This is a volume in
PROBABILITY AND MATHEMATICAL STATISTICS

Z. W. Birnbaum, founding editor
David Aldous, Y. L. Tong, series editors

A list of titles in this series appears at the end of this volume.

Identifiability in Stochastic Models

Characterization of Probability Distributions

B. L. S. PRAKASA RAO
Indian Statistical Institute
New Delhi, India

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Boston San Diego New York
London Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

Copyright © 1992 by Academic Press, Inc.
All rights reserved.
No part of this publication may be reproduced or
transmitted in any form or by any means, electronic
or mechanical, including photocopy, recording, or
any information storage and retrieval system, without
permission in writing from the publisher.

ACADEMIC PRESS, INC.
1250 Sixth Avenue, CA 92101

United Kingdom Edition published by
ACADEMIC PRESS LIMITED
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Prakasa Rao, B. L. S., date
Identifiability in stochastic models : characterization of
probability distributions / B. L. S. Prakasa Rao
p. cm. — (Probability and mathematical statistics)
Incl.udes bibliographical references and index.
ISBN 0-12-564015-3 (alk. paper)
1. Distribution (Probability theory) 2. Stochastic analysis.
I. Title. II. Series.
QA273.6.P73 1992
519.2'4—dc20 91-40038
CIP

Printed in the United States of America
92 93 94 95 EB 9 8 7 6 5 4 3 2 1

To
the three Great Ladies
my grandmother Parimi Moorthemma
my mother Bhagavatula Saradamba
my wife Bhagavatula Vasanta
for
their love and affection

Preface

The problem of identifiability is basic to all statistical methods and data analysis and it occurs in diverse areas such as reliability theory, survival analysis, econometrics, etc., where stochastic modeling is widely used. In many fields, the objective of the investigator's interest is not just the population or the probability distribution of an observable random variable but the physical structure or model leading to the probability distribution. Identification problems arise when observations can be explained in terms of one of several available models. In many problems of parameteric statistical inference, it is assumed that the family of probability distributions is completely known but for a set of unknown parameters. Any statistical procedure developed for estimation of these parameters is meaningful only if the unknown parameters are identifiable. The theory of competing risks in survival analysis is another area where identifiability is essential for the validity of the statistical procedures developed.

Identification problems in econometrics deal with the possibility of drawing inferences from observed samples obtained from an underlying theoretical structure. An important aspect of econometric theory involves derivation of conditions under which a given structure is identifiable. Lack of identification is a reflection of lack of sufficient information to discriminate between alternative structures. As Koopmans and Reiersol (1950) point out, the identification problem is "a general and fundamental problem arising in many fields of inquiry, as a concomitant of the scientific procedure that postulates the existence of a structure." However, they caution that "....the temptation to specify models in such a way as to produce identifiability of relevant characteristics is (should be) resisted." Another area where the problem of identifiability occurs is in the modeling of mixtures of populations. Mixtures of distributions are used quite frequently in building stochastic models in the biological and physical sciences. Identifiability of the mixing distribution is of paramount importance for modeling in this context. Mathematics dealing with the problem of identifiability *per se* is closely related to the so-called branch of "characterization problems" in probability theory. Summarization of statistical data without losing information is one of the fundamental objectives of statistical analysis. More precisely, the problem is to determine whether the knowledge of a possibly smaller set of functions of several random components is sufficient to determine the behaviour of a larger set of individual random components. Here the problem of identifiability consists in identifying the component distributions from the joint distributions of some functions of them.

The major motivation for writing this book is to bring together relevant material on identifiability as it occurs in diverse fields mentioned at the beginning as well as to discuss some new results on identifiability or characterization of probability distributions not found elsewhere. The idea for writing this book arose during a short visit in 1986 to Oklahama State University at the invitation of Professor I.I. Kotlarski. Professor Kotlarski is a major contributor for the material discussed in the first five chapters. It is a pleasure to thank Professor Kotlarski for his interest in this project.

As with all my earlier books, the Indian Statistical Institute has continued its support for this academic venture as well. I am grateful for its support. Thanks are due to V.P. Sharma for TeXing the manuscript on the word processor in an excellent manner in spite of the innumerable changes made during the TeXing process. My children Gopi, Vamsi and Venu and my wife Vasanta are now familiar with my idiosyncrasies after watching me work over four books and they put up with them. Thanks are due to them.

B.L.S. Prakasa Rao
New Delhi
January, 1992

Chapter 1

Introduction

Suppose X and Y are independent normally distributed random variables. Then $Z = X + Y$ is also normally distributed. Cramér (1936) proved that the converse is true, that is, if the sum Z of two independent random variables X and Y has a normal distribution, then both X and Y have to be normally distributed. On the other hand, if X and Y are independent standard normal random variables, then the ratio $U = X/Y$ has a Cauchy distribution. However the converse is not true as noted by Mauldon (1956). In other words, it is possible for X and Y to be independent and not normally distributed and yet $U = X/Y$ could have a Cauchy distribution. The following example due to Steck (1958) illustrates this situation. Another example is given in Laha (1958).

Example 1.1 : Suppose X and Y are independent and identically distributed (i.i.d.) random variables with the symmetric density function

$$f(x) = \frac{\sqrt{2}}{\pi} \frac{x^2}{1 + x^4}, -\infty < x < \infty. \tag{1.1}$$

We leave it to the reader to check that $U = X/Y$ has the standard Cauchy distribution. It is easy to see that U can also be written in the form $U = (1/Y)/(1/X)$ where $1/Y$ and $1/X$ are i.i.d. random variables with the symmetric density function

$$f^*(x) = \frac{\sqrt{2}}{\pi}\frac{1}{1+x^4}, -\infty < x < \infty. \tag{1.2}$$

Hence $U = X'/Y'$ has the standard Cauchy distribution when X' and Y' are i.i.d. with density function $f^*(x)$.

Laha (1959a,b) and Kotlarski (1960) gave a complete description of the family of all density functions f such that the quotient X/Y follows the standard Cauchy distribution whenever X and Y are i.i.d. with density f. A natural question now is to find additional conditions under which the normal distribution can be identified from the distribution of quotients of independent random variables. Kotlarski (1967) proved the following result. Suppose X, Y and Z are independent real-valued random variables with density functions symmetric about zero. Denote $U = X/Z$ and $V = Y/Z$. Then X, Y and Z are normally distributed with a common variance σ^2 if and only if the bivariate random vector (U, V) follows the bivariate Cauchy density given by

$$f_{U,V}(u,v) = \frac{1}{2\pi}\frac{1}{(1+u^2+v^2)^{3/2}}, \quad -\infty < u, v < \infty. \tag{1.3}$$

We will come back to the proof of this theorem later in this book.

What is to be noted above is that even though the distribution of the ratio $U = X/Y$ of two independent random variables X and Y does not determine the distributions of X and Y, the situation changes completely if we consider the joint distribution of two ratios $U = X/Z$ and $V = Y/Z$ where X, Y and Z are three independent random variables. Kotlarski's result indicates that if the joint distribution is bivariate Cauchy, then X, Y and Z are normally distributed under some technical assumptions.

Let us consider the problem in a more general framework.

Suppose $(\mathcal{X}, \mathcal{B})$ is a measurable space and $\mathcal{P}$ is a family of probability measures on $(\mathcal{X}, \mathcal{B})$. Let $Y = f(X)$ be a measurable map from $(\mathcal{X}, \mathcal{B})$ into $(\mathcal{Y}, \tau)$. Let Q_P^Y be the probability measure induced by Y on $(\mathcal{Y}, \tau)$ when P is the probability measure on $(\mathcal{X}, \mathcal{B})$. We are concerned with mappings $f(\cdot)$ such that Q_P^Y is the same for all $P \in \mathcal{P}$ denoted by $Q_{\mathcal{P}}^Y$ and if for some probability measure P' on $(\mathcal{X}, \mathcal{B}), Q_{P'}^Y = Q_{\mathcal{P}}^Y$, then $P' \in \mathcal{P}$.

Example 1.2 (Kovalenko (1960)) : Suppose $X_1, X_2, \cdots, X_n, n \geq 3$ are independent and identically distributed random variables with density $p(x - \theta), -\infty < \theta < \infty$. Let

$$\boldsymbol{Y} = (X_1 - X_n, X_2 - X_n, \ldots, X_{n-1} - X_n). \tag{1.4}$$

Kovalenko (1960) has proved that the distribution of $\boldsymbol{Y}$ determines the characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx \tag{1.5}$$

to within a factor of the form $e^{i\gamma t}$ on every interval where $\phi(t) \neq 0$. In particular, if $\phi(t) \neq 0$ for all t, then the statistic $\boldsymbol{Y}$ determines the distribution of X_i up to location for $1 \leq i \leq n$. This conclusion also holds if $\phi(t)$ is analytic in some neighbourhood of zero (see Theorem 2.1.1 in Chapter 2).

Example 1.3 (Zinger (1956)) : Let $\theta = (\mu, \sigma), -\infty < \mu < \infty, \sigma > 0$ and

$$p(x, \theta) = \frac{1}{\sigma} \phi \left(\frac{x - \mu}{\sigma} \right) \tag{1.6}$$

where ϕ is the standard normal density. Suppose $X_1, \ldots, X_n$ are independent and identically distributed random variables with density $p(x, \theta)$. Define

$$\boldsymbol{Y} = (Y_1, \ldots, Y_n) \tag{1.7}$$

where

$$Y_k = \frac{X_k - \overline{X}}{s}, \quad 1 \leq k \leq n \tag{1.8}$$

with $\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$ and $s^2 = \sum_{k=1}^{n} (X_k - \overline{X})^2$. It is easy to see that $\sum_{i=1}^{n} Y_i = 0$

and $\sum_{i=1}^{n} Y_i^2 = 1$. Hence the distribution of $\boldsymbol{Y}$ is concentrated on the set

$$\{\boldsymbol{y} : \sum_{i=1}^{n} y_i = 0, \ \sum_{i=1}^{n} y_i^2 = 1\} \tag{1.9}$$

which is of dimension $(n-2)$. It is known that the distribution of $\boldsymbol{Y}$ is uniform on a $(n-2)$-dimensional sphere when the density $p(x,\theta)$ is given by (1.6). Zinger (1956) proved the converse i.e., if the distribution of $\boldsymbol{Y}$ is uniform on a $(n-2)$-dimensional sphere, then the distributions of X_i, $1 \leq i \leq n$ are normal.

