x) = q_v \pi(y(v) \mid x(V \backslash v)) 1_{\{y(V \backslash v) = x(V \backslash v)\}}. \tag{19.4}$$

Suppose that the corresponding chain is irreducible and aperiodic. To prove that π is the stationary distribution, we check for the detailed balance equations. We must have for all states $x, y \in \Lambda^V$ that differ only by the phase at site v,

$$\pi(x) P(X_{n+1} = y \mid X_n = x) = \pi(y)\, P(X_{n+1} = x \mid X_n = y),$$

that is, in view of (19.4), for all $v \in V$,

$$\pi(x)\, q_v\, \pi(y(v) \mid x(V \backslash v)) = \pi(y)\, q_v\, \pi(x(v) \mid x(V \backslash v)).$$

This is indeed so, since the last equality reduces to the identity

$$\pi(x)\, q_v \, \frac{\pi(y(v), x(V\backslash v))}{P(X(V\backslash v) = x(V\backslash v))} = \pi(y(v), x(V\backslash v))\, q_v \, \frac{\pi(x)}{P(X(V\backslash v) = x(V\backslash v))}.$$

EXAMPLE 19.1.5: ISING MODEL, TAKE 3: WHAT MAGNETS DO. In the Ising model, the local characteristic at site v depends only on $x(\mathcal{N}_v)$. The Gibbs sampler is a "natural" sampler, in that it is an idealization of what happens in nature as physicists understand it. In a piece of ferromagnetic material, for instance, the spins are randomly changed according to the local specification. When nature decides to update the orientation of a dipole, it does so according to the law of statistical mechanics. It computes the local energy

$$U(x(v), x(\mathcal{N}_v)) = x(v) \left(\frac{J}{k} \sum_{w \sim v} x(w) + \frac{H}{k} \right)$$

for each of the two possible spins, that is $U_+ = U(+1, x(\mathcal{N}_v))$ and $U_- = U(-1, x(\mathcal{N}_v))$, and takes the corresponding orientation with a probability proportional to e^{-U_+} and e^{-U_-}, respectively, according to the fundamental law of statistical mechanics (the so-called Gibbs principle).

EXAMPLE 19.1.6: PERIODIC GIBBS SAMPLER. In practice, the updated sites are not chosen at random, but instead in a well-determined order $v(1), v(2), \ldots, v(N)$, where $\{v(i)\}_{1 \le i \le N}$ is an enumeration of all the sites of V, called a *scanning policy*. The sites are visited in this order periodically. The state of the random field after the n-th sweep is $Z_n = X_{nN}$, where X_k denotes the image before the kth update time. At time k, site $v(k \bmod N)$ is updated to produce the new image X_{k+1}. If $X_k = x$ and $v(k \bmod N) = v$, then $X_{k+1} = (y(v), x(V\backslash v))$ with probability $\pi(y(v) \mid x(V\backslash v))$. The Gibbs distribution π is stationary for $\{X_k\}_{k \ge 0}$, in the sense that if $P(X_k = \cdot) = \pi$, then $P(X_{k+1} = \cdot) = \pi$. In particular, π is a stationary distribution of the irreducible aperiodic Markov chain $\{Z_n\}_{n \ge 0}$, and $\lim_{n \uparrow \infty} P(Z_n = \cdot) = \pi$.

The transition matrix $\mathbf{P}$ of $\{Z_n\}_{n \ge 0}$ is

$$\mathbf{P} = \prod_{k=1}^{N} \mathbf{P}_{v(k)}\,, \tag{19.5}$$

where $\mathbf{P}_v = \left\{ p_{xy}^v \right\}_{x,y \in \Lambda^V}$, and the entry p_{xy}^v of $\mathbf{P}_v$ is nonzero if and only if $y(V\backslash v) = x(V\backslash v)$, and then

$$p_{xy}^v = \frac{e^{-\mathcal{E}(y(v), x(V\backslash v))}}{\sum_{\lambda \in \Lambda} e^{-\mathcal{E}(\lambda, x(V\backslash v)}}\,. \tag{19.6}$$

This expression will be used to produce a geometric rate of convergence of the periodic Gibbs sampler, namely,

$$|\mu^T \mathbf{P}^n - \pi| \;\le\; \frac{1}{2}|\mu - \pi|(1 - e^{-N\Delta})^n \,, \tag{19.7}$$

where $\Delta = \sup_{v\in V} \delta_v$ and

$$\delta_v = \sup\{|\mathcal{E}(x) - \mathcal{E}(y)| \;;\; x(V\backslash v) = y(V\backslash v)\} \,.$$

By (15.10),

$$|\mu^T \mathbf{P}^n - \pi| \;\le\; \frac{1}{2}|\mu - \pi|\delta(\mathbf{P})^n \,.$$

It follows that for any transition matrix $\mathbf{P}$ on a finite state space E,

$$\delta(\mathbf{P}) = 1 - \inf_{i,j\in E} \sum_{k\in E} p_{ik} \wedge p_{jk} \;\le\; 1 - |E| \left(\inf_{i,j\in E} p_{ij} \right) . \tag{19.8}$$

If we define $m_v(x) = \inf\{\mathcal{E}(y);\, y(V\backslash v) = x(V\backslash v)\}$, it follows from (19.6) that

$$p^v_{xy} = \frac{\exp\{-(\mathcal{E}(y(v), x(V\backslash v)) - m_v(x))\}}{\sum_{z(v)\in\Lambda} \exp\{-(\mathcal{E}(z(v), x(V\backslash v)) - m_v(x))\}} \ge \frac{e^{-\delta_v}}{|\Lambda|} \,,$$

and therefore, from (19.5),

$$\min_{x,y\in\Lambda^V} p_{xy} \ge \prod_{k=1}^{N} \frac{e^{-\delta_{v(k)}}}{|\Lambda|} \ge \frac{e^{-N\Delta}}{|\Lambda|^N} \,.$$

Using (19.8),

$$\delta(\mathbf{P}) \le 1 - |\Lambda|^N \frac{e^{-N\Delta}}{|\Lambda|^N} = 1 - e^{-N\Delta} \,,$$

and (19.7) follows.

EXAMPLE 19.1.7: BIRTH-AND-DEATH POINT PROCESS. Consider the point process model of Subsection 9.1.3. The Gibbs sampling procedure is the following. Choose v uniformly in V, and replace the phase $x(v)$ at site v by $y(v)$ chosen at random in $\{0, 1\}$ according to the probability $\pi(\cdot \,|\, x(V\backslash v))$. Therefore, if $x(v) = 0$ and $y(v) = 1$, there is a "birth" at site v, whereas the situation $x(v) = 1$ and $y(v) = 0$ corresponds to a "death" at site v.

EXAMPLE 19.1.8: PROPERLY COLOURED GRAPHS. The phase space Λ consists of a finite number of "colours" labeled from 1 to q. We describe a Markov chain $\{X_n\}_{n\ge 0}$ taking its values in the subset F of $E := \Lambda^V$ consisting of the "properly coloured" configurations, that is configurations x such that $x(v) \neq x(w)$ whenever $v \sim w$. We start from a properly coloured configuration X_0. Suppose at time n the state is x. We then choose uniformly at random a site v, and then choose uniformly at random a colour in the set of colours allowable at v in configuration x, that is

$$A_v(x) := \{j \in \{1, 2, \ldots, q\} \,;\, j \neq x(w) \text{ for all } w \text{ such that } w \sim v\} \,.$$

The new state at time $n+1$ is then y which is equal to x except for the new colour j at site v. This chain is irreducible if there are at least three colours, which we henceforth assume. The non-null elements of the transition matrix are

$$p_{xy} = \frac{1}{|V|} \times \frac{1}{|A_v(x)|},$$

where x and y differ only in the colour at site v. Note that for such "adjacent" configurations, $A_v(x) = A_v(y)$, and therefore $p_{xy} = p_{yx}$. This implies in particular that the uniform distribution (on F) is the stationary distribution of this chain.

19.1.3 Variance of Monte Carlo Estimators

We now consider the problem of evaluating expectations with respect to the target distribution by ergodic estimates. In Theorem 15.1.9, we obtained the formula

$$v(f, \mathbf{P}, \pi) := 2 \langle f, \mathbf{Z}f \rangle_\pi - \langle f, (I + \Pi)f \rangle_\pi \tag{19.9}$$

giving the asymptotic variance

$$v(f, \mathbf{P}, \pi) = \lim_{n\to\infty} \frac{1}{n} \operatorname{Var}_\mu \left(\sum_{k=1}^{n} f(X_k) \right).$$

Here $\{X_n\}$ is an ergodic HMC with finite state space E, transition matrix $\mathbf{P}$, and stationary distribution π, and

$$\mathbf{Z} = (I - \mathbf{P} + \Pi)^{-1}, \tag{19.10}$$

where $\Pi = \mathbf{1} \cdot \pi^T$, is the fundamental matrix.

Consider *reversible* transition matrices, such as those corresponding to the MCMC simulation algorithms. One may be interested in designing the best simulation algorithm in the sense that $v(f, \mathbf{P}, \pi)$ is to be minimized with respect to $\mathbf{P}$, uniformly in f, and of course for a fixed π. The following result answers the question in general terms.

Theorem 19.1.9 *(Peskun, 1973) Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be reversible ergodic transition matrices on the finite state space E, with the same stationary distribution π. If $\mathbf{P}_1$ has all its off-diagonal terms greater than or equal to the corresponding off-diagonal terms of $\mathbf{P}_2$, then*

$$v(f, \mathbf{P}_1, \pi) \le v(f, \mathbf{P}_2, \pi)$$

for all $f : E \to \mathbb{R}$.

Proof. Let $k, \ell \in E$ with $k \neq \ell$. From (19.9) we have

$$\frac{\partial}{\partial p_{k\ell}} v(f, \mathbf{P}, \pi) = 2 \left\langle f, \frac{\partial \mathbf{Z}}{\partial p_{k\ell}} f \right\rangle_\pi.$$

From $\mathbf{Z}\mathbf{Z}^{-1} = I$, it follows that $\left(\frac{\partial}{\partial p_{k\ell}}\mathbf{Z}\right)\mathbf{Z}^{-1} + \mathbf{Z}\left(\frac{\partial}{\partial p_{k\ell}}\mathbf{Z}^{-1}\right) = 0$, and therefore

$$\frac{\partial \mathbf{Z}}{\partial p_{k\ell}} = -\mathbf{Z}\frac{\partial \mathbf{Z}^{-1}}{\partial p_{k\ell}}\mathbf{Z},$$

so that

$$\frac{\partial}{\partial p_{k\ell}} v(f, \mathbf{P}, \pi) = -2\left\langle f, \left(\mathbf{Z}\frac{\partial \mathbf{Z}^{-1}}{\partial p_{k\ell}}\mathbf{Z}\right) f\right\rangle_{\pi}.$$

Since $\mathbf{P}$ is autoadjoint in $\ell^2(\pi)$, so is $\mathbf{Z}$, and therefore

$$\frac{\partial}{\partial p_{k\ell}} v(f, \mathbf{P}, \pi) = -2\left\langle \mathbf{Z}f, \left(\frac{\partial \mathbf{Z}^{-1}}{\partial p_{k\ell}}\right)\mathbf{Z}f\right\rangle_{\pi} = -2(\mathbf{Z}f)^T d(\Pi)\frac{\partial \mathbf{Z}^{-1}}{\partial p_{k\ell}}\mathbf{Z}f.$$

Now, from (19.10),

$$\frac{\partial \mathbf{Z}^{-1}}{\partial p_{k\ell}} = -\frac{\partial \mathbf{P}}{\partial p_{k\ell}},$$

and therefore

$$\frac{\partial}{\partial p_{k\ell}} v(f, \mathbf{P}, \pi) = 2(\mathbf{Z}f)^T d(\Pi)\frac{\partial \mathbf{P}}{\partial p_{k\ell}}\mathbf{Z}f.$$

Observe that since $\mathbf{P}$ is a stochastic matrix and $(\mathbf{P}, \pi)$ is reversible, the free parameters are $(p_{k\ell}; k < \ell)$. In view of the reversibility condition, the only non-null elements of $d(\Pi)\frac{\partial \mathbf{P}}{\partial p_{k\ell}}$ are the (ℓ, ℓ), (ℓ, k), (k, ℓ), and (k, k) elements, respectively equal to $-\pi(k), +\pi(k), +\pi(k)$, and $-\pi(k)$. Therefore, $d(\Pi)\frac{\partial \mathbf{P}}{\partial p_{k\ell}}$ is a negative definite symmetric matrix, and

$$\frac{\partial}{\partial p_{k\ell}} v(f, \mathbf{P}, \pi) \leq 0\,,$$

from which the conclusion follows. □

EXAMPLE 19.1.10: OPTIMALITY OF METROPOLIS. In the so-called Hastings algorithms

$$p_{ij} = q_{ij}\frac{s_{ij}}{1 + t_{ij}},$$

where t_{ij} depends on the candidate-generating matrix Q and π only. We would like to find the best MCMC algorithm in the Hastings class where Q is fixed. We have observed that from the constraints $\leq \alpha_{ij} \in (0, 1)$ and the required symmetry of $\{s_{ij}\}_{i,j\in E}$,

$$s_{ij} \leq 1 + \min(t_{ij}, t_{ji})\,,$$

with equality for the Metropolis algorithm. It follows from Theorem 19.1.9 that the Metropolis algorithm is optimal with respect to asymptotic variance in the class of Hastings algorithms with fixed candidate-generating matrix Q.

It is interesting to compare a given MCMC algorithm corresponding to a reversible pair $(\mathbf{P}, \pi)$ to independent sampling for which $\mathbf{P} = \pi$. From the variance point of

view, it follows from (19.9) that an MCMC algorithm based on $\mathbf{P}$ performs better than independent sampling uniformly in f if and only if

$$\langle f, \mathbf{Z}f \rangle_\pi \leq \langle f, f \rangle_\pi \tag{19.11}$$

for all $f : E \to \mathbb{R}$.

From (19.9), $\langle f, \mathbf{Z}f \rangle_\pi \geq 0$ for all f, and we have already observed that Z is self-adjoint in $\ell^2(\pi)$. Therefore its eigenvalues are real and nonnegative. Condition (19.11) is equivalent to the fact that these eigenvalues are smaller than or equal to 1. Therefore, in view of (19.10), (19.11) is equivalent to $\mathbf{P} - \Pi$ having all its characteristic roots negative or null.

EXAMPLE 19.1.11: BARKER SAMPLING AND INDEPENDENT SAMPLING. The trace of a matrix is by definition the sum of its diagonal elements. For a stochastic matrix it is therefore the sum of its elements minus the sum of its off-diagonal elements. In particular, tr $(\mathbf{P}) = r - \sum_{i>j}(p_{ij} + p_{ji})$. Since tr $(\Pi) = 1$, we have

$$\text{tr } (\mathbf{P} - \Pi) = r - 1 - \sum_{i>j}(p_{ij} + p_{ji}).$$

One can verify that for Barker's algorithm

$$\min(q_{ij}, q_{ji}) \leq p_{ij} + p_{ji} \leq \max(q_{ij}, q_{ji})$$

with equality if Q is symmetric. Therefore, in the case where Q is symmetric,

$$\text{tr } (\mathbf{P} - \Pi) = r - 1 + \sum_{i>j} q_{ij} \geq \frac{1}{2}(r - 2).$$

Thus, if $r \geq 2$, the sum of the characteristic roots of $\mathbf{P} - \Pi$ is positive, which implies that at least one characteristic root is positive.

Therefore Barker's algorithm is *not uniformly better* than independent sampling. This does not mean that Barker's algorithm cannot perform better than independent sampling for a specific f. Moreover, and more importantly, the fact that an MCMC algorithm does not perform as well as independent sampling is not too alarming, since MCMC algorithms are used when independent sampling cannot be implemented.

We now give a lower bound for the asymptotic variance of any MCMC estimator. Let $(\mathbf{P}, \pi)$ be a reversible pair, where $\mathbf{P}$ is irreducible. Its r (real) eigenvalues are ordered as follows:

$$\lambda_1 = 1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_r \geq -1.$$

For a given f, the formula

$$v(f, \mathbf{P}, \pi) = \sum_{j=1}^{r} \frac{1 + \lambda_j}{1 - \lambda_j} |\langle f, v_j \rangle_\pi|^2$$

obtained in Theorem 15.1.9 fully accounts for the interaction between f and $\mathbf{P}$, in terms of the asymptotic variance of the ergodic estimate of $\langle f \rangle_\pi$. Since the function $x \to \frac{1+x}{1-x}$ is increasing in $(0,1]$, and λ_2 is the second largest eigenvalue of $\mathbf{P}$, the worst (maximal) value of the performance index

$$\gamma(f, \mathbf{P}, \pi) = \frac{v(f, \mathbf{P}, \pi)}{\mathrm{Var}_\pi(f)} = \frac{\sum_{j=2}^r \frac{1+\lambda_j}{1-\lambda_j} |\langle f, v_j \rangle_\pi|^2}{\sum_{j=2}^r |\langle f, v_j \rangle_\pi|^2} \tag{19.12}$$

is attained for $f = v_2$, and is then equal to

$$\gamma(\mathbf{P}, \pi) = \frac{1+\lambda_2}{1-\lambda_2}. \tag{19.13}$$

Let $M(\pi)$ be the collection of irreducible transition matrices $\mathbf{P}$ such that the pair $(\mathbf{P}, \pi)$ is reversible, and denote by $\lambda_2(\mathbf{P})$ the second largest eigenvalue of $\mathbf{P}$. Assume that

$$\pi(1) \le \pi(2) \le \cdots \le \pi(r).$$

In particular, $0 < \pi(1) \le \frac{1}{2}$.

19.1.4 Monte Carlo Proof of Holley's Inequality

Recall the statement of Theorem 9.3.5. Let P and P' be probabilities on E, P strictly positive, such that

$$P'(x \vee y)P(x \wedge y) \ge P'(x)P(y) \text{ for all } x, y \in E\,. \tag{$\star\star$}$$

Then for any increasing set $A \subseteq E$, we have Holley's inequality:

$$P'(A) \ge P(A)\,.$$

Proof. We first give a Metropolis algorithm generating a probability measure P on E. The corresponding HMC $\{X_n\}_{n\ge 0}$ evolves as follows. If $X_n = x$, select an index $L_n = \ell$ uniformly at random in $\{1, 2, \ldots, L\}$ and select $Y_n = y_\ell$ uniformly at random in $\{0, 1\}$. Let y be identical to x except perhaps for the ℓ-th bit, equal to the just selected y_ℓ. This defines the candidate-generating matrix. Then, with (acceptance) probability $\min\left(\frac{P(y)}{P(x)}, 1\right)$, let $X_{n+1} = y$, otherwise let $X_{n+1} = x$. The acceptance is implemented by a random variable U_n uniformly distributed on $[0,1]$: acceptance of the candidate y is decided if and only if $U_n \le \min\left(\frac{P(y)}{P(x)}, 1\right)$. The sequences $\{L_n\}_{n\ge 0}$, $\{Y_n\}_{n\ge 0}$ and $\{U_n\}_{n\ge 0}$ are IID and mutually independent.

Note that if the support $S(P)$ of P is an increasing set (necessarily containing $\mathbf{1}$), the HMC so defined is irreducible on $S(P)$. Its stationary distribution is P.

Remember that condition $(\star\star)$ is equivalent to the following:

$$x \ge y \Longrightarrow P'(x+\ell)P(y) \ge P'(x)P(y+\ell) \text{ for all } x, y \in E^0_\ell, \text{ all } \ell\,, \tag{$\dagger\dagger$}$$

which implies in particular that the support of P' is increasing.

We define two HMC $\{X_n\}_{n\geq 0}$ and $\{X'_n\}_{n\geq 0}$ evolving in parallel according to the above Metropolis algorithm and using the same random sequences $\{L_n\}_{n\geq 0}$, $\{Y_n\}_{n\geq 0}$ and $\{U_n\}_{n\geq 0}$. Therefore, both HMC's are defined on the same probability space, say $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$. The initial states are $X_0 = \mathbf{0}$ and $X'_0 = \mathbf{1}$. We will show that $X'_n \geq X_n$ for all $n \geq 0$. In particular, for any increasing set A, $\widetilde{P}(X'_n \in A) \geq \widetilde{P}(X_n \in A)$. Therefore, passing to the limit as $n \uparrow \infty$, $P'(A) \geq P(A)$.

It remains to prove that if at a given step n, $X_n = x \geq X'_n = x'$, then at the next step, $X_{n+1} = y \geq X'_{n+1} = y'$. We can assume that $x_\ell = x'_\ell = 1 - Y_n$ since otherwise the inequality is obvious. Two possibilities remain to be examined:

(i) If $Y_n = 1$, then $x, x' \in E^0_\ell$ and the inequality is satisfied if and only if

$$\frac{P'(x' + \ell)}{P'(x')} \geq \frac{P(x + \ell)}{P(x)} ,$$

which is guaranteed by (††) because $x' \geq x$.

(ii) If $Y_n = 0$, let $y, y' \in E^0_\ell$ be such that $x = y + \ell$ and $x' = y' + \ell$. The inequality is satisfied if and only if

$$\frac{P'(y' + \ell)}{P'(y')} \geq \frac{P(y + \ell)}{P(y)} ,$$

which is guaranteed by (††) because $y' \geq y$. □

19.2 Simulated Annealing

19.2.1 The Search for a Global Minimum

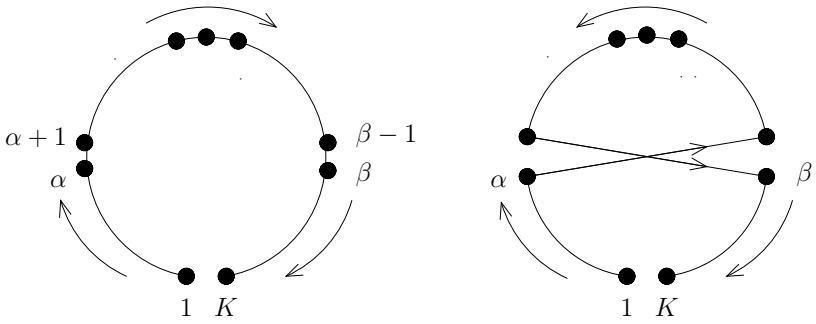
Let E be a finite set. A function U defined on this set and with real values, called the *cost function*, is to be minimized. More precisely, one is looking for any element $i_0 \in E$ minimizing the cost function. With a slight abuse of terminomogy, this element (and not the corresponding value of the cost function) is called a *global minimum*. When the set E is large, the combinatorial algorithms have a tendency to be trapped at a local minimum, as we shall see. The stochastic algorithm called simulated annealing (it is supposed to imitate the metallurgical process of the same name used to augment the strength of steel) claims to find a cure to this. Actual simulated annealing optimization methods will not be discussed in this section, which concentrates on the basic theory, and gives the opportunity to present an application of the ergodicity theory of non-homogeneous Markov chains.

Deterministic Descent Algorithms

The so-called descent algorithms define for each $i \in E$ a subset $N(i) \subset E \backslash \{i\}$, called the *neighborhood* of i, and proceed iteratively as follows. If at a given stage of the algorithm i is examined and not retained as a solution, at the next stage a candidate $j \in N(i)$ chosen according to a rule specific to each algorithm, and the values of the cost function are compared. If $U(i) \leq U(j)$, the procedure stops and i is the retained solution. Otherwise a new candidate $k \in N(j)$ is examined and

compared to j, and so on. The algorithm eventually comes to a stop and produces a solution, since E is finite. However, this solution is usually not optimal, due to the possible existence of local minima. In many situations, local optima exist and the algorithms become trapped at one of these local minima.

EXAMPLE 19.2.1: THE TRAVELING SALESMAN. A salesman must find the shortest route visiting exactly once each of the K cities of his business tour of the day. Here E is the set of the $K!$ admissible routes and $U(i)$ is the length of route i. One popular choice for the neighborhood $N(i)$ of route i is the collection of all the routes j obtained from i by a two-change, as in the figure.



The sizes of the neighborhoods are therefore reasonable in comparison to the size of the total search space. Note also that the computation of the new value of the cost function involves only four intercity distances.

The collection $\{N(i), i \in E\}$ is called a *neighborhood structure*. If for all pairs of state $i, j \in E$ there exists a *path* from i to j, that is, a sequence of states $i_1, \ldots, i_m \in E$ such that $i_1 \in N(i), i_2 \in N(i_1), \ldots, j \in N(i_m)$, the neighborhood structure is called communicating. This is the case for the 2-change neighborhood structure of the traveling salesman problem.

Stochastic Descent Algorithms

The basic idea of stochastic combinatorial optimization is to leave a possibility to escape from a local minimum trap. A canonical form of the stochastic descent algorithm is as follows. Let $Q = \{q_{ij}\}$ be an irreducible transition matrix on E. Also, for each parameter value T, and all states $i, j \in E$, let $\alpha_{ij}(T)$ be a probability. Calling X_n the current solution at stage n, the process $\{X_n\}_{n \geq 0}$ is a homogeneous Markov chain with state space E and transition matrix $\mathbf{P}(T)$ with general off-diagonal term

$$p_{ij}(T) = q_{ij}\alpha_{ij}(T). \tag{19.14}$$

We assume the chain irreducible. (For the Metropolis sampler, if Q is irreducible and U is not a constant, then $\mathbf{P}(T)$ is irreducible and aperiodic for all $T > 0$; see Exercise 19.3.3.) It is positive recurrent, since the state space is finite. Therefore, it has a unique stationary distribution $\pi(T)$.

One possible choice of the candidate-generating matrix Q consists in first choosing a communicating neighborhood structure and taking $q_{ij} > 0$ only if $i = j$ or $j \in N(i)$. The matrix Q is then irreducible. Conversely, one can associate with an irreducible transition matrix Q a communicating neighborhood structure defined by $N(i) = \{j;\, j \neq i,\, q_{ij} > 0\}$.

EXAMPLE 19.2.2: METROPOLIS, TAKE 2. Suppose that the current not retained candidate is i. At the next stage, a new candidate j is selected with probability q_{ij} and is retained with probability

$$\alpha_{ij}(T) = e^{-\frac{(U(j)-U(i))^+}{T}}, \tag{19.15}$$

where T is a positive constant; otherwise, j is rejected. The rule (19.15) gives a chance to the solution j when it is worse than i. This tolerance decreases as the deviation from i, measured by $U(j) - U(i)$, increases.

Suppose that the matrix Q is symmetric. With this special structure, the stationary distribution $\pi(T)$ does not depend on Q and is given by

$$\pi_i(T) = \frac{e^{-U(i)/T}}{\sum_{k\in E} e^{-U(k)/T}}.$$

Denote by H the set of global minima of the cost function. Then clearly, $\pi_i(T)$ is maximal on $i \in H$. But there is more: as $T \downarrow 0$, $\pi_i(T)$ tends to the uniform distribution on H. To see this, let m be the minimum value of the cost function, and write the right-hand side, after division of its numerator and denominator by $e^{-\frac{m}{T}}$, as

$$\frac{e^{-\frac{(U(i)-m)}{T}}}{|H| + \sum_{k\notin H} e^{-\frac{(U(k)-m)}{T}}}.$$

This observation suggests the following heuristic procedure. Start the algorithm with the value $T = a_0$ of the parameter, and wait a sufficiently long time for the chain to get close to its stationary regime. Then set $T = a_1 < a_0$ and again wait for the steady state. Then set $T = a_2 < a_1$, etc. At the kth change of the parameter T, the chain will be close to the stationary regime $\pi(a_k)$, and therefore if $\lim_{k\uparrow\infty} a_k = 0$, one expects that for large n, X_n will be with very high probability in H, the set of global minima.

However, for this to happen, the times in between the parameter changes must be sufficiently long for the chain to come close to the stationary distribution corresponding to the current value of the parameter. What is "sufficiently long"?

Simulated annealing algorithms all have a cooling schedule, that is, a sequence $\{T_n\}_{n\geq 0}$ of positive temperatures decreasing to 0 and controlling the transition rates of $\{X_n\}_{n\geq 0}$. At time n, $P(X_{n+1} = j \mid X_n = i) = p_{ij}(T_n)$. The question becomes: *How slowly* must the temperature converge to zero so that the distribution of X_n converges to the uniform distribution on H? Theoretical answers will be given in section 19.2.2.

19.2.2 Cooling Schedules

Slow Cooling

We begin with an example, and then proceed to the general theory.

EXAMPLE 19.2.3: ANNEALED GIBBS SAMPLER. (Geman and Geman, 1984) We use periodic scanning as in Example 19.1.6, except that at the n-th sweep, we introduce a temperature T_n. Therefore $\{Z_n\}_{n\geq 0}$ is a non-homogeneous MC, the transition matrix at time n being

$$\mathbf{P}(n) = \prod_{k=1}^{N} \mathbf{P}_{v(k)}^{T_n},$$

where the (x, y)-entry of $\mathbf{P}_v^T$ is

$$\frac{\exp\left\{-\frac{1}{T}U(y(v), x(V\backslash v))\right\}}{\sum_{\lambda\in\Lambda}\exp\left\{-\frac{1}{T}U(\lambda, x(V\backslash v))\right\}}$$

if $y = (y(v), x(V\backslash v))$. The bound of Example 19.1.6 gives

$$\delta(\mathbf{P}(n)) \leq 1 - e^{-\frac{N\Delta}{T_n}}.$$

In particular, by the block criterion of weak ergodicity,

$$\sum_{n=1}^{\infty} e^{-\frac{N\Delta}{T_n}} = \infty \tag{19.16}$$

is a sufficient condition of weak ergodicity. Now, $\mathbf{P}(n)$ has the stationary distribution

$$\pi_{T_n}(x) = \frac{e^{-\frac{1}{T_n}U(x)}}{Z_{T_n}}.$$

Also, for all $x \in \Lambda^V$, $\lim_{T\downarrow 0}\pi_T(x) = \frac{1}{|H|}$ if $x \in H$ and is 0 otherwise, where $H = \{x \in \Lambda^V; U(x) = \min\}$. Moreover, it can be shown that for $x \in H$, the quantity $\pi_T(x)$ increases as $T \downarrow 0$, whereas for $x \notin H$, it eventually decreases, and this guarantees that

$$\sum_{n=1}^{\infty} |\pi_{T(n+1)} - \pi_{T_n}| < \infty.$$

Therefore, by Theorem 15.2.12, if $T_n \downarrow 0$ in such a way that (19.16) is respected, then the non-homogeneous MC $\{Z_n\}_{n\geq 0}$ is strongly ergodic, with a limit distribution that is uniform on H.

The general results on non-homogeneous Markov chain of Section 15.2.2 will be applied to the simulated annealing algorithm corresponding to the transition matrix $\mathbf{P}(T)$ given by (19.14).

The transition matrix $\mathbf{P}(T)$ is assumed *uniformly irreducible* for sufficiently small $T \in (0,1]$. This means that for all ordered pair of states (i,j), there is a $\mathbf{P}(T)$-path from i to j which is independent of $T \in (0,c]$ for some $c > 0$. This is always satisfied in practice. For instance, for the Metropolis or Barker samplers, it suffices that Q be irreducible and that U be not a constant (see Exercise 19.3.3).

Define

$$d = \inf\{q_{ij}; \quad j \neq i, p_{ij}(T) > 0\},$$

a positive quantity, since the state space is finite.

The crucial assumption is the following: There exists $T^* \in (0,1]$ such that on $(0,T^*]$,

$$\alpha_{ij}(T) \downarrow 0 \text{ as } T \downarrow 0 \text{ if } U(j) > U(i), \tag{19.17}$$

$$\alpha_{ij}(T) \uparrow 1 \text{ as } T \downarrow 0 \text{ if } U(j) < U(i), \tag{19.18}$$

and

$$\lim_{T \downarrow 0} \alpha_{ij}(T) > 0 \text{ exists if } U(i) = U(j). \tag{19.19}$$

Assumptions (19.17) and (19.18) imply, in particular, that in the vicinity of 0, the functions $\alpha_{ij}(T)$ are monotonic if $U(i) \neq U(j)$. Define for each $T \in (0,1]$

$$\underline{\alpha}(T) = \inf_{i \in E, j \in N(i)} \alpha_{ij}(T). \tag{19.20}$$

Assumptions (19.17)–(19.19) imply that in the vicinity of 0,

$$\inf_{i \in E, j \neq i} \alpha_{ij}(T) = \inf_{\substack{i \in E, j \in N(i) \\ U(j) > U(i)}} \alpha_{ij}(T),$$

and therefore, in the vicinity of 0, $\underline{\alpha}(T)$ is *decreasing* to zero.