Example 1.4 (Prohorov (1965)) : Let $\theta = (\mu, \sigma), -\infty < \mu < \infty, \sigma > 0$ and

$$p(x,\theta) = \frac{1}{\sigma}\, p(\frac{x-\mu}{\sigma}) \tag{1.10}$$

where $p(\cdot)$ is a symmetric density function in the sense that $p(x) = p(-x)$, bounded and satisfies Cramér's condition

$$\int_{-\infty}^{\infty} e^{hx} p(x)dx < \infty \tag{1.11}$$

in a neighbourhood of zero. Suppose $X_1, X_2, \ldots, X_n, n \geq 6$ are i.i.d. with density $p(x,\theta)$. Define

$$\boldsymbol{Z}^* = \left[\left(\frac{Y_4 - Y_3}{Y_2 - Y_1}\right)^2, \left(\frac{Y_6 - Y_5}{Y_2 - Y_1}\right)^2\right] \tag{1.12}$$

where Y_k is as defined by (1.8). Let

$$Z_1^* = \frac{(Y_4 - Y_3)^2}{(Y_2 - Y_1)^2}, \ Z_2^* = \frac{(Y_6 - Y_5)^2}{(Y_2 - Y_1)^2} \ .$$

It is easy to see that

$$Z_1^* = \frac{(X_4 - X_3)^2}{(X_2 - X_1)^2}, \ Z_2^* = \frac{(X_6 - X_5)^2}{(X_2 - X_1)^2} \ .$$

Suppose $p'(\cdot)$ is another symmetric density possibly different from p such that the distribution of $\boldsymbol{Z}^*$ under $p(\cdot)$ is the same as the distribution of $\boldsymbol{Z}^*$

under $p'(\cdot)$. Let

$$\begin{aligned} \boldsymbol{W} &= (\log Z_1^*,\ \log Z_2^*) \\ &= (\log(X_4 - X_3)^2 - \log(X_2 - X_1)^2,\ \log(X_6 - X_5)^2 - \log(X_2 - X_1)^2). \end{aligned}$$

It can be checked that Cramér's condition is satisfied by the distribution of $\log(X_2 - X_1)^2$ under the density $p(\cdot)$ as well as under $p'(\cdot)$. Furthermore the distribution of $\boldsymbol{W}$ is the same under $p(\cdot)$ and $p'(\cdot)$. An application of the result given in Example 1.2 shows that the distribution of $\log(X_2 - X_1)^2$ is determined up to shift and hence the distribution of $(X_2 - X_1)^2$ up to scale. But the distribution of $(X_2 - X_1)$ is symmetric. Hence the distribution of $X_2 - X_1$ is also determined up to scale. If the density $p(\cdot)$ is standard normal, then $X_2 - X_1$ is also normal under the density $p'(\cdot)$ and, by Cramér's theorem, it follows that X_1 and X_2 are (independent) normally distributed random variables. In general, for a symmetric density $p(\cdot)$, the distribution of X_1 is determined by the distribution of $X_2 - X_1$ uniquely to within a shift parameter.

The type of problems discussed above may be termed as problems of identification of families of distributions of some random variables from some functions of them. Several problems of this kind are investigated in Chapter 2 to Chapter 5.

Other types of identifiability problems arise in econometrics, reliability or survival analysis and other areas where stochastic modeling is of paramount importance. Since stochastic modeling is modeling certain phenomena through a probability structure or probability distribution, the problems of identification that come up in stochastic modeling are similar to those discussed above. For instance, suppose a random variable X is distributed normally with mean $\mu_1 - \mu_2$ and variance 1 where μ_1 and μ_2 are real. It is clear that there is no way to estimate μ_1 and μ_2 separately using X and that the parameters μ_1 and μ_2 are not identifiable. However,

$\mu_1 - \mu_2$ is estimable and in fact X is the unique uniformly minimum variance unbiased estimator of $\mu_1 - \mu_2$. There are an infinite number of pairs (μ_1, μ_2) which give rise to the same value $\mu_1 - \mu_2$. Let us consider another example – of a regression model. Let

$$Y_1 = \alpha_0 + \alpha_1\eta_1 + \varepsilon_1 ,$$

and

$$Y_2 = \beta_0 + \beta_1 Y_1 + \varepsilon_2$$

where $\alpha_0, \alpha_1, \beta_0$ and β_1 are unknown parameters and η_1, ε_1 and ε_2 are random variables with $E(\varepsilon_1) = 0$ and $E(\varepsilon_2) = 0$. Suppose Y_1 is not observable but Y_2 is. Then

$$Y_2 = \gamma_0 + \gamma_1\eta_1 + \varepsilon_3$$

where

$$\gamma_0 = \beta_0 + \alpha_0\beta_1, \gamma_1 = \beta_1\alpha_1, \varepsilon_3 = \varepsilon_2 + \beta_1\varepsilon_1.$$

From the general theory on linear models, it follows that γ_0 and γ_1 are identifiable (estimable) under some reasonable assumptions on the random variables η_1, ε_1 and ε_2. However β_0, α_0 and β_1 are not identifiable individually in general. In problems of statistical inference, estimation of a parameter is not meaningful unless it is identifiable. The problem of identifiability occurs in reliability and survival analysis. Suppose an individual is subject to two possible causes of death (or two types of terminal illness). Let X_i be the lifetime of the individual exposed to cause i alone for $i = 1, 2$. In general $X_i, i = 1, 2$ are not observable but $Y = \min(X_1, X_2)$ is observable. Does the distribution of Y identify the distributions of X_1 and X_2? Mixtures of distributions are used in building probability models in the biological and physical sciences. In order to devise statistical procedures for inferential aspects, an important problem is identifiability of the mixing distribution. The problem of identifiability for these types of stochastic models is discussed in Chapters 6 to 8.

Chapter 2

Identifiability of Distributions of Random Variables Based on Some Functions of Them

In this chapter we consider characterization of distributions of independent random variables from the joint distribution of some functions of them. For instance, if X, Y and Z are three independent random variables, we would like to know conditions under which the joint distribution of $U = g(X, Y, Z)$ and $V = h(X, Y, Z)$ determine either the individual distributions of X, Y and Z or the family to which they belong when $g(\cdot)$ and $h(\cdot)$ are specified. $g(\cdot)$ and $h(\cdot)$ could be linear or nonlinear functions or they could be the maximum and minimum functions, and so on.

2.1 Identifiability by Sums (or Ratios)

Let X_1, X_2 and X_3 be three independent real-valued random variables.

Define

$$\begin{aligned} Z_1 &= X_1 - X_3, \\ Z_2 &= X_2 - X_3 . \end{aligned} \tag{2.1}$$

The following result was proved by Kotlarski (1967).

Theorem 2.1.1 : If the characteristic function of (Z_1, Z_2) does not vanish, then the joint distribution of (Z_1, Z_2) determines the distributions of X_1, X_2, X_3 up to a change of the location.

Proof : Let $\phi(t_1, t_2)$ denote the characteristic function (c.f.) of (Z_1, Z_2) and $\phi_k(t)$ be the c.f. of X_k for $1 \leq k \leq 3$. Then

$$\begin{aligned} \phi(t_1, t_2) &= E\{\exp[i(t_1 Z_1 + t_2 Z_2)]\} \\ &= E\{\exp[i(t_1(X_1 - X_3) + t_2(X_2 - X_3))]\} \\ &= E\{\exp[i(t_1 X_1 + t_2 X_2 - (t_1 + t_2) X_3)]\} \\ &= \phi_1(t_1)\phi_2(t_2)\phi_3(-t_1 - t_2) \end{aligned} \tag{2.2}$$

by the independence of $X_i, 1 \leq i \leq 3$. Since $\phi(t_1, t_2) \neq 0$ for all t_1 and t_2 by hypothesis, it follows that $\phi_i(t) \neq 0$ for all t for $1 \leq i \leq 3$.

Let Y_1, Y_2, Y_3 be another set of three independent random variables with characteristic functions $\psi_i(t), 1 \leq i \leq 3$ respectively satisfying the conditions in Theorem 2.1.1. Let

$$\begin{aligned} W_1 &= Y_1 - Y_3 , \\ W_2 &= Y_2 - Y_3, \end{aligned} \tag{2.3}$$

and $\psi(t_1, t_2)$ be the characteristic function of (W_1, W_2). Suppose that the joint distributions of (Z_1, Z_2) and (W_1, W_2) are the same. Then

$$\phi(t_1, t_2) = \psi(t_1, t_2), \qquad -\infty < t_1, t_2 < \infty \tag{2.4}$$

and it follows from (2.2) that

$$\phi_1(t_1)\phi_2(t_2)\phi_3(-t_1-t_2)$$
$$= \psi_1(t_1)\psi_2(t_2)\psi_3(-t_1-t_2), \qquad -\infty < t_1, t_2 < \infty. \tag{2.5}$$

Furthermore $\phi_i(t) \neq 0$ and $\psi_i(t) \neq 0$ for $1 \leq i \leq 3$ for all t by hypothesis. Let

$$\gamma_i(t) = \psi_i(t)/\phi_i(t), \qquad 1 \leq i \leq 3. \tag{2.6}$$

Observe that $\gamma_i(\cdot), 1 \leq i \leq 3$ are continuous complex-valued functions with $\gamma_i(0) = 1, 1 \leq i \leq 3$ satisfying the equation

$$\gamma_1(t_1)\gamma_2(t_2)\gamma_3(-t_1-t_2) = 1, -\infty < t_1, t_2 < \infty. \tag{2.7}$$

Let $t_1 = t$ and $t_2 = 0$ in (2.7). Then

$$\gamma_1(t)\gamma_3(-t) = 1, \quad -\infty < t < \infty. \tag{2.8}$$

Let $t_2 = t$ and $t_1 = 0$. Then

$$\gamma_2(t)\gamma_3(-t) = 1, \quad -\infty < t < \infty\,. \tag{2.9}$$

Substituting for $\gamma_1(t)$ and $\gamma_2(t)$ in terms of $\gamma_3(t)$ in (2.7), it follows that

$$\gamma_3(t_1+t_2) = \gamma_3(t_1)\gamma_3(t_2), \quad -\infty < t_1, t_2 < \infty \tag{2.10}$$

with $\gamma_3(0) = 1$. It is known that the only measurable solution of this equation is

$$\gamma_3(t) = e^{ct} \tag{2.11}$$

where c is a complex number. Hence, it follows from (2.8) and (2.9) that

$$\gamma_1(t) = \gamma_2(t) = \gamma_3(t) = e^{ct}. \tag{2.12}$$

Relation (2.6) implies that

$$\psi_j(t) = \phi_j(t)e^{ct}, \quad 1 \leq j \leq 3\,. \tag{2.13}$$

Since $\psi_j(t) = \overline{\psi_j(-t)}$ and $\phi_j(t) = \overline{\phi_j(-t)}$, being characteristic functions, it follows that $c = i\beta$ where β is a real number. Therefore

$$\psi_j(t) = \phi_j(t)e^{i\beta t}, \quad 1 \leq j \leq 3 \tag{2.14}$$

where β is a real number. From the uniqueness theorem for characteristic functions, it follows that X_j and $Y_j - \beta$ have the same distribution for $1 \leq j \leq 3$. This proves that the distributions of X_1, X_2, X_3 are determined up to a change of location. ■

Remarks 2.1.1 : If, in Theorem 2.1.1, $E(X_3)$ exists and is preassigned, then the distributions of X_1, X_2, X_3 are uniquely determined from the distribution of $(X_1 - X_3, X_2 - X_3)$. If the characteristic function of (Z_1, Z_2) in Theorem 2.1.1 is infinitely divisible, then the conclusion of Theorem 2.1.1 holds since the characteristic function of an infinitely divisible law is nonvanishing.

Remarks 2.1.2 : A slight variation of Theorem 2.1.1 for location parameter families is given in Prohorov (1965). Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2$ and $\boldsymbol{X}_3$ are independent and identically distributed ℓ-dimensional random vectors $\boldsymbol{X}_j = (X_j^{(1)}, \ldots, X_j^{(\ell)})$ with density $p(\boldsymbol{x}, \boldsymbol{\theta}) = p(\boldsymbol{x} - \boldsymbol{\theta})$. Further assume that $\boldsymbol{\theta} \in \Theta$ which is a k-dimensional subspace of R^ℓ. Without loss of generality, assume that $\Theta = \{\boldsymbol{\theta} \in R^\ell : \theta_{k+1} = \cdots = \theta_\ell = 0\}$. Further suppose that Cramér's condition holds, that is,

$$E_0[e^{(\boldsymbol{h}, \boldsymbol{X})}] = \int_{R^\ell} e^{(\boldsymbol{h}, \boldsymbol{x})} p(\boldsymbol{x}) d\boldsymbol{x} < \infty \tag{2.15}$$

for $\boldsymbol{h}$ in a neighbourhood of zero in R^ℓ.

Theorem 2.1.2 : Let $\boldsymbol{X}_3' = (X_3^{(1)}, \ldots, X_3^{(k)}, 0, \ldots, 0)$ and define

$$\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$$

where

$$\boldsymbol{Y}_1 = \boldsymbol{X}_1 - \boldsymbol{X}_3', \boldsymbol{Y}_2 = \boldsymbol{X}_2 - \boldsymbol{X}_3'.$$

Then the distribution of $\boldsymbol{Y}$ does not depend on θ and the distribution of $\boldsymbol{Y}$ determines the distribution of $\boldsymbol{X}_1$ up to shift . In fact, the distribution of $\boldsymbol{X}_1$ belongs to the family $\{p(\boldsymbol{x} - \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$.

Remarks 2.1.3 : The conclusion in Theorem 2.1.2 also holds under the condition that the common characteristic function $\phi(\boldsymbol{t})$ of $\boldsymbol{X}_j$ is nonzero for all $\boldsymbol{t} \in R^\ell$ instead of (2.15).

Remarks 2.1.4 : An analogue of Theorem 2.1.2 holds for scale parameter families in multidimensions. Suppose $\boldsymbol{X}_j, 1 \leq j \leq 3$, are i.i.d. ℓ-dimensional random vectors with density

$$p(\boldsymbol{x}, \theta) = \frac{1}{\theta^\ell} p(\frac{x_1}{\theta}, \ldots, \frac{x_\ell}{\theta}), \quad 0 < \theta < \infty.$$

Let $\boldsymbol{X}_j = (X_j^{(1)}, \ldots, X_j^{(\ell)})$. Consider the 2ℓ - dimensional random vector

$$\boldsymbol{V}_j = (\log |X_j^{(1)}|, \ldots, \log |X_j^{(\ell)}|, \text{sgn } X_j^{(1)}, \ldots, \text{sgn } X_j^{(\ell)}).$$

The density of $\boldsymbol{V}_j$ is of the form

$$q(\boldsymbol{v}, \phi) = q(v^{(1)} - \phi, \ldots, v^{(\ell)} - \phi, v^{(\ell+1)}, \ldots, v^{(2\ell)})$$

where $\phi = \log \theta$. Define

$$\boldsymbol{V}'_j = (\log |X_j^{(1)}|, \ldots, \log |X_j^{(\ell)}|, 0, \ldots, 0)$$

and

$$\boldsymbol{Y} = (\boldsymbol{V}_1 - \boldsymbol{V}'_3, \boldsymbol{V}_2 - \boldsymbol{V}'_3).$$

Prohorov (1965) proved the following theorem as a consequence of Theorem 2.1.2.

Theorem 2.1.3 : Suppose $p(\boldsymbol{x})$ is bounded and satisfies Cramér's condition (2.15). Then the distribution of $\boldsymbol{Y}$ does not depend on θ and the distribution of $\boldsymbol{Y}$ determines the distribution of $\boldsymbol{X}_1$ up to scale. In fact the distribution of $\boldsymbol{X}_1$ belongs to the scale parameter family

$$\{\frac{1}{\theta^\ell} p(\frac{x_1}{\theta}, \ldots, \frac{x_\ell}{\theta}), 0 < \theta < \infty\}.$$

Let us now consider an extension of Theorem 2.1.1 to linear forms.

Suppose X_1, X_2 and X_3 are three independent real-valued random variables. Consider two linear forms

$$Z_1 = a_1X_1 + a_2X_2 + a_3X_3, \tag{2.16}$$

$$Z_2 = b_1X_1 + b_2X_2 + b_3X_3, \tag{2.17}$$

such that $a_i : b_i \neq a_j : b_j$ for $i \neq j$. Rao (1971) proved the following result.

Theorem 2.1.4 : If the characteristic function of (Z_1, Z_2) does not vanish, then the distribution of (Z_1, Z_2) determines the distributions of X_1, X_2, X_3 up to a change of location.

The proof of this theorem rests on the following lemmas and corollaries due to Rao (1966, 1967).

Lemma 2.1.1 : Suppose $\gamma_1, \gamma_2, \ldots, \gamma_n$ are continuous complex-valued functions defined on the real line. If there exist distinct nonzero reals $c_1, c_2, \ldots, c_n$ such that

$$\sum_{i=1}^{n} \gamma_i(t + c_i u) = A(t|u) + B(u|t) \tag{2.18}$$

where $A(x|y)$ and $B(x|y)$ are polynomials in x of degree less than or equal to a and b respectively for any fixed y, then the $\gamma_i(t), 1 \leq i \leq n$, are polynomials of degree less than or equal to $a + b + n$.

Corollary 2.1.1 : Suppose, in (2.18),

$$A(t|u) = A(u) \text{ and } B(u|t) = B(t) \tag{2.19}$$

where $A(\cdot)$ and $B(\cdot)$ are continuous functions. Then the $\gamma_i(t), A(t)$ and $B(t)$ are all polynomials of degree less than or equal to n.

Lemma 2.1.2 : Suppose the expression on the right side in the equation (2.18) is of the form

$$A(t) + B(u) + P_k(t, u) \tag{2.20}$$

where $A(t)$ and $B(u)$ are continuous functions and $P_k(t,u)$ is a polynomial of degree k in t for fixed u and in u for fixed t. Then $\gamma_i(t), A(t)$ and $B(t)$ are all polynomials of degree less than or equal to max (n,k).

Lemma 2.1.3 : If the right side of (2.18) consists only of $P_k(t,u)$ as given in (2.20), then the $\gamma_i(t)$ are polynomials of degree less than or equal to $\max(n-2,k)$.

We refer the reader to Rao (1966, 1967) for proofs of these and related results (cf. Kagan *et al.* (1973)). Let us now prove Theorem 2.1.4.

Proof of Theorem 2.1.4 : Let $\phi_i(t)$ be the c.f. of $X_i, 1 \leq i \leq 3$. Since the c.f. of (Z_1, Z_2) does not vanish, it follows that $\phi_i(t) \neq 0$ for all t and for $1 \leq i \leq 3$. Let $\eta_i(t) = \log\ \phi_i(t)$ denote the continuous branch of the logarithm of the c.f. $\phi_i(t)$ with $\eta_i(0) = 0$. Suppose $\psi_i(t), 1 \leq i \leq 3$ is another set of possible characteristic functions for $X_i, 1 \leq i \leq 3$ satisfying the hypothesis. Let $\zeta_i(t) = \log\ \psi_i(t)$ as before and define

$$\gamma_i(t) = \eta_i(t) - \zeta_i(t), \quad -\infty < t < \infty. \tag{2.21}$$

Since the characteristic functions of (Z_1, Z_2) are the same for the choice $\phi_i, 1 \leq i \leq 3$, as well as $\psi_i, 1 \leq i \leq 3$, it follows that

$$\gamma_1(a_1 t + b_1 u) + \gamma_2(a_2 t + b_2 u) + \gamma_3(a_3 t + b_3 u) = 0 \tag{2.22}$$

for all t, u real. Since $a_i : b_i \neq a_j : b_j$ for $i \neq j, 1 \leq i, j \leq 3$ by hypothesis, the equation (2.22) can be written in one of the following forms depending on the values of a_i and b_i :

$$\begin{aligned}
&\text{(i)} \quad \gamma_1(t + c_1 u) + \gamma_2(t + c_2 u) + \gamma_3(t + c_3 u) = 0, \\
&\qquad\ c_1 \neq c_2 \neq c_3 \neq 0\ ; \\
&\text{(ii)} \quad \gamma_1(t + c_1 u) + \gamma_2(t + c_2 u) = A(t), c_1 \neq c_2 \neq 0\ ; \\
&\text{or} \\
&\text{(iii)} \quad \gamma_1(t + cu) = A(t) + B(u), c \neq 0\ .
\end{aligned} \tag{2.23}$$

An application of Lemmas 2.1.2 and 2.1.3 implies that each $\gamma_k(t)$ must be linear in t and hence

$$\phi_k(t) = \psi_k(t) \exp[\alpha_k t + \beta_k], \quad -\infty < t < \infty \tag{2.24}$$

where α_k and β_k are constants. Since ϕ_k and ψ_k are characteristic functions, it follows that $\beta_k = 0$ and $\alpha_k = id_k$ where d_k is real. Hence, for $1 \leq k \leq 3$,

$$\phi_k(t) = \psi_k(t) e^{id_k t}, \quad -\infty < t < \infty\,. \tag{2.25}$$

This proves the theorem. ■

Remarks 2.1.5 : The assumption in Theorem 2.1.1 that the characteristic function of (Z_1, Z_2) does not vanish can be replaced by the assumption that $X_k, 1 \leq k \leq 3$, have *analytic* characteristic functions. Since $\phi(0) = 1$ for any characteristic function, $\phi(t) \neq 0$ for t in a neighbourhood of zero. All the arguments given in the proof of Theorem 2.1.1 will be valid for t complex inside the region $\{t : |t| < t_0\}$, for some $t_0 > 0$ where the characteristic functions do not vanish. Because of the analyticity of the characteristic functions, the relation (2.14) will be valid for the whole real line. Similar remarks hold for Theorem 2.1.4 as the conclusions in Lemmas 2.1.1 to 2.1.3 continue to hold in regions $|t| < t_0, |u| < u_0$, if the corresponding equations hold in those regions.

Remarks 2.1.6 : If the assumption about the nonvanishing property of the characteristic function of (Z_1, Z_2) is omitted, then the conclusion of Theorem 2.1.1 does not hold, as shown by the following example .

Example 2.1.1 : Let $X_i, i = 1, 2$, and $Y_i, i = 1, 2$, be independent random variables with the characteristic functions $\phi_i, i = 1, 2$ and $\psi_i, i = 1, 2$ respectively given by

$$\phi_1(t) = \phi_2(t) = \psi_1(t) = \psi_2(t) = \begin{cases} 0 & \text{if} \quad |t| > 1 \\ 1 - |t| & \text{for} \quad |t| \leq 1\,. \end{cases} \tag{2.26}$$

Let X_3 be a random variable independent of X_1 and X_2 with the characteristic function

$$\phi_3(t) = \begin{cases} 0 & \text{if} \quad |t| > 2 \\ 1 - \frac{|t|}{2} & \text{if} \quad |t| \leq 2 \end{cases} \tag{2.27}$$

and Y_3 be another random variable independent of Y_1 and Y_2 with the characteristic function

$$\psi_3(t) = 1 - \frac{|t|}{2} \text{ for } |t| \leq 2, \psi_3(t+4) = \psi_3(t). \tag{2.28}$$

It is easy to see that

$$\phi_1(t_1)\phi_2(t_2)\phi_3(-t_1 - t_2) = \psi_1(t_1)\psi_2(t_2)\psi_3(-t_1 - t_2) \tag{2.29}$$

for all t_1 and t_2. Clearly $\psi_3(t)$ and $\phi_3(t)$ are not equal and $\psi_3(t)$ is not of the form $\phi_3(t)e^{i\delta t}$ for any real δ. Hence (X_1, X_2, X_3) and (Y_1, Y_2, Y_3) are sets of independent random variables such that the distribution of X_3 and the distribution of Y_3 do not just differ by location but are completely different, and yet the joint distribution of $(X_1 - X_3, X_2 - X_3)$ is the same as that of $(Y_1 - Y_3, Y_2 - Y_3)$.

Remarks 2.1.7 : Sasvari (1986) and Sasvari and Wolff (1986) improved the result in Theorem 2.1.1. They showed that if any two of the characteristic functions of $X_i, 1 \leq i \leq 3$ are analytic or have no zeroes, then the distribution of $(X_1 - X_3, X_2 - X_3)$, determines the distributions of X_1, X_2 and X_3 up to a change of location. Bondesson (1974) proved that Theorem 2.1.1 holds if either $\phi_i, 1 \leq i \leq 3$ or $\psi_i, 1 \leq i \leq 3$ in (2.5) have "no gaps" (cf. Lemma 4.4 of Bondesson (1974)).