Theorem 19.2.4 *Let $\{\mathbf{P}(T)\}_{T \in (0,1]}$ satisfy the above assumptions. Let $\{T_n\}_{n \geq 0}$ be a sequence of numbers in $(0,1]$ decreasing to zero as $n \to \infty$. Then if*

$$\sum_{k=0}^{\infty} (\underline{\alpha}(T_{kN}))^N = \infty, \tag{19.21}$$

$\{\mathbf{P}(T_n)\}_{n \geq 0}$ is weakly ergodic.

Proof. Slightly change the notation, letting $\mathbf{P}(n) := \mathbf{P}(T_n)$. The uniform irreducibility assumption guarantees the existence, for all ordered pair of states (i,j), of a path $i_0 = i, i_1, ..., i_N = j$ such that

$$p_{i_j, i_{j+1}}(kN + j, kN + j + 1) = p_{i_j, i_{j+1}}(T_{kN+j}) > 0.$$

But $p_{kl}(T) > 0$ implies $p_{kl}(T) \geq d\underline{\alpha}(T)$, and therefore

$$p_{i_j, i_{j+1}}(kN + j, kN + j + 1) \geq d\underline{\alpha}(T_{kN+j}).$$

Since $\underline{\alpha}(T)$ is, in the vicinity of 0, monotone decreasing, then for sufficiently large k

$$p_{i_j,i_{j+1}}(kN+j, kN+j+1) \geq d\underline{\alpha}(T_{(k+1)N}),$$

and therefore

$$p_{ij}(kN,(k+1)N) \geq d^N \left(\underline{\alpha}(T_{(k+1)N})\right)^N.$$

Therefore, in view of (7.3) of Chapter 6,

$$1-\delta(\mathbf{P}(kN,(k+1)N)) \geq d^N \left(\underline{\alpha}(T_{(k+1)N})\right)^N.$$

Therefore, (19.21) implies

$$\sum_{k=1}^{\infty}(1-\delta(\mathbf{P}(kN,(k+1)N))) = \infty,$$

and the conclusion follows from the block criterion. □

EXAMPLE 19.2.5: METROPOLIS, TAKE 3. The acceptance probabilities of the Metropolis sampler are

$$\alpha_{ij}(T) = e^{(U(j)-U(i))^+/T}.$$

We see that conditions (19.17)–(19.19) are satisifed. We have

$$\underline{\alpha}(T) = \inf_{\substack{j\in N(i)\\ U(i)<U(j)}} e^{-\{U(j)-U(i)\}/T},$$

and therefore, with $\Delta := \sup\{U(j)-U(i); j\in N(i)\}$,

$$\underline{\alpha}(T) \geq e^{-\frac{\Delta}{T}}.$$

It follows that

$$\sum_{k=0}^{\infty}\{\underline{\alpha}(T_{kN})\}^N \geq \sum_{k=0}^{\infty} e^{-\frac{N\Delta}{T_{kN}}}.$$

For a cooling schedule $\{T_k\}_{k\geq 0}$ satisfying

$$T_k \geq \frac{N\Delta}{\log(k)}, \tag{19.22}$$

we see that

$$\sum_{k=1}^{\infty}\{\underline{\alpha}(T_{kN})\}^N \geq \sum_{k=1}^{\infty}\frac{1}{kN} = \infty,$$

and therefore $\{\mathbf{P}(T_n)\}_{n\geq 1}$ is weakly ergodic.

Therefore, in view of Theorem 15.2.11, $\{\mathbf{P}(T_n)\}$ is strongly ergodic. As shown in Example 19.2.2, the limiting probability vector puts all its mass uniformly on the set H of global minima. Therefore a cooling schedule satisfying (19.22) guarantees convergence in distribution to the set of global minima.

Fast Cooling

We shall now see the effects of fast cooling. Denote by $\mathbf{P}(\lim)$ the transition matrix corresponding to the limit case $T \downarrow 0$. In particular, $p_{ij}(\lim) = 0$ if $U(i) < U(j)$. Call R_1 the recurrent communication class of some global minimum, and R_2 the recurrent communication class of some strictly local minimum. Note that R_1 only contains global minima, and in particular, R_1 and R_2 are disjoint. Define

$$\bar{\alpha}(2, T) = \sup_{i \in R_2, j \in N(i)} \alpha_{ij}(T). \tag{19.23}$$

Since for $j \in R_2$,

$$\sum_{\ell \in R_2} p_{j,\ell}(T_k) = 1 - \sum_{\substack{\ell \notin R_2 \\ j \in N(i)}} q_{j\ell} \alpha_{j\ell}(T_k) \geq 1 - \bar{\alpha}(2, T_k),$$

the probability of staying in R_2 forever is bounded from below by $\prod_{k=1}^{\infty}(1 - \bar{\alpha}(2, T_k))$. This infinite product is strictly positive if $\sum_{k=1}^{\infty} \bar{\alpha}(2, T_k) < \infty$. Therefore, if the chain has at least one strictly local minimum, then under the condition

$$\sum_{k=1}^{\infty} \bar{\alpha}(2, T_k) < \infty, \tag{19.24}$$

the probability that it stays eternally in R_2 is strictly positive. In particular, since no globally optimal solution is in R_2, with positive probablity the algorithm will never visit a globally optimal state.

EXAMPLE 19.2.6: METROPOLIS, TAKE 4. Suppose that

$$\delta_2 := \inf\{U(j) - U(i); i \in R_2, j \notin R_2, j \in N(i)\} > 0\,.$$

Since $\bar{\alpha}(2, T) \leq e^{-\frac{\delta_2}{T}}$, we have

$$\sum_{k=0}^{\infty} \bar{\alpha}(2, T_k) \leq \sum_{k=1}^{\infty} e^{-\delta_2 / T_k}\,.$$

Therefore, if the cooling schedule satisfies

$$T_k \leq \frac{\delta_2 - \alpha}{\log k}$$

for some $\alpha > 0$ such that $\delta_2 - \alpha > 0$, we have

$$\sum_{k=1}^{\infty} \bar{\alpha}(2, T_k) \leq \sum_{k=1}^{\infty} e^{-(\log k)(1+\epsilon)},$$

where $1 + \epsilon = \frac{1}{1 - \frac{\alpha}{\delta_2}}$ (and therefore $\epsilon > 0$). Thus

$$\sum_{k=1}^{\infty} \bar{\alpha}(2, T_k) \leq \sum_{k=1}^{\infty} \frac{1}{k^{1+\epsilon}} < \infty,$$

which implies that the cooling schedule does not yield convergence in distribution to the uniform distribution on the set of global minima.

For the simulated annealing algorithm based on the Metropolis sampler, there exists a constant γ such that a necessary and sufficient of convergence, whatever the initial state, is

$$\sum_{k=1}^{\infty} e^{-\frac{\gamma}{T_k}} = \infty.$$

In particular, a logarithmic cooling schedule $T_k = \frac{a}{\log(k+1)}$ yields convergence if and only if $a \geq \gamma$ (Hajek, 1998).

The results of convergence given in the present section are of theoretical and qualitative interest only. Practical algorithms use faster than logarithmic schedules on a finite horizon. The theory and the performance evaluation of these algorithms is outside the scope of this book.

Books for Further Information

[van Laarhoven and Aarts, 1987]. [Liu, 2001]. See also the review article [Diaconis, 2009].

19.3 Exercises

Exercise 19.3.1. EIGENSTRUCTURE OF METROPOLIS
(Liu, 1995) Let π and p be two strictly positive probability distributions on $E = \{1, 2, \ldots, r\}$, and let $w(i) := \frac{\pi(i)}{p(i)}$. The Metropolis algorithm corresponding to the candidate-generating matrix Q given by $q_{ij} = p_j$ for all $i, j \in E$ has the transition matrix $\mathbf{P}$ given by

$$p_{ij} = p(j) \min\left(1, \frac{w(j)}{w(i)}\right),$$

for $i \neq j$. Assume that the states of E are ordered in such a way that

$$w(1) \geq w(2) \geq \cdots \geq w(r).$$

Verify that the eigenvalues λ_k and the corresponding right-eigenvectors v_k, $1 \leq k \leq r$, of $\mathbf{P}$ are $\lambda_1 = 1$, $v_1 = \mathbf{1}$, and for $k \geq 1$,

$$\begin{aligned} \lambda_{k+1} &= \sum_{j \geq k} \pi(j) \left(\frac{1}{w(j)} - \frac{1}{w(k)}\right), \\ v_{k+1} &= \left(0, \ldots, 0, \sum_{\ell=k+1}^{r} \pi(\ell), -\pi(k), \ldots, -\pi(k)\right)^T, \end{aligned}$$

where the first $k-1$ entries of v_{k+1} are null. What is the potential value of this knowledge in a sampling context?

Exercise 19.3.2. Rate of convergence of a Metropolis algorithm
Define the probability distribution π on $E = \{1, \ldots, r\}$ by

$$\pi(j) = \frac{\theta^{j-1}(1-\theta)}{1-\theta^r}$$

where $\theta \in (0, 1)$. Apply the Metropolis algorithm with candidates randomly generated, that is, $p(j) = \frac{1}{r}$. Give a bound for $d_V(\delta_r \mathbf{P}^n, \pi)^2$. (Of course, this is a pure classroom exercise.)

Exercise 19.3.3. Irreducibility of the Barker sampling chain
Show that for both the Metropolis and Barker samplers, if Q is irreducible and U is not a constant, then $\mathbf{P}(T)$ is irreducible and aperiodic for all $T > 0$.

Exercise 19.3.4. The modified random walk
Consider the usual random walk on a graph. Its stationary distribution is in general non-uniform. We wish to modify it so as to obtain an HMC with uniform stationary distribution. Now accept a transition from vertex i to vertex j of the original random walk with probability α_{ij}. Find one such acceptance probability depending only on $d(i)$ and $d(j)$ that guarantees that the corresponding Monte Carlo Markov chain admits the uniform distribution as stationary distribution.

Chapter 20

Convergence Rates

20.1 Reversible Transition Matrices

20.1.1 A Characterization of Reversibility

For an ergodic Markov chain, one may take the value at a "large" time n, as a sample of the stationary distribution. The accuracy of the sample is measured in terms of the distance in variation between the sample and the target distribution. The following sections are devoted to the obtention of convergence speeds of an ergodic HMC to its stationary distribution, and in particular, of bounds of the second largest eigenvalue modulus of its transition matrix. This is done primarily for reversible HMC's, since most Monte Carlo Markov chains (see Chapter 19) are of this type.

The main result of Perron and Frobenius is that convergence to steady state of an ergodic finite state space HMC is geometric, with relative speed equal to the second-largest eigenvalue modulus (SLEM). Even if there are a few interesting models, especially in biology, where the eigenstructure of the transition matrix can be extracted, this situation remains nevertheless exceptional. The added structure of reversible transition matrices allows to push the analysis further and avoids recourse to the Perron–Fröbenius theorem. For convenience, we recall the definition of reversibility, and introduce a slight change in the terminology.

Definition 20.1.1 *Let* $\mathbf{P}$ *be a transition matrix and* π *a strictly positive probability vector on* E*. The pair* $(\mathbf{P}, \pi)$ *is called reversible if the detailed balance equations (6.8) are satisfied.*

It will also be assumed that the state space is finite, say $E := \{1, 2, \ldots, r\}$, and that $\mathbf{P}$ is irreducible. This implies in particular that π is the unique stationary distribution, that $\pi > 0$, and that $\mathbf{P}$ is positive recurrent. For short, we shall sometimes say: "$\mathbf{P}$ is reversible".

Let $\ell^2(\pi)$ be the real vector space $\mathbb{R}^r$ endowed with the scalar product

$$\langle x, y \rangle_\pi := \sum_{i \in E} x(i) y(i) \pi(i)$$

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_20

and the corresponding norm $\|x\|_\pi := \left(\sum_{i\in E} x(i)^2\pi(i)\right)^{\frac{1}{2}}$. We shall write

$$\langle x\rangle_\pi := \sum_i \pi(i)x(i) = \langle x, 1\rangle_\pi$$

for the *mean* of x with respect to π. The variance of x with respect to π is

$$\mathrm{Var}_\pi(x) := \sum_i \pi(i)x(i)^2 - \left(\sum_i \pi(i)x(i)\right)^2 = \|x\|_\pi^2 - \langle x\rangle_\pi^2 \,.$$

Similarly to $\ell^2(\pi)$, $\ell^2(\frac{1}{\pi})$ is defined as the real vector space $\mathbb{R}^r$ endowed with the scalar product

$$\langle x, y\rangle_{\frac{1}{\pi}} := \sum_{i\in E} x(i)y(i)\frac{1}{\pi(i)} \,.$$

Theorem 20.1.2 *The pair $(\mathbf{P}, \pi)$ is reversible if and only if $\mathbf{P}$ is self-adjoint in $\ell^2(\pi)$, that is,*

$$\langle \mathbf{P}x, y\rangle_\pi = \langle x, \mathbf{P}y\rangle_\pi \tag{20.1}$$

for all $x, y \in \ell^2(\pi)$.

Proof. Suppose $(\mathbf{P}, \pi)$ is reversible. Then

$$\begin{aligned}\langle \mathbf{P}x, y\rangle_\pi &= \sum_{i\in E}\left\{\left(\sum_{j\in E} p_{ij}x(j)\right) y(i)\pi(i)\right\} \\ &= \sum_{i,j\in E} \pi(i)p_{ij}\, x(j)y(i) = \sum_{i,j\in E} \pi(j)p_{ji}\, y(i)x(j) \\ &= \sum_{j\in E}\left\{x(j)\left(\sum_{i\in E} p_{ji}y(i)\right)\pi(j)\right\} = \langle x, \mathbf{P}y\rangle_\pi \,.\end{aligned}$$

Conversely, suppose $\mathbf{P}$ self-adjoint in $\ell^2(\pi)$. Let δ_k be the k-th vector of the canonical basis of $\mathbb{R}^r$ (the only non-null entry, 1, is in the k-th position). Then the detailed balance equation (6.8) follows from (20.1) with the choice $x = \delta_i$, $y = \delta_j$. □

Reversibility of $(\mathbf{P}, \pi)$ is equivalent to the fact that

$$\mathbf{P}^* := D^{\frac{1}{2}}\mathbf{P}D^{-\frac{1}{2}}$$

is a symmetric matrix, where

$$D = D(\pi) := \mathrm{diag}\{\pi(1), \ldots, \pi(r)\} \,. \tag{20.2}$$

More explicitly,

$$p^*_{ij} = p_{ij}\frac{\sqrt{\pi(i)}}{\sqrt{\pi(j)}} \,.$$

Note that

$$x^T D y = \langle x, y \rangle_\pi \ . \tag{20.3}$$

Since $\mathbf{P}^*$ is symmetric, its eigenvalues are real, it is diagonalizable, and the sets of right- and left-eigenvectors are the same.

Choose an orthonormal basis of $\mathbb{R}^r$ formed of right-eigenvectors $w_1, \ldots, w_r$ associated, respectively, with the eigenvalues $\lambda_1, \ldots, \lambda_r$. Define u and v by

$$w = D^{-\frac{1}{2}} u, \quad w = D^{\frac{1}{2}} v,$$

where w is a right- (and therefore left-) eigenvector of $\mathbf{P}^*$, corresponding to the eigenvalue λ. In particular,

$$u = Dv \, . \tag{20.4}$$

The matrices $\mathbf{P}$ and $\mathbf{P}^*$ have the same eigenvalues, and moreover, v (resp., u) is a right-eigenvector (resp., left-eigenvector) of $\mathbf{P}$ corresponding to the eigenvalue λ.

Orthonormality (with respect to the usual Euclidean norm) of the collection $\{w_1, \ldots, w_r\}$ is equivalent to orthonormality in $\ell^2(\pi)$ of $\{v_1, \ldots, v_r\}$, that is,

$$\langle v_i, v_j \rangle_\pi = \delta_{ij}.$$

Similarly, $\{u_1, \ldots, u_r\}$ is an orthonormal collection in $\ell^2(\frac{1}{\pi})$:

$$\langle u_i, u_j \rangle_{\frac{1}{\pi}} = \delta_{ij}.$$

The eigenvectors u_1 and v_1 may always be chosen as follows

$$u_1 = \pi \, , \qquad v_1 = \mathbf{1} \, .$$

Since $\{v_1, \ldots, v_r\}$ is also a basis of $\mathbb{R}^r$, any vector $x \in \mathbb{R}^r$ can be expressed as $x = \sum_{i \in E} \alpha_i v_i$. In particular, $\langle x, v_j \rangle_\pi = \alpha_j$, and therefore

$$x = \sum_{j=1}^{r} \langle x, v_j \rangle_\pi \, v_j . \tag{20.5}$$

Similarly,

$$x^T = \sum_{j=1}^{r} \langle x, u_j \rangle_{\frac{1}{\pi}} \, u_j^T . \tag{20.6}$$

The variance of x with respect to π is

$$\mathrm{Var}_\pi(x) = \sum_{j=2}^{r} | \langle x, v_j \rangle_\pi |^2 \, . \tag{20.7}$$

For all n, $\mathbf{P}^n v_j = \lambda_j^n v_j$, and therefore

$$\mathbf{P}^n x = \sum_{j=1}^{r} \lambda_j^n \langle x, v_j \rangle_\pi \, v_j . \tag{20.8}$$

Similarly,

$$x^T \mathbf{P}^n = \sum_{j=1}^{r} \lambda_j^n \langle x, u_j \rangle_{\frac{1}{\pi}} u_j^T. \tag{20.9}$$

From (20.8), (20.3), and (20.4), we obtain $\mathbf{P}^n x = \sum_{j=1}^{r} \lambda_j^n v_j u_j^T x$, and we therefore retrieve the representation (6.11) for $A = \mathbf{P}$. From (20.8),

$$\mathbf{P}^n x - \langle x \rangle_\pi \mathbf{1} = \sum_{j=2}^{r} \lambda_j^n \langle x, v_j \rangle_\pi v_j. \tag{20.10}$$

20.1.2 Convergence Rates in Terms of the SLEM

Theorem 20.1.3 *Defining* $\pi_{min} := \min_{k \in E} \pi(k)$,

$$\max_{i \in E} d_V(p_{i\cdot}(n), \pi) \leq \frac{\rho^n}{2\pi_{min}}.$$

Proof. From (20.10), for all $i \in E$,

$$p_{ik}(n) - \pi(k) = \sum_{j=2}^{r} \lambda_j^n v_j(i) v_j(k) \pi(k). \tag{20.11}$$

Therefore,

$$\begin{aligned} d_V(p_{i\cdot}(n), \pi) &\leq \frac{1}{2} \sum_{k=1}^{r} \left| \sum_{j=2}^{r} \lambda_j^n v_j(i) v_j(k) \pi(k) \right| \\ &\leq \frac{1}{2} \sum_{k=1}^{r} \max_{\ell \in E} \left(\sum_{j=2}^{r} \lambda_j^n |v_j(i)||v_j(\ell)| \right) \pi(k) \\ &= \frac{1}{2} \max_{\ell \in E} \left(\sum_{j=2}^{r} \lambda_j^n |v_j(i)||v_j(\ell)| \right). \end{aligned}$$

Therefore, denoting by ρ the SLEM of $\mathbf{P}$

$$d_V(p_{i\cdot}(n), \pi) \leq \frac{1}{2} \max_{\ell \in E} \left(\sum_{j=2}^{r} |v_j(i)||v_j(\ell)| \right) \rho^n.$$

By Schwarz's inequality,

$$\begin{aligned} \sum_{j=2}^{r} |v_j(i)||v_j(\ell)| &\leq \left(\sum_{j=2}^{r} v_j(i)^2 \right)^{\frac{1}{2}} \left(\sum_{j=2}^{r} v_j(\ell)^2 \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{j=1}^{r} v_j(i)^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{r} v_j(\ell)^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Now, from (20.5) with $x = \delta_i$, we have that $\delta_i = \sum_{j=1}^r v_j(i)\pi(i)v_j$. Writing this equality for the i-th coordinate gives $1 = \sum_{j=1}^r v_j(i)^2\pi(i)$, and therefore

$$\left(\sum_{j=1}^r v_j(i)^2\right)^{\frac{1}{2}} \left(\sum_{j=1}^r v_j(\ell)^2\right)^{\frac{1}{2}} \leq (\pi(i)\pi(\ell))^{-\frac{1}{2}} .$$

□

The constant before ρ^n is often too large. Maybe if we start from a specific state i with high probability, the bound can be improved. This is done in the next theorem.

Theorem 20.1.4 *Let* $\mathbf{P}$ *be a reversible irreducible transition matrix on the finite state space* $E = \{1, \ldots, r\}$*, with the stationary distribution* π*. Then for all* $n \geq 1$ *and all* $i \in E$*,*

$$d_V(\delta_i^T\mathbf{P}^n, \pi)^2 \leq \frac{p_{ii}(2)}{2\pi(i)}\rho^{2n-2}, \tag{20.12}$$

where ρ *is the* SLEM *of* $\mathbf{P}$*.*

Proof. From (20.10) and (20.7), we have that

$$\|\mathbf{P}^n x - \langle x\rangle_\pi \mathbf{1}\|_\pi^2 = \sum_{j=2}^r |\lambda_j|^{2n}|\langle x, v_j\rangle_\pi|^2 \leq \rho^{2n}\mathrm{Var}_\pi(x). \tag{20.13}$$

Now, by reversibility and Schwarz's inequality in $\ell^2(\pi)$,

$$\begin{aligned}
\left|\sum_{j\in E} p_{ij}x(j)\right|^2 &= \left|\sum_{j\in E} p_{ji}\frac{\pi(j)}{\pi(i)}x(j)\right|^2 \leq \left(\sum_{j\in E}\frac{p_{ji}}{\pi(i)}|x(j)|\pi(j)\right)^2 \\
&\leq \left(\sum_{j\in E} x(j)^2\pi(j)\right)\left(\sum_{j\in E}\left(\frac{p_{ji}}{\pi(i)}\right)^2\pi(j)\right) \\
&= \left(\sum_{j\in E} x(j)^2\pi(j)\right)\left(\sum_{j\in E}(p_{ji}p_{ij})\frac{1}{\pi(i)}\right) = \left(\sum_{j\in E} x(j)^2\pi(j)\right)\frac{p_{ii}(2)}{\pi(i)},
\end{aligned}$$

that is,

$$\left|\sum_{j\in E} p_{ij}x(j)\right|^2 \leq \frac{p_{ii}(2)}{\pi(i)}\|x\|_\pi^2.$$

With $x = \mathbf{P}^{n-1}y - \langle y\rangle_\pi \mathbf{1}$, this gives, in view of (20.13),

$$\begin{aligned}
\left|\sum_{j=1}^r p_{ij}(n)y(j) - \sum_{i=1}^r \pi(j)y(j)\right|^2 &\leq \frac{p_{ii}(2)}{\pi(i)}\|\mathbf{P}^{n-1}y - \langle y\rangle_\pi \mathbf{1}\|_\pi^2 \\
&\leq \frac{p_{ii}(2)}{\pi(i)}\mathrm{Var}_\pi(y)\rho^{2n-2}.
\end{aligned}$$

The result then follows from the following alternative expression of the distance in variation (Exercise 16.4.3):

$$d_V(\alpha,\beta) = \frac{1}{2}\sup\left(\sum_{i=1}^{r}\alpha(i)y(i) - \sum_{i=1}^{r}\beta(i)y(i);\ \sup|y(i)| = 1\right),$$

and the observation that if y is such that $\sup|y(i)| \le 1$, then $\mathrm{Var}_\pi(y) \le 1$. □

For the next estimate, we shall need to define the χ^2-contrast of α with respect to β,

$$\chi^2(\alpha;\beta) := \sum_{i\in E}\frac{(\alpha(i)-\beta(i))^2}{\beta(i)}.$$

Note that

$$\chi^2(\alpha;\pi) = \|\alpha-\pi\|^2_{\frac{1}{\pi}}. \tag{20.14}$$

Also

$$4d_V(\alpha,\beta)^2 \le \chi^2(\alpha;\beta), \tag{20.15}$$

as follows from Schwarz's inequality:

$$\begin{aligned}\left(\sum_{i\in E}|\alpha(i)-\beta(i)|\right)^2 &= \left(\sum_{i\in E}\left|\frac{\alpha(i)}{\beta(i)}-1\right|\beta(i)^{\frac{1}{2}}\beta(i)^{\frac{1}{2}}\right)^2\\ &\le \sum_{i\in E}\left(\frac{\alpha(i)}{\beta(i)}-1\right)^2\beta(i) = \sum_{i\in E}\frac{1}{\beta(i)}(\alpha(i)-\beta(i))^2.\end{aligned}$$

Theorem 20.1.5 *Let $\mathbf{P}$ be a reversible irreducible transition matrix on the finite state space $E = \{1,\dots,r\}$, with the stationary distribution π. Then for any probability distribution μ on E, and for all $n \ge 1$,*

$$\|\mu^T\mathbf{P}^n - \pi^T\|_{\frac{1}{\pi}} \le \rho^n\|\mu-\pi\|_{\frac{1}{\pi}}. \tag{20.16}$$

Also, for $n \ge 1$, all $i \in E$, and all $A \subset E$,

$$|\delta_i^T\mathbf{P}^n(A) - \pi^T(A)| \le \left(\frac{1-\pi(i)}{\pi(i)}\right)^{\frac{1}{2}}\min\left(\pi(A)^{\frac{1}{2}},\frac{1}{2}\right)\rho^n, \tag{20.17}$$

where ρ is the SLEM *of $\mathbf{P}$. In particular,*

$$4d_V(\delta_i^T\mathbf{P}^n,\pi)^2 \le \frac{1-\pi(i)}{\pi(i)}\rho^{2n}. \tag{20.18}$$

Proof. Recall that $u_1 = \pi$, and therefore $\langle\mu-\pi,u_1\rangle_{\frac{1}{\pi}} = \sum_{i\in E}(\mu(i)-\pi(i)) = 0$. Therefore, by (20.9), and denoting by α_j the quantity $\langle\mu-\pi,u_j\rangle_{\frac{1}{\pi}}$, we obtain

$$\begin{aligned}\|(\mu-\pi)^T\mathbf{P}^n\|^2_{\frac{1}{\pi}} &= \sum_{j=2}^{r}\alpha_j^2\lambda_j^{2n}\|u_j\|^2_{\frac{1}{\pi}} = \sum_{j=2}^{r}\alpha_j^2\lambda_j^{2n}\\ &\le \rho^{2n}\sum_{j=2}^{r}\alpha_j^2 = \rho^{2n}\|\mu-\pi\|^2_{\frac{1}{\pi}},\end{aligned}$$

and (20.16) follows, since $\pi^T\mathbf{P}^n = \pi^T$.

Define $\mu_n^T := \delta_i^T\mathbf{P}^n$. By Schwarz's inequality,

$$\begin{aligned}|\mu_n(A)-\pi(A)|^2 &= \left|\sum_{\ell\in A}\left(\frac{\mu_n(\ell)}{\pi(\ell)}-1\right)\pi(\ell)\right|^2\\ &\le \left(\sum_{\ell\in A}\left(\frac{\mu_n(\ell)}{\pi(\ell)}-1\right)^2\pi(\ell)\right)\pi(A) \le \left(\sum_{\ell\in E}\left(\frac{\mu_n(\ell)}{\pi(\ell)}-1\right)^2\pi(\ell)\right)\pi(A)\\ &= \|\delta_i^T\mathbf{P}^n-\pi^T\|^2_{\frac{1}{\pi}}\pi(A) \le \rho^{2n}\|\delta_i-\pi\|^2_{\frac{1}{\pi}}\pi(A),\end{aligned}$$

where the last inequality uses (20.16). But, as simple calculations reveal,

$$\|\delta_i-\pi\|^2_{\frac{1}{\pi}} = \frac{1-\pi(i)}{\pi(i)}, \tag{20.19}$$

and therefore

$$|\delta_i^T\mathbf{P}^n(A)-\pi^T(A)| \le \left(\frac{1-\pi(i)}{\pi(i)}\right)^{\frac{1}{2}}\pi(A)^{\frac{1}{2}}\rho^n. \tag{20.20}$$

Now,

$$|\mu_n(A)-\pi(A)|^2 \le d_V(\mu_n,\pi)^2 \le \frac{1}{4}\chi^2(\mu_n;\pi).$$

But, by (20.16), (20.14), and (20.19)

$$\chi^2(\mu_n;\pi) = \|\delta_i^T\mathbf{P}^n-\pi^T\|^2_{\frac{1}{\pi}} \le \rho^{2n}\|\delta_i-\pi\|^2_{\frac{1}{\pi}} = \rho^{2n}\frac{1-\pi(i)}{\pi(i)}.$$

Therefore,

$$|\delta_i^T\mathbf{P}^n(A)-\pi^T(A)| \le \left(\frac{1-\pi(i)}{\pi(i)}\right)^{\frac{1}{2}}\frac{1}{2}\rho^n. \tag{20.21}$$

Combining (20.20) and (20.21) gives (20.17). Inequality (20.18) then follows since $d_V(\alpha,\beta) = \sup_{A\subset E}|\alpha(A)-\beta(A)|$. □

20.1.3 Rayleigh's Spectral Theorem

The results of the previous section are useful once a bound for the SLEM is available. Coming back to the eigenvalues of $\mathbf{P}$, we know that $\lambda_1 = 1$ is one of them, with multiplicity 1. This eigenvalue corresponds to the unique right-eigenvector v_1 such that $\|v_1\|_\pi = 1$, namely $v_1 = \mathbf{1}$. Moreover, the eigenvalues of $\mathbf{P}$ are all in the closed unit disk of $\mathbb{C}$, and in the reversible case of interest in this section, they are real. Therefore, with proper ordering,

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_r \geq -1. \tag{20.22}$$

Note that this order is different from the one adopted in (6.12) for the statement of the Perron–Frobenius theorem. In (20.22), λ_2 is the second-largest eigenvalue (SLE), whereas in (6.12) it was the eigenvalue with the second-largest *modulus*. The strict inequality $\lambda_1 > \lambda_2$ expresses the fact that λ_1 is the unique eigenvalue equal to 1. We also know from the Perron–Fröbenius theorem that the only eigenvalue of modulus 1 and not equal to 1, in this case -1, occurs if and only if the chain is periodic of period $d = 2$. In particular, in the reversible case, the period cannot exceed 2.

It will be convenient to consider the matrix $I - \mathbf{P}$, called the Laplacian of the HMC. Its eigenvalues are $\beta_i = 1 - \lambda_i$ $(1 \leq i \leq r)$ and therefore

$$0 = \beta_1 < \beta_2 \leq \cdots \leq \beta_r \leq 2\,.$$

Clearly, a right-eigenvector of $I - \mathbf{P}$ corresponding to $\beta_i = 1 - \lambda_i$ is v_i, a right-eigenvector of $\mathbf{P}$ corresponding to λ_i.

The Dirichlet form $\mathcal{E}_\pi(x, x)$ associated with a reversible pair $(\mathbf{P}, \pi)$ is defined by

$$\mathcal{E}_\pi(x, x) := \langle (I - \mathbf{P})x, x \rangle_\pi\,.$$

We shall keep in mind that $\mathcal{E}_\pi(x, x)$ also depends on $\mathbf{P}$.

We have

$$\mathcal{E}_\pi(x, x) = \frac{1}{2} \sum_{i,j \in E} \pi(i) p_{ij} (x(j) - x(i))^2 \tag{20.23}$$

$$= \sum_{i<j} \pi(i) p_{ij} (x(j) - x(i))^2\,. \tag{20.24}$$

Proof.

$$\begin{aligned} \langle (I - \mathbf{P})x, x \rangle_\pi &= \sum_{i,j \in E} \pi(i) p_{ij} x(i)(x(i) - x(j)) \\ &= \sum_{i,j \in E} \pi(j) p_{ji} x(j)(x(j) - x(i)) \\ &= \sum_{i,j \in E} \pi(i) p_{ij} x(j)(x(j) - x(i))\,, \end{aligned}$$

where the second equality is obtained by a change of indexation, and the third uses the reversibility of $(\mathbf{P}, \pi)$. Expressing $\mathcal{E}_\pi(x, x)$ as the half-sum of the second and last terms in the above chain of equalities yields (20.23). Equality (20.25) then follows from the detailed balance equations $\pi(i)p_{ij} = \pi(j)p_{ji}$. □

An analogous (and simpler) computation gives

$$\mathrm{Var}_\pi(x) = \frac{1}{2} \sum_{i,j \in E} \pi(i)\pi(j)(x(j) - x(i))^2\,. \tag{20.25}$$

The next result gives a characterization of the second-largest eigenvalue (SLE) λ_2, or equivalently of $\beta_2 = 1 - \lambda_2$.

Theorem 20.1.6 *Let* $\mathbf{P}$ *be an irreducible transition matrix on the finite state space* $E = \{1, 2, \ldots, r\}$*, and let* π *be its stationary distribution. If* $(\mathbf{P}, \pi)$ *is reversible,*

$$\beta_2 = \inf \left\{ \frac{\mathcal{E}_\pi(x,x)}{Var_\pi(x)} ;\ Var_\pi(x) \neq 0 \quad \right\} .$$

Remark 20.1.7 Condition $\mathrm{Var}_\pi(x) \neq 0$ just says that x is not, as a function, a constant. Said otherwise, it is not of the form $x = c\mathbf{1}$ for some $c \in \mathbb{R}$.

Proof. First observe from (20.23) that the ratio $\frac{\mathcal{E}_\pi(x,x)}{\mathrm{Var}_\pi(x)}$ is invariant by translation since

$$\mathcal{E}_\pi(x,x) = \mathcal{E}_\pi(x - c\mathbf{1}, x - c\mathbf{1}) \text{ and } \mathrm{Var}_\pi(x - c\mathbf{1}) = \mathrm{Var}_\pi(x) \tag{20.26}$$

for any real number c, and invariant by scaling (when replacing x by cx where c is a non-null real number). Therefore, we may restrict attention to the case where the variance is 1 and the mean is null. From (20.8), $(I - \mathbf{P})x = \sum_{j=1}^r \beta_j \langle x, v_j \rangle_\pi v_j$, and therefore

$$\mathcal{E}_\pi(x,x) = \sum_{j=1}^r \beta_j |\langle x, v_j \rangle_\pi|^2 .$$

Also from (20.7),

$$\langle x, x \rangle_\pi = \sum_{j=1}^r |\langle x, v_j \rangle_\pi|^2 = 1$$

and

$$\langle x, v_1 \rangle_\pi = \langle x, \mathbf{1} \rangle_\pi = \langle x \rangle_\pi = 0 .$$

Therefore

$$\begin{aligned} \mathcal{E}_\pi(x,x) &= \sum_{i=2}^r \sum_{j=2}^r \beta_j |\langle x, v_j \rangle_\pi|^2 \\ &\leq \beta_2 \sum_{j=2}^r |\langle x, v_j \rangle_\pi|^2 = \beta_2 . \end{aligned}$$

The inequality becomes an equality when $x = v_2$ since $\mathcal{E}_\pi(v_2, v_2) = \beta_2$. □

Remark 20.1.8 The second largest eigenvalue (SLE) is not in general the second largest eigenvalue modulus (SLEM). Both an upper bound of λ_2 (the SLE and a lower bound of the smallest eigenvalue λ_r are needed in order to obtain a bound for the SLEM. Note that the lazy Markov chain (see Example 6.2.5) associated with a reversible Markov chain has all its eigenvalues, equal to $1 + \lambda_i$ $(1 \leq i \leq r)$ non-negative (see (20.22)), and the SLEM is then equal to the SLE.

20.2 Bounds for the SLEM

20.2.1 Bounds via Rayleigh's Characterization

Next theorem gives a method based on Rayleigh's theorem to obtain an upper bound of λ_2 and a lower bound of λ_r.

Theorem 20.2.1 *(a) If $A > 0$ is such that for all $x \in \mathbb{R}^r$,*

$$Var_\pi(x) \leq A\mathcal{E}_\pi(x,x)\,, \tag{20.27}$$

then, denoting by λ_2 the SLE *of $\mathbf{P}$, $\lambda_2 \leq 1 - \frac{1}{A}$.*

(b) If there exists $B > 0$ such that for all $x \in \mathbb{R}^r$,

$$\langle \mathbf{P}x, x\rangle_\pi + \|x\|_\pi^2 \geq B\|x\|_\pi^2, \tag{20.28}$$

then $\lambda_j \geq -1 + B$ $(1 \leq j \leq r)$.

Proof. (a) It follows from Theorem 20.1.6 that $\beta_2 \geq 1/A$.

(b) Taking $x = v_j$ in (20.28) and using the fact that $\mathbf{P}v_j = \lambda_j v_j$ gives $\lambda_j + 1 \geq B$. □

The following is a useful consequence of Rayleigh's characterization of the second largest eigenvalue.

Theorem 20.2.2 *Consider two reversible* HMC*'s on the same finite state space $E = \{1, 2, \ldots, r\}$ and let $(\mathbf{P}, \pi)$, and $(\widetilde{\mathbf{P}}, \widetilde{\pi})$, be their respective transition matrices and stationary distributions. Suppose that there exists two positive constants A and B such that for all $i \in E$, all $x \in E^r$,*

$$\pi(i) \leq A\widetilde{\pi}(i) \text{ and } \mathcal{E}_{\widetilde{\mathbf{P}}}(x,x) \leq B\mathcal{E}_{\mathbf{P}}(x,x)\,.$$

Then, $\widetilde{\beta}_2 \leq AB\beta_2$.

Proof. The quantity $\|x - c\mathbf{1}\|^2$ is minimized for $c = \langle x\rangle_\pi$ and is then equal to $\mathrm{Var}_\pi(x)$. In particular, for $c = \langle x\rangle_{\widetilde{\pi}}$,

$$\begin{aligned}
\mathrm{Var}_\pi(x) &\leq \|x - \langle x\rangle_{\widetilde{\pi}}\|^2 \\
&= \sum_i \pi(i)(x(i) - \langle x\rangle_{\widetilde{\pi}})^2 \\
&\leq A\sum_i \widetilde{\pi}(i)(x(i) - \langle x\rangle_{\widetilde{\pi}})^2 \\
&= A\mathrm{Var}_{\widetilde{\pi}}(x)\,.
\end{aligned}$$

Therefore

$$\frac{1}{\mathrm{Var}_{\widetilde{\pi}}(x)} \le A \frac{1}{\mathrm{Var}_{\pi}(x)} \text{ and } \mathcal{E}_{\widetilde{\pi}}(x,x) \le B \mathcal{E}_{\pi}(x,x)\,,$$

so that

$$\frac{\mathcal{E}_{\widetilde{\pi}}(x,x)}{\mathrm{Var}_{\widetilde{\pi}}(x)} \le AB \frac{\mathcal{E}_{\pi}(x,x)}{\mathrm{Var}_{\pi}(x)}\,.$$

Minimizing over the non-null x's yields the announced inequality. □

EXAMPLE 20.2.3: BARKER AND METROPOLIS ALGORITHMS. Theorem 20.2.2 will be applied to the comparison of the Barker and Metropolis reversible Markov chains with the same stationary distribution

$$\pi(i) = \widetilde{\pi}(i) = \frac{e^{-U(i)}}{Z}\,,$$

where $U : E \to \mathbb{R}$. For Barker's algorithm, with $z := \frac{e^{-U(j)}}{e^{-U(i)}}$,

$$p_{ij} = \frac{1}{1+z}\,,$$

whereas for the Metropolis algorithm,

$$\widetilde{p}_{ij} = 1 \wedge z\,.$$

It follows that

$$\frac{1}{2} \le \frac{\mathcal{E}_{\mathbf{P}}(x,x)}{\mathcal{E}_{\widetilde{\mathbf{P}}}(x,x)} \le 1$$

and therefore $\beta_2 \le \widetilde{\beta}_2 \le 2\beta_2$.

Weighted Paths

The next two results give an upper bound and a lower bound in terms of the geometry of the transition graph.

In the transition graph associated with $\mathbf{P}$, we shall denote a directed edge $i \to j$ by e, and call $e^- = i$ and $e^+ = j$ its initial vertex and end vertex respectively. Define for any such directed edge e,

$$Q(e) = \pi(i) p_{ij}\,. \tag{20.29}$$

For each ordered pair of *distinct* states (i,j), select arbitrarily one and only one path from i to j (that is, a sequence $i, i_1, \ldots, i_m, j$ such that $p_{ii_1} p_{i_1 i_2} \cdots p_{i_m j} > 0$) which does not use the same edge twice. Let Γ be the collection of paths so selected. For a path $\gamma_{ij} \in \Gamma$, let

$$|\gamma_{ij}|_Q := \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)} = \frac{1}{\pi(i) p_{ii_1}} + \frac{1}{\pi(i_1) p_{i_1 i_2}} + \cdots + \frac{1}{\pi(i_m) p_{i_m j}}\,.$$

Define the *Poincaré coefficient*

$$\kappa = \kappa(\Gamma) := \max_e \sum_{\gamma_{ij} \ni e} |\gamma_{ij}|_Q \pi(i) \pi(j)\,.$$

Theorem 20.2.4 *(Diaconis and Strook, 1991) Let* $\mathbf{P}$ *be an irreducible transition matrix on the finite state space* E*, with stationary distribution* π*, and assume* $(\mathbf{P}, \pi)$ *to be reversible. Denoting by* λ_2 *its* SLE*,*

$$\lambda_2 \leq 1 - \frac{1}{\kappa}\,.$$

Proof. If suffices to show that (20.27) holds for $A = \kappa$. For this, write

$$\begin{aligned}\mathrm{Var}_\pi(x) &= \frac{1}{2}\sum_{i,j\in E}(x(i)-x(j))^2\pi(i)\pi(j)\\ &= \frac{1}{2}\sum_{i,j\in E}\left\{\sum_{e\in\gamma_{ij}}\frac{1}{Q(e)^{\frac{1}{2}}}Q(e)^{\frac{1}{2}}(x(e^-)-x(e^+))\right\}^2\pi(i)\pi(j).\end{aligned}$$

By Schwarz's inequality, this quantity is bounded above by

$$\begin{aligned}&\frac{1}{2}\sum_{i,j\in E}\left\{|\gamma_{ij}|_Q\sum_{e\in\gamma_{ij}}Q(e)(x(e^-)-x(e^+))^2\right\}\pi(i)\pi(j)\\ =\ &\frac{1}{2}\sum_e\left\{Q(e)(x(e^-)-x(e^+))^2\left[\sum_{\gamma_{ij}\ni e}|\gamma_{ij}|_Q\pi(i)\pi(j)\right]\right\} \leq \mathcal{E}_\pi(x,x)\kappa(\Gamma).\end{aligned}$$

□

We now proceed to obtain a lower bound. For each state i, select exactly one closed path σ_i from i to i that does not pass twice through the same edge, and with an *odd* number of edges (for this to be possible, we assume that $\mathbf{P}$ is aperiodic), and let Σ be the collection of paths so selected. For a path $\sigma_i \in \Sigma$, let

$$|\sigma_i|_Q = \sum_{e\in\sigma_i}\frac{1}{Q(e)}\,.$$

Define

$$\alpha = \alpha(\Sigma) = \max_e \sum_{\sigma_i\ni e}|\sigma_i|_Q\pi(i)\,.$$

Theorem 20.2.5 *(Diaconis and Strook, 1991) Let* $\mathbf{P}$ *be an irreducible and aperiodic transition matrix on the finite state space* E*, with stationary distribution* π*, and assume* $(\mathbf{P}, \pi)$ *to be reversible. Then*

$$\lambda_r \geq -1 + \frac{2}{\alpha}\,.$$

Proof. It suffices to prove (20.28) with $B = \frac{2}{\alpha}$. For this, we use the easily established identity

$$\frac{1}{2}\sum_{i,j\in E}(x(i)+x(j))^2\pi(i)p_{ij} = \langle \mathbf{P}x, x\rangle_\pi + \|x\|_\pi^2 . \qquad (\star)$$

If σ_i is a path from i to i with an odd number of edges, of the form $\sigma_i = (i_0 = i, i_1, i_2, \ldots, i_{2m}, i)$, then

$$\begin{aligned} x(i) &= \frac{1}{2}\{(x(i_0)+x(i_1)) - (x(i_1)+x(i_2)) + \cdots + (x(i_{2m})+x(i))\} \\ &= \frac{1}{2}\sum_{e\in\sigma_i}(-1)^{n(e)}(x(e^+)+x(e^-)), \end{aligned}$$

where $n(e) = k$ if $e = (i_k, i_{k+1}) \in \sigma_i$. Therefore,

$$\|x\|_\pi^2 = \sum_{i\in E}\frac{\pi(i)}{4}\left\{\sum_{e\in\sigma_i}\frac{1}{Q(e)^{\frac{1}{2}}}Q(e)^{\frac{1}{2}}(-1)^{n(e)}(x(e^+)+x(e^-))\right\}^2,$$

and by Schwarz's inequality, this quantity is smaller than or equal to

$$\begin{aligned} &\sum_{i\in E}\left\{\frac{\pi(i)}{4}|\sigma_i|_Q\sum_{e\in\sigma_i}(x(e^+)+x(e^-))^2Q(e)\right\} \\ &\quad= \frac{1}{4}\sum_e\left\{(x(e^+)+x(e^-))^2Q(e)\sum_{\sigma_i\ni e}|\sigma_i|_Q\pi(i)\right\} \\ &\quad\le \frac{\alpha}{4}\sum_e(x(e^-)+x(e^+))^2Q(e). \end{aligned}$$

Therefore, in view of $(\star)$,

$$\|x\|_\pi^2 \le \frac{\alpha}{2}\left\{\|x\|_\pi^2 + \langle \mathbf{P}x, x\rangle_\pi\right\},$$

and this is the announced inequality. □

EXAMPLE 20.2.6: RANDOM WALK ON A GRAPH. For the random walk on a graph $G = (V, \mathcal{E})$, recall that the stationary distribution is $\pi(i) = \frac{d_i}{2|\mathcal{E}|}$, where d_i is the degree of vertex i and $|\mathcal{E}|$ is the number of edges. We first apply the bound of Theorem 20.2.4. For any edge e, $Q(e) = \frac{1}{2|\mathcal{E}|}$. Denoting by $|\gamma|$ the length of a path γ,

$$|\gamma_{ij}|_Q = |\gamma_{ij}| \times 2|\mathcal{E}| .$$

Therefore

$$\begin{aligned} \kappa &= \max_e \sum_{\gamma_{ij}\ni e}|\gamma_{ij}| \times 2|\mathcal{E}|\frac{d_i d_j}{4|\mathcal{E}|^2} \\ &\le \max_e \sum_{\gamma_{ij}\ni e}|\gamma_{ij}| \times \frac{d_{max}^2}{2|\mathcal{E}|} \end{aligned}$$

where d_{max} is the maximum degree of a vertex. Therefore

$$\kappa(\Gamma) \leq \frac{1}{2|\mathcal{E}|} d_{max}^2 K \,,$$

where

$$K := \max_e |\{\gamma \in \Gamma; e \in \gamma\}| \times \max\{|\gamma|; \gamma \in \Gamma\} \,.$$

Finally

$$\lambda_2 \leq 1 - \frac{2|\mathcal{E}|}{d^2|K} \,. \tag{20.30}$$

Similar calculations give for the bound in Theorem 20.2.5

$$\lambda_r \geq -1 + \frac{2}{d_{max}|\sigma|b} \,, \tag{20.31}$$

where $|\sigma| = \max |\sigma_i|$, and $b := \max_e |\{\sigma \in \Sigma; e \in \sigma\}|$.

Bottleneck Bound

This bound concerns finite state space irreducible reversible transition matrices $\mathbf{P}$. It is in terms of flows on the transition graph.

For a non-empty set $B \subset E$, define the *capacity* of B,

$$\pi(B) := \sum_{i \in B} \pi(i) \,,$$

and the *edge flow* out of B,

$$Q\left(B, \overline{B}\right) := \sum_{i \in B, j \in \overline{B}} \pi(i) \, p_{ij} \,.$$

Note that $Q\left(B, \overline{B}\right) = Q\left(\overline{B}, B\right)$ and that $0 \leq Q\left(B, \overline{B}\right) \leq \pi(B) \leq 1$. For non-empty B, define the bottleneck ratio of B:

$$\Phi(B) := \frac{Q\left(B, \overline{B}\right)}{\pi(B)} \,.$$

The bottleneck ratio of the pair $(\mathbf{P}, \pi)$ is

$$\Phi^* := \inf\left(\Phi(B); 0 < |B| < |E|, \pi(B) \leq \frac{1}{2}\right) \,. \tag{20.32}$$

EXAMPLE 20.2.7: For the pure random walk on $G = (V, \mathcal{E})$,

$$\pi(i) p_{ij} = \frac{d_i}{2|\mathcal{E}|} \, d_i = \frac{1}{2|\mathcal{E}|}$$

if $\langle i, j \rangle$ is an edge, $= 0$ otherwise. In this case, defining the internal boundary ∂B to be the set of states $i \in B$ that are connected to an element of $\overline{B}$ by an edge,

$$\Phi(B) = \frac{|\partial B|}{\sum_{i \in B} d_i} \,.$$

Theorem 20.2.8 *Cheeger's inequality:*

$$1 - 2\Phi^* \leq \lambda_2 \leq 1 - \frac{1}{2}(\Phi^*)^2 .$$

Proof. (Jerrum and Sinclair, 1989)

(a) Apply Rayleigh's spectral theorem,

$$1 - \lambda_2 \leq \frac{\mathcal{E}_\pi (x, x)}{\|x\|_\pi^2}$$

for any nontrivial vector x such that $\langle x \rangle_\pi = 0$. Select $B \subset E$ such that $\pi(B) \leq \frac{1}{2}$, and define

$$x(i) = \begin{cases} 1 - \pi(B) & \text{if } i \in B, \\ -\pi(B) & \text{if } i \notin B. \end{cases}$$

Then $\langle x \rangle_\pi = 0$ and $\|x\|_\pi^2 = \pi(B)(1 - \pi(B))$. Also,

$$\begin{aligned} \mathcal{E}_\pi (x, x) &= \frac{1}{2} \sum_{ij} (x(i) - x(j))^2 \pi(i) p_{ij} \\ &= \frac{1}{2} \sum_{i \in B} (\cdots) \sum_{j \notin B} (\cdots) + \frac{1}{2} \sum_{i \notin B} (\cdots) \sum_{j \in B} (\cdots) \\ &= \frac{1}{2} Q(\overline{B}, B) + \frac{1}{2} Q(B, \overline{B}) = Q(\overline{B}, B). \end{aligned}$$

Therefore,

$$1 - \lambda_2 \leq \frac{Q(\overline{B}, B)}{\pi(B)(1 - \pi(B))} \leq 2 \, \frac{Q(\overline{B}, B)}{\pi(B)}.$$

This being true for all B such that $\pi(B) \leq \frac{1}{2}$, we have, by definition of Φ^*,

$$1 - \lambda_2 \leq 2 \, \Phi^*.$$

(b) Let u be a left-eigenvector of $\mathbf{P}$ associated with an eigenvalue $\lambda \neq 1$. In particular, u is orthogonal to π, the left-eigenvector associated with the eigenvalue $\lambda_1 = 1$, and therefore u has positive as well as negative entries. The same is true for x defined by

$$x(i) = \frac{u(i)}{\pi(i)}.$$

Assume without loss of generality that for some k $(1 \leq k \leq r)$

$$x(1) \geq \cdots \geq x(k) > 0 \geq x(k+1) \geq \cdots \geq x(r),$$

and that $\pi(B) \leq \frac{1}{2}$ for $B := \{1, \ldots, k\}$ (if necessary, change the order of the states, and for the last assumption, change u into $-u$). Let

$$y(i) := \frac{u(i)}{\pi(i)} \, 1_{\{u(i) > 0\}} .$$

We have $u^T(I-\mathbf{P}) = u^T(1-\lambda)$, and therefore

$$u^T(I-\mathbf{P})\,y = (1-\lambda)\,u^T y = (1-\lambda)\sum_{i\in B}\pi(i)\,y(i)^2. \tag{20.33}$$

Also,

$$\begin{aligned} u^T(I-\mathbf{P})\,y &= \sum_{i\in B}\sum_{j=1}^{r}(\delta_{ji}-p_{ji})\,u(j)\,y(i) \\ &\geq \sum_{i\in B}\sum_{j\in B}(\delta_{ji}-p_{ji})\,u(j)\,y(i), \end{aligned}$$

since the missing terms $-p_{ji}u(j)\,y(i)$ corresponding to $i\in B$ and $j\notin B$ are positive or null. Therefore,

$$u^T(I-\mathbf{P})y \geq \langle y,(I-\mathbf{P})y\rangle_\pi,$$

and by (20.33), (20.23) and reversibility (Theorem 20.1.2),

$$1-\lambda \geq \frac{\sum_{i<j}\pi(i)\,p_{ij}\,(y(i)-y(j))^2}{\sum_{i\in B}\pi(i)\,y(i)^2}.$$

From $(a+b)^2 \leq 2(a^2+b^2)$, we obtain

$$\sum_{i<j}\pi(i)\,p_{ij}\,(y(i)+y(j))^2 \leq 2\sum_{i<j}\pi(i)\,p_{ij}\left(y(i)^2+y(j)^2\right),$$

and, by reversibility,

$$\begin{aligned} \sum_{i<j}\pi(i)\,p_{ij}\left(y(i)^2+y(j)^2\right) &= \sum_{i<j}\pi(i)\,p_{ij}y(i)^2+\sum_{i<j}\pi(j)\,p_{ji}y(j)^2 \\ &= \sum_{i\neq j}\pi(i)\,p_{ij}y(i)^2 \leq \sum_{i\in B}\pi(i)\,y(i)^2. \end{aligned}$$

Therefore

$$1-\lambda \geq \frac{\sum_{i<j}\pi(i)\,p_{ij}\,(y(i)-y(j))^2}{\sum_{i\in B}\pi(i)\,y(i)^2}\quad\frac{\sum_{i<j}\pi(i)\,p_{ij}\,(y(i)+y(j))^2}{2\sum_{i\in B}\pi(i)\,y(i)^2}.$$

By Schwarz's inequality and identity $a^2-b^2=(a-b)(a+b)$,

$$\begin{aligned} &\left(\sum_{i<j}\pi(i)\,p_{ij}\left(y(i)^2-y(j)^2\right)\right)^2 \\ &\qquad\leq \left(\sum_{i<j}\pi(i)\,p_{ij}\,(y(i)-y(j))^2\right)\left(\sum_{i<j}\pi(i)\,p_{ij}\,(y(i)+y(j))^2\right), \end{aligned}$$

and therefore

$$1-\lambda \geq \frac{1}{2}\left(\frac{\sum_{i<j}\pi(i)\,p_{ij}\left(y(i)^2-y(j)^2\right)}{\sum_{i\in B}\pi(i)\,y(i)^2}\right)^2. \qquad (\dagger)$$

Define $B_\ell=\{1,\ldots,\ell\}$. We have

$$\begin{aligned}\sum_{i<j}\pi(i)\,p_{ij}\left(y(i)^2-y(j)^2\right) &= \sum_{i<j}\pi(i)\,p_{ij}\left(\sum_{i\leq l<j}\left(y(\ell)^2-y(\ell+1)^2\right)\right)\\ &= \sum_{\ell=1}^{k}\left(y(\ell)^2-y(\ell+1)^2\right)\sum_{i\in B_\ell, j\notin B_\ell}\pi(i)\,p_{ij}\\ &= \sum_{\ell=1}^{k}\left(y(\ell)^2-y(\ell+1)^2\right)\;F(B_\ell).\end{aligned}$$

Since for $1\leq\ell\leq k$, $\pi(B_\ell)\leq\pi(B)\leq\frac{1}{2}$, we have $F(B_\ell)\geq\Phi^*\,\pi(B_\ell)$. Therefore,

$$\begin{aligned}\sum_{i<j}\pi(i)\,p_{ij}\left(y(i)^2-y(j)^2\right) &\geq \Phi^*\sum_{\ell=1}^{k}\left(y(\ell)^2-y(\ell+1)^2\right)\;\pi(B_\ell)\\ &= \Phi^*\sum_{\ell=1}^{k}\left\{\left(y(\ell)^2-y(\ell+1)^2\right)\;\sum_{i=1}^{\ell}\pi(i)\right\}\\ &= \Phi^*\sum_{i=1}^{k}\left\{\pi(i)\;\left(\sum_{\ell=i}^{k}\left(y(\ell)^2-y(\ell+1)^2\right)\right)\right\}\\ &= \Phi^*\sum_{i\in B}\pi(i)\,y(i)^2.\end{aligned}$$

Therefore, from (†)

$$1-\lambda\geq\frac{(\Phi^*)^2}{2}.$$

□

EXAMPLE 20.2.9: THE CYCLIC GRAPH. The vertices are n points uniformly distributed on the unit circle, and the n edges are those linking the neighbouring vertices. Take n odd. For any B, $Q(B,\overline{B})=\frac{1}{n}$ and one may easily check that Φ^* is achieved by any set B of $\frac{n-1}{2}$ consecutive vertices, and then

$$\Phi^*=\frac{2}{n-1}.$$

Therefore

$$\lambda_2\leq 1-\frac{2}{(n-1)^2}.$$

It can be verified that the bound in Example 20.2.6 gives in this special case

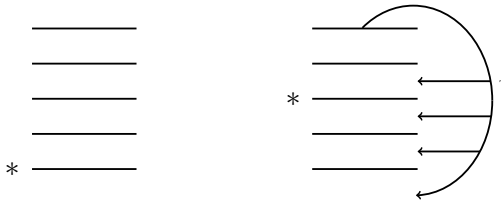
$$\lambda_2\leq 1-\frac{8n}{(n-1)^2(n+1)}$$

and is therefore of the same order but with a better constant. It turns out that in this case the exact eigenvalues are available (Diaconis, 1988): $\cos(2\pi\frac{j}{n})$ $(0 \leq j \leq n-1)$, and therefore $\lambda_2 = 1 - \frac{2\pi^2}{n^2} + O(\frac{1}{n^4})$. The Poincaré bound is therefore comparable, up to a factor π^2, to the actual spectral gap.

20.2.2 Strong Stationary Times

Strong stationary times, by which exact sampling for the stationary distribution of a positive recurrent HMC can be achieved, will be defined right after the two following examples.

EXAMPLE 20.2.10: TOP TO RANDOM CARD SHUFFLING, TAKE 1. (Aldous and Diaconis, 1987) The title refers to a method of shuffling a deck of N cards whereby the top card of the deck is removed and placed at random in the deck, and the procedure is repeated *ad infinitum*.



This defines an irreducible HMC $\{X_n\}_{n\geq 0}$, where a state is a permutation of the deck. In other words, it is a random walk on the group $\mathcal{S}_N$ of permutations on the set of N cards. Its stationary distribution is the uniform ditribution (Example 6.2.11). (Alternatively, use symmetry, and to make symmetry more apparent, arrange the cards in a circle rather than in a deck.) Denote by $\star$ the card originally at the bottom. If there are j cards below $\star$ at time n, the $j!$ possible arrangements of these cards are equally likely, as the following inductive argument shows. The statement is true for $n = 0$. Suppose it is true for some $n \geq 0$, then it is true for $n+1$. In fact two events can take place at time $n+1$. Either the top card is placed above $\star$ in which case the claim is trivially true, or it is placed under $\star$ and it is also true since inserting at a random position an element in a random permutation of j elements results in a random permutation of $j+1$ elements.

Let τ_j be the jth time a card is inserted below $\star$. If there are N cards, then, at time τ_{N-1}, card $\star$ has reached the top. Let $\tau = \tau_{N-1} + 1$. Since for $j \leq N-1$, at time τ_j all the $j!$ arrangements of the j cards below $\star$ are equally likely, the distribution of X_τ is uniform.

Note that τ is a X_0^n-stopping time.

EXAMPLE 20.2.11: LAZY WALK ON THE HYPERCUBE, TAKE 2. In Example 6.2.6, the lazy random walk on the N-dimensional hypercube $E = \{0,1\}^N$ was described distributionwise by the recurrence equation $X_{n+1} = f(X_n, Z_{n+1})$ where $\{Z_n\}_{n\geq 1}$ is an IID sequence of random variables uniformly distributed on $\{1,\ldots,N\}$ independent of the initial state X_0. More precisely, $Z_n = (U_n, B_n)$ where the sequence $\{(U_n, B_n)\}_{n\geq 1}$ is IID and uniformly distributed on $\{1,2,\ldots,N\} \times \{0,1\}$. The position at time $n+1$ is that of X_n except that the bit in position U_{n+1} is replaced by B_{n+1}.

Define a random time τ to be the first time for which the set $\{U_1, U_2, \ldots, U_n\}$ contains all the elements of $\{1,2,\ldots,N\}$. Because at this time all the coordinates have been replaced by independent fair bits, the distribution at time τ is the uniform distribution, that is, the stationary distribution.

This time, however, τ is not a X_0^n-stopping time. It is a randomized X_0^n-stopping time in the sense of the next definition.

Definition 20.2.12 *Let $\{X_n\}_{n\geq 0}$ be a* HMC *with the representation as in Theorem 6.1.4. A random time τ with values in $\bar{\mathbb{N}}$ is called a randomized X_0^n-stopping time if, for all $k \in \mathbb{N}$, the event $\{\tau = k\}$ is expressible in terms of $X_0, Z_1, \ldots, Z_k$.*

The times τ of Examples 20.2.10 and 20.2.11 are randomized stopping times (Exercise 20.4.13).

If τ is a randomized X_0^n-stopping, for all $m, n \geq 0$ and for all $i, j \in E$,

$$P\left(X_{m+n} = j \,|X_n = i, \tau \leq n\right) = p_{ij}\left(m\right).$$

Indeed, $\{\tau \leq n\}$ is expressible in terms of $X_0, Z_1, \ldots, Z_n$, and is therefore independent of X_{m+n} given $X_n = j$. Similar formulas, formally identical to the case where τ is a usual, non-randomized, X_0^n-stopping time, hold true and will be used in the calculations below.

Definition 20.2.13 *(Fill, 1991; Aldous and Diaconis, 1987; Diaconis and Fill, 1991) A randomized X_0^n-stopping time τ with respect to the* HMC *$\{X_n\}_{n\geq 0}$ admitting a unique stationary distribution π is called a strong stationary time of this* HMC *iff it is almost surely finite and*

(α) X_τ is distributed according to π and is independent of τ.

If the requirement of independence of X_τ and τ is dropped, τ is simply called a stationary time. The times τ of Examples 20.2.10 and 20.2.11 are strong stationary times (Exercise 20.4.13).

In the above definition, condition (α) is equivalent to either one of the following two conditions:

(β) For all $i \in E$ and all $n \geq 0$,

$$P\left(X_n = i \,|\tau = n\right) = \pi\left(i\right).$$

(γ) For all $i \in E$ and all $n \geq 0$,

$$P\left(X_n = i \,|\tau \leq n\right) = \pi\left(i\right).$$

The reader is invited to provide the proof (Exercise 20.4.14).

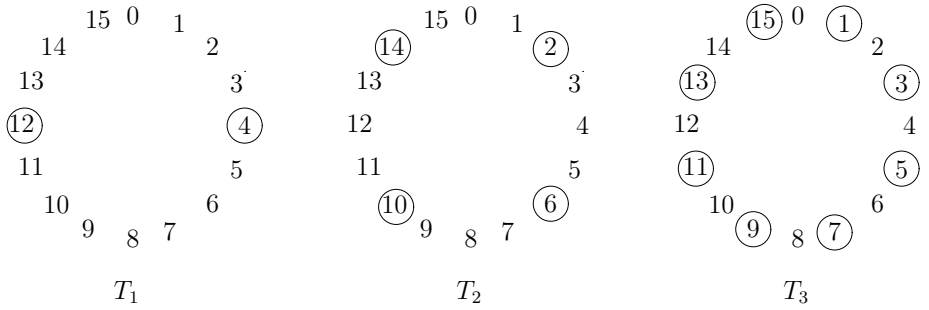
Also, if either (α), or (β), or (γ), holds, then $\{X_{\tau+n}\}_{n\geq 0}$ is a stationary HMC with the transition matrix $\mathbf{P}$ and is independent of τ. To check this, just write

$$\begin{aligned}
&P(\tau = k, X_\tau = i_0, X_{\tau+1} = i_1, \ldots, X_{\tau+n} = i_n) \\
&\quad = P(\tau = k, X_k = i_0, X_{k+1} = i_1, \ldots, X_{k+n} = i_n) \\
&\quad = P(\tau = k, X_k = i_0) P(X_{k+1} = i_1, \ldots, X_{k+n} = i_n | \tau = k, X_k = i_0) \\
&\quad = P(\tau = k) \pi(i_0) P(X_{k+1} = i_1, \ldots, X_{k+n} = i_n | X_k = i_0) \\
&\quad = P(\tau = k) P_\pi(X_k = i_0, X_{k+1} = i_1, \ldots, X_{k+n} = i_n) \\
&\quad = P(\tau = k) P_\pi(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n).
\end{aligned}$$

The announced result then follows.

Convergence Rates via Strong Stationary Times

EXAMPLE 20.2.14: LAZY WALK ON THE CIRCLE, TAKE 2. (Diaconis and Fill, 1991) Let $\{X_n\}_{n\geq 0}$ be a symmetric random walk on $E = \mathbb{Z}_d$, the integers modulo d, identified with d points on the circle (see the figure below). It moves one step in either direction or remains still, each motion with probability $\frac{1}{3}$.



This chain is clearly ergodic with the uniform probability on E. A strong stationary time can be constructed as follows in the case $d = 2^a$. We treat the case $d = 2^4 = 16$ for definiteness.