It is easy to extend Theorem 2.1.1 to n independent real random variables, in the following form.

Theorem 2.1.5 : Let $X_i, 1 \leq i \leq n$, be n independent real-valued random variables and define

$$Z_i = X_i - X_n, \quad 1 \leq i \leq n-1. \tag{2.30}$$

Suppose the characteristic function of $\boldsymbol{Z} = (Z_1, \ldots, Z_{n-1})$ does not vanish. Then the distribution of $\boldsymbol{Z}$ determines the distributions of $X_1, ..., X_n$ up to change of location.

Remarks 2.1.8 : Rao (1971) extended Theorem 2.1.4 to p linear functions $Z_i, 1 \leq i \leq p$, of n independent random variables X_i. He obtained conditions sufficient for determining the *smallest* number p of linear functions $\boldsymbol{Z}_i, 1 \leq i \leq p$, such that their joint distribution specifies the distribution of each random variable $X_i, 1 \leq i \leq n$, up to a change of location. He showed that

$$\frac{p(p-1)}{2} < n \leq \frac{p(p+1)}{2} \ .$$

For details, see Rao (1971) (cf. Kagan *et al.* (1973)).

Remarks 2.1.9 : Theorem 2.1.1 can be extended to n-dimensional random vectors $\boldsymbol{X}_k$. Rao (1971) proved the following theorem.

Theorem 2.1.6 : Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2$ and $\boldsymbol{X}_3$ are independent n-dimensional random vectors. Consider two linear functions

$$\begin{aligned} \boldsymbol{Z}_1 &= \boldsymbol{A}_1\boldsymbol{X}_1 + \boldsymbol{A}_2\boldsymbol{X}_2 + \boldsymbol{A}_3\boldsymbol{X}_3 \ , \\ \boldsymbol{Z}_2 &= \boldsymbol{B}_1\boldsymbol{X}_1 + \boldsymbol{B}_2\boldsymbol{X}_2 + \boldsymbol{B}_3\boldsymbol{X}_3 \end{aligned} \tag{2.31}$$

such that

(i) $\boldsymbol{A}_i$ is either zero or a nonsingular matrix and only one of $\boldsymbol{A}_i$ is zero for any i,

(ii) $\boldsymbol{B}_i$ is either zero or a nonsingular matrix and only one of $\boldsymbol{B}_i$ is zero for any i,

(iii) $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are not simultaneously zero for any i, and

(iv) the matrix $\boldsymbol{B}_i\boldsymbol{A}_j^{-1} - \boldsymbol{B}_j\boldsymbol{A}_i^{-1}$ is nonsingular when defined.

If the characteristic function of $(\boldsymbol{Z}_1, \boldsymbol{Z}_2)$ does not vanish, then the distributions of the $\boldsymbol{X}_i$ are determined up to a change of location.

This theorem follows as a consequence of extensions of Lemmas 2.1.1 to

2.1.3 to the multivariate case. For details, see Rao (1971). We will consider more general results dealing with random elements taking values in Hilbert space later in this book.

Theorem 2.1.1 can be rephrased in terms of ratios instead of sums in the following way.

Theorem 2.1.7 : Suppose X_1, X_2, X_3 are three independent positive random variables. Let $Y_1 = X_1/X_3$ and $Y_2 = X_2/X_3$. If the characteristic function of $(\log Y_1, \log Y_2)$ does not vanish, then the distribution of (Y_1, Y_2) determines the distributions of X_1, X_2, X_3 up to a change of scale.

Proof: This theorem follows immediately from Theorem 2.1.1 since log X_k, $k = 1, 2, 3$, satisfy the assumptions of Theorem 2.1.1. ∎

Remarks 2.1.10 : The positivity condition on the random variables X_k, $1 \le k \le 3$, in Theorem 2.1.7 can be replaced by the conditions that the random variables X_k have distributions symmetric about the origin and that $P(X_k = 0) = 0$ for $1 \le k \le 3$.

Applications

Theorem 2.1.8(Characterization of the normal distribution): Let X_1, X_2, X_3 be three independent random variables symmetrically distributed about the origin with $P(X_k = 0) = 0, 1 \le k \le 3$. Let (Y_1, Y_2) be defined by

$$Y_1 = \frac{X_1}{X_3} \text{ and } Y_2 = \frac{X_2}{X_3}. \tag{2.32}$$

A necessary and sufficient condition for the independent random variables $X_k, 1 \le k \le 3$, to be normally distributed with a common variance σ^2 is that the joint density of (Y_1, Y_2) is the bivariate Cauchy density given by

$$g_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2\pi(1 + y_1^2 + y_2^2)^{3/2}}, \quad -\infty < y_1, y_2 < +\infty. \tag{2.33}$$

Proof : Let $\phi_k(t)$ denote the characteristic function of $\log|X_k|$. If X_k has a normal distribution with mean 0 and variance σ^2, then

$$\begin{aligned}\phi_k(t) &= E[\exp\{it\log|X_k|\}] \\ &= (\sqrt{2}\,\sigma)^{it}\,\Gamma((1+it)/2)\pi^{-\frac{1}{2}} \qquad (2.34)\end{aligned}$$

and hence the characteristic function of $(\log|Y_1|, \log|Y_2|)$ is given by

$$\begin{aligned}\phi(t_1,t_2) &= \phi_1(t_1)\phi_2(t_2)\phi_3(-t_1-t_2) \qquad (2.35)\\ &= \pi^{-\frac{3}{2}}\Gamma(\frac{1+it_1}{2})\Gamma(\frac{1+it_2}{2})\Gamma(\frac{1-i(t_1+t_2)}{2})\;. \qquad (2.36)\end{aligned}$$

Note that $\phi(t_1,t_2)$ is nonvanishing for all t_1 and t_2. It can be checked that the characteristic function of $(\log|Y_1|, \log|Y_2|)$ is given by (2.36) whenever (Y_1, Y_2) has joint density given by (2.33). Hence the distributions of X_i are *determined* up to change of scale by Theorem 2.1.7 and Remarks 2.1.10. If the X_i are normally distributed with mean 0 and variance σ^2, then one is led to the equation (2.36). Hence the random variables $X_i, 1 \le i \le 3$, have to be normally distributed with mean zero and the same variance σ^2. ■

Theorem 2.1.9 (Characterization of the gamma distribution): Let X_1, X_2, X_3 be three independent positive random variables. Define

$$Y_1 = \frac{X_1}{X_3} \text{ and } Y_2 = \frac{X_2}{X_3}. \qquad (2.37)$$

A necessary and sufficient condition for X_k to have a gamma distribution with parameters p_k and $\alpha, 1 \le k \le 3$ is that the joint density of (Y_1, Y_2) is the bivariate beta density given by

$$\begin{aligned}g(y_1,y_2) &= \frac{\Gamma(p_1+p_2+p_3)}{\Gamma(p_1)\Gamma(p_2)\Gamma(p_3)}\frac{y^{p_1-1}y_2^{p_2-1}}{(1+y_1+y_2)^{p_1+p_2+p_3}}\,, \quad y_1>0, y_2>0 \qquad (2.38)\\ &= 0 \qquad\qquad \text{otherwise}\;.\end{aligned}$$

Proof : Let $\phi_k(t)$ denote the characteristic function of $\log X_k$. If X_k has the gamma distribution with parameters p_k and α, then

$$\begin{aligned}\phi_k(t) &= E[\exp(it\ \log\ X_k)] \\ &= \alpha^{-it}\frac{\Gamma(p_k+it)}{\Gamma(p_k)} \qquad (2.39)\end{aligned}$$

and hence the characteristic function of $(\log Y_1, \log Y_2)$ is

$$\begin{aligned}\phi(t_1,t_2) &= \phi_1(t_1)\phi_2(t_2)\phi_3(-t_1-t_2)\\ &= \frac{\Gamma(p_1+it_1)}{\Gamma(p_1)}\frac{\Gamma(p_2+it_2)}{\Gamma(p_2)}\frac{\Gamma(p_3-it_1-it_2)}{\Gamma(p_3)}.\end{aligned} \tag{2.40}$$

It can be checked that the characteristic function of $(\log Y_1, \log Y_2)$, whenever (Y_1, Y_2) has the joint density (2.38), is also given by the expression on the right side of (2.40). An application of Theorem 2.1.7 gives the result. ■

Theorem 2.1.10 (Characterization of the gamma distribution) : Let X_1, X_2, X_3 be three independent positive random variables and let (U_1, U_2) be defined by

$$U_1 = \frac{X_1}{X_1+X_2}, \; U_2 = \frac{X_1+X_2}{X_1+X_2+X_3}. \tag{2.41}$$

A necessary and sufficient condition for X_k to be gamma-distributed with parameters p_k and $\alpha, 1 \le k \le 3$, is that U_1 and U_2 are independent beta–distributed random variables , U_1 with parameters (p_1, p_2) and U_2 with parameters (p_1+p_2, p_3).

Theorem 2.1.11 (Another characterization of the normal distribution) : Let X_1, X_2, X_3 be independent random variables symmetrically distributed about the origin and satisfying the condition
$P(X_k = 0) = 0, 1 \le k \le 3$. Let

$$V_1 = \frac{X_1}{\sqrt{X_1^2+X_2^2}} \text{ and } V_2 = \frac{\sqrt{X_1^2+X_2^2}}{\sqrt{X_1^2+X_2^2+X_3^2}}. \tag{2.42}$$

A necessary and sufficient condition for X_k to be normally distributed with a common variance σ^2 for $1 \le k \le 3$ is that V_1 and V_2 are independent and V_1, V_2 are distributed according to the densities

$$f_1(v) = \begin{cases} \frac{1}{\pi\sqrt{1-v^2}} & \text{if } |v| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.43}$$

and

$$f_2(v) = \begin{cases} \frac{v}{\sqrt{1-v^2}} & \text{if } 0 < v < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.44}$$

respectively.

For the proofs of Theorems 2.1.10 to 2.1.11, see Kotlarski (1967). Related results characterizing the chi-square distribution and the normal distribution using the Student's t distribution are given in Kotlarski (1966a,b).

Suppose that X_0 and X_1 are independent identically distributed random variables distributed according to the chi-square distribution with n degrees of freedom. It is known that

$$Y = \frac{\sqrt{n}}{2} \frac{X_1 - X_0}{\sqrt{X_1 X_0}} \tag{2.45}$$

has the t distribution with n degrees of freedom (cf. Cacoullos (1965)). The problem is to find out whether the chi-square distribution can be characterized by this property. The answer is "no." There are independent positive random variables identically distributed with a distribution different from the chi-square distribution for which Y follows the t distribution. However, suppose there are three independent random variables X_0, X_1, X_2 and let

$$Y_1 = \frac{\sqrt{n}}{2} \frac{X_1 - X_0}{\sqrt{X_1 X_0}}, Y_2 = \frac{\sqrt{n}}{2} \frac{X_2 - X_0}{\sqrt{X_2 X_0}}. \tag{2.46}$$

Note that

$$Y_1 = \frac{\sqrt{n}}{2}(\sqrt{Z_1} - \frac{1}{\sqrt{Z_1}}), Y_2 = \frac{\sqrt{n}}{2}(\sqrt{Z_2} - \frac{1}{\sqrt{Z_2}}) \tag{2.47}$$

where

$$Z_1 = \frac{X_1}{X_0} \text{ and } Z_2 = \frac{X_2}{X_0}. \tag{2.48}$$

Kotlarski (1966a) proved that, using a suitable distribution for the random vector (Y_1, Y_2), one can characterize the chi-square distribution of the random variables X_0, X_1, X_2.