Starting from 0, let T_1 be the first time either state 4 or 12 is visited. Clearly, X_{T_1} is uniformly distributed on $\{4, 12\}$ and is independent of T_1. Next, let T_2 be the first time after T_1 when the chain visits the states at distance 2 from X_{T_1}. Then X_{T_2} is uniformly distributed on $\{2, 6, 10, 14\}$ and is independent of T_2. Time T_3 is now the first time after T_2 when the chain hits a state at distance 1 from X_{T_2}.

Then X_{T_3} is uniformly distributed on the odd numbers $\{1,3,5,7,9,11,13,15\}$ and is independent of T_3. Finally, let T be the first time after T_3 where the chain makes a clockwise move or stays still. We can take T as the desired strong stationary time, since it is independent of X_T, and X_T is uniform on E.

For $d = 2^a$, the distances successively traveled are $\pm 2^{a-2}, \pm 2^{a-3}, \ldots, \pm 1 = \pm 2^{a-a}$. The mean time to travel at distance b of this symmetric walk is $\frac{3}{2}b^2$. The last step from T_{a-1} to $T_a = T$ takes $\frac{3}{2}$ time units on average. Therefore

$$E_0[T_a] = \frac{3}{2}(2^{2a-4} + \ldots + 4 + 1) = \frac{3}{2}2^{2a}(2^{-4} + 2^{-6} + \ldots + 2^{-2(a-1)} + 2^{-2a}).$$

Therefore, for $a \geq 2$,

$$E_0[T_a] \leq \frac{3}{16}2^{2a} = \frac{3}{16}d^2.$$

By Markov's inequality,

$$P_0(T_a > n) \leq \frac{E_0[T_a]}{n} \leq \frac{3}{16}\frac{d^2}{n},$$

and therefore, since the result would be the same for any state,

$$d_V(\mu^T \mathbf{P}^n, \pi^T) \leq \frac{3}{16}\frac{d^2}{n}.$$

In this case, $t_{mix}(\varepsilon) \leq \frac{3}{16}\frac{d^2}{\varepsilon}$.

The tail of the distribution of a strong stationary time gives a bound for the rate of convergence in variation of an ergodic HMC. This is the content of Theorem 20.2.16 below. For this, it is convenient to use the notion of separation distance below.

Let α and β be two probability distributions on the denumerable space E. The *separation* of α from β, denoted by $s(\alpha;\beta)$, is defined by

$$s(\alpha;\beta) = \max_{i \in E}\left(1 - \frac{\alpha(i)}{\beta(i)}\right).$$

Note that $0 \leq s(\alpha;\beta) \leq 1$. (For the lower bound, observe that one cannot have $\alpha(i) > \beta(i)$ for all i.)

Theorem 20.2.15 *Let α and β be two probability distributions on the denumerable space E. Then*

$$d_V(\alpha;\beta) \leq s(\alpha;\beta).$$

Proof. Recall that $d_V(\alpha;\beta) = \sum_{i;\,\beta(i)>\alpha(i)}(\beta(i) - \alpha(i))$. But the latter sum equals

$$\sum_{i;\,\beta(i)>\alpha(i)} \beta(i)\left(1 - \frac{\alpha(i)}{\beta(i)}\right) \leq \left(\sum_{i;\,\beta(i)>\alpha(i)} \beta(i)\right) s(\alpha;\beta) \leq s(\alpha;\beta).$$

□

Theorem 20.2.16 *Let* $\mathbf{P}$ *be transition matrix of an irreducible positive recurrent* HMC $\{X_n\}_{n\geq 0}$ *with the stationary distribution* π*. If* τ *is a strong stationary time of the chain with initial distribution* μ*, then*

$$s\left(\mu^T\mathbf{P}^n;\pi^T\right) \leq P(\tau > n)\,.$$

Proof. By (γ) after Definition 20.2.13,

$$P\left(X_n = i\right) \geq P\left(X_n = i, \tau \leq n\right) = \left(1 - P\left(\tau > n\right)\right)\pi\left(i\right).$$

Therefore, for all i,

$$P\left(\tau > n\right) \geq 1 - \frac{P\left(X_n = i\right)}{\pi\left(i\right)}.$$

□

Theorem 20.2.17 *Let* τ *be a strong stationary time of the* HMC $\{X_n\}_{n\geq 0}$*. Then* τ *is also a stationary coupling time of the same chain.*

Proof. For each $m \geq 0$, define on $\{\tau = m\}$ the process $\left\{Y_n^{(m)}\right\}_{n\geq m}$ by

$$Y_n^{(m)} = X_n \text{ if } n \geq m.$$

Since for $n \geq m$, by definition of a strong stationary time, $P\left(X_n = i,\, \tau = m\right) = \pi\left(i\right)P\left(\tau = m\right)$, we see that, conditionally on $\{\tau = m\}$, $\left\{Y_n^{(m)}\right\}_{n\geq m}$ is a stationary HMC. It can be extended to a stationary HMC $\left\{Y_n^{(m)}\right\}_{n\geq 0}$. Letting $Y_n := Y_n^{(m)}$ on $\{\tau = m\}$, we obtain an HMC $\{Y_n\}_{n\geq 0}$ that is stationary and such that $X_n = Y_n$ for $n \geq \tau$. □

20.2.3 Reversibilization

Suppose $\mathbf{P}$ is an ergodic transition matrix on the finite state space $E = \{1, 2, \ldots, r\}$, with stationary distribution π. This time, $(\mathbf{P}, \pi)$ is not assumed reversible. What can be done to catch up with the results obtained above for the reversible case?

Consider the transition matrix $\tilde{\mathbf{P}}$ of the time-reversed chain, defined by

$$\tilde{p}_{ij} := \frac{\pi(j)p_{ji}}{\pi(i)}\,,$$

or, in compact form, with $D = D(\pi)$ as in (20.2),

$$\tilde{\mathbf{P}} := D^{-1}\mathbf{P}^T D. \tag{20.34}$$

The matrix $M = M(\mathbf{P}) := \mathbf{P}\tilde{\mathbf{P}}$, that is

$$M = \mathbf{P}D^{-1}\mathbf{P}^T D \tag{20.35}$$

is reversible with respect to its stationary distribution π. To prove this assertion, we have to verify that $M^* := D^{\frac{1}{2}} M D^{-\frac{1}{2}}$ is symmetric, that is

$$D^{\frac{1}{2}} M D^{-\frac{1}{2}} = \left(D^{\frac{1}{2}} M D^{-\frac{1}{2}}\right)^T .$$

(The right-hand side is

$$D^{-\frac{1}{2}} M^T D^{\frac{1}{2}} = D^{-\frac{1}{2}} D\mathbf{P}D^{-1}\mathbf{P}^T D^{\frac{1}{2}} = D^{\frac{1}{2}}\mathbf{P}D^{-1}\mathbf{P}^T D^{\frac{1}{2}}$$

whereas the left-hand side is $D^{\frac{1}{2}}\mathbf{P}D^{-1}\mathbf{P}^T D D^{-\frac{1}{2}} = D^{\frac{1}{2}}\mathbf{P}D^{-1}\mathbf{P}^T D^{\frac{1}{2}}$.)

The eigenvalues of M are real, and all belong to $[-1, +1]$. In fact, they all belong to the interval $[0, 1]$. To see this, observe that M has the same eigenvalues as $D^{\frac{1}{2}} M D^{-\frac{1}{2}}$, and that the latter matrix is $(D^{\frac{1}{2}}\mathbf{P}D^{-\frac{1}{2}})(D^{\frac{1}{2}}\mathbf{P}D^{-\frac{1}{2}})^T$, a symmetric definite non-negative matrix. In particular, the SLE is the SLEM.

The matrix M given by (20.35) is the *multiplicative reversibilization* of $\mathbf{P}$. See Exercise 20.4.12 for another type of reversibilization.

Theorem 20.2.18 *(Fill, 1991) Let $\gamma_1 = \gamma_1(M)$ be the second-largest eigenvalue of $M = \mathbf{P}\tilde{\mathbf{P}}$, where $\mathbf{P}$ is an ergodic transition matrix on the finite state space E. Then for any probability distribution ν on E,*

$$|\nu^T\mathbf{P}^n - \pi^T|^2 \le \gamma_1(M)^n \chi^2(\nu;\pi)\,. \tag{20.36}$$

Inequality (20.36) is called the χ^2-contrast bound.

Proof. The following identity (Mihaïl, 1989) will be needed:

$$\text{Var}_{\,\pi}(x) = \text{Var}_{\,\pi}(\tilde{\mathbf{P}}x) + \langle (1-M)x, x\rangle_\pi\,. \tag{20.37}$$

It is proven as follows. First, from (20.26), if we let $\hat{x} = x - \langle x\rangle_\pi \mathbf{1}$, then

$$\begin{aligned}\langle (I-M)x, x\rangle_\pi &= \langle (I-M)\hat{x}, \hat{x}\rangle_\pi = \|\hat{x}\|_\pi^2 - \langle M\hat{x}, \hat{x}\rangle_\pi \\ &= \|\hat{x}\|_\pi^2 - \langle \mathbf{P}\tilde{\mathbf{P}}\hat{x}, \hat{x}\rangle_\pi = \|\hat{x}\|_\pi^2 - \|\tilde{\mathbf{P}}\hat{x}\|_\pi^2,\end{aligned}$$

where the fact that $\tilde{\mathbf{P}}$ is the adjoint of $\mathbf{P}$ in $\ell^2(\pi)$ was taken into account. The identity (20.37) follows since $\|\hat{x}\|_\pi^2 = \text{Var}_{\,\pi}(x)$ and $\|\tilde{\mathbf{P}}\hat{x}\|_\pi^2 = \text{Var}_{\,\pi}(\tilde{\mathbf{P}}x)$.

Now let $\chi_n^2 := \chi^2(\nu^T\mathbf{P}^n;\pi)$ and $\rho_n(i) := \frac{(\nu^T\mathbf{P}^n)(i)}{\pi(i)}$. One verifies by inspection that $\text{Var}_{\,\pi}(\rho_n) = \chi_n^2$ and $\tilde{\mathbf{P}}\rho_n = \rho_{n+1}$. Therefore, from (20.37),

$$\chi_n^2 = \chi_{n+1}^2 + \langle (1-M)\rho_n, \rho_n\rangle_\pi\,.$$

By Rayleigh's spectral theorem,

$$\langle (1-M)\rho_n, \rho_n\rangle_\pi \ge (1-\gamma_1(M))\,\text{Var}_{\,\pi}(\rho_n) = (1-\gamma_1(M))\chi_n^2\,,$$

and therefore $\chi_{n+1}^2 \le \gamma_1\chi_n^2$, from which it follows that $\chi_n^2 \le \gamma_1^n\chi_0^2$. But by (20.15), $d_V(\nu,\pi)^2 \le \chi^2(\nu;\pi)$, and this finishes the proof. □

20.3 Mixing Times

20.3.1 Basic Definitions

For a positive recurrent HMC with transition matrix $\mathbf{P}$ and stationary distribution π, define for all $n \geq 0$

$$d(n) := \max_{i \in E} d_V\left(\delta_i \mathbf{P}^n, \pi\right), \quad \overline{d}(n) := \max_{i,j \in E} d_V\left(\delta_i \mathbf{P}^n, \delta_j \mathbf{P}^n\right). \tag{20.38}$$

These quantities are equivalent in the sense that

$$d(n) \leq \overline{d}(n) \leq 2d(n).$$

Proof.

The right-most inequality follows from the triangle inequality

$$d_V\left(\delta_i^T \mathbf{P}^n, \delta_j^T \mathbf{P}^n\right) \leq d_V\left(\delta_i^T \mathbf{P}^n, \pi\right) + d_V\left(\delta_j^T \mathbf{P}^n, \pi\right).$$

Now, for all $k \geq 0$, $\pi(k) = \sum_j \pi(j) p_{jk}(n)$, and therefore

$$\begin{aligned}
\pi(A) &= \sum_{k \in A} \sum_j \pi(j) p_{jk}(n) = \sum_j \pi(j) \left(\sum_{k \in A} p_{jk}(n) \right) \\
&= \sum_j \pi(j) P_j(X_n \in A) = \sum_j \pi(j) \delta_j^T \mathbf{P}^n(A).
\end{aligned}$$

Therefore

$$\begin{aligned}
d_V\left(\delta_i^T \mathbf{P}^n, \pi\right) &= \sup_{A \subseteq E} \left(\delta_i^T \mathbf{P}^n(A) - \pi(A)\right) \\
&= \sup_{A \subseteq E} \left| \sum_j \pi(j) \left(\delta_i^T \mathbf{P}^n(A) - \delta_j^T \mathbf{P}^n(A)\right) \right| \\
&\leq \sup_{A \subseteq E} \sum_j \pi(j) \left| \delta_i^T \mathbf{P}^n(A) - \delta_j^T \mathbf{P}^n(A) \right| \\
&= \sum_j \pi(j) \sup_{A \subseteq E} \left| \delta_i^T \mathbf{P}^n(A) - \delta_j^T \mathbf{P}^n(A) \right| \\
&= \sum_j \pi(j) d_V\left(\delta_i^T \mathbf{P}^n, \delta_j^T \mathbf{P}^n\right) \leq d_V\left(\delta_k^T \mathbf{P}^n, \delta_j^T \mathbf{P}^n\right),
\end{aligned}$$

for any $k \in E$. Hence the left-most inequality. □

Define the following mixing times. For $\varepsilon > 0$,

$$t_{mix}(\varepsilon) := \inf\{n \geq 0 \,;\, d(n) \leq \varepsilon\}, \tag{20.39}$$

and

$$t_{mix} := t_{mix}(1/4). \tag{20.40}$$

Lemma 20.3.1 *The function $\overline{d}$ is sub-multiplicative, that is, for all integers m, n:*

$$\overline{d}(n+m) \leq \overline{d}(n) \times \overline{d}(m)\,.$$

Proof. Let Y and Z be two random variables with respective distributions $\delta_i^T \mathbf{P}^n$ and $\delta_j^T \mathbf{P}^n$, and realizing maximal coupling for these distributions, that is

$$d_V(\delta_i^T \mathbf{P}^n, \delta_j^T \mathbf{P}^n) = P(Y \neq Z)\,.$$

Observe that

$$p_{i,k}(n+m) = \sum_\ell p_{i,\ell}(n) p_{\ell,k}(m) = \sum_\ell P(Y=\ell) p_{\ell,k}(m) = E\left[p_{Y,k}(m)\right]$$

and similarly, $p_{j,k}(n+m) = E\left[p_{Z,k}(m)\right]$. Therefore,

$$p_{i,k}(n+m) - p_{j,k}(n+m) = E\left[p_{Y,k}(m) - p_{Z,k}(m)\right]$$

and

$$\begin{aligned} &d_V(\delta_i^T \mathbf{P}^{n+m}, \delta_j^T \mathbf{P}^{n+m}) \\ &\quad = \frac{1}{2}\sum_k |p_{i,k}(n+m) - p_{j,k}(n+m)| = \frac{1}{2}\sum_k \left|E\left[p_{Y,k}(m) - p_{Z,k}(m)\right]\right| \\ &\quad \leq E\left[\frac{1}{2}\sum_k |p_{Y,k}(m) - p_{Z,k}(m)|\right] = E\left[d_V\left(p_{Y,\cdot}(m), p_{Z,\cdot}(m)\right)\right]. \end{aligned}$$

The quantity under expectation is null if $Y = Z$ and is in any case bounded by $\overline{d}(n)$. Therefore

$$\begin{aligned} d_V(\delta_i^T \mathbf{P}^{n+m}, \delta_j^T \mathbf{P}^{n+m}) &\leq E\left[\overline{d}(n) 1_{Y\neq Z}\right] \\ &= \overline{d}(n) P(Y \neq Z) = \overline{d}(n) d_V(\delta_i^T \mathbf{P}^n, \delta_j^T \mathbf{P}^n)\,. \end{aligned}$$

Maximizing over i, j yields the announced result. □

When N is an integer, by Lemma 20.3.1,

$$d(N t_{mix}(\varepsilon)) \leq \overline{d}(N t_{mix}(\varepsilon)) \leq \overline{d}(t_{mix}(\varepsilon))^N \leq (2\varepsilon)^N\,.$$

In particular, with $\varepsilon = \frac{1}{4}$,

$$d(N t_{mix}) \leq 2^{-N}.$$

With $N = N(\varepsilon) := \lceil \log_2 \varepsilon^{-1} \rceil$, $2^{-N} \leq \varepsilon$, and therefore, $d(N t_{mix}) \leq \varepsilon$, which implies that

$$t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}. \tag{20.41}$$

20.3.2 Upper Bounds via Coupling

We now show how to compute mixing times via coupling. Recall that two random sequences $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ with values in the same set E are said to couple at time τ if $n \geq \tau$ implies that $X_n = Y_n$. By Theorem 16.1.12, $d_V(X_n, Y_n) \leq P(\tau \geq n)$. Applying this inequality to the coupling time of two HMC's with the same transition matrix $\mathbf{P}$ with initial states i and j respectively, we have that

$$d_V(\delta_i \mathbf{P}^n, \delta_j \mathbf{P}^n) \leq P_{i,j}(\tau \geq n)\,.$$

By Markov's inequality,

$$P_{i,j}(\tau \geq n) \leq \frac{E_{i,j}[\tau]}{n}\,.$$

Therefore

Theorem 20.3.2

$$d(n) \leq \max_{i,j\in E} P_{i,j}(\tau \geq n) \leq \max_{i,j\in E} \frac{E_{i,j}[\tau]}{n}\,.$$

EXAMPLE 20.3.3: LAZY WALK ON THE CIRCLE, TAKE 1. This is by definition a lazy random walk on the graph consisting of N points regularly placed on a circle with an edge between each pair of adjacent vertices. The stationary distribution is the uniform distribution. We construct a Markovian coupling $\{X_n, Y_n\}_{n\geq 0}$ in the following way. At time n, supposing $X_n \neq Y_n$, a fair coin is tossed. If heads, the first particle moves one step in the direction chosen at random by means of another fair coin tossed independently and the other particle stays still. If tails, the second particle moves one step in the direction chosen at random by means of another fair coin tossed independently and the other particle stays still. The two particles make identical moves as soon as they collide for the first time. Calling D_n the distance between the two particles, $\{D_n\}_{n\geq 0}$ is a symmetric random walk on $\{0, 1, \ldots, N\}$ with absorbing states 0 and N. The coupling time is the first time τ where the symmetric random walk is absorbed at 0 or N. Therefore $E_{i,j}[\tau] = k(N-k)$ where k is the distance between i and j, and $d(n) \leq \max_{i,j\in E} \frac{E_{i,j}[\tau]}{n} \leq \frac{N^2}{4n}$. The right-hand side equals $\frac{1}{4}$ for $n = N^2$, therefore $t_{mix} \leq N^2$.

EXAMPLE 20.3.4: TOP TO RANDOM CARD SHUFFLING, TAKE 2. (Aldous and Diaconis, 1987) In the notation of take 1 of this example,

$$\tau = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau - \tau_{N-1})\,,$$

where $\tau - \tau_{N-1} = 1$. At time τ_i, card $\star$ has i cards below it, and the probability that the current top card is inserted below $\star$ is therefore $\frac{i+1}{N}$. Therefore, $\tau_{i+1} - \tau_i$ is geometric:

$$P(\tau_{i+1} - \tau_i = k) = \frac{i+1}{N}\left(1 - \frac{i+1}{N}\right)^{k-1}\,.$$

Consider now the following problem: Sample uniformly with replacement an urn containing N balls, and denote by V the number of draws until each ball has been sampled at least once. Let V_i be the number of draws until i distinct balls have been sampled at least once. We have the identity

$$V = (V - V_{N-1}) + \cdots + (V_2 - V_1) + V_1 \,.$$

Once i distinct balls have been drawn at least once, there is a probability $\frac{N-i}{N}$ of sampling a ball not previously sampled. Therefore, $V_i - V_{i-1}$ is geometric:

$$P(V_i - V_{i-1} = k) = \frac{N-i}{N}\left(1 - \frac{N-i}{N}\right)^{k-1} \,.$$

In particular, τ and V have the same distribution. For each ball b, let A_b be the event that ball b was not drawn in the first $m = N\ \log\,(N)\ +\ cN$ draws, $c \geq 0$. We have

$$P(V > m) = P(\cup_b A_b) = N(1 - \frac{1}{N})^m \leq Ne^{-\frac{m}{N}} = Ne^{-\log(N)-c} = e^{-c} \,.$$

Therefore

$$d(N\ \log\,(N)\ +\ cN) \leq (P(\tau > N\ \log\,(N)\ +\ cN) \leq e^{-c} \,,$$

where $d(k) = d_V(\mu^T \mathbf{P}^k, \pi^T)$. In particular, $t_{mix}(\varepsilon) \leq N \log N - \log(\varepsilon)\, N$.

20.3.3 Lower Bounds

Consider an irreducible ergodic HMC on the finite state space E, with transition matrix $\mathbf{P}$, and with a uniform stationary distribution π. Let $d_+(i)$ be the out-degree of state i, that is the number of directed edges in the transition graph out of vertex i: $d_+(i) := |\{j \in E\,;\, p_{ij} > 0\}|$, and let $d_{+,max} := \max_{i\in E} d_+(i)$. Therefore, starting from any state, the maximum number of states accessible in n steps is at most $d^n_{+,max}$. The distribution of X_n is therefore concentrated on a subset of E with at most $d^n_{+,max}$ elements. In particular, for any state i

$$d_V(\delta_i^T \mathbf{P}^n, \pi) \geq \frac{1}{|E|}(|E| - d^n_{+,max}) \,.$$

In particular, if $d^n_{+,max} \leq (1-\varepsilon)|E|$, that is, if $n \leq \frac{\log((1-\varepsilon)|E|)}{\log d_{+,max}}$, then $d(n) \geq \varepsilon$. This implies that

$$t_{mix}(\varepsilon) \geq \frac{\log\,((1-\varepsilon)|E|)}{\log d_{+,max}} \,.$$

EXAMPLE 20.3.5: RANDOM WALK ON A GRAPH. Let d_{max} be the maximal degree of the graph $G = (V, \mathcal{E})$. For the random walk on this graph, $d_{+,max} = d_{max}$ and therefore $t_{mix}(\varepsilon) \geq \frac{\log((1-\varepsilon)|E|)}{\log d_{max}}$.

From the directed transition graph of an irreducible ergodic HMC on the finite state space E, with transition matrix $\mathbf{P}$, construct a graph whose vertex set is E and with an edge linking i and j if and only if $p_{ij} + p_{ji} > 0$. The diameter D of the chain is by definition the diameter of this graph, that is the maximal graph distance between two vertices. If i_0 and j_0 are two states at the maximal graph distance D, then $\delta_{i_0}\mathbf{P}^{\lfloor (D-1)/2 \rfloor}$ and $\delta_{i_0}\mathbf{P}^{\lfloor (D-1)/2 \rfloor}$ have disjoint support, and therefore $\overline{d}(\lfloor (D-1)/2 \rfloor) = 1$. In particular, for any $\varepsilon < \frac{1}{2}$,

$$t_{mix}(\varepsilon) \geq \left\lfloor \frac{D-1}{2} \right\rfloor .$$

For the next result, recall definition (20.32) of the bottleneck ratio.

Theorem 20.3.6 *For an ergodic* HMC *with transition matrix* $\mathbf{P}$ *and bottleneck ratio* Φ^**,*

$$t_{mix} \geq \frac{1}{4\Phi^*} .$$

Proof. Denote by π_B the restriction of π to the set $B \subset E$, and by ρ_B the probability π conditioned by B:

$$\pi_B(A) = \pi(A \cap B),\ A \subseteq B, \text{ and } \rho_B(A) = \frac{\pi(A \cap B)}{\pi(B)},\ A \subseteq E .$$

We have

$$\begin{aligned} \pi(B) d_V\left(\rho_B \mathbf{P}, \rho_B\right) &= \pi(B) \sum_{i; \rho_B \mathbf{P}(i) \geq \rho_B(i)} \left(\rho_B \mathbf{P}(i) - \rho_B(i)\right) \\ &= \sum_{i; \pi_B \mathbf{P}(i) \geq \pi_B(i)} \left(\pi_B \mathbf{P}(i) - \pi_B(i)\right). \end{aligned}$$

Since $\pi_B(i) = 0$ on $\overline{B}$, and $\pi_B(i) = \pi(i)$ on B,

$$\pi_B \mathbf{P}(i) = \sum_{j \in B} \pi_B(j) p_{ji} \leq \sum_{j \in E} \pi(j) p_{ji} = \pi(i)$$

and therefore, if $i \in B$,

$$\pi_B \mathbf{P}(i) \leq \pi_B(i) ,$$

and for $i \notin B$, since $\pi_B(i)$ is then null,

$$\pi_B \mathbf{P}(i) \geq 0 = \pi_B(i) .$$

Therefore, from

$$\pi(B) d_V\left(\rho_B \mathbf{P}, \rho_B\right) = \sum_{i \in \overline{B}} \left(\pi_B \mathbf{P}(i) - \pi_B(i)\right) .$$

Because $\pi_B(i) = 0$ outside B, the right-hand side reduces to

$$\sum_{i\in\overline{B}}\sum_{j\in B}\pi(j)p_{ji} = Q(B,\overline{B}),$$

and, dividing by $\pi(B)$

$$d_V\left(\rho_B\mathbf{P},\rho_B\right) = \Phi(B)\,.$$

Now, since for all $n \geq 0$, $d_V\left(\rho_B\mathbf{P}^{n+1},\rho_B\mathbf{P}^n\right) \leq d_V\left(\rho_B\mathbf{P},\rho_B\right)$,

$$d_V\left(\rho_B\mathbf{P}^{n+1},\rho_B\mathbf{P}^n\right) \leq \Phi(B)\,.$$

By the triangle inequality applied to the sum $\rho_B\mathbf{P}^n - \rho_B = \sum_{k=0}^{n-1}\rho_B\mathbf{P}^{k+1} - \rho_B\mathbf{P}^k$,

$$d_V\left(\rho_B\mathbf{P}^n,\rho_B\right) \leq n\Phi(B)\,. \qquad (\star)$$

If $\pi(B) \leq \frac{1}{2}$,

$$d_V(\rho_B,\pi) \geq \pi(\overline{B}) - \rho_B(\overline{B}) = \pi(\overline{B}) = 1-\pi(B) \geq \frac{1}{2}\,.$$

By the triangle inequality

$$\frac{1}{2} \leq d_V(\rho_B,\pi) \leq d_V(\rho_B,\rho_B\mathbf{P}^n) + d_V(\rho_B\mathbf{P}^n,\pi)\,.$$

Letting $n = t_{mix}$, and using $(\star)$,

$$\frac{1}{2} \leq t_{mix}\Phi(B) + \frac{1}{4}$$

from which the result follows. □

EXAMPLE 20.3.7: TOP TO RANDOM CARD SHUFFLING, TAKE 3. (Aldous and Diaconis, 1987) We prove that for any $\varepsilon > 0$, there exists a constant c_0 such that for all $c \geq c_0$, for sufficiently large N,

$$d(N\log N - cN) \geq 1-\varepsilon\,.$$

In particular, there exists a constant c such that for sufficiently large N,

$$t_{mix} \geq N\log N - cN\,.$$

Proof. Let A_j denote the event that the original bottom j cards are in their relative original order. Denote by σ_0 the original configuration of the deck.

Let τ_j be the time it takes for the j-th card from the bottom to reach the top of the deck, and let $\tau_{j,i}$ the time it takes for this card to pass from position i to position $i+1$ (positions are counted from the bottom up). Then

$$\tau_j = \sum_{i=j}^{N-1}\tau_{j,i}\,.$$

The $\tau_{j,i}$'s ($j \le i \le N-1$) are independent geometric random variables with parameter $p := \frac{i}{N}$. In particular, $E[\tau_{j,i}] = \frac{N}{i}$ and $\mathrm{Var}(\tau_{j,i}) \le \frac{N^2}{i^2}$, and therefore

$$E[\tau_j] = \sum_{i=j}^{N-1} \frac{N}{i} \ge N(\log N - \log j - 1)\,,$$

and

$$\mathrm{Var}(\tau_j) \le N^2 \sum_{i=j}^{\infty} \frac{1}{i(i+1)} \le \frac{N^2}{j-1}\,.$$

From these bounds and Chebyshev's inequality,

$$\begin{aligned} P(\tau_j < N\log N - cN) &\le P(\tau_j - E[\tau_j] < -N(c - \log j - 1)) \\ &\le P(|\tau_j - E[\tau_j]\,| > N(c - \log j - 1)) \\ &\le \frac{\mathrm{Var}(\tau_j)}{N^2(c-\log j - 1)^2} \\ &\le \frac{\frac{N^2}{j-1}}{N^2(c-\log j-1)^2} \\ &\le \frac{1}{j-1} \times \frac{1}{N^2(c-\log j-1)^2} \le \frac{1}{j-1} \end{aligned}$$

(provided that $c \ge \log j + 2$ for the last inequality).

If $\tau_j \ge N\log N - cN$, the original j bottom cards are still in their original relative order, and therefore, for $c \ge \log j + 2$,

$$\delta_{\sigma_0}\mathbf{P}^{N\log N - cN}(A_j) \ge P(\tau_j \ge N\log N - cN) \ge 1 - \frac{1}{j-1}\,.$$

Now for the stationary distribution π, here the uniform distribution on the set of permutations $\mathcal{S}_N$, $\pi(A_j) = \frac{1}{j!} \le \frac{1}{j-1}$. Therefore, for $c \ge \log j + 2$,

$$d(N\log N - cN) \ge d_V\left(\delta_{\sigma_0}\mathbf{P}^{N\log N - cN}, \pi\right) \ge \delta_{\sigma_0}\mathbf{P}^{N\log N - cN}(A_j) - \pi(A_j) \ge 1 - \frac{2}{j-1}\,.$$

With $j = e^{c-2}$, if $N \ge e^{c-2}$,

$$d(N\log N - cN) \ge 1 - \frac{2}{e^{c-2}-1}\,.$$

Denoting by $g(c)$ the right-hand side of the above inequality, we have that

$$\liminf_{N\uparrow\infty} d(N\log N - cN) \ge g(c)\,,$$

where $\lim_{c\uparrow\infty} g(c) = 1$. □

Summarizing in rough terms the results of Examples 20.3.4 and 20.3.7: in order to shuffle a deck of N cards by the top-to-random method, "$N\log N$ shuffles suffice, but no less".

Books for Further Information

The problem of finding convergence rates for Markov chains is of central importance in applications, for instance in Monte Carlo sampling. It has therefore received considerable attention and generated a vast literature. This chapter is an introduction to this area and the reader is directed for more theory and examples to [Aldous and Fill, 2002, 2014] and [Levin, Peres, and Wilmer, 2009]. Examples where eigenvalues are known exactly are given in [Diaconis, 1988], a useful reference for the comparison of bounds. [Karlin, 1968] and [Karlin and Taylor, 1975] have a number of examples in biology.

20.4 Exercises

Exercise 20.4.1. The χ^2 distance in terms of the eigenstructure
Show that

$$\chi^2(p_{i\cdot}(n);\pi(\cdot)) = \sum_{j=2}^{r} \lambda_j^{2n} v_j(i)^2,$$

where v_j is the jth right-eigenvector associated with the reversible ergodic pair $(\mathbf{P},\pi)$, and λ_j is the corresponding eigenvalue.

Exercise 20.4.2. A characterization of the SLE
Let $\{X_n\}_{n\geq 0}$ be a stationary HMC corresponding to $(\mathbf{P},\pi)$. Show that the SLE λ_2 of $\mathbf{P}$ is equal to the maximum of the correlation coefficient between $f(X_0)$ and $f(X_1)$ among all real-valued functions f such that $E[f(X_0)] = 0$.

Exercise 20.4.3. Another Poincaré type coefficient
(Sinclair, 1990) Prove the version of Theorem 20.2.4 where Poincaré's coefficient κ is replaced by

$$\widetilde{\kappa} = \max_e Q(e)^{-1} \sum_{\gamma_{ij}, e\in\gamma_{ij}} |\gamma_{ij}|\pi(i)\pi(j),$$

where $|\gamma|$ is the length of path γ. In the pure random walk case of Example20.2.6 compare with the Poincaré type bound of Theorem 20.2.4.

Exercise 20.4.4. The star
Consider the random walk on the connected graph, the "star", with one central vertex connected to n outside vertices. Check that the corresponding transition matrix has eigenvalues $+1$, 0 and -1, where 0 has multiplicity $n-1$. What is the period? To eliminate periodicity, make it a lazy walk with holding probability $p_{ii} = \beta$. Show that eigenvalues the eigenvalues are now $+1$, β and $2\beta - 1$, where β has multiplicity $n-1$. For small α, compare the exact SLEM with the bound of Theorem 20.2.4.