The results discussed in Theorems 2.1.1, 2.1.4 and 2.1.6 only indicate or give sufficient conditions under which the joint distribution of two or several linear forms determine the distributions of the individual summands up to change of location. But no method has been given to explicitly determine the distributions of individual summands if the joint distribution of suitable linear forms is known. We now consider this problem.

Remarks 2.1.11 (Explicit determination of the distributions of the individual summands) : Let X_0, X_1, X_2 be independent real-valued random variables with characteristic functions ϕ_0, ϕ_1, ϕ_2 respectively. Assume that ϕ_0, ϕ_1, ϕ_2 are nonvanishing everywhere. Define

$$Y_1 = X_0 + X_1 \text{ and } Y_2 = X_0 + X_2. \tag{2.49}$$

Let $\psi(t_1, t_2)$, the characteristic function of (Y_1, Y_2), be known. Then

$$\psi(t_1, t_2) = \phi_0(t_1 + t_2)\phi_1(t_1)\phi_2(t_2), \quad -\infty < t_1, t_2 < \infty. \tag{2.50}$$

Clearly $\psi(t_1, t_2)$ is nonvanishing. Let $t_2 = 0$. Then the equation (2.50) gives

$$\phi_0(t_1)\phi_1(t_1) = \psi(t_1, 0), \quad -\infty < t_1 < \infty. \tag{2.51}$$

Let $t_1 = 0$ in (2.50). Then we have

$$\phi_0(t_2)\phi_2(t_2) = \psi(0, t_2), \quad -\infty < t_2 < \infty. \tag{2.52}$$

Relations (2.50) to (2.52) show that

$$\begin{aligned} &\phi_0(t_1 + t_2)\phi_1(t_1)\phi_2(t_2)\psi(t_1, 0)\psi(0, t_2) \\ &= \psi(t_1, t_2)\phi_0(t_1)\phi_1(t_1)\phi_0(t_2)\phi_2(t_2) \end{aligned} \tag{2.53}$$

and hence

$$\phi_0(t_1 + t_2) = \frac{\psi(t_1, t_2)}{\psi(t_1, 0)\psi(0, t_2)}\phi_0(t_1)\phi_0(t_2) \tag{2.54}$$

for t_1, t_2 real. Let $\psi_i(t) = \log \phi_i(t)$ be the continuous branch of the logarithm of $\phi_i(\cdot)$ with $\psi_i(0) = 0$. Then it follows that

$$\psi_0(t_1' + t_2) = \log \frac{\psi(t_1', t_2)}{\psi(t_1', 0)\psi(0, t_2)} + \psi_0(t_1') + \psi_0(t_2) \tag{2.55}$$

for all t_1' and t_2 real. Integrating on both sides of the equation (2.55) with respect to t_1' over the interval $[0, t_1]$, it can be checked that

$$\begin{aligned}\int_0^{t_1} \psi_0(t_1' + t_2)dt_1' &= \int_0^{t_1} \log \frac{\psi(t_1', t_2)}{\psi(t_1', 0)\psi(0, t_2)} dt_1' \\ &+ \int_0^{t_1} \psi_0(t_1')dt_1' + \int_0^{t_1} \psi_0(t_2)dt_1'. \quad (2.56)\end{aligned}$$

Let $t = t_1' + t_2$ in the integral on the left hand side of (2.56). Then we have

$$\begin{aligned}\int_{t_2}^{t_1+t_2} \psi_0(t)dt &= \int_0^{t_1} \log \frac{\psi(t_1', t_2)}{\psi(t_1', 0)\psi(0, t_2)} dt_1' \\ &+ \int_0^{t_1} \psi_0(t)dt + t_1\psi_0(t_2). \quad (2.57)\end{aligned}$$

Rewriting (2.55) in the form

$$\psi_0(t_1 + t_2') = \log \frac{\psi(t_1, t_2')}{\psi(t_1, 0)\psi(0, t_2')} + \psi_0(t_1) + \psi_0(t_2') \quad (2.58)$$

and integrating on both sides of this equation with respect to t_2' over the interval $[0, t_2]$, it can be shown that

$$\begin{aligned}\int_{t_1}^{t_1+t_2} \psi_0(t)dt &= \int_0^{t_2} \log \frac{\psi(t_1, t_2')}{\psi(t_1, 0)\psi(0, t_2')} dt_2' \\ &+ \int_0^{t_2} \psi_0(t)dt + t_2\psi_0(t_1). \quad (2.59)\end{aligned}$$

Equating (2.57) and (2.59), we have

$$\begin{aligned}t_1\psi_0(t_2) - t_2\psi_0(t_1) &= \int_0^{t_2} \log \frac{\psi(t_1, t_2')}{\psi(t_1, 0)\psi(0, t_2')} dt_2' \\ &- \int_0^{t_1} \log \frac{\psi(t_1', t_2)}{\psi(t_1', 0)\psi(0, t_2)} dt_1' \quad (2.60)\end{aligned}$$

for all t_1, t_2. Dividing both sides of the equation (2.60) by $t_1 t_2$, we have

$$\begin{aligned}\frac{\psi_0(t_2)}{t_2} - \frac{\psi_0(t_1)}{t_1} &= \frac{1}{t_1 t_2}[\int_0^{t_2} \log \frac{\psi(t_1, t_2')}{\psi(t_1, 0)\psi(0, t_2')} dt_2' \\ &- \int_0^{t_1} \log \frac{\psi(t_1', t_2)}{\psi(t_1', 0)\psi(0, t_2)} dt_1'] \quad (2.61)\end{aligned}$$

for $-\infty < t_1, t_2 < \infty, t_1 t_2 \neq 0$. Let $t_2 = t$ and $t_1 \to 0$. Assume that $m_0 = E(X_0)$ is finite and that the interchange of the limit and integral sign is permitted in the following computations. Then, we have

$$\lim_{t\to 0}\frac{\psi_0(t)}{t} = im_0 \tag{2.62}$$

and, from (2.61),

$$\begin{aligned}\frac{\psi_0(t)}{t} &= im_0 + \frac{1}{t}\lim_{t_1\to 0}[\int_0^t \frac{1}{t_1}\log\frac{\psi(t_1,v)}{\psi(t_1,0)\psi(0,v)}dv \\ &\quad -\frac{1}{t_1}\int_0^{t_1}\log\frac{\psi(u,t)}{\psi(u,0)\psi(0,t))}du] \\ &= im_0 + \frac{1}{t}\lim_{t_1\to 0}[\int_0^t \frac{1}{t_1}\log\frac{\psi(t_1,v)}{\psi(t_1,0)\psi(0,v)}dv] \\ &\quad -\log\frac{\psi(0,t)}{\psi(0,0)\psi(0,t)} \\ &= im_0 + \frac{1}{t}\lim_{t_1\to 0}[\int_0^t \frac{1}{t_1}\log\frac{\psi(t_1,v)}{\psi(t_1,0)\psi(0,v)}dv] \\ &= im_0 + \frac{1}{t}\int_0^t \frac{\partial}{\partial u}[\log\frac{\psi(u,v)}{\psi(u,0)\psi(0,v)}]|_{u=0}dv. \end{aligned}\tag{2.63}$$

Hence

$$\psi_0(t) = im_0 t + \int_0^t \frac{\partial}{\partial u}[\log\frac{\psi(u,v)}{\psi(u,0)\psi(0,v)}]_{u=0}\,dv\,. \tag{2.64}$$

Using this formula for $\psi_0(t)$, one can compute $\phi_0(t)$ and hence $\phi_1(t)$ and $\phi_2(t)$ by the relations

$$\phi_1(t) = \frac{\psi(t,0)}{\phi_0(t)},\ \phi_2(t) = \frac{\psi(0,t)}{\phi_0(t)},\quad -\infty < t < \infty. \tag{2.65}$$

Relations (2.64) and (2.65) give explicit formulae for computing the characteristic functions of X_0, X_1 and X_2 given the characteristic function of $(X_0 + X_1, X_0 + X_2)$.

The results given above are due to Kotlarski .

2.2 Identifiability by Maxima

Let X_0, X_1 and X_2 be independent real-valued random variables. Define

$$Y_1 = X_0 V X_1 \text{ and } Y_2 = X_0 V X_2 \tag{2.66}$$

where aVb denotes $\max(a,b)$. It is of interest to know whether the joint distribution of (Y_1, Y_2) determines the individual distributions of X_0, X_1 and X_2.

Theorem 2.2.1 : The joint distribution of (Y_1, Y_2) *uniquely* determines the distributions of X_0, X_1 and X_2 provided the supports of X_0, X_1 and X_2 are the same.

Proof : Let F_i and F_i^* denote alternate possibilities for the distribution functions of $X_i, i = 0, 1, 2$. Let the joint distribution of (Y_1, Y_2) be denoted by $G(y_1, y_2)$. Then, for $-\infty < y_1 \le y_2 < +\infty$,

$$\begin{aligned} G(y_1, y_2) &= P(Y_1 \le y_1, Y_2 \le y_2) \\ &= P(X_0 \le y_1, X_1 \le y_1, X_0 \le y_2, X_2 \le y_2) \\ &= P(X_0 \le y_1, X_1 \le y_1, X_2 \le y_2) \\ &= F_0(y_1)F_1(y_1)F_2(y_2) \end{aligned} \tag{2.67}$$

by the independence of X_0, X_1 and X_2. Since F_i^* is the alternate possible distribution for $X_i, i = 0, 1, 2$, it follows that

$$F_0(y_1)F_1(y_1)F_2(y_2) = F_0^*(y_1)F_1^*(y_1)F_2^*(y_2) \tag{2.68}$$

for $-\infty < y_1 \le y_2 < \infty$. Let $y_2 \to \infty$. Then it follows that

$$F_0(y_1)F_1(y_1) = F_0^*(y_1)F_1^*(y_1), \quad -\infty < y_1 < \infty. \tag{2.69}$$

Relations (2.68) and (2.69) show that

$$F_2(y_2) = F_2^*(y_2) \tag{2.70}$$

for all $-\infty < y_2 < \infty$ provided $F_0(y_1)F_1(y_1) > 0$. Note that the support of F_0F_1 is the same as the support of $F_0^*F_1^*$ from (2.69). Let us now choose $-\infty < y_2 \le y_1 < \infty$ and compute

$$\begin{aligned} G(y_1,y_2) &= P(Y_1 \le y_1, Y_2 \le y_2) \\ &= P(X_0 \le y_1, X_1 \le y_1, X_0 \le y_2, X_2 \le y_2) \\ &= P(X_0 \le y_2, X_1 \le y_1, X_2 \le y_2) \\ &= F_0(y_2)F_1(y_1)F_2(y_2) \\ &= F_0(\min(y_1,y_2))F_1(y_1)F_2(y_2). \end{aligned} \tag{2.71}$$

This relation leads to the equation

$$F_0(y_2)F_1(y_1)F_2(y_2) = F_0^*(y_2)F_1^*(y_1)F_2^*(y_2) \tag{2.72}$$

for $-\infty < y_2 \le y_1 < \infty$. Let $y_1 \to \infty$. Then

$$F_0(y_2)F_2(y_2) = F_0^*(y_2)F_2^*(y_2) \tag{2.73}$$

for $-\infty < y_2 < \infty$. Hence, from (2.72) and (2.73), we have

$$F_1(y_1) = F_1^*(y_1) \tag{2.74}$$

whenever $-\infty < y_1 < \infty$ provided $F_0(y_2)F_2(y_2) > 0$. Note again that the support of F_0F_2 is the same as the support of $F_0^*F_2^*$ from (2.73). Since the supports of F_0, F_1 and F_2 are all the same, it can be seen from (2.68), (2.70) and (2.74) that

$$F_i(y) = F_i^*(y), i = 0,1,2 \tag{2.75}$$

over the common support of X_0, X_1, X_2. Hence the distribution of (Y_1, Y_2) uniquely determines the distributions of X_0, X_1 and X_2. ∎

Remarks 2.2.1 : It is known that $Y_1 = X_0 V X_1$ alone cannot determine the distributions of X_0 and X_1 uniquely unless X_0 and X_1 are i.i.d. random variables. For a discussion on this topic, see Section 7.3.