Exercise 20.4.5. Random walk on a binary tree
Consider a random walk on a graph, where the graph is now a full binary tree of depth L.

(i) Show that the second largest eigenvalue λ_2 satisfies

$$\lambda_2 \leq 1 - \frac{1}{9L2^{L-1}}.$$

(ii) Explain why formula (20.31) does not apply directly. Show that

$$\lambda_2 \geq 1 - \left(2(2^L - 1)\left(1 - \frac{1}{2^{L+1} - 2}\right)\right)^{-1},$$

which is equivalent for large L to $1 - 2^{-L-1}$. Hint: Use Rayleigh's characterization with x as follows: $x(i) = 0, 1$, or -1 according to whether i is the root, a vertex on the right of the tree, or one on the left.

Exercise 20.4.6. POINCARÉ TYPE BOUND FOR THE RANDOM WALK ON A CUBE
Consider the random walk on the N-dimensional cube. Apply the Poincaré type bound of Theorem 20.2.5 with paths γ_x leading from x to y by changing the coordinates of x when they differ from that of y, inspecting the coordinates from left to right. Show that

$$\lambda_2 \leq 1 - \frac{2}{N^2}.$$

(In this example, the exact eigenvalues are available: $1 - \frac{2j}{N}$ with multiplicity $\binom{N}{j}$ $(0 \leq j \leq N)$, and therefore

$$\lambda_2 = 1 - \frac{2}{N}.$$

Therefore, the bound is off by a factor N.)

Exercise 20.4.7. $d(n)$ AND $\overline{d}(n)$
Refer to Definition 20.38 and denote by $\mathcal{P}(E)$ the collection of all probability distributions on E. Prove that

$$d(n) = \sup_{\mu \in \mathcal{P}(E)} d_V(\mu^T \mathbf{P}^n, \pi)$$

and

$$\overline{d}(n) = \sup_{\mu, \nu \in \mathcal{P}(E)} d_V(\mu^T \mathbf{P}^n, \nu^T \mathbf{P}^n).$$

Exercise 20.4.8. RANDOM WALK ON THE HYPERCUBE, TAKE 1
In Example 20.2.11 prove that $t_{mix}(\varepsilon) \leq N \log N - \log(\varepsilon)\, N$. Compare with the top to random card shuffle of Example 20.3.4. (Hint: the coupon collector.)

Exercise 20.4.9. RANDOM WALK ON A GROUP
Consider the random walk $\{X_n\}_{n \geq 0}$ on a group G (defined by (6.2.11) and the lines following it) with increment measure μ and transition matrix $\mathbf{P}$. Let $\{\hat{X}_n\}_{n \geq 0}$ be another random walk on G, this time corresponding to the increment measure $\hat{\mu}$, the symmetric of μ (that is, for all $g \in G$, $\hat{\mu}(g) = \mu(g^{-1})$). Let $\hat{\mathbf{P}}$ be the

corresponding transition matrix. Then, for all $n \geq 1$, denoting by π the common stationary distribution of the above two chains (equal the uniform distribution on G), we have that

$$d_V(\delta_e^T \mathbf{P}^n, \pi) = d_V(\delta_e^T \hat{\mathbf{P}}^n, \pi)$$

Exercise 20.4.10. Move-to-front policy
A professor has his N books on a bookshelf. His books are equally interesting for his research, so that when he decides to take one from his library, it is in fact chosen at random. The thing is that, being lazy or perhaps too busy, he does not put the book back to where it was, but instead at the beginning of the collection, in front of the other books. The arrangement on the shelf can be represented by a permutation σ of $\{1, 2, \ldots, N\}$, and the evolution of the arrangement is therefore an HMC on the group of permutations $\mathcal{S}_N$.

(i) Show that this chain is irreducible and ergodic, and admits the uniform distribution as stationary distribution.

(ii) Inspired by the top-to-random card shuffle Example 20.3.4 and Exercise 20.4.9, show that $t_{mix}(\varepsilon) \leq N \log N - \log(\varepsilon)\, N$.

Exercise 20.4.11. Mixing time of random walk on the binary tree
Consider the random walk on the rooted binary tree of depth k whose number of edges is therefore $N = 2^{k+1} - 1$. Show that its mixing time satisfies the lower bound

$$t_{mix} \geq \frac{N-2}{2}\,.$$

(Hint: consider the set $B \subset E$ consisting of the direct descendent of the root to the right, v_R, and of all the descendents of v_R.)

Exercise 20.4.12. Additive reversibilization
The *additive reversibilization* of $\mathbf{P}$ is, by definition, the matrix $A = A(\mathbf{P}) := \frac{1}{2}(\mathbf{P} + \tilde{\mathbf{P}})$, that is

$$A := \frac{1}{2}(\mathbf{P} + D^{-1}\mathbf{P}^T D)\,. \tag{20.42}$$

Show that this matrix is indeed reversible with respect to π.

Exercise 20.4.13. Strong stationary times
Prove that the times τ in Examples 20.2.10 and 20.2.11 are strong stationary times.

Exercise 20.4.14. Characterizations of strong stationary times
Show that condition (α) of Definition 20.2.13 is equivalent to either one of the following two conditions:

(β) For all $i \in E$ and all $n \geq 0$,

$$P\left(X_n = i \,|T = n\right) = \pi\left(i\right)\,.$$

(γ) For all $i \in E$ and all $n \geq 0$,

$$P\left(X_n = i \,|T \leq n\right) = \pi\left(i\right) .$$

Exercise 20.4.15. SEPARATION DISTANCE
Let $\mathcal{P}\left(E\right)$ be the collection of probability distributions on the countable set E. Show that for all α, $\beta \in M_p\left(E\right)$,

$$s\left(\alpha;\beta\right) = \inf\left\{s \geq 0\,;\ \alpha = \left(1-s\right)\beta + s\gamma,\, \gamma \in \mathcal{P}\left(E\right)\right\},$$

where s denotes the separation pseudo-distance.

Exercise 20.4.16. MIXING TIME OF THE REVERSED RANDOM WALK ON A GROUP
The situation is that prevailing in Theorem 6.2.11. Let now μ be a not necessarily symmetric probability distribution on G, and define its inverse $\hat{\mu}$ by

$$\hat{\mu}(g) = \mu(g^{-1}) .$$

Define the HMC $\{\hat{X}_n\}_{n\geq 0}$ by

$$\hat{X}_{n+1} = \hat{Z}_{n+1} * \hat{X}_n$$

where $\{\hat{Z}_n\}_{n\geq 1}$ is an IID sequence with values in G and distribution $\hat{\mu}$, independent of the initial state $\hat{X}_0$. It turns out that the forward HMC $\{X_n\}_{n\geq 0}$ and the backward HMC $\{\hat{Z}_n\}_{n\geq 1}$ have the same mixing times: $t_{mix} = \hat{t}_{mix}$.

Chapter 21

Exact Sampling

21.1 Backward Coupling

21.1.1 The Propp–Wilson Algorithm

The classical Monte Carlo Markov chain method of Chapter 19 provides an approximate sample of a probability distribution π on a finite state space E. Chapter 20 gives ways of measuring the accuracy of such an approximate sample in terms of its variation distance from the target distribution. The goal is now to construct an *exact* sample of π, that is, a random variable Z such that $P(Z = i) = \pi(i)$ for all $i \in E$. The following algorithm (Propp and Wilson, 1993) is based on a coupling idea. One starts as usual from an *ergodic* transition matrix $\mathbf{P}$ with stationary distribution π, just as in the classical MCMC method.

The algorithm is based on a representation of $\mathbf{P}$ in terms of a recurrence equation, that is, for a given function f and an IID sequence $\{Z_n\}_{n\geq 1}$ independent of the initial state, the chain satisfies the recurrence

$$X_{n+1} = f(X_n, Z_{n+1})\,. \tag{21.1}$$

The algorithm constructs a family of HMC's with transition matrix $\mathbf{P}$ with the help of a unique IID sequence of random vectors $\{Y_n\}_{n\in\mathbb{Z}}$, called the *updating sequence*, where $Y_n = (Z_{n+1}(1), \cdots, Z_{n+1}(r))$ is a r-dimensional random vector, and where the coordinates $Z_{n+1}(i)$ have a common distribution, that of Z_1. For each $N \in \mathbb{Z}$ and each $k \in E$, a process $\{X_n^N(k)\}_{n\geq N}$ is defined recursively by:

$$X_N^N(k) = k,$$

and, for $n \geq N$,

$$X_{n+1}^N(k) = f(X_n^N(k), Z_{n+1}(X_n^N(k)).$$

(Thus, if the chain is in state i at time n, it will be at time $n+1$ in state $j = f(i, Z_{n+1}(i)$.) Each of these processes is therefore an HMC with the transition matrix $\mathbf{P}$. Note that for all $k, \ell \in E$, and all $M, N \in \mathbb{Z}$, the HMC's $\{X_n^N(k)\}_{n\geq N}$ and $\{X_n^M(\ell)\}_{n\geq M}$ use at any time $n \geq \max(M, N)$ the same updating random vector Y_{n+1}.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_21

If, in addition to the independence of $\{Y_n\}_{n\in\mathbb{Z}}$, the components $Z_{n+1}(1)$, $Z_{n+1}(2)$, ..., $Z_{n+1}(r)$ are, for each $n \in \mathbb{Z}$, independent, we say that the updating is *componentwise independent*.

Definition 21.1.1 *The random time*

$$\tau^+ = \inf\{n \geq 0; X_n^0(1) = X_n^0(2) = \cdots = X_n^0(r)\}$$

is called the forward coupling time. The random time

$$\tau^- = \inf\{n \geq 1; X_0^{-n}(1) = X_0^{-n}(2) = \cdots = X_0^{-n}(r)\}$$

is called the backward coupling time.

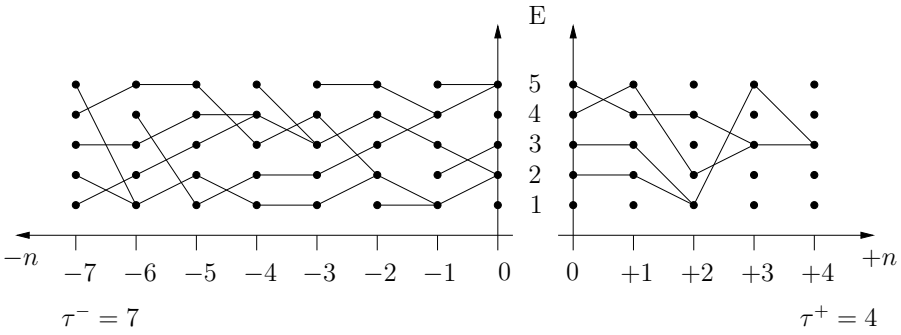


Figure 1. Backward and forward coupling

Thus, τ^+ is the first time at which the chains $\{X_n^0(i)\}_{n\geq 0}$, $1 \leq i \leq r$, coalesce.

Lemma 21.1.2 *When the updating is componentwise independent, the forward coupling time τ^+ is almost surely finite.*

Proof. Consider the (immediate) extension of Theorem 16.2.1 to the case of r independent HMC's with the same transition matrix. It cannot be applied directly to our situation, because the chains are not independent. However, the probability of coalescence in our situation is bounded below by the probability of coalescence in the completely independent case. To see this, first construct the independent chains model, using r independent IID componentwise independent updating sequences. The difference with our model is that we use too many updatings. In order to construct from this a set of r chains as in our model, it suffices to use for two chains the same updatings as soon as they meet. Clearly, the forward coupling time of the so modified model is smaller than or equal to that of the initial completely independent model. □

Let $\tau := \tau^-$. Let

$$Z = X_0^{-\tau}(i).$$

(This random variable is independent of i. In Figure 1, $Z = 2$.) Then,

Theorem 21.1.3 *With a componentwise independent updating sequence, the backwardward coupling time τ is almost surely finite. Also, the random variable Z has the distribution π.*

Proof. We shall show at the end of the current proof that for all $k \in \mathbb{N}$, $P(\tau \leq k) = P(\tau^+ \leq k)$, and therefore the finiteness of τ follows from that of τ^+ proven in the last lemma. Now, since for $n \geq \tau$, $X_0^{-n}(i) = Z$,

$$\begin{aligned} P(Z = j) &= P(Z = j, \tau > n) + P(Z = j, \tau \leq n) \\ &= P(Z = j, \tau > n) + P(X_0^{-n}(i) = j, \tau \leq n) \\ &= P(Z = j, \tau > n) - P(X_0^{-n}(i) = j, \tau > n) + P(X_0^{-n}(i) = j) \\ &= P(Z = j, \tau > n) - P(X_0^{-n}(i) = j, \tau > n) + p_{ij}(n) \\ &= A_n - B_n + p_{ij}(n)\,. \end{aligned}$$

But A_n and B_n are bounded above by $P(\tau > n)$, a quantity that tends to 0 as $n \uparrow \infty$ since τ is almost surely finite. Therefore

$$P(Z = j) = \lim_{n \uparrow \infty} p_{ij}(n) = \pi(j).$$

It remains to prove the equality of the distributions of the forwards and backwards coupling time. For this, select an arbitrary integer $k \in \mathbb{N}$. Consider an updating sequence constructed from a *bona fide* updating sequence $\{Y_n\}_{n \in \mathbb{Z}}$, by replacing $Y_{-k+1}, Y_{-k+2}, \ldots, Y_0$ by $Y_1, Y_2, \ldots, Y_k$. Call τ' the backwards coupling time in the modified model. Clearly τ and τ' have the same distribution.

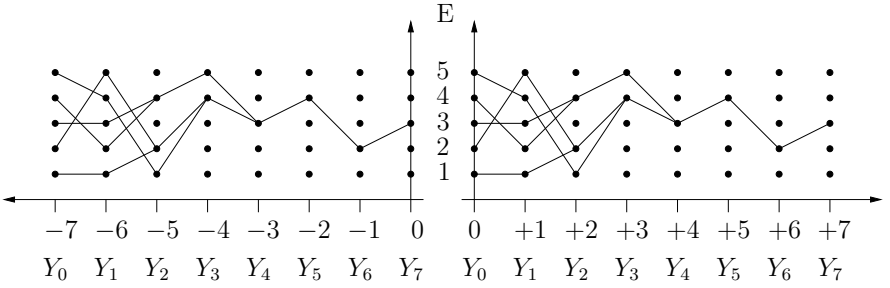


Figure 2. $\tau^+ \leq k$ implies $\tau' \leq k$

Suppose that $\tau^+ \leq k$. Consider in the modified model the chains starting at time $-k$ from states $1, \ldots, r$. They coalesce at time $-k + \tau^+ \leq 0$ (see Figure 2), and consequently $\tau' \leq k$. Therefore $\tau^+ \leq k$ implies $\tau' \leq k$, so that

$$P(\tau^+ \leq k) \leq P(\tau' \leq k) = P(\tau \leq k).$$

Now, suppose that $\tau' \leq k$. Then, in the modified model, the chains starting at time $k - \tau'$ from states $1, \ldots, r$ must at time $-k + \tau^+ \leq 0$ coalesce at time k. Therefore (see Figure 3), $\tau^+ \leq k$. Therefore $\tau' \leq k$ implies $\tau^+ \leq k$, so that

$$P(\tau \leq k) = P(\tau' \leq k) \leq P(\tau^+ \leq k).$$

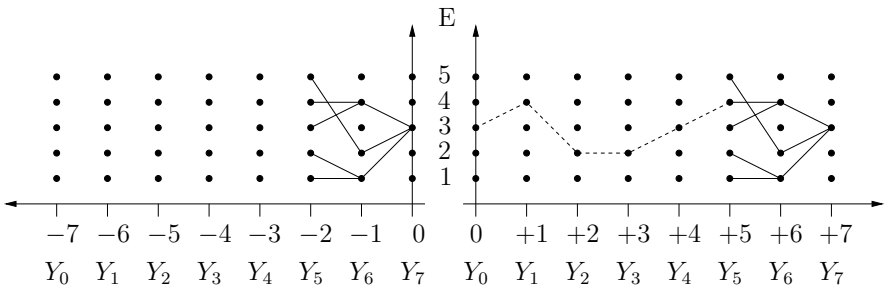


Figure 3. $\tau' \leq k$ implies $\tau^+ \leq k$

□

Note that the coalesced value at the forward coupling time is not a sample of π (see Exercise 21.4.1).

21.1.2 Sandwiching

The above exact sampling algorithm is often prohibitively time-consuming when the state space is large. However, if the algorithm required the coalescence of *two*, instead of r processes, then it would take less time. The Propp and Wilson algorithm does this in a special, yet not rare, case.

It is now assumed that there exists a partial order relation on E, denoted by $\preceq$, with a minimal and a maximal element (say, respectively, 1 and r), and that we can perform the updating in such a way that for all $i, j \in E$, all $N \in \mathbb{Z}$, all $n \geq N$,

$$i \preceq j \Rightarrow X_n^N(i) \preceq X_n^N(j).$$

However we do not require componentwise independent updating (but the updating vectors sequence remains IID). The corresponding sampling procedure is called the *monotone Propp–Wilson algorithm*.

Define the backwards *monotone* coupling time

$$\tau_m = \inf\{n \geq 1; X_0^{-n}(1) = X_0^{-n}(r)\}.$$

Theorem 21.1.4 *The monotone backwards coupling time τ_m is almost surely finite. Also, the random variable $X_0^{-\tau_m}(1)$ $(= X_0^{-\tau_m}(r))$ has the distribution π.*

Proof. We can use most of the proof of Theorem 21.1.3. We need only to prove independently that τ^+ is finite. This is the case because τ^+ is dominated by the first time $n \geq 0$ such that $X_n^0(r) = 1$, and the latter is finite in view of the recurrence assumption. □

Monotone coupling will occur with representations of the form (21.1) such that for all z,

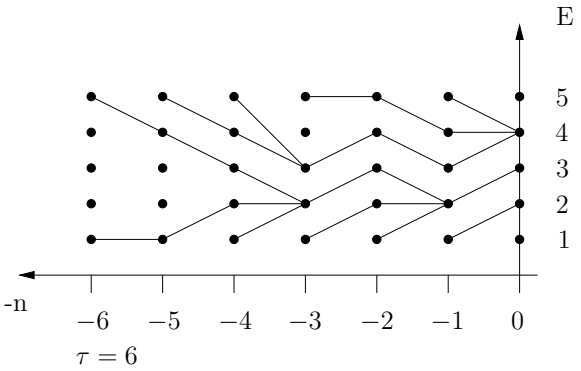


Figure 4. Monotone Propp–Wilson algorithm

$$i \preceq j \Rightarrow f(i,z) \preceq f(j,z),$$

and if for all $n \in \mathbb{Z}$, all $i \in \{1,\dots,r\}$,

$$Z_{n+1}(i) = Z_{n+1}.$$

EXAMPLE 21.1.5: A DAM MODEL. We consider the following model of a dam reservoir. The corresponding HMC, with values in $E = \{0,2,\dots,r\}$, satisfies the recurrence equation

$$X_{n+1} = \min\{r, \max(0, X_n + Z_{n+1})\},$$

where, as usual, $\{Z_n\}_{n\geq 1}$ is IID. In this specific model, X_n is the content at time n of a dam reservoir with maximum capacity r, and $Z_{n+1} = A_{n+1} - c$, where A_{n+1} is the input into the reservoir during the time period from n to $n+1$, and c is the maximum release during the same period. The updating rule is then monotone.

The Impatient Simulator

The average number of trials $E[\tau_-]$ needed for a naive use of the Propp–Wilson algorithm may be forbidding, and an impatient simulator could be tempted to fix a large value for the number of steps he is ready to perform before giving up, and start new attempts until he obtains coalescence within the precribed limit of time. This will introduce a bias (see Exercise 21.4.3). What else can we do to accelerate the procedure?

It is recommended that instead of trying the times -1, -2, etc., one uses successive restarting times of the form $\alpha^r T_0$. Let k be the first k for which $\alpha^k T_0 \geq \tau_-$. The number of simulation steps used is $2\left(T_0 + \alpha T_0 + \cdots + \alpha^k T_0\right)$ (the factor 2 accounts for the fact that we are running two chains), that is

$$2T_0\left(\frac{\alpha^{k+1}-1}{\alpha-1}\right) < 2T_0\left(\frac{\alpha^2}{\alpha-1}\right)\alpha^{k-1} \leq 2\tau_-\frac{\alpha^2}{\alpha-1}$$

steps, where we have assumed that $T_0 \leq \tau_-$. In the best case, supposing we are informed of the exact value of τ_- by some oracle, the number of steps is $2\tau_-$. The ratio of the worst to best cases is $\frac{\alpha^2}{\alpha-1}$, which is minimized for $\alpha = 2$. This is why one usually suggests to start the successive attempts of backward coalescence at times of the form $-2^k T_0$ $(k \geq 0)$.

We shall now relate the average backward recurrence time to the mixing time of the chain.

Let $(E, \preceq)$ be a partially ordered set. A subset A of E is called a chain if $(A, \preceq)$ is totally ordered. Denote by $\ell := \ell(E)$ the length of the longest chain. For instance, if $E = \{0,1\}^N$, and if $\preceq$ is defined by

$$(x_1, \ldots, x_N) \preceq (y_1, \ldots, y_N) \Longleftrightarrow x_i \leq y_i \, (1 \leq i \leq N) \, ,$$

$\ell = N$ (start with the maximal element $(1, \ldots, 1)$ and successively change the 1's into 0's until you reach the minimal state $(0, \ldots, 0)$).

Theorem 21.1.6

$$\frac{P(\tau_+ > k)}{\ell} \leq \overline{d}(k) \leq P(\tau_+ > k) \, .$$

Proof. Let $h(x)$ denote the length of the longest chain whose maximal element is x. In the example, it is the Hamming weight of x, that is, the number of 1's in it. If $X_0^k(1) \neq X_0^k(r)$, then $h(X_0^k(1)) + 1 \leq h(X_0^k(r))$, and if $X_0^k(1) = X_0^k(r)$, then $h(X_0^k(1)) \leq h(X_0^k(r))$. Therefore

$$1_{\{X_0^k(1) \neq X_0^k(r)\}} \leq h(X_0^k(r)) - h(X_0^k(1)) \, .$$

In particular,

$$P(\tau_+ > k) = P(X_0^k(1) \neq X_0^k(r)) \leq E\left[h(X_0^k(r)) - h(X_0^k(1))\right] \, .$$

Denoting by ρ_i^k the distribution $\delta_i^T \mathbf{P}^k$ of the chain with initial state i at time k

$$\begin{aligned} E\left[h(X_0^k(r)) - h(X_0^k(1))\right] &= E_{\rho_r^k}\left[h(X_0)\right] - E_{\rho_1^k}\left[h(X_0)\right] \\ &\leq d_V(\rho_r^k, \rho_1^k)\left(\max h(x) - \min h(x)\right) \leq \ell \overline{d}(k) \, . \end{aligned}$$

This proves the first inequality. For the second, observe that the event that two chains starting in arbitrary initial distributions μ and ν will disagree at time k implies that $\tau_+ > k$. Therefore $d_V(\mu\mathbf{P}^k, \nu\mathbf{P}^k) \leq P(\tau_+ > k)$ and the last inequality follows since $\overline{d}(k) := \sup_{\mu,\nu} d_V(\mu\mathbf{P}^k, \nu\mathbf{P}^k)$. □

The next theorem states that the function $k \to P(\tau_+ > k)$ is submultiplicative.

Theorem 21.1.7 *Let k_1 and k_2 be integer-valued random variables. Then*

$$P(\tau_+ > k_1 + k_2) \leq P(\tau_+ > k_1) P(\tau_+ > k_2) \, .$$

Proof. Exercise 21.4.4. □

Lemma 21.1.8

$$kP(\tau_+ > k) \le E[\tau_+] \le \frac{k}{P(\tau_+ \le k)}\,.$$

Proof. The first inequality is just Markov's inequality. By the telescope formula,

$$E[\tau_+]P(\tau_+ \ge 0) + P(\tau_+ \ge 1) + \cdots = \sum_{i=0}^{\infty} (P(\tau_+ > ik+1) + \cdots + P(\tau_+ > (i+1)k))$$

$$\le 1 + \sum_{i=0}^{\infty} kP(\tau_+ > ik)\,.$$

By submultiplicativity of $k \to P(\tau_+ > k)$, $P(\tau_+ > ik) \le P(\tau_+ > k)^i$. Therefore,

$$E[\tau_+] \le k \sum_{i=0}^{\infty} P(\tau_+ > k)^i = k \frac{1}{1 - P(\tau_+ > k)} = k \frac{1}{P(\tau_+ \le k)}\,.$$

□

Define the mixing time of the chain T_{mix} to be the first time k such that $\overline{d}(k) \le \frac{1}{e}$. Recall that $k \to \overline{d}(k)$ is submultiplicative, and therefore, after $k = T_{mix}(1 + \log \ell)$ steps,

$$\overline{d}(k) \le \overline{d}(T_{mix})^{1+\log \ell} = \left(\frac{1}{e}\right)^{1+\log \ell} \frac{1}{e \times e^{\log \ell}} = \frac{1}{e\ell}\,.$$

By Theorem 21.1.6,

$$P(\tau_+ > k) \le \overline{d}(k) \times \ell \le \frac{1}{e}\,.$$

Therefore, in view of Lemma 21.1.8

$$E[\tau_+] \le \frac{k}{P(\tau_+ \le k)} \le \frac{k}{1 - 1/e} \le 2k = 2T_{mix}(1 + \log \ell)\,.$$

Suppose we make m independent experiments, called the reference experiments, resulting in the IID forward coalescence time sequence $T_1, \ldots, T_m$. We now would like to have an idea of the time of forward coalescence τ_+ of the experiment (independent of the reference experiments) we are about to perform. By the submultiplicativity property of Theorem 21.1.7,

$$P(\tau_+ > T_1 + \ldots + T_m) \le P(\tau_+ > T_1) \times \cdots \times P(\tau_+ > T_m) \le P(\tau_+ > T_1)^m\,.$$

By symmetry, since τ_+ and T_1 are independent and identically distributed, $P(\tau_+ > T_1) = \frac{1}{2}$, and therefore

$$P(\tau_+ > T_1 + \ldots + T_m) \le \frac{1}{2^m}\,.$$

Recalling that the forward and backward coalescence times have the same distribution, we also have that

$$P(\tau_- > T_1 + \ldots + T_m) \le \frac{1}{2^m}\,.$$

21.2 Boltzmann Sampling

21.2.1 The Boltzmann Distribution

The goal in this section is similar to that of Section 21.1: to sample a given probability distribution. But now the distribution is simpler, since it is a uniform distribution over a certain class of elements of a given "size", for instance binary trees with a given number of vertices. In the previous section, one of the difficulties resided in the evaluation of the partition function. Here things are different because there exist efficient combinatorial methods for counting the elements of complex collection of objects. But there remains the resource consuming task of generating a uniform sample in a large set. As for Monte Carlo methods, the brute force method of the inverse is not applicable because of the precision required (the probabilities concerned are minuscule) and of the difficulty of encoding the objects in question. The Boltzmann sampling method exploits the classical combinatorial analysis methods that take advantage of the recursive description, when available, of the collection of objects in question.

Unlabeled Models

Consider a denumerable class $\mathcal{C}$ of "objects", the objects γ therein having a "size" denoted by $|\gamma|$. Let $\mathcal{C}_n \subseteq \mathcal{C}$ denote the collection of objects of size n and let $C_n := |\mathcal{C}_n|$ denote its cardinality, henceforth assumed finite for all n. A primary concern is uniform sampling of an object of $\mathcal{C}_n$ for predetermined n. That is, we seek a random device that generates a random element $Y \in \mathcal{C}_n$ such that $P(Y = \gamma) = \frac{1}{C_n}$ for all $\gamma \in \mathcal{C}_n$. For a start, we shall not be so ambitious. Instead, we shall provide a sampler that generates a random object $Y \in \mathcal{C}$ in such a way that for all n

$$P(Y = \gamma \mid |Y| = n) = \frac{1}{C_n}\,. \tag{21.2}$$

In other words, if the sampler produces an object of size n, it is selected uniformly among all the objects of size n. If we need a uniform sample of $\mathcal{C}_n$ for predetermined n, it suffices to do rejection sampling: produce samples until you find one of the required size n. Of course, the (random) number N of trials needed for this sample may be forbiddingly large. This important issue will be discussed later.

The family, indexed by $x \in \mathbb{R}_+\backslash\{0\}$, of probability distributions on $\mathcal{C}$

$$\pi_x(\gamma) = \frac{1}{C(x)} x^{|\gamma|} \qquad (\gamma \in \mathcal{C})\,, \tag{21.3}$$

where

$$C(x) = \sum_{\gamma \in \mathcal{C}} x^{|\gamma|}$$

is the normalizing constant, is called an ordinary Boltzmann model for $\mathcal{C}$. Since $\sum_{\gamma\in\mathcal{C}} \equiv \sum_n \sum_{\gamma\in\mathcal{C}_n}$, an alternative expression of the normalizing factor is

$$C(x) = \sum_n C_n x^n\,. \tag{21.4}$$

The function C is the ordinary generating function (OGF) of $\mathcal{C}$. The admissible values of the parameter x are those in $(0, R_C)$ where R_C is the radius of convergence of the series (21.4).

Two particular collections need to be distinguished right away since they intervene in most recursive descriptions of collections of objects. The "empty collection" $\mathcal{E}$ (with 0 element) and the collection $\mathcal{Z}$ with only one element, of respective OGF's $E(x) \equiv 1$ and $Z(x) = x$.

The distribution (21.3) is such that any random element with this distribution satisfies the requirement (21.2).

Proof. We have that

$$P_x(Y = \gamma, |Y| = n) = \frac{1}{C(x)} x^n$$

and

$$P_x(|Y| = n) = \frac{1}{C(x)} \sum_{\gamma \in \mathcal{C}, |\gamma| = n} x^n = \frac{1}{C(x)} \sum_{\gamma \in \mathcal{C}_n} x^n = \frac{1}{C(x)} C_n x^n \,,$$

and therefore

$$P_x(Y = \gamma \mid |Y| = n) = \frac{P_x(Y = \gamma, |Y| = n)}{P_x(|Y| = n)} = \frac{1}{C_n} \,.$$

□

Labeled Models

Consider a class $\mathcal{C}$ of objects such that any object of size n can be considered as an assembly of n atoms labeled by positive integers, with the restriction that there is no repetition of the labels. If the labels of an object of size n all belong to $\{1, 2, \ldots, n\}$, one says that the object is tightly (or well) labeled. Otherwise it is called a loosely (or weakly) labeled object. In the sequel, the following convention is adopted: the atoms are assumed distinguishable from one another, and a given labeling sequence refers to a fixed order of presentation of the atoms. For instance if the atoms of an object of size 4 are presented in the order A, B, C, D, the labeling $(3, 2, 4, 1)$ says that A has the label 3, B the label 2, C the label 4 and D the label 1.

EXAMPLE 21.2.1: CYCLIC PERMUTATIONS. Here $\mathcal{C} = \mathcal{K}$, the collection of all cyclic permutations, where the objects in $\mathcal{K}$ of size n are the cyclic permutations of $\{1, 2, \ldots, n\}$. A cyclic permutation of size n is represented by a collection of n points (the atoms) regularly spaced on the unit circle. These points are numbered from 1 to n. Reading the labels clockwise starting from any point gives a cyclic permutation. For instance, with $n = 4$, the atoms being called A,B,C,D and placed in the clockwise order on the circle, the labeling $3, 1, 2, 4$ gives the cyclic permutation $(3, 1, 2, 4)$. Note that the cyclic permutations are well labeled.

One may conventionally describe a cyclic permutation by choosing to start systematically from the atom labeled 1, here B, which gives $(1, 2, 4, 3)$, the same cyclic permutation as $(3, 1, 2, 4)$. With this convention, it is immediate to count the number of cyclic permutations of size n: $K_n = (n-1)!$. Unfortunately, the series $\sum_n (n-1)! x^n$ is divergent for all $x > 0$.

The just mentioned OGF divergence problem is the principal motivation for the following new definition, that of an exponential Boltzmann sampler, for which the probability of drawing the object $\gamma \in \mathcal{C}$ is now

$$\hat{\pi}_x(\gamma) = \frac{1}{\hat{C}(x)} \frac{x^{|\gamma|}}{|\gamma|!} \qquad (\gamma \in \mathcal{C}), \tag{21.5}$$

where

$$\hat{C}(x) = \sum_{\gamma \in \mathcal{C}} \frac{x^{|\gamma|}}{|\gamma|!}$$

is the normalizing constant. Alternatively, since $\sum_{\gamma \in \mathcal{C}} \equiv \sum_n \sum_{\gamma \in \mathcal{C}_n}$,

$$\hat{C}(x) = \sum_n \frac{C_n}{n!} x^n . \tag{21.6}$$

The function $\hat{C}$ is called the exponential generating function (EGF) of $\mathcal{C}$. The admissible values of the parameter x are those in $(0, \hat{R}_C)$ where $\hat{R}_C$ is the radius of convergence of the series (21.6). In the example of cyclic permutations,

$$\hat{K}(x) = \sum_n (n-1)! \frac{x^n}{n!} = \sum_n \frac{x^n}{n} = \log\left(\frac{1}{1-x}\right)$$

when $0 < x < \hat{R}_K = 1$.

As in the labeled case, one verifies that the distribution (21.5) satisfies requirement (21.2). Indeed,

$$P_x(Y = \gamma, |Y| = n) = \frac{1}{\hat{C}(x)} \frac{x^n}{n!}$$

and

$$P_x(|Y| = n) = \frac{1}{\hat{C}(x)} \sum_{\gamma \in \mathcal{C}, |\gamma| = n} \frac{x^n}{n!} = \frac{1}{\hat{C}(x)} \sum_{\gamma \in \mathcal{C}_n} \frac{x^n}{n!} = \frac{1}{\hat{C}(x)} C_n \frac{x^n}{n!} ,$$

and therefore

$$P_x(Y = \gamma \mid |Y| = n) = \frac{\hat{P}_x(Y = \gamma, |Y| = n)}{\hat{P}_x(|Y| = n)} = \frac{1}{C_n} .$$

21.2.2 Recursive Implementation of Boltzmann Samplers

We are now going to describe, for a variety of classes $\mathcal{C}$ of objects, a specific operation that delivers a random variable $Y \in \mathcal{C}$ distributed according to the distribution π_x. This operation, called the Boltzman sampler of $\mathcal{C}$ will be denoted by $BS_x(\mathcal{C})$.

Unlabeled Samplers

Some interesting collections of objects are built either by means of elementary operations (disjoint union, cartesian product, ...) on simpler collections $\mathcal{A}, \mathcal{B}, \ldots$, or via a recursive process. The object of this section is to construct the corresponding Boltzmann samplers from the Boltzmann samplers of the collections intervening in their construction.

Let $\mathcal{A}$ and $\mathcal{B}$ be two denumerable collections of objects such that A_n and B_n are finite for all n.

Disjoint union Let $\mathcal{C} := \mathcal{A} + \mathcal{B}$ where $\mathcal{A}$ and $\mathcal{B}$ are considered disjoint (that is, if the "same" object belongs to both collections, it appears twice in $\mathcal{C}$). We have that $C_n = A_n + B_n$, and therefore

$$C(x) = A(x) + B(x)\,.$$

An element γ in $\mathcal{C}$ is either an element $\alpha \in \mathcal{A}$ or an element $\beta \in \mathcal{B}$. The probability that a random element Y selected according to the probability distribution (21.3) is $\alpha \in \mathcal{A}$ is

$$\pi_{\mathcal{C},x}(\alpha) = \frac{x^{|\alpha|}}{A(x)+B(x)} = \frac{x^{|\alpha|}}{A(x)}\left(\frac{A(x)}{A(x)+B(x)}\right)$$

with a similar expression for $\pi_{\mathcal{C},x}(\gamma)$. In particular, the probability that a random element Y is in $\mathcal{A}$ (resp., $\mathcal{B}$) is

$$\pi_{\mathcal{C},x}(\mathcal{A}) = \frac{A(x)}{A(x)+B(x)} \qquad (\text{resp., } \pi_{\mathcal{C},x}(\mathcal{B}) = \frac{B(x)}{A(x)+B(x)})\,.$$

Therefore the Boltzmann model for $\mathcal{C}$ is the mixture of the Boltzmann models for $\mathcal{A}$ and $\mathcal{B}$ with respective mixing probabilities $\frac{A(x)}{A(x)+B(x)}$ and $\frac{B(x)}{A(x)+B(x)}$. Denote by $BS_x(\mathcal{C})$ the Boltzmann sampler for $\mathcal{C}$, that is, the operation that delivers a random $Y \in \mathcal{C}$ according to the Boltzmann distribution (21.3). In this case, it can be decomposed as follows: Draw a Bernoulli variable, equal to 1 with probability $\frac{A(x)}{A(x)+B(x)}$, and to 0 with probability $\frac{B(x)}{A(x)+B(x)}$. If 1, call $BS_x(\mathcal{A})$, else, call $BS_x(\mathcal{B})$. This is symbolized by

$$BS_x(\mathcal{C}) = \left(\text{Bern } \left(\frac{A(x)}{A(x)+B(x)}\right) \longrightarrow BS_x(\mathcal{A}) \mid BS_x(\mathcal{B})\right)\,. \tag{21.7}$$

In general, the notation

$$BS_x(\mathcal{C}) = (\text{Bern } (p_1, p_2, \cdots, p_k) \longrightarrow BS_x(\mathcal{A}_1) \mid \cdots \mid BS_x(\mathcal{A}_k))$$

tells us that the Boltzmann generator of $\mathcal{C}$ consists of two steps: (1) choose an index $i \in \{1, 2, \ldots, k\}$ with probabiliy p_i, and (2) call the Boltzmann generator of $\mathcal{A}_i$. To make this notation consistent with that used in (21.7), we add the notational convention Bern $(p, 1-p) \equiv$ Bern (p).

Cartesian product Consider now the collection $\mathcal{C} := \mathcal{A} \times \mathcal{B}$ of ordered pairs from $\mathcal{A}$ and $\mathcal{B}$. Take for a size function

$$|\gamma| = |(\alpha, \beta)| := |\alpha| + |\beta| .$$

We have

$$C(x) = \sum_{\alpha,\beta} x^{|\alpha|+|\beta|} = \sum_{\alpha,\beta} x^{|\alpha|}\, x^{|}\beta| = \left(\sum_{\alpha} x^{|\alpha|}\right)\left(\sum_{\beta} x^{|}\beta|\right) .$$

Therefore

$$C(x) = A(x)B(x)$$

and

$$\pi_{\mathcal{C},x}((\alpha,\beta)) = \frac{x^{|\alpha|}}{A(x)}\frac{x^{|\beta|}}{B(x)} = \pi_{\mathcal{A},x}((\alpha))\pi_{\mathcal{B},x}((\beta)) .$$

The Boltzmann sampler for $\mathcal{C}$ therefore calls independently the Boltzmann samplers for $\mathcal{A}$ and $\mathcal{B}$, which give respectively the values α and β, and produces the value $\gamma = (\alpha, \beta)$. This is symbolized by

$$BS_x(\mathcal{C}) = (BS_x(\mathcal{A})\,;\, BS_x(\mathcal{B})) .$$

The next collection of objects can be defined as a sum, or recursively.

Sequences Let now $\mathcal{C} := \mathcal{A}^*$ be the collection of finite sequences of elements from $\mathcal{A}$ (including the empty sequence). It is represented by the symbolic recursive equation

$$\mathcal{C} = \mathcal{E} + \mathcal{A} \times \mathcal{C} . \tag{21.8}$$

According to the rule for disjoint union and cartesian product, the recursive description (21.8) gives for the generating function the equation $C = 1 + AC$, so that

$$C(x) = \frac{1}{1 - A(x)} .$$

Applying the rules for the sum $\mathcal{A}_1 + \mathcal{B}_1$ where $\mathcal{A}_1 = \mathcal{E}$ and $\mathcal{A}_2 = \mathcal{A} \times \mathcal{C}$ of respective generating functions $A_1(x) = E(x) = 1$ and $A_2(x) = A(x)C(x) = \frac{A(x)}{1-A(x)}$, we have

$$BS_x(\mathcal{C}) = \left(\text{Bern } \left(\frac{A_1(x)}{A_1(x) + A_2(x)}\right) \longrightarrow BS_x(\mathcal{A}_1) \mid BS_x(\mathcal{A}_2)\right) ,$$

or, since $\frac{A_1(x)}{A_1(x)+A_2(x)} = 1 - A(x)$,

$$BS_x(\mathcal{C}) = (\text{Bern } (1 - A(x)) \longrightarrow BS_x(\mathcal{E}) \mid BS_x(\mathcal{A} \times \mathcal{C})) .$$

This means that with probability $1 - A(x)$, the Boltzmann sampler outputs the empty sequence and stops, and that with probability $A(x)$, it calls $BS_x(\mathcal{A}\times\mathcal{C})$. The Boltzmann sampler $BS_x(\mathcal{A} \times \mathcal{C})$ issues a random element of $\mathcal{A}$ selected according to the corresponding Boltzmann distribution concatenated with a random element of $\mathcal{C}$. This is clearly a recursive process. It will eventually stop, namely when a call of $BS_x(\mathcal{C})$ results in the empty sequence.

Note that the recursive process terminates in a geometric time of mean $\frac{1}{1-A(x)}$, and therefore the Boltzmann generator for a finite sequence of elements of $\mathcal{A}$ can be

described as follows: Draw a geometric random variable of parameter $A(x)$, and given the value k of this variable, produce k successive independent elements of $\mathcal{A}$ sampled from the Boltzmann distribution associated with $\mathcal{A}$. This is symbolized by

$$BS_x(\mathcal{A}^*) = (\text{Geom } (A(x)) \longrightarrow BS_x(\mathcal{A})) \ .$$

This notation is a particular case of the following one

$$(Y \longrightarrow BS_x(\mathcal{A}))$$

which means that if the value of the integer-valued random variable Y is k, then k independent calls of the Boltzmann sampler for $\mathcal{A}$ are done.

We have just seen how a recursive definition of a collection of objects gives a corresponding recursive sampling procedure. The Boltzmann samplers of collections of objects that are defined via elementary operations and recursive procedures can be constructed using the corresponding elementary operations and recursive procedures.

EXAMPLE 21.2.2: BINARY TREES. A recursive representation of the collection $\mathcal{B}$ of finite binary trees is as follows

$$\mathcal{B} = \mathcal{Z} + (\mathcal{Z} \times \mathcal{B} \times \mathcal{B}) \, ,$$

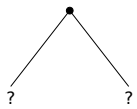
where $\mathcal{Z}$ is the collection consisting of a single object, here the tree with just one vertex. The interpretation is as follows. A binary tree consists, either of a single vertex, or of a vertex with two branches stemming from it and at the extremities of which are two finite binary trees. The size of a binary tree is the number of its vertices. Therefore $Z(x) = x$ and $B(x) = x + xB(x)^2$, that is,

$$B(x) = \frac{1 - \sqrt{1 - 4x^2}}{2x} \, .$$

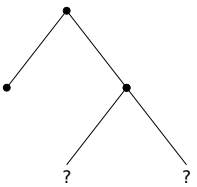
(The other choice of the root leads to $C(0) = C_0 = \infty$.) The Boltzmann sampler therefore has the symbolic form

$$BS_x(\mathcal{B}) = \left(\text{Bern } \left(\frac{x}{B(x)} \right) \longrightarrow \mathcal{Z} \mid (\mathcal{Z} \, ; \, BS_x(\mathcal{B}) \, ; \, BS_x(\mathcal{B})) \right) \, .$$

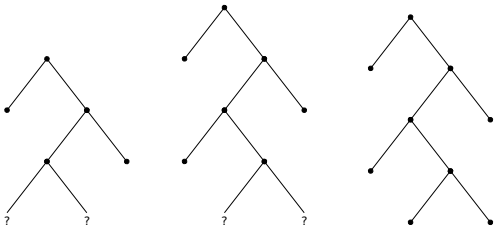
The following figures show in a special case the succession of operations. The first call of the sampler resulted (probability $1 - \frac{x}{B(x)}$) in an element of $(\mathcal{Z} \, ; \, BS_x(\mathcal{B}) \, ; \, BS_x(\mathcal{B}))$, that is, a point (the element of $\mathcal{Z}$) plus two indeterminate trees each represented in the figure by a question mark (?).



Each question mark will be replaced by a finite binary tree, each time obtained by calling $BS_x(\mathcal{B})$. The next figure shows that the ? on the left has been replaced by a vertex, and the ? on the right by a vertex and two indeterminate trees.



The process continues until there are no indeterminate trees left.



Remark 21.2.3 It is clear at this point that a by-product of the Boltzmann sampling method is the obtention of the generating function of the sequence C_n ($n \geq 1$) from which the sequence itself can, at least theoretically, be extracted. However, historically, the method of combinatorial analysis exploiting the recursive description of a collection of objects in order to compute its cardinality came first.

Labeled Samplers

Disjoint sum For the disjoint union $\mathcal{C} = \mathcal{A} + \mathcal{B}$ the same computations as those of the unlabeled case lead to

$$\hat{C}(x) = \hat{A}(x) + \hat{B}(x).$$

and therefore

$$BS_x(\mathcal{C}) = \left(\text{Bern} \left(\frac{\hat{A}(x)}{\hat{A}(x) + \hat{B}(x)} \right) \longrightarrow BS_x(\mathcal{A}) \mid BS_x(\mathcal{B}) \right). \tag{21.9}$$

Labeled product We shall now define the labeled product $\mathcal{A} \star \mathcal{B}$ of two labeled collections $\mathcal{A}$ and $\mathcal{B}$. It is the union

$$\bigcup_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}} \alpha \star \beta \,,$$

where, by definition, $\alpha \star \beta$ is the set of order-consistent relabelings of the cartesian product (α, β). An example will perhaps best explain the notion of order-consistent relabeling.

Suppose that α and β are of sizes k and m respectively. The size of (α, β) is therefore $n = k + m$. Take for instance $k = 3$ and $m = 2$. The object α consists

of 3 atoms, call them BLUE, WHITE and RED, which receive labels from 1 to 3, say BLUE= 3, WHITE= 1 and RED= 2, so that the corresponding object is well labeled by the sequence $(3, 2, 1)$. Similarly, the object β contains 2 atoms, say MOUSE and DUCK, which receive labels (numbers) from 1 to 2, say MOUSE= 1 and DUCK= 2, so that the corresponding object is well labeled by the sequence $(1, 2)$. Now the object (α, β) consists of 5 objects, BLUE, WHITE, RED, MOUSE and DUCK that must be well labeled from 1 to 5. This implies a relabeling of the basic atoms. This relabeling is said to be order-consistent (with the original labeling) if it respects inside of each element α and β the relative order of the initial labeling. For instance, the objects appearing in the same order as previously, that is BLUE, WHITE, RED, MOUSE and DUCK, the label $(5, 4, 1)(2, 3)$ is order-consistent with the original labeling $(3, 2, 1)(1, 2)$, whereas $(1, 3, 4)(5, 2)$ is not.

It is important to note that the elements of $\mathcal{A} \star \mathcal{B}$ are of the same nature as those of $\mathcal{A} \times \mathcal{B}$. They are not of the form $\alpha \star \beta$ as the notation unfortunately wrongly suggests.

There are exactly $\binom{k+m}{k} = \binom{n}{k}$ consistent relabelings, and therefore the number of objects of size n in $\mathcal{C} = \mathcal{A} \star \mathcal{B}$ is

$$C_n = \sum_{k=0}^{n} \binom{n}{k} A_k B_{n-k} \,.$$

In particular, an elementary computation shows that

$$\hat{C}(x) = \hat{A}(x)\hat{B}(x) \,.$$

The probability of drawing an element from $\mathcal{A} \star \mathcal{B}$ of size $|\gamma| = |\alpha| + |\beta|$ is

$$\begin{aligned} \frac{x^{|\gamma|}}{|\gamma|!\hat{C}(x)} &= \frac{x^{|\alpha|}}{|\alpha|!\hat{A}(x)} \times \frac{x^{|\beta|}}{|\beta|!\hat{B}(x)} \times \frac{|\alpha|!|\beta|!}{(|\alpha| + |\beta|)!} \\ &= \frac{x^{|\alpha|}}{|\alpha|!\hat{A}(x)} \times \frac{x^{|\beta|}}{|\beta|!\hat{B}(x)} \times \frac{1}{\frac{(|\alpha|+|\beta|)!}{|\alpha|!|\beta|!}} \,. \end{aligned}$$

This reads as follows: randomly select an element $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$, by selecting independently $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$ according to the respective distributions $\frac{x^{|\alpha|}}{|\alpha|!\hat{A}(x)}$ and $\frac{x^{|\beta|}}{|\beta|!\hat{B}(x)}$, an operation that is symbolized by

$$BS_x(\mathcal{A} \times \mathcal{B}) = (BS_x(\mathcal{A})\,;\, BS_x(\mathcal{B}))\,,$$

and then select randomly and uniformly an element (α', β') in $\alpha \star \beta$ (whose cardinality is, remember, $\frac{(|\alpha|+|\beta|)!}{|\alpha|!|\beta|!}$).

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be labeled collections. The associativity property

$$\mathcal{A} \star (\mathcal{B} \star \mathcal{C}) = (\mathcal{A} \star \mathcal{B}) \star \mathcal{C}$$

is easily checked, and both sides define the labeled collection $\mathcal{A} \star \mathcal{B} \star \mathcal{C}$. This definition is extended to an arbitrary number of collections: $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_k$.

Calculations similar to those of the case $k = 2$ give for the Boltzmann sampler of $\mathcal{A}_1 \star \mathcal{A}_2 \star \cdots \star \mathcal{A}_k$ the following two-step procedure. First obtain the individual samples of $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_k$:

$$BS_x(\mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_k) = (BS_x(\mathcal{A}_1)\,;\, BS_x(\mathcal{A}_2)\,;\, \ldots\,;\, BS_x(\mathcal{A}_k))\,,$$

which gives an element $(\alpha_1, \alpha_2, \ldots \alpha_k)$ of $\mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_k$, and obtain a random element of $(\alpha_1 \star \alpha_2 \star \cdots \star \alpha_k)$ by a random order-consistent relabeling chosen uniformly among all the $\frac{|\alpha_1|+\cdots+|\alpha_k|}{|\alpha_1|!\cdots|\alpha_k|!}$ order-consistent relabelings.

This random order-consistent relabeling is the final phase of Boltzmann sampling and its mention is generally omitted.

Labeled sequences The k-th labeled power of $\mathcal{A}$ is, by definition, the labeled product $\mathcal{A} \star \mathcal{A} \star \cdots \star \mathcal{A}$ (k times). It is also denoted by $\mathcal{A}^{\star k}$. An element thereof is called a labeled k-sequence of elements of $\mathcal{A}$. The collection of labeled k-sequences of elements of $\mathcal{A}$ will be denoted by $Set_k(\mathcal{A})$.

Denote by $Set(\mathcal{A})$ the collection of sequences of elements from the labeled collection $\mathcal{A}$:

$$Set(\mathcal{A}) = \mathcal{E} + \mathcal{A} + \mathcal{A} \star \mathcal{A} + \mathcal{A} \star \mathcal{A} \star \mathcal{A} + \cdots\,,$$

where $\mathcal{E}$ is the empty collection. Therefore, $Set(\mathcal{A})$ is the solution of the recursive equation $\mathcal{C} = \mathcal{E} + \mathcal{A} \star \mathcal{C}$. Application of the union rule and the labeled product rules yields for the EGF of the collection of labeled sequences the expression

$$\hat{C}(x) = \sum_{k=0}^{\infty} \hat{A}(x)^k = \frac{1}{1 - \hat{A}(x)}\,.$$

In particular, the same samplers as in the unlabeled case apply, in recursive form:

$$BS_x(\mathcal{C}) = \Big(\text{Bern}\ \Big(1 - \hat{A}(x)\Big) \longrightarrow BS_x(\mathcal{E}) \mid BS_x(\mathcal{A} \times \mathcal{C})\Big)\,,$$

or in non-recursive form:

$$BS_x(\mathcal{A}^*) = \Big(\text{Geom}\ (\hat{A}(x)) \longrightarrow BS_x(\mathcal{A})\Big)\,.$$

If the value of the geometric random variable is k, this gives an element $(\alpha_1, \ldots, \alpha_k) \in \mathcal{A}^k$, and a random element from $\alpha_1 \star \cdots, \star\alpha_k$ has to be drawn uniformly among the order-consistent relabelings of $(\alpha_1, \ldots, \alpha_k)$).

Labeled sets Informally, a labeled k-set of elements of the labeled collection $\mathcal{A}$ is a labeled k-sequence of elements of $\mathcal{A}$, but two k-sequences that differ only by a permutation of their components represent the same labeled k-set. Formally, the collection $Set_k(\mathcal{A})$ of labeled k-sets of elements of the labeled collection $\mathcal{A}$ is the quotient of $Seq_k(\mathcal{A})$ by the equivalence relation that identifies two k-sequences that differ only by a permutation of their components. The EGF of $\mathcal{C} = Seq_k(\mathcal{A})$ is

$$\hat{C}_k(x) = \sum_n C_{k,n} \frac{x^n}{n!}$$

where $C_{k,n}$ is the number of k-sets of size n. Since a labeled k-set is associated with exactly $k!$ k-sequences, $C_{k,n} = \frac{1}{k!}\widetilde{C}_{k,n}$ where $\widetilde{C}_{k,n}$ is the number of k-sequences of size n. Therefore

$$\hat{C}_k(x) = \sum_n \frac{1}{k!}\widetilde{C}_{k,n}\frac{x^n}{n!}\,,$$

that is:

$$\hat{C}_k(x) = \frac{1}{k!}\hat{A}(x)^k\,.$$

The collection of all labeled k-sets of elements of the labeled collection $\mathcal{A}$ is

$$Set(\mathcal{A}) := \mathcal{E} + \mathcal{A} + Set_2(\mathcal{A}) + \cdots = \bigcup_{k\geq 0} Set_k(\mathcal{A})\,.$$

The EGF of $\mathcal{C} = Set(\mathcal{A})$ is therefore

$$\hat{C}(x) = \sum_{k\geq 0}\frac{1}{k!}\hat{A}(x)^k = e^{\hat{A}(x)}\,,$$

and the probability that a set in the collection $\mathcal{C} := Set(\mathcal{A})$ has k components is therefore

$$\frac{1}{\hat{C}(x)}\sum_n C_{k,n}\frac{x^n}{n!} = \frac{1}{\hat{C}(x)}\frac{1}{k!}\hat{A}(x)^k = e^{-\hat{A}(x)}\frac{\hat{A}(x)^k}{k!}\,.$$

The Boltzmann sampler of $\mathcal{C} := Set(\mathcal{A})$ is therefore

$$BS_x(\mathcal{C}) = \Big(Poi(\hat{A}(x)) \longrightarrow BS_x(\mathcal{A})\Big)\,.$$

It will be convenient to introduce the collection

$$Set_{\geq m}(\mathcal{A}) := \bigcup_{k\geq m} Set_k(\mathcal{A})$$

where m is a positive integer. The EGF of $\mathcal{C} = Set_{\geq m}(\mathcal{A})$ is

$$\hat{C}(x) = \sum_{k\geq m}\frac{1}{k!}\hat{A}(x)\,,$$

and the probability that a set in the collection $\mathcal{C} := Set_{\geq m}(\mathcal{A})$ has k components is therefore

$$\frac{\frac{1}{k!}\hat{A}(x)}{\sum_{k\geq m}\frac{1}{k!}\hat{A}(x)}\,.$$

This is the distribution of a Poisson random variable Y of parameter $\lambda = \hat{A}(x)$ conditioned by $Y \geq m$, denoted $Poi_{\geq m}(\lambda)$. The corresponding Boltzmann sampler is

$$BS_x(\mathcal{C}) = \Big(Poi_{\geq m}(\hat{A}(x)) \longrightarrow BS_x(\mathcal{A})\Big)\,,$$

where $Poi_{\geq m}(\lambda)$ represents the distribution of a Poisson random variable Y of parameter λ conditioned by $Y \geq m$.

Labeled cycles Informally, a labeled k-cycle of elements of the labeled collection $\mathcal{A}$ is a labeled k-sequence of elements of $\mathcal{A}$, but two k-sequences that differ only by a cyclic shift of their elements represent the same k-cycle. This definition is analogous to that of labeled k-sets. The collection of labeled k-cycles is denoted by $Cyc_k(\mathcal{A})$, and the collection of labeled cycles by

$$Cyc(\mathcal{A}) := \mathcal{E} + \mathcal{A} + Cyc_2(\mathcal{A}) + \dots .$$

Exercise 21.4.10 asks you to find the corresponding Boltzmann generator.

From the elementary operations above, one can construct more complex objects. The relevant notion is that of specifiability.

Definition 21.2.4 *A* specifiable *labeled collection of objects is one that can be finitely specified, possibly in a recursive way, from finite sets by means of disjoint unions, Cartesian products as well as the sequence, set and cycle constructions.*

EXAMPLE 21.2.5: SURJECTIONS. A surjection from a finite non-empty set A to a finite non-empty set B is a mapping $\phi : A \to B$ such that $\phi(A) = B$, that is each element of B is the image by ϕ of at least one element of A. Identifying A and B with $\{1, 2, \dots, n\}$ and $\{1, 2, \dots, r\}$ respectively, where $n \geq r \geq 1$, a surjection ϕ from A to B (called a r-surjection since the range A has cardinality r) can be represented by the ordered r-tuple of subsets of $\{1, 2, \dots, n\}$

$$(\phi^{-1}(1), \phi^{-1}(2), \dots, \phi^{-1}(r)) .$$

For instance, the 4-surjection ϕ defined by

$$\phi(1) = 3, \phi(2) = 1, \phi(3) = 2, \phi(4) = 1, \phi(5) = 3, \phi(6) = 4, \phi(7) = 4, \phi(8) = 1, \phi(9) = 2 ,$$

is represented by the following 4-tuple of subsets of $\{1, 2, \dots, 9\}$

$$(\{2, 4, 8\}, \{3, 9\}, \{1, 5\}, \{6, 7\}) .$$

The collection of all r-surjections is therefore

$$\mathcal{R}^{(r)} = Seq_r(Set_{\geq 1}(\mathcal{Z})) .$$

In particular its EGF is

$$R^{(r)}(x) = (e^x - 1)^r$$

whose coefficients are given by

$$R_n^{(r)} = \sum_{j=1} r \binom{r}{j} (-1)^j (r - j)^n .$$

An r-partition of the non-empty set $A = \{1, 2, \dots, n\}$ is a partition of A in r non-empty subsets, called blocks. The collection $\mathcal{S}^r$ of r-partitions is

$$\mathcal{S}^r = Set_r(Set_{\geq 1}(\mathcal{Z})),.$$

The relation between an r-partition and an r-surjection is the following: an r-partition corresponds to a group of $r!$ r-surjections deriving from one another by a permutation of the r possible values. The EGF of $\mathcal{S}^r$ is therefore obtained by dividing by $r!$ the EGF of $\mathcal{R}^r$:

$$S^{(r)}(x) = \frac{1}{r!}(e^x - 1)^r\,.$$

EXAMPLE 21.2.6: PARTITIONS. The collection $\mathcal{S}$ of all set partitions is defined formally by

$$\mathcal{S} := \mathcal{E} + \mathcal{S}^1 + \mathcal{S}^2 + \cdots \mathcal{S}^n + \ldots$$

or, equivalently,

$$\mathcal{S} = Set\left(Set_{\geq 1}(\mathcal{Z})\right)\,,$$

where $\mathcal{Z}$ is the collection consisting of a unique object, of size 1, and $\mathcal{E}$ is the "empty partition", the unique partition of $\varnothing$. The generating function of $\mathcal{S}$ is

$$\hat{S}(x) = \sum_{r=1}^{\infty} \frac{1}{r!}(e^x - 1)^r = e^{e^x - 1}\,.$$

The corresponding generator consists (the detailed proof is asked in Exercise 21.4.9) of the following procedure in three steps. First choose the number K of blocks according to a Poisson distribution of parameter $e^x - 1$. Then, if $K = k$, draw k independent variables $Y_1, \ldots, Y_k$ from a Poisson distribution conditioned by ≥ 1 which represent the sizes of the blocks. This gives the shape of the partition, that is, the number and respective sizes of the block. The size of the set obtained being n, a random permutation of the n atoms completes the construction (this is the usual final random order-consistent relabeling).

We quote the following result, which is one of the key advantages of Boltzmann sampling.

Theorem 21.2.7 *(Duchon, Flajolet, Louchard and Schaeffer, 2004) Let $\mathcal{C}$ be a specifiable labeled collection with an* EGF *of convergence radius $\hat{R}_C$. Assume as given an oracle that provides the finite collection of exact values at any $x \in (0, \hat{R}_C)$ of the* EGF*'s intervening in the specification of $\mathcal{C}$. Then the Boltzmann generator of $\mathcal{C}$ has a complexity (measured in terms of the number of usual real-arithmetic operations: add, substract, multiply and divide) that is linear in the size of the output object.*

The general proof is omitted as the conclusion can in many examples be verified directly. Of course a similar result is true for unlabeled collections.

21.2.3 Rejection Sampling

We now turn to the initial problem, that of sampling uniformly an object of a given size. Some preliminary computations concerning the random variable $N := |Y|$ representing the size of the object selected from the Boltzmann distribution. The generating function of this random variable is

$$\sum_n P_x(N=n)z^n = \sum_n \frac{1}{C(x)} C_n x^n z^n = \frac{C(zx)}{C(x)}\,.$$

Straighforward computations based on the identities

$$E_x[N] = \left(\frac{\partial}{\partial z}\frac{C(zx)}{C(x)}\right)_{z=1},\ E_x[N(N-1)] = \left(\frac{\partial^2}{\partial z^2}\frac{C(zx)}{C(x)}\right)_{z=1}$$

give

$$E_x[N] = x\frac{C'(x)}{C(x)},\ E_x[N^2] = \frac{xC'(x)+x^2C''(x)}{C(x)}\,.$$

Also, it can be readily checked that

$$V_x(N) := \mathrm{Var}(N) = x\frac{d}{dx}E_x[N]\,.$$

In particular, $x \to E_x[N]$ is a strictly increasing function of the parameter x if $\mathcal{C}$ does not reduce to some $\mathcal{C}_{n_0}$.

Approximate Size Sampling Performance

If the objective is to sample uniformly $\mathcal{C}_n \subset \mathcal{C}$, the collection of objects of size n, for a given size n or sizes near n, it seems natural to tune the Boltzmann sampler to a value x_n such that

$$E_{x_n}[N] = n\,,$$

or equivalently, x_n is a root in $(0, R_C)$ of

$$n = x\frac{C'(x)}{C(x)}\,. \tag{21.10}$$

The function $x \in (0, R_C) \to E_x[N]$ is a strictly increasing function and we shall assume that

$$\lim_{x\uparrow R_C-} E_x[N] = \infty\,, \tag{21.11}$$

which guarantees the existence of x_n for all n. This choice of the parameter will lead to samples that have an average size n, but the variance may be large and rejection sampling may require a forbidding number of trials until one finds a sample of size exactly n.

Suppose now that we accept a relative tolerance ε, in other words that we are happy with values of N that lie in the interval $[n(1-\varepsilon), n(1+\varepsilon)]$. By Markov's inequality,

$$P_x(|N-n| \geq n\varepsilon) \leq \frac{E\left[(N-n)^2\right]}{n^2\varepsilon^2}\,.$$

Plugging in the value $x = x_n$ for which $n = E_x[N]$, this inequality reads

$$P_x(n(1-\varepsilon) \leq N \leq n(1+\varepsilon)) \leq \frac{V_x(N)}{E_x[N]}\frac{1}{\varepsilon^2}\,.$$

Therefore:

Theorem 21.2.8 *(Duchon, Flajolet, Louchard and Schaeffer, 2004) Under condition (21.11) and*

$$\lim_{x\uparrow R_C-} \frac{V_x(N)}{E_x[N]} = 0\,, \tag{21.12}$$

we have that for any $\varepsilon > 0$, with x_n a solution of (21.10), the probability of obtaining a sample of size in $[n(1-\varepsilon), n(1+\varepsilon)]$ tends to 1 as $n \uparrow \infty$.

EXAMPLE 21.2.9: PARTITIONS. For the class $\mathcal{S}$ of partitions, $S(x) = e^{e^x-1}$ is an entire function ($R_C = \infty$) and

$$E_x[N] = xe^x \text{ and } V_x(N) = x(x+1)e^x\,,$$

and therefore conditions (21.11) and (21.12) are satisfied. Since x_n is determined by the implicit equation $n = x_n e^{x_n}$, $x_n \sim \log n - \log\log n$ and the variance $V_{x_n}(N) \sim \sqrt{n\log n}$. Therefore, the probability that a sample falls outside of the tolerance interval $[n(1-\varepsilon), n(1+\varepsilon)]$ is smaller than a quantity equivalent to $\frac{1}{\varepsilon^2}\sqrt{\frac{\log n}{n}}$.

Exact Size Sampling Performance

The following result of analysis is stated without proof.

Suppose that the generating function C satisfies conditions (21.11) and (21.11), and has in addition the following properties.