The results of this section are due to Kotlarski.

2.3 Identifiability by Minima

A result analogous to Theorem 2.2.1 holds for minima of random variables.

Theorem 2.3.1 : Let X_0, X_1 and X_2 be three independent random variables. Define

$$Y_1 = X_0 \Lambda X_1 \text{ and } Y_2 = X_0 \Lambda X_2 \tag{2.76}$$

where $a\Lambda b$ denotes $\min(a, b)$. Suppose the distribution functions F_0, F_1 and F_2 of X_0, X_1, X_2 respectively satisfy the conditions

$$F_i(a) = 1, F_i(w) < 1 \text{ for } w < a, i = 0, 1, 2 \tag{2.77}$$

for some $a \leq +\infty$. Then the joint distribution of (Y_1, Y_2) *uniquely* determines the distributions of X_0, X_1 and X_2.

Proof : This theorem can be derived either as a consequence of Theorem 2.2.1 or directly. Let $\overline{F}_i = 1 - F_i$. It is easy to check that

$$P(Y_1 > y_1, Y_2 > y_2) = \overline{F}_0(y_1 V y_2)\overline{F}_1(y_1)\overline{F}_2(y_2) \tag{2.78}$$

for all y_1 and y_2 and the rest of the proof is similar to that of Theorem 2.2.1. ■

Remarks 2.3.1 (Explicit determination of the component distributions): Given the joint distribution $G_{Y_1,Y_2}(y_1, y_2)$ of (Y_1, Y_2) in Theorem 2.2.1, one can explicitly write down the distributions of F_0, F_1 and F_2. In fact, it is easy to check that

$$F_0(z) = \frac{G_{Y_1,Y_2}(z,\infty)G_{Y_1,Y_2}(\infty,z)}{G_{Y_1,Y_2}(z,z)}, \tag{2.79}$$

$$F_1(x) = \frac{G_{Y_1,Y_2}(x,x)}{G_{Y_1,Y_2}(\infty,x)} F_2(y) = \frac{G_{Y_1,Y_2}(y,y)}{G_{Y_1,Y_2}(y,\infty)} \tag{2.80}$$

using the relation (2.71).

Example 2.3.1 : Let X_0, X_1 and X_2 be independent positive random variables whose distribution functions satisfy the conditions $F(+0) = 0$,

$0 < F(w) < 1$ for $w > 0$; that is, the support of F is $[0,\infty)$. Define

$$Y_1 = X_0 \Lambda X_1 \text{ and } Y_2 = X_0 \Lambda X_2. \tag{2.81}$$

Suppose that

$$P(Y_1 > y_1, Y_2 > y_2) = \begin{cases} \exp(-ay_1 - by_2 - c\max(y_1, y_2)) & \\ & \text{if } y_1 > 0, y_2 > 0 \\ \exp(-(a+c)y_1) & \text{if } y_1 > 0, y_2 \leq 0 \\ \exp(-(b+c)y_2) & \text{if } y_1 \leq 0, y_2 > 0 \\ 1 & \text{if } y_1 \leq 0, y_2 \leq 0. \end{cases} \tag{2.82}$$

Then all the components X_0, X_1 and X_2 are exponentially distributed with positive parameters a, b and c respectively. This result follows from Theorem 2.3.1. It is easy to check from the definition of (Y_1, Y_2) in Theorem 2.3.1 that

$$\begin{aligned} H(y_1, y_2) &= P(Y_1 > y, Y_2 > y_2) \\ &= P(X_0 > y_1 V y_2) P(X_1 > y_1) P(X_2 > y_2) \\ &= \overline{F}_0(y_1 V y_2) \overline{F}_1(y_1) \overline{F}_2(y_2) \end{aligned} \tag{2.83}$$

and, given $H(y_1, y_2)$, one can find $\overline{F}_0, \overline{F}_1$ and $\overline{F}_2$ from $H(\cdot,\cdot)$ by the following relations :

$$\overline{F}_0(z) = \frac{H(z,-\infty)H(-\infty,z)}{H(z,z)} \ , \tag{2.84}$$

$$\overline{F}_1(x) = \frac{H(x,x)}{H(-\infty,x)} \tag{2.85}$$

and

$$\overline{F}_2(y) = \frac{H(y,y))}{H(y,-\infty)} \ . \tag{2.86}$$

It is easy to show that X_0, X_1 and X_2 have exponential densities when H is given by (2.82), using the relations (2.84) and (2.85).

The results in this section are due to Kotlarski.

2.4 Identifiability by Maximum and Minimum

Let X_0, X_1 and X_2 be independent random variables. Define

$$Y_1 = X_0 \Lambda X_1 \text{ and } Y_2 = X_0 V X_2 \ . \tag{2.87}$$

Theorem 2.4.1 : Let F_i be the distribution function of $X_i, i = 0, 1, 2$. Suppose that, for some *fixed* a, b, x_0, q satisfying $-\infty \le a < x_0 < b \le +\infty$, $0 < q < 1$,

$$\begin{aligned} &F_1(x) < 1, x < b; F_1(b-0) = 1 \text{ (if } b \in R), \\ &F_2(y) > 0, y > a; F_2(a+0) = 0 \text{ (if } a \in R), \\ &F_0(a+0) = 0, F_0(b-0) = 1, F_0(x_0) = q \end{aligned} \tag{2.88}$$

and F_0 is strictly increasing in (a, b). Then the joint distribution of (Y_1, Y_2) *uniquely* determines the distributions F_0, F_1 and F_2.

Proof : For $-\infty < y_1 \le y_2 < \infty$,

$$\begin{aligned} P(Y_1 > y_1, Y_2 \le y_2) &= P(X_0 > y_1, X_1 > y_1, X_0 \le y_2, X_2 \le y_2) \\ &= P(y_1 < X_0 \le y_2, X_1 > y_1, X_2 \le y_2) \\ &= (F_0(y_2) - F_0(y_1))\overline{F}_1(y_1)F_2(y_2). \end{aligned} \tag{2.89}$$

Suppose $\{F_0^*, F_1^*, F_2^*\}$ is another set of distribution functions for $\{X_0, X_1, X_2\}$ satisfying the conditions in the theorem such that the distributions of (Y_1, Y_2) under $\{F_i\}$ as well as $\{F_i^*\}$ are the same. Then, for $-\infty < y_1 \le y_2 < \infty$,

$$\begin{aligned} &[F_0^*(y_2) - F_0^*(y_1)]\overline{F_1^*}(y_1)F_2^*(y_2) \\ &\quad = [F_0(y_2) - F_0(y_1)]\overline{F}_1(y_1)F_2(y_2). \end{aligned} \tag{2.90}$$

Let $y_2 \to +\infty$ in (2.89). Then

$$\overline{F_0^*}(y_1)\overline{F_1^*}(y_1) = \overline{F}_0(y_1)\overline{F}_1(y_1), \ -\infty < y_1 < \infty. \tag{2.91}$$

Let $y_1 \to -\infty$ in (2.89). Then

$$F_0^*(y_2)F_2^*(y_2) = F_0(y_2)F_2(y_2), \ -\infty < y_2 < \infty. \tag{2.92}$$

Combining the relations (2.89) to (2.91), we have

$$[F_0^*(y_2) - F_0^*(y_1)]\overline{F_1^*}(y_1)F_2^*(y_2)\overline{F}_0(y_1)\overline{F}_1(y_1)F_0(y_2)F_2(y_2)$$
$$= [F_0(y_2) - F_0(y_1)]\overline{F}_1(y_1)F_2(y_2)\overline{F}_0^*(y_1)\overline{F}_1^*(y_1)F_0^*(y_2)F_2^*(y_2) \quad (2.93)$$

for $-\infty < y_1 \leq y_2 < \infty$. Applying the conditions (2.87), we have

$$\frac{F_0^*(y_2) - F_0^*(y_1)}{F_0(y_2) - F_0(y_1)} = \frac{\overline{F_0^*}(y_1)}{\overline{F}_0(y_1)}\frac{F_0^*(y_2)}{F_0(y_2)} \quad (2.94)$$

for $-\infty \leq a < y_1 < y_2 < b \leq \infty$. Since $F_0^*(x_0) = F_0(x_0) = q$, it follows that, for $-\infty \leq a < y \leq x_0$,

$$\frac{F_0^*(x_0) - F_0^*(y)}{F_0(x_0) - F_0(y)} = \frac{\overline{F}_0^*(y)}{\overline{F}_0(y)} \ . \quad (2.95)$$

It is easy to see that the relation (2.94) implies that

$$F_0^*(y) = F_0(y) \quad \text{for } -\infty < y \leq x_0 \ . \quad (2.96)$$

Similarly we can prove that

$$F_0^*(y) = F_0(y) \quad \text{for } x_0 \leq y < +\infty \ . \quad (2.97)$$

Relations (2.90) and (2.91) prove that

$$F_1^*(y) = F_1(y) \text{ and } F_2^*(y) = F_2(y) \quad (2.98)$$

completing the proof of the theorem. ∎

Remarks 2.4.2 (Explicit determination) : Given the joint distribution of (Y_1, Y_2), one can explicitly write down the distributions of X_0, X_1 and X_2. Let

$$\begin{aligned} H(u,v) &\equiv P(Y_1 > u, Y_2 \leq v) \\ &= \overline{F}_1(u)F_2(v)[F_0(v) - F_0(u)], \ -\infty < u < v < \infty. \end{aligned} \quad (2.99)$$

It can be checked that

$$\begin{aligned} F_0(z) &= \begin{cases} \frac{q[H(z,x_0)-H(-\infty,x_0)H(z,\infty)]}{qH(z,x_0)-H(-\infty,x_0)H(z,\infty)}, & z \leq x_0 \\ \frac{qH(x_0,\infty)H(-\infty,z)}{H(x_0,\infty)H(-\infty,z)-(1-q)H(x_0,z)}, & z \geq x_0 \end{cases} \quad (2.100) \\ \overline{F}_1(x) &= \frac{H(x,\infty)}{\overline{F}_0(x)}, F_2(y) = \frac{H(-\infty,y)}{F_0(y)}. \end{aligned}$$

where x_0 and q are as defined by (2.87). We do not give the details here.

The results in this section are due to Kotlarski (1978) .

2.5 Identifiability by Product and Minimum (or Maximum)

Let X_0, X_1 and X_2 be *positive* independent random variables. Define

$$Y_1 = X_0 \Lambda X_1 \text{ and } Y_2 = X_0 X_2 \ . \tag{2.101}$$

Theorem 2.5.1 : Suppose there exists $a_0 > 0$ such that the distribution functions F_0, F_1 and F_2 of X_0, X_1 and X_2 satisfy the conditions

$$F_i(x) < 1, i = 0, 1 \text{ for } x < a_0 \leq \infty. \tag{2.102}$$

Further suppose that there exists $\alpha_0 > 1$ such that $h_i(\alpha)$ $= E(X_i^\alpha) > 0$ and finite for $0 \leq \alpha \leq \alpha_0, i = 0, 2$ and in addition assume that there exists a fixed constant $q > 0$ such that $0 < E(X_0) = q < \infty$. Then the joint distribution of (Y_1, Y_2) *uniquely* determines the distributions of X_0, X_1 on the interval $(-\infty, a_0)$ and the moments $E(X_2^\alpha), 0 \leq \alpha \leq \alpha_0$.

Proof : Let χ_A denote the indicator function of a set A. Then, for any $0 \leq \alpha \leq \alpha_0$ and $-\infty < \beta < \infty$,

$$\begin{aligned} H(\alpha, \beta) &\equiv E[\chi_{(\beta,\infty)}(Y_1) Y_2^\alpha] \\ &= E[\chi_{(\beta,\infty)}(X_0 \Lambda X_1)(X_0 X_2)^\alpha] \\ &= E[\chi_{(\beta,\infty)}(X_0)\chi_{(\beta,\infty)}(X_1) X_0^\alpha X_2^\alpha] \\ &= E[\chi_{(\beta,\infty)}(X_0) X_0^\alpha] E[\chi_{(\beta,\infty)}(X_1)] E[X_2^\alpha] \\ &= \{\int_\beta^\infty x^\alpha \, dF_0(x)\} \overline{F}_1(\beta) h_2(\alpha). \end{aligned} \tag{2.103}$$

If F_0^*, F_1^* and F_2^* are alternate possibilities for the distribution functions of X_0, X_1 and X_2 respectively satisfying (2.101), then we have

$$H(\alpha, \beta) = \{\int_\beta^\infty x^\alpha dF_0^*(x)\} \overline{F_1^*}(\beta) h_2^*(\alpha) \tag{2.104}$$

where $h_2^*(\alpha) = E(X_2^\alpha)$ when X_2 has distribution F_2^*. Relations (2.102) and (2.103) imply that

$$\{\int_\beta^\infty x^\alpha \, dF_0^*(x)\}\overline{F_1^*}(\beta)h_2^*(\alpha)$$

$$= \{\int_\beta^\infty x^\alpha dF_0(x)\}\overline{F_1}(\beta)h_2(\alpha), 0 \le \alpha \le \alpha_0. \tag{2.105}$$

Let $\alpha = 0$. Then we have

$$\overline{F_0^*}(\beta)\overline{F_1^*}(\beta) = \overline{F_0}(\beta)\overline{F_1}(\beta), -\infty < \beta < \infty. \tag{2.106}$$

Let $\beta = 0$ in (2.104). Then we have

$$h_0^*(\alpha)h_2^*(\alpha) = h_0(\alpha)h_2(\alpha), 0 \le \alpha \le \alpha_0.$$

Relations (2.104) and (2.105) lead to the equation

$$\{\int_\beta^\infty x^\alpha dF_0^*(x)\}\overline{F_1^*}(\beta)h_2^*(\alpha)\overline{F_0}(\beta)\overline{F_1}(\beta)h_0(\alpha)h_2(\alpha)$$

$$= \{\int_\beta^\infty x^\alpha dF_0(x)\}\overline{F_1}(\beta)h_2(\alpha)\overline{F_0^*}(\beta)\overline{F_1^*}(\beta)h_0^*(\alpha)h_2^*(\alpha),$$

$$0 \le \alpha \le \alpha_0, \qquad -\infty < \beta < \infty. \tag{2.107}$$

Under the condition (2.101), $\overline{F_i}(\beta)$ and $\overline{F_i^*}(\beta)$ are positive for $i = 0, 1$ when $-\infty < \beta < a_0 \le \infty$ and hence

$$\frac{\int_\beta^\infty x^\alpha dF_0^*(x)}{\overline{F_0^*}(\beta)h_0^*(\alpha)} = \frac{\int_\beta^\infty x^\alpha dF_0(x)}{\overline{F_0}(\beta)h_0(\alpha)}, 0 \le \alpha \le \alpha_0, -\infty < \beta < a_0. \tag{2.108}$$

Since $EX_0 = q < \infty$ is the same under both F_0 and F_0^* by hypothesis, it follows that $h_0(1) = h_0^*(1) = q$. Hence

$$\frac{\int_\beta^\infty x dF_0^*(x)}{\overline{F_0^*}(\beta)} = \frac{\int_\beta^\infty x dF_0(x)}{\overline{F_0}(\beta)}, \ -\infty < \beta < a_0 \tag{2.109}$$

from (2.107) or equivalently

$$\frac{\int_\beta^\infty x d\overline{F_0^*}(x)}{\overline{F_0^*}(\beta)} = \frac{\int_\beta^\infty x d\overline{F_0}(x)}{\overline{F_0}(\beta)}, \ -\infty < \beta < a_0 \tag{2.110}$$

where $\bar{F} = 1 - F$. Integrating by parts on both sides of (2.109), we have

$$\frac{[x\overline{F_0^*}(x)]_\beta^\infty - \int_\beta^\infty \overline{F_0^*}(x)dx}{\overline{F_0^*}(\beta)} = \frac{[x\overline{F}_0(x)]_\beta^\infty - \int_\beta^\infty \overline{F_0}(x)dx}{\overline{F_0}(\beta)} . \tag{2.111}$$

Observe that $\lim_{x\to+\infty} x[1-F_0(x)] = 0$ when $E_{F_0}(X_0) = \int_{-\infty}^{\infty} xdF_0(x)$ is finite. Hence, we have

$$\frac{-\beta\overline{F_0^*}(\beta) - \int_\beta^\infty \overline{F_0^*}(x)dx}{\overline{F_0^*}(\beta)} = \frac{-\beta\overline{F_0}(\beta) - \int_\beta^\infty \overline{F_0}(x)dx}{\overline{F_0}(\beta)} \tag{2.112}$$

which leads to the equation

$$\frac{\int_\beta^\infty \overline{F_0^*}(x)dx}{\overline{F_0^*}(\beta)} = \frac{\int_\beta^\infty \overline{F_0}(x)dx}{\overline{F_0}(\beta)}, -\infty < \beta < a_0 \tag{2.113}$$

or equivalently

$$\frac{\overline{F_0^*}(\beta)}{\int_\beta^\infty \overline{F_0^*}(x)dx} = \frac{\overline{F_0}(\beta)}{\int_\beta^\infty \overline{F_0}(x)dx}, \quad -\infty < \beta < a_0. \tag{2.114}$$

Therefore

$$\log \int_\beta^\infty \overline{F_0^*}(x)dx = \log \int_\beta^\infty \overline{F_0}(x)dx + c, -\infty < \beta < a_0 \tag{2.115}$$

for some constant c. Hence

$$\int_\beta^\infty \overline{F_0^*}(x)dx = d \int_\beta^\infty \overline{F_0}(x)dx, -\infty < \beta < a_0 \tag{2.116}$$

for some constant d. Taking derivatives with respect to β on both sides, we have

$$\overline{F_0^*}(\beta) = d\,\overline{F_0}(\beta), -\infty < \beta < a_0 \ . \tag{2.117}$$

Let $\beta = 0$. Then $\overline{F_0}(0) = \overline{F_0^*}(0) = 1$ by hypothesis and hence $d = 1$ which proves that

$$F_0^*(\beta) = F_0(\beta), -\infty < \beta < a_0. \tag{2.118}$$

Relation (2.105) will imply that

$$F_1^*(\beta) = F_1(\beta), \quad -\infty < \beta < a_0 \tag{2.119}$$

and (2.104) shows that

$$h_2^*(\alpha) = h_2(\alpha), \quad 0 \leq \alpha \leq \alpha_0. \tag{2.120}$$

This completes the proof of Theorem 2.5.1. ∎

Remarks 2.5.1 (Explicit determination) : One can explicitly write down the distributions F_0, F_1 and the function h_2, given $H(\alpha, \beta)$ defined by (2.102).

In fact, let $H_1(\beta)$ be defined by

$$\frac{H_1'(\beta)}{H_1(\beta)} = \frac{H(1,0)H(0,\beta)}{H(1,0)H(0,\beta)\beta - h_0(1)H(1,\beta)} \tag{2.121}$$

where $h_0(1) = E(X_0)$ is specified. Then

$$\overline{F_0}(z) = \frac{H_1'(z)}{H_1'(0)}, \overline{F_1}(y) = \frac{H(0,y)}{\overline{F}_0(y)}, h_2(\alpha) = \frac{H(\alpha,0)}{h_0(\alpha)} \ . \tag{2.122}$$

We will not discuss the details here. The results presented here are due to Kotlarski .

A result analogous to Theorem 2.5.1 holds identifying the probability distributions through the product and the maximum. We will state the result without proof. The result is due to Kotlarski.

Theorem 2.5.2 : Let X_0, X_1 and X_2 be independent positive random variables. Define

$$Y_1 = X_0 V X_1 \text{ and } Y_2 = X_0 X_2 \ . \tag{2.123}$$

Suppose the distribution functions F_0, F_1 and F_2 of X_0, X_1 and X_2 satisfy the conditions

$$F_i(z) > 0 \text{ for } z > 0, i = 0, 1, 2 \ .$$

Further suppose that $E[X_i^\alpha] = h_i(\alpha)$ is finite and positive for $0 \leq \alpha \leq \alpha_0, \alpha_0 > 1$ for $i = 0, 2$ and $E(X_0) = q < \infty$ is a fixed positive constant.

Then the joint distribution of (Y_1, Y_2) uniquely determines the distributions of X_0, X_1 and the moments $E(X_2^\alpha), 0 \leq \alpha \leq \alpha_0$.

Let

$$\begin{aligned} H(\alpha,\beta) &= E[Y_2^\alpha \chi_{(-\infty,\beta)}(Y_1)] \\ &= h_2(\alpha)F_1(\beta)\int_0^\beta z^\alpha dF_0(z). \end{aligned} \tag{2.124}$$

Then

$$F_0(z) = \exp\{-\int_z^\infty \frac{H(0,u)h_2(1)}{uH(0,u)h_2(1) - H(1,u)} du\}, \tag{2.125}$$

$$F_1(y) = \frac{H(0,y)}{F_0(y)}, h_2(\alpha) = \frac{H(\alpha,\infty)}{h_0(\alpha)},\ 0 < \alpha \le \alpha_0.$$

2.6 Identifiability by Sum and Maximum (or Minimum)

Let X_0, X_1 and X_2 be independent random variables and

$$Y_1 = X_0 + X_1 \text{ and } Y_2 = X_0 V X_2. \tag{2.126}$$

Let $M_i(\alpha) = Ee^{\alpha X_i}$ for $i = 0,1$. Suppose that $M_i(\alpha)$ finite for $0 \le \alpha \le \alpha_0$ and $M_0(\alpha_0)$ is a given constant for $\alpha_0 \ne 0$. Further suppose that $F_i(y) > 0$ for $y > a \ge -\infty$ for $i = 0,2$ and

$$\lim_{z\to-\infty} e^{\alpha_0 z}F_0(z) = 0, 0 < \int_{-\infty}^\beta e^{\alpha_0 z}\, dF_0(z) < \infty.$$

Theorem 2.6.1 : Under the conditions stated above, the joint distribution of (Y_1, Y_2) *uniquely* determines the distributions of X_0, X_2 on the interval (a,∞) and the function $M_1(\alpha), 0 \le \alpha \le \alpha_0$.

Let

$$\begin{aligned} H(\alpha,\beta) &= E[e^{\alpha Y_1}\chi_{(-\infty,\beta)}(Y_2)] \\ &= M_1(\alpha)F_2(\beta)\int_{-\infty}^\beta e^{\alpha z} dF_0(z). \end{aligned} \tag{2.127}$$

Denote

$$\frac{H_1'(\beta)}{H_1(\beta)} = \frac{\alpha_0 H(\alpha_0,\infty)H(0,\beta)}{H(\alpha_0,\infty)H(0,\beta) - e^{\alpha_0\beta}M_0(\alpha_0)H(\alpha_0,\beta)}\ . \tag{2.128}$$

Then $F_0(z) = Ae^{-\alpha_0 z}H_1'(z)$ where A is constant so that $F_0(\infty) = 1$ and

$$F_2(y) = \frac{H(0,y)}{F_0(y)}, M_1(\alpha) = \frac{H(\alpha,\infty)}{M_0(\alpha)}. \tag{2.129}$$

An analogous result holds characterizing probability measures by sum and minimum. The result is due to Kotlarski.