(i) There exists a function $\delta : (0, R_C) \to (0, \pi)$ such that for $|\theta| < \delta(x)$ as $x \uparrow R_C-$,

$$C(xe^{i\theta}) \sim C(x)e^{i\theta E_x[N] + \frac{1}{2}\theta^2 V_x(N)}\,.$$

(ii) Uniformly as $x \uparrow R_C-$, for $\delta(x) \leq |\theta| \leq \pi$,

$$C(xe^{i\theta}) = o\left(\frac{C(x)}{V_x(N)}\right)\,.$$

Then,

$$\frac{C_n x_n}{C(x_n)} \sim \frac{1}{\sqrt{2\pi}x_n V_{x_n}(N)}\,.$$

Therefore (Duchon, Flajolet, Louchard and Schaeffer, 2004 [1]):

[1] This article also gives the references concerning the asymptotics of the coefficients of the power series development of $C(x)$.

Theorem 21.2.10 *Under the above conditions, the Boltzmann sampler $BS_{x_n}(\mathcal{C})$ succeeds to deliver a sample of size exactly n in a mean number of trials asymptotic to $\sqrt{2\pi}V_{x_n}(N)$. In particular, if $\mathcal{C}$ is specifiable, the overall cost of sampling is $O(n\sqrt{2\pi}V_{x_n}(N))$ on average.*

21.3 Exact Sampling of a Cluster Process

21.3.1 The Brix–Kendall Exact Sampling Method

(Brix–Kendall, 2002) In order to simplify the notation, denote the grid $\mathbb{Z}^2$ by V, a point in V being therefore of the form $v = (i, j)$. Let $X := \{X(v)\}_{v\in V}$ be a simple point process on V, that is, a random sequence with values in $\{0, 1\}$, called the "germ point process"[2], a "point" being a vertex v such that $X(v) = 1$. Let $Z := \{Z(v)\}_{v\in V}$ be a collection of integer-valued random variables. Consider now a family $\{Z_u\}_{u\in V}$ of independent copies of Z, this family being also independent of the basic simple point process X.

Define the cluster process with germ point process X and typical cluster Z to be the sequence $Y := \{Y_v\}_{v\in V}$ given by

$$Y_v = \sum_{u\in V} X_u Z_u(v - u)\,.$$

The random variable just displayed can a priori take infinite values. To avoid this, we impose the condition

$$\sum_{u\in V} E\left[X_u\right] E\left[Z(v - u)\right] < \infty \qquad (v \in V)\,. \tag{21.13}$$

We wish to obtain a sample of the cluster process Y on a finite "window" $C \subset V$. Note that it suffices to produce the positive values Y_v where $v \in C$. In principle, we have to generate for each point of the germ point process located at u a sample of Z_u. It is assumed that the clusters Z_u are easy to obtain, and therefore the problem that remains is the possibly infinite number of points of the germ point process X. However, we observe that the probability that a point of the germ point process located at u contributes to the cluster process inside the window C is

$$\alpha_u := P\left(\sum_{v\in C} Z(v - u) > 0\right)\,.$$

In particular, in view of assumption (21.13) and of the finiteness of the window C, the average number of germ points contributing to the cluster process Y in the window C, $\sum_{u\in V} E\left[X_u\right] p_u$, is finite, and in particular the number of contributing points is almost surely finite. This suggests the following procedure to sample the cluster point process on the finite window in two steps[3]:

[2]The problem considered in this section is the discrete version of the original one featuring point processes.

[3][Brix and Kendall, 2002]

Step 1. Generate a thinned version $\widetilde{X}$ of the germ point process X, that is a point process with the same distribution as X after independent thinning with thinning probability α_u.

Step 2. From each point of $\widetilde{X}$ located at u, generate a sample of $\{Z_u(v-u)\}_{v\in C}$ conditioned by $\sum_{v\in C} Z_u(v-u) > 0$.

The implementation of Step 2 consists in generating independent copies of Z_u until one of them satisfies condition $\sum_{v\in C} Z_u(v-u) > 0$.

Step 1 can be implemented in the case where the germ point process is a Bernoulli process, that is an independent sequence $\{X(v)\}_{v\in V}$ with $P(X_u = 1) = \beta_u$. The thinned germ point process $\widetilde{X}$ is then an independent sequence $\{\widetilde{X}(v)\}_{v\in V}$ with $P(\widetilde{X}_u = 1) = \beta_u\alpha_u = p_u$. There remains the task of generating such a sequence.

21.3.2 Thinning the Grid

We first deal with the one-dimensional case. A "point process" on the one-dimensional "grid" $\mathbb{N}$ is, by definition, a sequence $\{X_n\}_{n\in\mathbb{N}}$ of IID $\{0,1\}$-valued random variables, with the common distribution given by $P(X_n = 1) = p_n$ $(n \geq 0)$. (We are therefore "thinning the grid" $\mathbb{N}$, considered as a deterministic point process, with the thinning probability function p_n.) Suppose that $\sum_{n\geq 0} p_n < \infty$, which guarantees that the thinned grid has almost surely a finite number of points. In order to obtain a sample of this point process, one cannot just draw a random variable X_n for all points in succession, because there is *a priori* no stopping rule indicating that we have reached the last point, denoted by T, of the point process. We have to proceed otherwise. For this, note that

$$P(T = n) = P(X_n = 1, X_{n+1} = 0, X_{n+2} = 0, \ldots) = p_n \prod_{k\geq n+1} (1 - p_k) \qquad (\star)$$

and that, for $0 \leq k \leq n-1$,

$$\begin{aligned} P(X_k = 1 \mid T = n) &= \frac{P(X_k = 1, T = n)}{P(T = n)} \\ &= \frac{P(X_k = 1, X_n = 1, X_{n+1} = 0, X_{n+2} = 0, \ldots)}{P(X_n = 1, X_{n+1} = 0, X_{n+2} = 0, \ldots)} \\ &= \frac{P(X_k = 1)P(X_n = 1, X_{n+1} = 0, X_{n+2} = 0, \ldots)}{P(X_n = 1, X_{n+1} = 0, X_{n+2} = 0, \ldots)} = P(X_k = 1)\,. \end{aligned}$$

Therefore, in order to simulate the thinned grid, one may start by sampling a variable T with the distribution $(\star)$, and if $T = n$, set $X_n = 1, X_{n+1} = 0, X_{n+2} = 0, \ldots$ and for $0 \leq k \leq n-1$, sample X_k with the distribution $P(X_k = 1) = p_k$.

There is still an issue left aside in the presentation of the thinning procedure of the grid $\mathbb{N}$. Can we really sample T? In fact one needs a closed expression of the distribution of this variable, in particular of the infinite product $\prod_{k\geq n+1}(1-p_k)$. If this is not possible, we may be lucky enough to find a dominating distribution function $q_n \geq p_n$ such $\sum_n q_n < \infty$ and such that the infinite product $\prod_{k\geq n+1}(1-q_k)$

is computable. One would then sample the thinned grid with thinning probability function q_n. A point of this dominating grid located at k will be kept with probability p_k/q_k as a point of the desired sample.

For instance, try $q_n = 1 - e^{-\alpha_n}$ with $\sum_{n\geq 0} \alpha_n < \infty$ so that

$$\sum_{n\geq 0} q_n = \sum_{n\geq 0} 1 - e^{-\alpha_n} \leq \sum_{n\geq 0} \alpha_n < \infty\,.$$

The infinite products $\prod_{k\geq n+1}(1-q_k)$ should be computable, or equivalently, the sum $\sum_{n\geq 0}\alpha_n$ should be computable (and finite). This is the case if, for instance, $\alpha_n = C\frac{1}{n^2}$.

Thinning the two-dimensional grid $\mathbb{Z}^2$ is conceptually the same. Here the probability of keeping the point $v \in \mathbb{Z}^2$ is p_v, where it is assumed that $\sum_{v\in V} p_v < \infty$ whereby guaranteeing that the number of points of the thinned grid is finite. It suffices to apply bijectively $\mathbb{Z}^2$ on $\mathbb{N}$ by enumerating the points of $\mathbb{Z}^2$ as $\{v_n\}_{n\geq 0}$ (the function $n \to v_n$ is called a scanning of the grid). The rest is then obvious. The ordering of the points of V has an influence on the computation load (see Exercises 21.4.13 and 21.4.14).

Books for Further Information

[Levin, Peres, and Wilmer, 2009] for the Propp–Wilson algorithm, [Flajolet and Sedgewick, 2009] for Boltzmann sampling.

21.4 Exercises

Exercise 21.4.1. Forward coupling does not yield exact sampling
Refer to the Propp–Wilson algorithm. Show that the coalesced value at the forwards coupling time is not a sample of π. For a counterexample use the two-state HMC with $E = \{1, 2\}$, $p_{1,2} = 1$, $p_{2,2} = p_{2,1} = 1/2$.

Exercise 21.4.2. Monotone Propp–Wilson for the Ising model
Consider the classical Ising model of Example 9.1.7 with energy function $U(x) = \sum_{\langle v,w\rangle} x(v)x(w)$. Define on the state space $E = \{-1, +1\}^S$ the partial order relation $\preceq$ defined as follows: $x = (x(v), v \in V) \preceq y = (y(v), v \in V)$ if and only if for all $v \in V$, $x(v) = +1$ implies $y(v) = +1$. Show that the monotone Propp–Wilson algorithm of Section 21.1 can be applied.

Exercise 21.4.3. The impatient simulator
Find a very simple example showing that use of the Propp–Wilson algorithm by an impatient customer introduces a bias.

Exercise 21.4.4.
Prove Theorem 21.1.7.

Exercise 21.4.5. The binary tree
Let C_n be the number of binary trees of size n. Give a recursion equation linking $C_1, \ldots, C_n$.

Exercise 21.4.6. General graphs
Explain why the unlabeled collection $\mathcal{G}$ of general plane trees can be represented by the grammar $\mathcal{G} = \mathcal{Z} \times \mathcal{G}^*$. What is its EGF? Describe the corresponding Boltzmann sampler.

Exercise 21.4.7. Unary-binary trees
A collection of plane trees $\mathcal{V}$ is defined by the grammar $\mathcal{V} = \mathcal{Z} \times (\mathcal{E} + \mathcal{V} + \mathcal{V}^2)$. Describe the general form of the corresponding trees and give the corresponding Boltzmann sampler.

Exercise 21.4.8. Filaments
Consider the labeled class of objects $\mathcal{F}$ defined by $\mathcal{F} = Set(Seq_{\geq 1}(\mathcal{Z}))$. (A sample will look like a finite set whose elements are non empty sequences of points. A given sequence of points is represented by a segment (a "filament") whose length is the number of points in the sequence. The filaments are placed at random in space, and then represent an assembly of filaments floating freely in a liquid.)

What is the EGF of this model? Describe the corresponding Boltzmann sampler? Is Theorem 21.2.8 applicable?

Exercise 21.4.9. Partitions
Prove in detail the conclusions of Example 21.2.6.

Exercise 21.4.10. Labeled cycles
Show that the EGF of the collection $\mathcal{C} := Cyc(\mathcal{A})$ is

$$\hat{C}(x) = \log \frac{1}{1 - \hat{A}(x)}\,.$$

Deduce from this that

$$BS_x(\mathcal{C}) = \Big(Log(\hat{A}(x)) \longrightarrow BS_x(\mathcal{A}) \Big)$$

where $Log(\lambda)$ represents the log-law of parameter $\lambda < 1$ of distribution

$$P(Y = k) = \frac{1}{\log(1-\lambda)^{-1})} \frac{\lambda^k}{k}\,.$$

Exercise 21.4.11. Approximate Boltzmann sampling
Suppose that the conditions of Theorem 21.2.8 are satisfied. Defining $\sigma(x) := V_x(N)$, show that the average number of trials necessary to obtain a Boltzmann sample of size in the interval $[n(1-\varepsilon), n(1+\varepsilon)]$ is smaller than $\frac{1}{1-\frac{\sigma(x_n)}{n^2\varepsilon^2}}$.

Exercise 21.4.12. Exact Boltzmann sampling
Show that the EGF's of Example 21.2.6 (partitions) and Exercise 21.4.8 (filaments), respectively, satisfy the conditions of Theorem 21.2.10, and that the asymptotic average cost of exact sampling is, respectively, $O\left(n^{\frac{3}{2}}\sqrt{\log n}\right)$ and $O\left(n^{\frac{3}{2}}\right)$.

Exercise 21.4.13. Comparison of evanescence rates
Consider a sequence $\{X_n\}_{n\in\mathbb{N}}$ of IID $\{0,1\}$-valued random variables, with the common distribution given by $P(X_n = 1) = p_n$ $(n \geq 0)$, and such that $\sum_n p_n < \infty$. Let $\{\hat{X}_n\}_{n\in\mathbb{N}}$ be another sequence of IID $\{0,1\}$-valued random variables, with the common distribution given by $P(\hat{X}_n = 1) = \hat{p}_n$ $(n \geq 0)$. Let T and $\hat{T}$ be their respective vanishing time (for instance $T = \inf\{n \geq 0\,;\, \sum_{k\geq n} X_k = 0\}$). Show that if the sequence $\{\hat{p}_n\}_{n\in\mathbb{N}}$ is obtained by reordering the sequence $\{p_n\}_{n\in\mathbb{N}}$ in decreasing order, then $\hat{T}$ is stochastically smaller than T. (Hint: coupling.)

Exercise 21.4.14. Optimal thinning order
Apply the result of Exercise 21.4.13 to the problem of Subsection 21.3.2 where $v \to p_v$ is decreasing with the distance (say, euclidean) $|v|$ from the origin to v ($|v_1| \leq |v_2|$ implies $p_{v_1} \geq p_{v_2}$). Give an optimal scanning, that is a scanning that minimizes the average number of sites to be inspected.

Appendix A

Appendix

A.1 Some Results in Analysis

Infinite Products

Let $\{a_n\}_{n\geq 1}$ be a sequence of numbers of the interval $[0,1)$.

(a) If $\sum_{n=1}^{\infty} a_n < \infty$, then

$$\lim_{n\uparrow\infty} \prod_{k=1}^{n} (1-a_k) > 0.$$

(b) If $\sum_{n=1}^{\infty} a_n = \infty$, then

$$\lim_{n\uparrow\infty} \prod_{k=1}^{n} (1-a_k) = 0.$$

Proof. (a): For any numbers $c_1, \ldots, c_n$ in $[0,1)$, it holds that $(1-c_1)(1-c_2)\cdots(1-c_n) \geq 1-c_1-c_2-\cdots-c_n$ (proof by induction). Since $\sum_{n=1}^{\infty} a_n$ converges, there exists an integer N such that for all $n \geq N$,

$$a_N + \cdots + a_n < \frac{1}{2}.$$

Therefore, defining $\pi(n) = \prod_{k=1}^{n}(1-a_k)$, we have that for all $n \geq N$,

$$\frac{\pi(n)}{\pi(N-1)} = (1-a_N)\cdots(1-a_n) \geq 1-(a_N+\cdots+a_n) \geq \frac{1}{2}.$$

Therefore, the sequence $\{\pi(n)\}_{n\geq N}$ is a nonincreasing sequence bounded from below by $\frac{1}{2}\pi(N-1) > 0$, so that $\lim_{n\uparrow\infty} \pi(n) > 0$.

(b): Using the inequality $1-a \leq e^{-a}$ when $a \in [0,1)$, we have that $\pi(n) \leq e^{-a_1-a_2-\cdots-a_n}$, and therefore, if $\sum_{n=1}^{\infty} a_n = \infty$, $\lim_{n\uparrow\infty} \pi(n) = 0$. □

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6

Abel's Theorem

Lemma A.1.1 *Let* $\{b_n\}_{n\geq 1}$ *and* $\{a_n\}_{n\geq 1}$ *be two sequences of real numbers such that*

$$b_1 \geq b_2 \geq \cdots \geq b_n \geq 0,$$

and such that for some real numbers m *and* M*, and all* $n \geq 1$*,*

$$m \leq a_1 + \cdots + a_n \leq M.$$

Then, for all $n \geq 1$*,*

$$b_1 m \leq a_1 b_1 + \cdots + a_n b_n \leq b_1 M. \tag{A.1}$$

Proof. Let $s_n = a_1 + \cdots + a_n$, and use Abel's summation technique to obtain

$$\begin{aligned} a_1 b_1 + \cdots + a_n b_n &= b_1 s_1 + b_2 (s_2 - s_1) + \cdots + b_n (s_n - s_{n-1}) \\ &= s_1[b_1 - b_2] + \cdots + s_{n-1}[b_{n-1} - b_n] + s_n[b_n]. \end{aligned}$$

The bracketed terms are all nonnegative, and therefore replacing each s_i by its lower bound or upper bound yields the result. □

We recall without proof a standard result of calculus.

Lemma A.1.2 *The sum of a uniformly convergent series of continuous functions is a continuous function.*

Theorem A.1.3 *Let* $\{a_n\}_{n\geq 1}$ *be a sequence of real numbers such that the radius of convergence of the power series* $\sum_{n=0}^{\infty} a_n z^n$ *is* 1*. Suppose that the sum* $\sum_{n=0}^{\infty} a_n$ *is convergent. Then the power series* $\sum_{n=0}^{\infty} a_n x^n$ *is uniformly convergent in* $[0, 1]$ *and*

$$\lim_{x \uparrow 1} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n, \tag{A.2}$$

where $x \uparrow 1$ *means that* x *tends to* 1 *from below.*

Proof. It suffices to prove that $\sum_{n=0}^{\infty} a_n x^n$ is uniformly convergent in $[0, 1]$, since (A.2) then follows by Lemma A.1.2. Write $A_n^p = a_n + \cdots + a_p$. By convergence of $\sum_{n=0}^{\infty} a_n$, for all $\epsilon > 0$, there exists an integer $n_0 \geq 1$ such that $p \geq n \geq n_0$ implies $|A_n^p| \leq \epsilon$, and therefore, since for $x \in [0, 1]$, the sequence $\{x^n\}_{n\geq 0}$ is nonincreasing, Abel's lemma gives, for all $x \in [0, 1]$,

$$|a_n x^n + \ldots + a_p x^p| \leq \epsilon x^n \leq \epsilon,$$

from which uniform convergence follows. □

Theorem A.1.4 *Let $\{a_n\}_{n\geq 0}$ be a sequence of nonnegative real numbers such that the power series $\sum_{n=0}^{\infty} a_n z^n$ has a radius of convergence equal to 1. If*

$$\lim_{x\uparrow 1} \sum_{n=0}^{\infty} a_n x^n = a \leq \infty, \tag{A.3}$$

then

$$\sum_{n=0}^{\infty} a_n = a. \tag{A.4}$$

Proof. For $x \in [0,1)$, $\sum_{n=0}^{\infty} a_n x^n \leq \sum_{n=0}^{\infty} a_n$ (the a_n are nonnegative), and therefore by (A.3), $a \leq \sum_{n=0}^{\infty} a_n$. This proves the result when $a = \infty$.
We now suppose that $a < \infty$. From $\sum_{n=0}^{p} a_n = \lim_{x\uparrow 1} \sum_{n=0}^{p} a_n x^n$, we have that $\sum_{n=0}^{p} a_n \leq a < \infty$. Thus, $\sum_{n=1}^{p} a_n$ is a nondecreasing sequence, converging to some α, $\alpha \leq a < \infty$. By Abel's theorem, $\lim_{x\uparrow 1} \sum_{n=0}^{\infty} a_n x^n = \alpha$, and therefore $\alpha = a$ and $\sum_{n=0}^{\infty} a_n = a$. □

Dominated Convergence for Series

Theorem A.1.5 *Let $\{a_{nk}\}_{n\geq 1, k\geq 1}$ be an array of real numbers such that for some sequence $\{b_k\}_{k\geq 1}$ of nonnegative numbers satisfying $\sum_{k=1}^{\infty} b_k < \infty$, it holds that for all $n \geq 1, k \geq 1$, $|a_{nk}| \leq b_k$. If for all $k \geq 1$, $\lim_{n\uparrow\infty} a_{nk} = a_k$, then*

$$\lim_{n\uparrow\infty} \sum_{k=1}^{\infty} a_{nk} = \sum_{k=1}^{\infty} a_k \,.$$

Proof. Let $\epsilon > 0$ be fixed. Since $\sum_{k=1}^{\infty} b_k$ is a convergent series, one can find $M = M(\epsilon)$ such that $\sum_{k=M+1}^{\infty} b_k < \frac{\epsilon}{3}$. In particular, since $|a_{nk}| \leq b_k$ and therefore $|a_k| \leq b_k$, we have

$$\sum_{k=M+1}^{\infty} |a_{nk}| + \sum_{k=M+1}^{\infty} |a_k| \leq \frac{2\epsilon}{3}.$$

Now, for sufficiently large n,

$$\sum_{k=1}^{M} |a_{nk} - a_k| \leq \frac{\epsilon}{3}.$$

Therefore, for sufficiently large n,

$$\left| \sum_{k=1}^{\infty} a_{nk} - \sum_{k=1}^{\infty} a_k \right| \leq \sum_{k=1}^{M} |a_{nk} - a_k| + \sum_{k=M+1}^{\infty} |a_{nk}| + \sum_{k=M+1}^{\infty} |a_k| \leq \frac{\epsilon}{3} + \frac{2\epsilon}{3} = \epsilon.$$

□

Theorem A.1.6 *Let $\{a_{nk}\}_{n\geq 1,k\geq 1}$ be an array of nonnegative real numbers such that for all $k \geq 1$, the sequence $\{a_{nk}\}_{n\geq 1}$ is non-decreasing with limit $a_k \leq \infty$. Then*

$$\lim_{n\uparrow\infty} \sum_{k=1}^{\infty} a_{nk} = \sum_{k=1}^{\infty} a_k \, .$$

Proof. If $\sum_{k=1}^{\infty} a_k < \infty$, the result is a direct application of the dominated convergence theorem.
For the case $\sum_{k=1}^{\infty} a_k = \infty$, let $A > 0$ be fixed, and choose $M = M(A)$ such that $\sum_{k=1}^{M} a_k \geq 2A$. For sufficiently large n, $\sum_{k=1}^{M} (a_k - a_{nk}) \leq A$. Therefore, for sufficiently large n,

$$\sum_{k=1}^{\infty} a_{nk} \geq \sum_{k=1}^{M} a_k + \sum_{k=1}^{M} (a_{nk} - a_k) \geq 2A - A = A.$$

□

Theorem A.1.7 *Let $\{a_{nk}\}_{n\geq 1,k\geq 1}$ be an array of nonnegative real numbers. Then*

$$\sum_{k=1}^{\infty} \liminf_{n\uparrow\infty} a_{nk} \leq \liminf_{n\uparrow\infty} \sum_{k=1}^{\infty} a_{nk}.$$

Proof. By definition of lim inf, for fixed k,

$$z_{nk} := \inf(a_{nk}, a_{n+1,k}, \ldots)$$

increases, as $n \uparrow \infty$, to $\liminf_{n\uparrow\infty} a_{nk}$. Therefore, by monotone convergence,

$$\sum_{k=1}^{\infty} \liminf_{n\uparrow\infty} a_{nk} = \lim_{n\uparrow\infty} \uparrow \sum_{k=1}^{\infty} z_{nk}.$$

But since $z_{nk} \leq a_{nk}$,

$$\sum_{k=1}^{\infty} z_{nk} \leq \sum_{k=1}^{\infty} a_{nk},$$

and therefore

$$\lim_{n\uparrow\infty} \sum_{k=1}^{\infty} z_{nk} \leq \liminf_{n\uparrow\infty} \sum_{k=1}^{\infty} a_{nk}.$$

□

Tykhonov's Theorem

Theorem A.1.8 *Let* $\{x_n\}_{n\geq 0}$ *be a sequence of elements of* $[0,1]^{\mathbb{N}}$*, that is*

$$x_n = (x_n(0),\, x_n(1), \ldots),$$

where $x_n(k) \in [0,1]$ *for all* $k,\, n \in \mathbb{N}$*. There exists a strictly increasing sequence of integers* $\{n_l\}_{l\geq 0}$ *and an element* $x \in \{0,1\}^{\mathbb{N}}$ *such that*

$$\lim_{l\uparrow\infty} x_{n_l}(k) = x(k) \tag{A.5}$$

for all $k \in \mathbb{N}$*.*

Proof. Since the sequence $\{x_n(0)\}_{n\geq 0}$ is contained in the closed interval $[0,1]$, by the Boltzano–Weierstrass theorem, one can extract a subsequence $\left\{x_{n_0(l)}(0)\right\}_{l\geq 0}$ such that

$$\lim_{l\uparrow\infty} x_{n_0(l)}(0) = x(0)$$

for some $x(0) \in [0,1]$. In turn, one can extract from $\left\{x_{n_0(l)}(1)\right\}_{l\geq 0}$ a subsequence $\left\{x_{n_1(l)}(1)\right\}_{l\geq 0}$ such that

$$\lim_{l\uparrow\infty} x_{n_1(l)}(1) = x(1)$$

for some $x(1) \in [0,1]$. Note that

$$\lim_{l\uparrow\infty} x_{n_1(l)}(0) = x(0).$$

Iterating this process, we obtain for all $j \in \mathbb{N}$ a sequence $\left\{x_{n_j(l)}\right\}_{l\geq 0}$ that is a subsequence of each sequence $\left\{x_{n_0(l)}(1)\right\}_{l\geq 0}, \ldots, \left\{x_{n_{j-1}(l)}(1)\right\}_{l\geq 0}$ and such that

$$\lim_{l\uparrow\infty} x_{n_j(l)}(k) = x(k)$$

for all $k \leq j$, where $x(1), \ldots, x(j) \in [0,1]$. The diagonal sequence $n_l = n_l(l)$ then establishes (A.5). □

A.2 Greatest Common Divisor

Let $a_1, \ldots, a_k \in \mathbb{N}$ be such that $\max(a_1, \ldots, a_k) > 0$. Their greatest common divisor (GCD) is the largest positive integer dividing all of them. It is denoted by $\text{g.c.d}\,(a_1, \ldots, a_k)$. Clearly, removing all zero elements does not change the g.c.d, so that we may assume without loss of generality that all the a_k's are positive.

Let $\{a_n\}_{n\geq 1}$ be a sequence of positive integers. The sequence $\{d_k\}_{k\geq 1}$ defined by $d_k = \gcd(a_1, \ldots, a_k)$ is bounded below by 1 and is nonincreasing, and it therefore has a limit $d \geq 1$, a positive integer called the g.c.d of the sequence $\{a_n\}_{n\geq 1}$. Since the d_k's are integers, the limit is attained after a finite number of steps, and therefore there exists a positive integer k_0 such that $d = \gcd(a_1, \ldots, a_k)$ for all $k \geq k_0$.

Lemma A.2.1 *Let $S \subset \mathbb{Z}$ contain at least one nonzero element and be closed under addition and subtraction. Then S contains a least positive element a, and $S = \{ka\,;k \in \mathbb{Z}\}$.*

Proof. Let $c \in S,\ c \neq 0$. Then $c - c = 0 \in S$. Also $0 - c = -c \in S$. Therefore, S contains at least one positive element. Denote by a the smallest positive element of S. Since S is closed under addition and subtraction, S contains a, $a+a = 2a, \ldots$ and $0 - a = -a, 0 - 2a = -2a, \ldots$, that is, $\{ka\,;k \in \mathbb{Z}\} \subset S$.
Let $c \in S$. Then $c = ka + r$, where $k \in \mathbb{Z}$ and $0 \leq r < a$. Since $r = c - ka \in S$, we cannot have $r > 0$, because this would contradict the definition of a as the smallest positive integer in S. Therefore, $r = 0$, i.e., $c = ka$. Therefore, $S \subset \{ka\,;k \in \mathbb{Z}\}$. □

Lemma A.2.2 *Let $a_1, \ldots, a_k$ be positive integers with greatest common divisor d. There exist $n_1, \ldots, n_k \in \mathbb{Z}$ such that $d = \sum_{i=1}^{k} n_i a_i$.*

Proof. The set $S = \left\{\sum_{i=1}^{k} n_i a_i\ ;\ n_1, \ldots, n_k \in \mathbb{Z}\right\}$ is closed under addition and subtraction, and therefore, by Lemma A.2.1, $S = \{ka\,;k \in \mathbb{Z}\}$, where $a = \sum_{i=1}^{k} n_i a_i$ is the smallest positive integer in S.

Since d divides all the a_i's, d divides a, and therefore $0 < d \leq a$. Also, each a_i is in S and is therefore a multiple of a, which implies that $a \leq \text{g.c.d}\,(a_1, \ldots, a_k) = d$. Therefore, $d = a$. □

Theorem A.2.3 *Let d be the g.c.d of $A = \{a_n\,;n \geq 1\}$, a set of positive integers that is closed under addition. Then A contains all but a finite number of the positive multiples of d.*

Proof. We may assume without loss of generality that $d = 1$ (otherwise, divide all the a_n's by d). For some k, $d = 1 = \text{g.c.d}\,(a_1, \ldots, a_k)$, and therefore by Lemma A.2.2,

$$1 = \sum_{i=1}^{k} n_i a_i$$

for some $n_1, \ldots, n_k \in \mathbb{Z}$. Separating the positive from the negative terms in the latter equality, we have $1 = M - P$, where M and P are in A.
Let $n \in \mathbb{N}$, $n \geq P\,(P-1)$. We have $n = aP + r$, where $r \in [0, P-1]$. Necessarily, $a \geq P-1$, otherwise, if $a \leq P-2$, then $n = aP + r < P(P-1)$. Using $1 = M - P$, we have that $n = aP + r\,(M - P) = (a - r)\,P + rM$. But $a - r \geq 0$. Therefore, n is in A. We have thus shown that any $n \in \mathbb{N}$ sufficiently large (say, $n \geq P\,(P-1)$) is in A. □

A.3 Eigenvalues

The basic results of the theory of matrices relative to eigenvalues and eigenvectors will now be reviewed, and the reader is referred to the classical texts for the proofs.

Let A be a square matrix of dimension $r \times r$, with complex coefficients. If there exists a scalar $\lambda \in \mathbb{C}$ and a column vector $v \in \mathbb{C}^r$, $v \neq 0$, such that

$$Av = \lambda v \text{ (resp., } v^T A = \lambda v^T), \tag{A.6}$$

then v is called a right-eigenvector (resp., a left-eigenvector) associated with the eigenvalue λ. There is no need to distinguish between right and left-eigenvalues because if there exists a left-eigenvector associated with the eigenvalue λ, then there exists a right-eigenvector associated with the same eigenvalue λ. This follows from the facts that the set of eigenvalues of A is exactly the set of roots of the *characteristic equation*

$$\det(\lambda I - A) = 0 \tag{A.7}$$

where I is the $r \times r$ identity matrix, and that

$$\det(\lambda I - A) = \det(\lambda I - A^T).$$

The *algebraic multiplicity* of λ is its multiplicity as a root of the *characteristic polynomial* $\det(\lambda I - A)$.

If $\lambda_1, \cdots, \lambda_k$ are *distinct* eigenvalues corresponding to the right-eigenvectors $v_1, \cdots, v_k$ and the left-eigenvectors $u_1, \cdots, u_k$, then $v_1, \cdots, v_k$ are independent, and so are $u_1, \cdots, u_k$.

Call R_λ (resp., L_λ) the set of right-eigenvectors (resp., left-eigenvectors) associated with the eigenvalue λ, plus the null vector. Both L_λ and R_λ are vector subspaces of $\mathbb{C}^r$, and they have the same dimension, called the *geometric multiplicity* of λ. In particular, the largest number of independent right-eigenvectors (resp., left-eigenvectors) cannot exceed the sum of the geometric multiplicities of the distinct eigenvalues.

The matrix A is called *diagonalizable* if there exists a nonsingular matrix Γ of the same dimensions such that

$$\Gamma A \Gamma^{-1} = \Lambda, \tag{A.8}$$

where

$$\Lambda = \text{ diag } (\lambda_1, \cdots, \lambda_r)$$

for some $\lambda_1, \cdots, \lambda_r \in \mathbb{C}$, not necessarily distinct. It follows from (A.8) that with $U = \Gamma^T$, $U^T A = U^T \Lambda$, and with $V = \Gamma^{-1}$, $AV = V\Lambda = \Lambda V$, and therefore $\lambda_1, \cdots, \lambda_r$ are eigenvalues of A, and the ith row of $U^T = \Gamma$ (resp., the ith column of $V = \Gamma^{-1}$) is a left-eigenvector (resp., right-eigenvector) of A associated with the eigenvalue λ_i. Also, $A = V\Lambda U^T$ and therefore

$$A^n = V \Lambda^n U^T. \tag{A.9}$$

Clearly, if A is diagonalizable, the sum of the geometric multiplicities of A is exactly equal to r. It turns out that the latter is a sufficient condition of diagonalizability of A. Therefore, A is diagonalizable if and only if the sum of the geometric multiplicities of the distinct eigenvalues of A is equal to r.

EXAMPLE A.3.1: DISTINCT EIGENVALUES. By the last result, if the eigenvalues of A are distinct, A is diagonalizable. In this case, the diagonalization process can be described as follows. Let $\lambda_1, \cdots, \lambda_r$ be the r distinct eigenvalues and let $u_1, \cdots, u_r$ and $v_1, \cdots, v_r$ be the associated sequences of left and right-eigenvectors, respectively. As mentioned above, $u_1, \cdots, u_r$ form an independent collection of vectors, and so do $v_1, \cdots, v_r$. Define

$$U = [u_1 \cdots u_r], V = [v_1 \cdots v_r]. \tag{A.10}$$

Observe that if $i \neq j$, $u_i^T v_j = 0$. Indeed, $\lambda_i u_i^T v_j = u_i^T A v_j = \lambda_j u_i^T v_j$, which implies $(\lambda_i - \lambda_j) u_i^T v_j = 0$, and in turn $u_i^T v_j = 0$, since $\lambda_i \neq \lambda_j$ by hypothesis. Since eigenvectors are determined up to multiplication by an arbitrary non-null scalar, one can choose them in such a way that $u_i^T v_i = 1$ for all $i \in [1, r]$. Therefore,

$$U^T V = I, \tag{A.11}$$

where I is the $r \times r$ identity matrix. Also, by definition of U and V,

$$U^T A = \Lambda U^T, AV = \Lambda V. \tag{A.12}$$

In particular, by (A.11), $A = V \Lambda U^T$. From the last identity and (A.11) again, we obtain for all $n \geq 0$,

$$A^n = V \Lambda^n U^T, \tag{A.13}$$

that is,

$$A^n = \sum_{i=1}^{r} \lambda_i^n v_i u_i^T. \tag{A.14}$$

A.4 Kolmogorov's 0–1 Law

The discussion of this section is partly heuristic, but the proof of the 0-1 law follows closely the rigorous proof that the reader will find in a standard text on probability theory.

Let $\{X_n\}_{n \geq 1}$ be a sequence of discrete random variables taking their values in the denumerable space E. Define, for each $1 \leq m \leq n$, the sigma-field $\sigma(X_m, \ldots, X_n)$ to be the smallest sigma-field that contains all the events $\{X_\ell = i_\ell\}$ for all $m \leq \ell \leq n$, all $i_\ell \in E$. Equivalently, it is the smallest sigma-field that contains all the events of the form $\{X_m = i_m, \ldots, X_n = i_n\}$ for all $i_m, \ldots, i_m \in E$. Still equivalently, it is the smallest sigma-field that contains all the events of the form $\{(X_m, \ldots, X_n) \in C\}$ for all $C \in E^{m-n+1}$. In other words, this sigma-field contains

all the events that are expressible[1] in terms of $(X_m, \ldots, X_n)$. We shall simplify the notation and denote it by $\sigma(X_m^n)$.

For fixed $m \geq 1$, the union $\cup_{n=m}^{\infty} \sigma(X_m^n)$ is an algebra (not a sigma-field in general) denoted $\mathcal{A}(X_m^n)$. The smallest sigma-field containing $\mathcal{A}(X_m^n)$ is, by definition, the sigma-field $\sigma(X_m^\infty)$. The sigma-field $\sigma(X_1^\infty)$ contains the events that are expressible in terms of $X_1, X_2, \ldots$

The intersection of sigma-fields is a sigma-field. Therefore $\cap_{m\geq 1}\sigma(X_m^\infty)$ is a sigma-field, called the tail sigma-field of the sequence $\{X_n\}_{n\geq 1}$. Any event therein is called a tail event and does not depend on any finite number of terms of the stochastic sequence, say $X_1, \ldots, X_r$, because it belongs to $\sigma(X_{r+1}^\infty)$ and therefore is expressible in terms of $X_{r+1}, X_{r+2}, \ldots$. Typical tail events for a real-valued sequence are $\{\exists \lim_{n\uparrow\infty} X_n\}$, or $\{\exists \lim_{n\uparrow\infty} \frac{\sum_{k=1}^n X_k}{n}\}$. However, $\{\lim_{n\uparrow\infty} \sum_{k=1}^n X_k = 0\}$ is not a tail event, since it depends crucially on X_1, for example.

The following result is the Kolmogorov zero-one law.

Theorem A.4.1 *The tail sigma-field of a sequence $\{X_n\}_{n\geq 1}$ of independent random variables is trivial, that is, if A is a tail event, then $P(A) = 0$ or 1.*

Proof. The proof depends on the following lemma.

Lemma A.4.2 *Let $\mathcal{A}$ be an algebra generating the sigma-field $\mathcal{F}$ and let P be a probability on $\mathcal{F}$. To any event $B \in \mathcal{F}$ and any $\varepsilon > 0$, one can associate an event $A \in \mathcal{A}$ such that $P(A \triangle B) \leq \varepsilon$.*

Proof. The collection of sets

$$\mathcal{G} := \{B \in \mathcal{F};\ \forall \varepsilon > 0, \exists A \in \mathcal{A} \text{ with } P(A \triangle B) \leq \varepsilon\}$$

obviously contains $\mathcal{A}$. It is a sigma-field. Indeed, $\Omega \in \mathcal{A} \subseteq \mathcal{G}$ and stability of $\mathcal{G}$ by complementation is clear. Finally, let the B_n's $(n \geq 1)$ be in $\mathcal{G}$ and $\varepsilon > 0$ be given. By the sequential continuity of probability, there exists K such that $P(\cup_{n\geq 1} B_n - \cup_{n=1}^K B_n) \leq 2^{-1}\varepsilon$. Also there exist A_n's in $\mathcal{A}$ such that $P(A_n \triangle B_n) \leq 2^{-n-1}\varepsilon$. Therefore

$$P((\cup_{n=1}^K A_n) \triangle (\cup_{n=1}^K B_n)) \leq \sum_{n=1}^K 2^{-n-1}\varepsilon \leq \sum_{n\geq 1} 2^{-n-1}\varepsilon = 2^{-1}\varepsilon\,.$$

Finally:

[1] We rely on the intuition of the reader for the definition of "expressible" since the correct definition would require more care, as would be the case in a more advanced course in abstract probability. Let us for the time being say the following. A real random variable Z is "expressible" in terms of $\{X_n\}_{n\geq 1}$ if there exists a (measurable) function $f : E^\infty \to \mathbb{R}$ such that $Z = f(X_1, X_2, \ldots)$. An event A is "expressible" in terms of $\{X_n\}_{n\geq 1}$ if the random variable $Z := 1_A$ is expressible in terms of of $\{X_n\}_{n\geq 1}$. This definition of course assumes that you know the meaning of "measurable function".

$$P((\cup_{n=1}^{K} A_n) \triangle (\cup_{n\geq 1} B_n)) \leq \varepsilon\,.$$

The proof of stability of $\mathcal{G}$ by countable unions is completed since $\mathcal{A}$ is an algebra and therefore $\cup_{n=1}^{K} A_n \in \mathcal{A}$.

Therefore $\mathcal{G}$ is a sigma-field that contains $\mathcal{A}$ and in particular the sigma-field $\mathcal{F}$ generated by $\mathcal{A}$. □

We now return to the proof of Theorem A.4.1. Fix an arbitrary $\varepsilon > 0$. Let A be an event of the tail sigma-field. It is a fortiori an event of $\sigma(X_m^\infty)$, and since the latter is the smallest sigma-field containing the algebra $\cup_{n=1}^{\infty}\sigma(X_1^n)$, there exists (by Lemma A.4.2) $A_\varepsilon \in \cup_{n=1}^{\infty}\sigma(X_1^n)$ such that $P(A \triangle A_\varepsilon) \leq \varepsilon$. This A_ε is in some $\sigma(X_1^p)$, by definition of $\cup_{n=1}^{\infty}\sigma(X_1^n)$. The event A being in the tail sigma-field is in particular in $\sigma(X_{p+1}^\infty)$. Since the latter is the smallest sigma-field containing the algebra $\cup_{\ell=p+1}^{\infty}\sigma(X_{p+1}^\ell)$, there exists (by Lemma A.4.2) some $r > p$ and some $\tilde{A}_\varepsilon \in \sigma(X_{p+1}^r)$ such that $P(A \triangle \tilde{A}_\varepsilon) \leq \varepsilon$. Now, $\tilde{A}_\varepsilon$ and A_ε are independent, and therefore,

$$P(\tilde{A}_\varepsilon \cap A_\varepsilon) = P(\tilde{A}_\varepsilon)P(\cap A_\varepsilon)\,.$$

But the left-hand side and the right-hand side are by a proper choice of ε arbitrarily close from $P(A \cap A) = P(A)$ and $P(A)P(A) = P(A)^2$. Therefore $P(A) = P(A)^2$ and this is possible only if $P(A) = 0$ or $P(A) = 1$. □

A.5 The Landau Notation

In the so-called Landau notational system, $f(n) = O(g(n))$ means that there exists a positive real number M and an integer n_0 such that for $|f(n)| \leq M|g(n)|$ for all $n \geq n_0$; $f(n) = o(g(n))$ and $f(n) = \omega(g(n))$ mean repectively that $\lim_{n\uparrow\infty} \frac{|f(n)|}{|g(n)|} = 0$ and $\lim_{n\uparrow\infty} \frac{|f(n)|}{|g(n)|} = \infty$.

The notation $f(n) \ll g(n)$ will be used to mean that $f(n) = o(g(n))$. Of course, $f(n) \gg g(n)$ means that $g(n) \ll f(n)$. Also, the symbol $\omega(n)$ will represent a function increasing arbitrarily slowly to ∞. In particular, any power of an "omega function" is an omega function. This is why the reader will encounter equalities of the type $\omega(n)^2 = \omega(n)$, which of course must be interpreted symbolically.

Bibliography

Aldous, D. and P. Diaconis, "Shuffling cards and stopping times", American Mathematical Monthly, 93, 333–348, 1981.

Aldous, D. and P. Diaconis, "Strong uniform times and finite random walks", Adv. Appl. Math., 8, 69–97, 1987.

Aldous, D. and J.A. Fill, *Reversible Markov Chains and Random Walks on Graphs*, 2002, 2014 version at http://www.stat.berkeley.edu/∼aldous/RWG/book.html.

Alon, N. and J.H. Spencer, *The Probabilistic Method*, Wiley, 1991, 3rd ed. 2010.

Anantharam, V. and P. Tsoucas, "A proof of the Markov chain tree theorem", Statist. Probab. Lett., **8** (2), 189–192, 1989.

Anily, S. and A. Federgruen, "Simulated annealing methods with general acceptance probabilities", Journal of Applied Probability, 24, 657–667, 1987.

Ash, R.B., *Information Theory*, Interscience, 1965, Dover, 1990.

Asmussen, S., *Applied Probability and Queues*, Wiley, Chichester, 1987.

Asmussen, S. and P.W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, 2007.

Athreya, K. and P. Jagers, *Branching Processes*, Springer-Verlag, 1973.

Athreya, K. and P. Ney, *Branching Processes*, Springer-Verlag, 1972.

Azuma, K., "Weighted sums of certain dependent random variables", Tohoku Math. J., 19, 357–367, 1967.

Barbour, A.D., L. Holst, S. Janson and J.H. Spencer, *Poisson Approximations*, Oxford University Press, 1992.

Barker, A.A., "Monte Carlo calculations of the radial distribution functions for a proton–electron plasma", Austral. J. Phys., 18, 119–133, 1965.

Bartlett, M.S., "On theoretical models for competitive and predatory biological systems", *Biometrika*, 44, 1957.

Baxter, R.J., *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.

Besag, J., "Spatial interaction and the statistical analysis of lattice systems", Journal of the Royal Statistical Society, B-31, 192–236, 1974.

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6

Bodini, O., Fusy, E. and C. Pivoteau, "Random sampling of plane partitions", arXiv:0712.0111v1.

Bollobás, B., *Random Graphs*, Cambridge University Press, 2nd ed. 2010.

Bollobás, B. and O. Riordan, *Percolation*, Cambridge University Press, 2006.

Bollobás, B. and A. Thomason, "Hereditary and monotone properties of graphs" in *The Mathematics of Paul Erdös, II, Algorithms and Combinatorics, 14*, 70–78, Springer, 1997.

Brémaud, P., *Markov Chains*, Springer, 1999.

A. Brix and W.S. Kendall, "Simulation of cluster point processes without edge effects", Adv. Appl. Prob., **34**, 267–280 (2002).

Capetanakis, J.I., "Tree algorithm for packet broadcast channels", IEEE Transactions on Information Theory, IT-25, 505–515, 1979.

Chandra, A.K.P., Raghavan, P., Ruzzo, W.L., Smolenski, R. and P. Tiwari, "The electrical resistance of a graph captures its commute and cover times", Comp. Complexity, 6, 4, 312–340, 1996.

Chen, L.H.Y., "Poisson approximation for dependent trials", Annals of Probability, 3 (3): 534–545, 1975.

Cohn, H., "Finite non-homogeneous Markov chains: asymptotic behaviour", Adv. Appl. Probab., 8, 502–516, 1976.

Cohn, H., "Countable non-homogeneous Markov chains: asymptotic behaviour", Adv. Appl. Probab., 9, 542–552, 1977.

Cover, T. and J.A. Thomas, *Elements of Information Theory*, Wiley, 1991, 2nd ed. 2006.

Cover, T., "On the competitive optimality of Huffman codes", IEEE Trans Inf. Theory, 37, 1, 172-174, 1991.

Csiszár, I. and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.

Csiszár, I. and P.C. Shields, *Information Theory and Statistics*, now publishers, 2004.

Csiszár, I., "The method of types", IEEE Trans Inf. Theory, 2505–2523, 1998.

Dembo, A. and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, 1998 (2nd ed. 2010).

Diaconis, P., *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, California, 1988.

Diaconis, P., "The Markov chain Monte Carlo revolution", Bull. Amer. Math. Soc. 46 , 179–205, 2009.

Diaconis, P. and J.A. Fill, "Strong stationary times via a new form of duality", The Annals of Probability, 18, 4, 1483–1522, 1991.

Diaconis, P. and L. Saloff-Coste, "Comparison theorems for reversible Markov chains", The Annals of Probability, 3,3 696–730, 1993

Diaconis, P. and L. Saloff-Coste, "What do we know about the Metropolis algorithm?", J. Comp. System Sci., 57, 1, 20–36, 1998.

Diaconis, P. and M. Shahshahani, "Generating a random permutation with random permutations", Z. für W., 57, 2, 159–179, 1981.

Diaconis, P. and D. Stroock, "Geometric bounds for eigenvalues of Markov chains", The Annals of Applied Probability, 1, 1, 36–61, 1991.

Doeblin, W., "Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples", Bulletin Mathématique de la Société Roumaine des Sciences, 39, No.1, 57–115, No.2, 3–61, 1937.

Dobrushin, R.L., "Central limit theorems for non-stationary Markov chains II", Theory of Probability and its Applications, 1, 329–383 (English translation), 1956.

Dobrushin, R.L., "Existence of a phase transition in two- and three-dimensional Ising models", Theory of Probability and its Applications, 10, 193–213, 1965.

Dobrushin, R. L. , "Prescribing a system of random variables by conditional distributions", Theor. Probab. Appl. (Engl. Tr., 15, 453–486, 1970).

Doyle, P.G. and J.L. Snell, *Random Walks and Electrical Networks*, 2000. arXiv: math/0001057 v 1 [math. PR], 11 Jan 2000.

Duchon, P., Flajolet, P., Louchard, G. and G. Schaeffer, "Boltzmann samplers for the random generation of combinatorial structures", Combinatorics, Probability and Computing, **13**, 577–625, 2004.

Grimmett, G.R., "A Theorem on random fields", Bulletin of the London Mathematical Society, 81–84, 1973.

Elias, P., "Error-free coding", IRE Trans. Inf. Theory, IT-4, 29-37, 1954.

Erdös, P., Feller, W. and H. Howard, "A theorem on power series", Bull. Am. Math. Soc. 55, 201–204, 1949.

Erdös, P. and A. Rényi, "On the evolution of random graphs", Magyar Tud. Akad. Mat. Kutató Int. Közl. 5 (1960), 17–61.

Erdös, P. and A. Rényi, "On Random Graphs ", in Publ. Math. Debrecen 6, 290–297, 1959.

Fano, R.M., *Transmission of Information: A statistical theory of communications*, Wiley, New York, 1961.

Fayolle, G., "Etude du comportement d'un canal radio partagé entre plusieurs utilisateurs", Thèse de Docteur–Ingénieur, Université Paris 6, 1975.

Fayolle, G., Malishev, V.A. and M.V. Menshikov, *Topics in the Constructive Theory of Countable Markov Chains*, Cambridge University Press, 1995.

Feller, W. and S. Orey, "A renewal theorem", J. Math. Mech. 10, 619–624, 1961.

Fill, J.A., "Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with an application to the exclusion process", Annals of Applied Probability, 1, 1, 62–87, 1991.

Fill, J.A., "An interruptible algorithm for perfect sampling via Markov chains", Annals of Applied Probability, 8, 131–162, 1998.

Fishman, G.S., *Monte Carlo*, Springer, NY, 1996.

Flajolet, Ph., Zimmerman, P. and B.V. Cutsem, "A calculus for the random generation of labelled combinatorial structures, Theoretical Computer Science, 132(1-2), 1–35, 1994.

Flajolet, Ph., Fusy, E. and C. Pivoteau "Boltzmann sampling of unlabelled structures", in *Proceedings of the 4th Workshop on Analytic Algorithms and Combinatorics, ANALCO'07 (New Orleans)*, 201–211, SIAM, 2007.

Flajolet, Ph. and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.

Fortuin, C.M., "On the random cluster model, II. The percolation model", Physica, 58, 393–418, 1972.

Fortuin, C.M., "On the random cluster model, III. The simple random-cluster process", Physica, 59, 545–570, 1972.

Fortuin, C.M. and P.W. Kasteleyn, On the random cluster model, I. Introduction and relation to other models, Physica, 57, 536–564, 1972.

Fortuin, C.M., Kasteleyn, P.W. and J. Ginibre, "Correlation inequalities on some partially ordered sets", Communications on Mathematical Physics 22, 89-103, 1971.

Foss, S. and R.L. Tweedie, "Perfect simulation and backwards coupling", Stochastic Models, 14, 187–203, 1998.

Foster, F.G., "On the Stochastic Matrices Associated with Certain Queuing Processes", The Annals of Mathematical Statistics 24 (3), 1953.

Franceschetti, M. and R. Meester, *Random Networks for Communication*, Cambridge University Press, 2007.

Fréchet, M., *Recherches Théoriques Modernes sur le Calcul des Probabilités. Second livre.* Hermann, Paris, 1938.

Frieze, A. and M. Karonski, *Introduction to Random Graphs*, Camridge University Press, 2015.

Frigesi, A., C.-R. Hwang and L. Younès, "Optimal spectral properties of reversible stochastic matrices", Monte Carlo methods and the simulation of Markov random fields, The Annals of Applied Probability, 2, 3, 610–628, 1992.

Frobenius, G., "Über Matrizen aus nicht negativen Elementen", Sitzungsber. Knigl. Preuss. Akad. Wiss., 456–477, 1912.

Frobenius, G.,"Über Matrizen aus positiven Elementen, 1", Sitzungsber. Knigl. Preuss. Akad. Wiss.:, 471–476, 1908.

Frobenius, G., "Über Matrizen aus positiven Elementen, 2", Sitzungsber. Knigl. Preuss. Akad. Wiss., 514–518, 1909.

Gallager, R., *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.

Gantmacher, F., *The Theory of Matrices*, Volume 2, AMS Chelsea Publishing, 2000. (First ed. in 1959, with a different title: "Applications of the theory of matrices".)

Geman, D., *Random Fields and Inverse Problems in Imaging*. In: Lecture Notes in Mathematics, 1427, 113–193, Springer, Berlin, 1990.

Geman, S. and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE Transactions of Pattern Analysis and Machine Intelligence, 6, 721–741, 1984.

Gibbs, W., *Elementary Principles of Statistical Mechanics*, Yale University Press, 1902.

Gilbert, E. N., "Random graphs", Annals of Mathematical Statistics 30, 1141–1144, 1959.

Glauber, R., "Time-dependent statistics of the Ising model", J. Math. Physics, 4, 294–307.

Griffeath, D., "A Maximal coupling for Markov chains", Z. Wahrsch. Verw. Gebiete, 31, 95–106, 1975.

Griffeath, D., "Coupling methods for Markov chains", in: *Studies in Probability and Ergodic Theory*, Advances in Mathematics Supplementary Studies, vol. 2 (G.-C. Rota, ed.), Academic Press, New York, 1978.

Grimmett, G.R., "A Theorem on random fields", Bulletin of the London Mathematical Society, 81–84, 1973.

Grimmett, G., *Percolation*, Springer, 2nd ed. 1999.

Grimmett, G., *Probability on Graphs*, Cambridge University Press, 2010.

Grinstead, C.M. and J.L. Snell, *Introduction to Probability*, American Mathematical Society, 1997.

Häggström, O. and K. Nelander, "On exact simulation of Markov random fields using coupling from the past", Scand. J. Statist. 26, 395–411, 1999

Häggström, O., *Finite Markov Chains and Algorithmic Applications*, London Mathematical Society Student texts, 52, Cambridge University Press, 2002.

Hajek, B., "Cooling schedule for optimal annealing", Mathematics of Operations Research 13, 311–329, 1988.

Hajnal, J., "The ergodic properties of nonhomogeneous finite Markov chains", Proc. Cambridge Philos. Society, 52, 67–77, 1956.

Hajnal, J., "Weak ergodicity in nonhomogeneous Markov chains", Proc. Cambridge Philos. Society, 54, 236–246, 1958.

Hammersley, J.M. and P. Clifford, "Markov fields on finite graphs and lattices", unpublished manuscript, 1968.
http://www.statslab.cam.ac.uk/∼grg/books/hammfest/hamm-cliff.pdf.

Harris, T.E., *The Theory of Branching Processes*, Dover, 1989.

Harris, T.E., "A lower bound for the critical probability in a certain percolation process", Proc. Cambridge Math. Soc., 56, 13–20, 1960.

Hastings, W.K., "Monte Carlo sampling methods using Markov chains and their applications", Biometrika, 57, 97–109, 1970.

Hoeffding, W., "Probability inequalities for sums of bounded random variables", Journal of the American Statistical Association, 58, 13-30, 1963.

Holley, R., "Remarks on the FKG inequalities", Communications in Mathematical Physica, 227-231, 1974.

Huber, M., Perfect sampling using bounding chains, The Annals of Applied Probability, 14, 2, 734-753, 2004.

Huffman, D.A.," A method for the construction of minimal redundancy codes", Proc. IRE, 40, 1098-1101, 1952.

Iosifescu, I., *Finite Markov Chains and their Applications*, Wiley, Chichester, UK, 1980.

Isaacson, D.L. and R.W. Madsen, "Strongly ergodic behavior for nonstationary Markov processes", Annals of Probability, 1, 329-335, 1973.

Isaacson, D.L. and R.W. Madsen, *Markov Chains*, Wiley, NY, 1976.

Ising, E., "Beitrag zur Theorie des Ferromagnetismus", Zeitschrift für Physiks, 31, 253–258, 1925.

Janson, S., Luczak, T. and A. Rućinski, *Random Graphs*, Wiley, 2000.

Jerrum, M. and A. Sinclair, "Approximating the permanent", SIAM Journal of Computing, 18, 1149–1178, 1989.

Kac, M., *Random Walk and the Theory of Brownian Motion*, American Mathematical Monthly, 54, 369–391, 1947.

Kakutani, S., "Markov processes and the Dirichlet problem", Proc. Jap. Acad., **21**, 227–233, 1945.

Karlin, S., *A First Course in Stochastic Processes*, Academic Press, NY, 1966.

Karlin, S. and M.H. Taylor, *A First Course in Stochastic Processes*, Academic Press, NY, 1975.

Keilson, J., *Markov Chain Models, Rarity and Exponentiality*, Springer, NY, 1979.

Kemeny, J.G. and J.L. Snell, *Finite Markov chains*, Van Nostrand, 1960.

Kemeny, J.G., Snell, J.L. and A.W. Knapp, *Denumerable Markov chains*, Van Nostrand, 1960.

Kemeny, J.G., "Generalization of fundamental matrix", Linear Algebra and its Applications, 38, 193–206, 1991.

Kesten, H., *Percolation Theory for Mathematicians*, Birkhäuser, Boston, 1982.

Kinderman, R. and J.L. Snell, *Markov Random Fields and their Applications*, Contemporary Math., Vol. 1, Providence, RI: Amer. Math. Soc., 1980.

Kingman, J.F.C., *Regenerative phenomena*, Wiley, 1972.

Kirkpatrick, S., Gelatt, C.D. and M.P. Vecchi, "Optimization by Simulated Annealing", Science, 220, 671–680, 1982.

Klenke, A., *Probability Theory: A Comprehensive Course*, Universitext, Springer-Verlag, 2008 (2nd ed. 2014).

Knuth, D.E., *The Art of Computer Programming; Vol. I: Sorting and Searching*, Addison–Wesley, 1973.

Knuth, D.E. and A.C. Yao, "The complexity of random number generaion", in *Algorithms and Complexity: Recent Results and New Directions*, J.F. Traub ed., 357-428, Academic Press, 1976.

Kolmogorov, A., "Anfangsgründe der Theorie der Markoffschen ketten mit unendich vielen möglichen Zuständen", Rec. Math. Moskov (Mat. Sbornik), 1, 43, 607–610, 1936.

Kraft, L.G., *A device for quantizing, grouping and coding amplitude modulated pulses*, MS thesis Dept. El. Eng., MIT, 1949.

Lanford, O.E. and D. Ruelle, "Observables at infinity and states with short range correlations in statistical mechanics", Communications in Mathematical Physics, 13, 194–215, 1969.

Le Cam, L., "An approximation theorem for the Poisson binomial distribution", Pacific J. Math, **10**, 1181–1197, 1960.

Levin, D.A., Peres, Y. and E.L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, 2009.

Lindvall, T., *Lectures on the Coupling Method*, Wiley, NY, 1992.

Liu, J., "Eigenanalysis for a Metropolis sampling scheme with comparisons to rejection sampling and importance sampling", Statistics and Computing, 1995.

Liu, J., *Monte Carlo Strategies in Scientic Computing*, Springer Series in Statistics. Springer, 2001.

Lovasz, L., "Random walks on graphs: a survey", Combinatorics, Paul Erdös is eighty, 1–46, 1993.

Luczak, T., "On the equivalence of two basic models of random graphs", in *Proceedings of Random Graphs 87*, M. Karonski, J. Jaworski and A. Rucinski eds., 151–158, Wiley, Chichester, 1990.

Lyons, R. and Y. Peres, *Probability on Trees and Networks*, Cambridge University Press, 2016.

MacKay, D.J., *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

Markov, A.A., "Extension of the law of large numbers to dependent events" (in Russian), Bull. Soc. Phys. Math. Kazan, 2, 15, 155-156, 1906.

S. Meyn, S. and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Cambridge University Press, 2009 (first ed. by Springer–Verlag, 1993).

McEliece, R., *The Theory of Information and Coding*, Cambridge University Press, 2002.

McMillan, B., "Two inequalities implied by unique decipherability", IEEE Trans. Information Theory 2 (4): 115–116, 1956.

Metropolis, N., M.N. Rosenbluth, A.W. Rosenbluth, A.H. Teller and E. Teller, "Equations of state calculations by fast computing machines", J. Chem. Phys., 21, 1087–92, 1953.

Mihaïl, M., *Combinatorial aspects of expanders*, Ph.D. dissertation, Department of Computer Science, Harvard University, 1989.

Mitzenmacher, M. and E. Upfal, *Probability and Computing*, Cambridge University Press, 2005.

Montenegro, R. and P. Tetali, "Mathematical Aspects of Mixing Times in Markov Chains", Foundations and Trends in Theoretical Computer Science: Vol. 1: No. 3, pp 237–354, 2006.

Motwani, R. and P. Raghavan, *Randomized algorithms*, Cambridge University Press, 1995.

Nash-Williams, C.S.J.A., "Random walk and electric currents in networks", Proceedings of the Cambridge Philosophical Society, 55, 181–194, 1959.

Norris, J.R., *Markov chains*, Cambridge University Press, 1997.

Orey, S., *Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand, London, 1971.

Pakes, A.G., "Some conditions for ergodicity and recurrence of Markov chains", Oper. Res. 17, 1058-1061, 1969.

Perron, O., "Zur Theorie der Matrices", Mathematische Annalen 64 (2): 248–263, 1907.

Peskun, P.H., "Optimum Monte Carlo sampling using Markov chains", Biometrika, 60, 3, 607–612, 1973.

Peierls, R., "On Ising's model of ferromagnetism", Proceedings of the Cambridge Philosophical Society, 32, 477–481, 1936.

Pitman, J., "Uniform rates of convergence for Markov chain transition probabilities", Z. für W., 29, 193-227, 1974.

Pitman, J., "On coupling of Markov chains", Z. für W., 35, 4, 315–322, 1976.

Polya, G., "Über eine Aufgabe betreffend die Irrfahrt im Strassennetz", Math. Ann., **84**, 149–160, 1921.

Potts, R.B., "Some generalized order-disorder transformations", Proceedings of the Cambridge Philosophical Society, 48, 106–109, 1952.

Propp, J.G. and D.B. Wilson, "Exact sampling with coupled Markov chains and applications to statistical mechanics", Rand. Struct. Algs, 9, 223-252, 1996.

Rabin, M.O, "Probabilistic algorithm for testing primality", Journal of Number Theory, 12, 1, 128–138, 1980.

Rabin, M.O, *Fingerprinting by random polynomials*, Center for Research in Computing Technology Harvard University Report TR-15-81, 1981.

Revuz, D., *Markov Chains*, North-Holland, Amsterdam, 1984.

B.D. Ripley, F.P. Kelly, "Markov point processes", *Journal of the London Mathematical Society*, 15, 188–192, 1977.

Rom, R. and M. Sidi, *Multiple Access Protocols, Performance and Analysis*, Springer, NY, 1990.

Sanov, I.N., "On the probability of large deviations of random variables", Mat. Sbornik 42, 1144, 1957.

Schwartz, J.T., "Fast probabilistic algorithms for verification of polynomial identities", Journal of the ACM, **27**, 701–717, 1980.

Seneta, E., *Nonnegative Matrices and Markov Chains*, 2nd edition, Springer, NY, 1981.

Shamir, E. and J.H. Spencer, "Sharp concentration of the chromatic number in random graphs $\mathcal{G}(n,p)$", Combinatorica, **7**, 1, 121–129, 1987.

Shannon, C.E., "A mathematical theory of communication", Bell Syst. Tech. J., 27, 379–423, 623–656, 1948.

Shannon, C.E. and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.

Sinclair, A. and M. Jerrum, "Approximate counting, uniform generation and rapidly mixing Markov chains", Inform. and Comput., **82**, 93–133, 1989.

Sinclair, A., "Improved bounds for mixing rates of Markov chains on combinatorial structures", Tech. Report, Dept. Computer Science, U. of Edinburgh, 1990.

Sinclair, A., *Randomness and Computation*, Lecture notes CS271 Fall 2011, http://www.cs.berkeley.edu/ sinclair/cs271/f11.html

Smith, W.L., "Regenerative stochastic processes", Proc. R. Soc. A 232, 6–31, 1955.

Tsybakov, B.S. and V.A. Mikhailov, "Random multiple packet access: Part-and-try algorithm", Probl. Information Transmission, 16, 4, 305–317, 1980.

Tunstall, B.K., *Synthesis of noiseless compression codes*, Ph.D. Dissertation, Georgia Institute of Technology, 1967.

van Laarhoven, P.J., and E.H. Aarts, *Simulated Annealing: Theory and Applications*, Springer, 1987.

Welch, T., "A Technique for High-Performance Data Compression", Computer 17, 6, 8–19, 1984.

Werner, W., *Percolation et Modèle d'Ising*, Cours Spécialisés, vol. 16, Société Mathématique de France, 2009.

Williams, D., *Probability and Martingales*, Cambridge University Press, 1991.

Winkler, G., *Image Analysis, Random Fields and Dynamical Monte Carlo Methods*, Springer, NY, 1995.

Yan, D. and H. Mukai, "Stochastic discrete optimization", SIAM J. on Control and Optimization, 30, 3, 549–612, 1992.

Zippel, R.E., "Probabilistic algorithms for sparse polynomials", *Proceedings* EUROSAM *1979*, Springer Lecture Notes in Computer Science, **72**, 216–226.

Ziv, J. and A. Lempel, "Compression of individual sequences via variable-rate coding", IEEE Transactions on Information Theory, 24, 5, 530–536, 1978.

Index

P. Brémaud, *Discrete Probability Models and Methods,*
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6