Let X_0, X_1 and X_2 be independent random variables and

$$Y_1 = X_0 + X_1 \text{ and } Y_2 = X_0 \Lambda X_2. \tag{2.130}$$

Let $M_i(\alpha) = Ee^{\alpha X_i}, i = 0,1$. Suppose that $M_i(\alpha)$ is finite for $0 \leq \alpha \leq \alpha_0, \alpha_0 > 0$ and $M_0(\alpha_0)$ is a fixed constant. Further suppose that the distribution functions F_i of $X_i, i = 0,1,2$ satisfy the conditions $F_i(x) < 1$ for $x < a_0, a_0 \leq \infty$ and

$$\lim_{z \to \infty} e^{\alpha_0 z}\overline{F_0}(z) = 0.$$

Theorem 2.6.2 : Under the conditions stated above, the joint distribution of (Y_1, Y_2) uniquely determines the distributions of $X_i, i = 0,2$ on the interval $(-\infty, a_0)$ and the function $M_1(\alpha), 0 \leq \alpha \leq \alpha_0$.

Let

$$\begin{aligned} H(\alpha,\beta) &= E[e^{\alpha Y_1}\chi_{(\beta,\infty)}(Y_2)] \\ &= M_1(\alpha)\overline{F_2}(\beta)\int_\beta^\infty e^{\alpha z}dF_0(z). \end{aligned} \tag{2.131}$$

Denote

$$\frac{H_1'(\beta)}{H_1(\beta)} = \frac{\alpha_0 H(\alpha_0,0)H(0,\beta)}{H(\alpha_0,\beta)H(0,\beta) - M_0(\alpha_0)H(\alpha_0,\beta)e^{-\alpha_0\beta}}\ . \tag{2.132}$$

Then $\overline{F_0}(z) = Ae^{-\alpha_0 z}H_1'(z)$ where A is determined from $F_0(\infty) = 1$. Furthermore

$$\overline{F_2}(y) = \frac{H(0,y)}{\overline{F_0}(y)}, M_1(\alpha) = \frac{H(\alpha,0)}{M_0(\alpha)}. \tag{2.133}$$

The proofs of Theorems 2.6.1 and 2.6.2 are left as exercises for the reader. The results stated here are due to Kotlarski.

2.7 Identifiability by Product and Sum

Let X_0, X_1 and X_2 be *positive* random variables and define

$$Y_1 = X_1 X_0 \text{ and } Y_2 = X_2 + X_0. \tag{2.134}$$

Assume that $h_i(\alpha) = E[e^{\alpha X_i}] < \infty, i = 0, 1$ for $0 \leq \alpha \leq \alpha_0, \alpha_0 > 0$.

Theorem 2.7.1 : Suppose $q = E(X_0)$ exists and is a fixed positive constant. Then the joint distribution of (Y_1, Y_2) *uniquely* determines the distributions of X_0 and X_2 and the function $E(e^{\alpha X_1}), 0 \leq \alpha \leq \alpha_0$.

Let

$$\begin{aligned} H(\alpha,\beta) &= EY_1^\alpha e^{i\beta Y_2} \\ &= h_1(\alpha)\phi_2(\beta)\int_0^\infty z^\alpha e^{i\beta z} dF_0(z) \end{aligned} \tag{2.135}$$

where $h_1(\alpha) = E[e^{\alpha X_1}], \phi_2(\beta) = E[e^{i\beta X_2}]$ and $F_0(\cdot)$ is the distribution function of X_0. Denote

$$\frac{H_1'(\beta)}{H_1(\beta)} = \frac{iqH(1,\beta)}{H(1,0)H(0,\beta)} \ . \tag{2.136}$$

Then

$$\phi_0(\beta) = \frac{H_1(\beta)}{H_1(0)}, \phi_2(\beta) = \frac{H(0,\beta)}{\phi_0(\beta)}, h_1(\alpha) = \frac{H(\alpha,0)}{h_0(\alpha)} \ . \tag{2.137}$$

We omit the proof of this result due to Kotlarski.

Remarks 2.7.1 : Since the function $h_1(\alpha) = E(e^{\alpha X_1})$ is determined for $0 \leq \alpha \leq \alpha_0, \alpha_0 > 0$ and X_1 is a *positive* random variable, it follows that the moment-generating function of X_1 is determined in a neighbourhood of the origin and hence the distribution of X_1 is determined in addition to the distributions of X_0 and X_2 in Theorem 2.7.1.

2.8 Identifiability by Maxima of Several Random Variables

Let $X_1, X_2, \ldots, X_n$ be independent *positive* random variables with distribution functions $F_1, F_2, \ldots, F_n$ respectively. Suppose that $F_j(x) > 0$ for all $x > 0, 1 \leq j \leq n$. Define

$$\begin{aligned} Y_1 &= \max(a_1X_1, \ldots, a_nX_n), \\ Y_2 &= \max(b_1X_1, \ldots, b_nX_n) \end{aligned} \tag{2.138}$$

where $a_i > 0, b_i > 0$ for $1 \leq i \leq n$ and $a_i : b_i \neq a_j : b_j$ for $1 \leq i \neq j \leq n$.

Theorem 2.8.1 : The joint distribution of (Y_1, Y_2) *uniquely* determines the distributions of $X_j, 1 \leq j \leq n$.

Proof : Let F_j^* be an alternative possible distribution of X_j for $1 \leq j \leq n$. It is easy to see that

$$\begin{aligned} H(t,s) &= P(Y_1 \leq t, Y_2 \leq s) \\ &= \Pi_{j=1}^n F_j(\frac{t}{a_j} \Lambda \frac{s}{b_j}), 0 \leq t, s < \infty. \end{aligned} \tag{2.139}$$

Since F_j^* is an alternative distribution, it follows that

$$\Pi_{j=1}^n F_j(\frac{t}{a_j} \Lambda \frac{s}{b_j}) = \Pi_{j=1}^n F_j^*(\frac{t}{a_j} \Lambda \frac{s}{b_j}), 0 \leq t, s < \infty. \tag{2.140}$$

Let $v_j(t) = \log\ F_j(\frac{t}{b_j}) - \log\ F_j^*(\frac{t}{b_j})$. The equation (2.139) can be written in the form

$$\sum_{j=1}^{n} v_j(c_j t \Lambda s) = 0, \quad 0 \leq t, s < \infty \tag{2.141}$$

where the $c_j = \frac{b_j}{a_j}$ are pairwise distinct. Without loss of generality, assume that $0 < c_1 < c_2 < \cdots < c_n$. Let $t > 0$ and $s = \tau t$ where $c_{n-1} < \tau < c_n$. Then the equation (2.140) can be written in the form,

$$\sum_{j=1}^{n-1} v_j(c_j t) + v_n(\tau t) = 0, \quad 0 < t < \infty. \tag{2.142}$$

This equation proves that $v_n(\cdot)$ is constant on the interval $(c_{n-1}t, c_n t)$ for any $t > 0$. Since $t > 0$ is arbitrary, it follows that $v_n(\cdot)$ is constant on $(0, \infty)$. Since $v_j(t) \to 0$ as $t \to +\infty$, it follows that $v_n(t) = 0$ for $t > 0$. Repeating this process it is easy to see that

$$v_j(t) = 0, 1 \leq j \leq n-1 \ . \tag{2.143}$$

This implies that

$$F_j(\frac{t}{b_j}) = F_j^*(\frac{t}{b_j}), 0 < t < \infty, 1 \leq j \leq n \tag{2.144}$$

from the definition of $v_j(\cdot)$. Since t is arbitrary, it follows that

$$F_j(t) = F_j^*(t), \ \ 1 \leq j \leq n, 0 < t < \infty \, . \tag{2.145}$$

This proves the theorem. ■

The next example indicates that the conclusion of the theorem does not hold if $\{X_i\}$ are random variables taking positive and negative values with positive probability.

Example 2.8.1 : Let X_1 and X_2 be independent identically distributed random variables with distribution function $F(x)$ where

$$\begin{aligned} F(x) &> 0 \text{ for all } x \in R, \\ \frac{F(-0)}{F(+0)} &= \alpha, 0 < \alpha < 1. \end{aligned} \tag{2.146}$$

Then the distribution of (Y_1, Y_2) where

$$Y_1 = \max(X_1, X_2), Y_2 = \max(X_1, \beta X_2)$$

with $\beta > 0, \beta \neq 1$ does not determine the distributions of the random variables X_1 and X_2. This can be seen as follows. Let X_1' and X_2' be independent random variables with distribution functions F_1^* and F_2^* respectively where

$$F_1^*(x) = \begin{cases} F(x) & \text{if} \quad x > 0 \\ \alpha F(x) & \text{if} \quad x \leq 0 \end{cases} \tag{2.147}$$

and

$$F_2^*(x) = \begin{cases} F(x) & \text{if } x > 0 \\ \frac{1}{\alpha}F(x) & \text{if } x \leq 0\,. \end{cases} \tag{2.148}$$

Define

$$Y_1' = \max(X_1', X_2'), Y_2' = \max(X_1', \beta X_2'), \beta > 0, \beta \neq 1.$$

It is easy to check that the joint distribution of (Y_1', Y_2') is the same as that of (Y_1, Y_2). However, the distributions of X_i and X_i' are different for $i = 1, 2$.

The following result holds if $X_1, X_2, \ldots, X_n$ are independent random variables with distribution functions $F_1, F_2, \ldots, F_n$ respectively, where $F_j(x) > 0$ for all $x \in R$ and $P(X_j = 0) = 0$ for $1 \leq j \leq n$.

Theorem 2.8.2 : Under the conditions stated above, the joint distribution of (Y_1, Y_2) defined by (2.137) *uniquely* determines the distributions of $X_j, 1 \leq j \leq n$.

Proof : As in the proof of Theorem 2.8.1, we have

$$\sum_{j=1}^{n} v_j(c_j t \Lambda s) = 0, \; -\infty < t, s < \infty \tag{2.149}$$

where $c_j = \frac{a_j}{b_j}$ are pairwise distinct and $0 < c_1 < \cdots < c_n$. It follows from the arguments given in Theorem 2.8.1 that $v_j(t) = 0$ for $t > 0$. Suppose $t < 0$. Let $s = \tau t, \tau \in (c_1, c_2)$. Then, the equation (2.140) takes the form

$$v_1(\tau t) + \sum_{j=2}^{n} v_j(c_j t) = 0. \tag{2.150}$$

Hence $v_1(\cdot)$ is constant on the interval $(c_2 t, c_1 t)$. Since $t < 0$ is arbitrary, it follows that $v_1(t) = 0$ on $(-\infty, 0)$. Note that v_1 is continuous at $x = 0$. Hence $v_1(0) = 0$. Therefore $v_1(t) = 0$ for all t. By induction, it follows that $v_j(t) = 0$ for all $t, 1 \leq j \leq n$ and hence $F_j = F_j^*$ for $1 \leq j \leq n$. This completes the proof of the theorem. ∎

The results in this section are due to Klebanov (1973b).

2.9 Identifiability by Random Sums

Let X_0, X_1 and X_2 be independent random variables and $Y_1 = X_0 + X_1, Y_2 = X_0 + X_2$. We have proved in Section 2.1 that the distribution of (Y_1, Y_2) uniquely determines the distributions of X_0, X_1 and X_2 up to shift provided the characteristic functions of $X_k, k = 0, 1, 2$, do not vanish. We now study results of a similar type involving random sums of random variables.

Theorem 2.9.1 : Let $N, X_i, Y_i, i \geq 1$ be independent random variables nondegenerate at zero where N is a nonnegative integer-valued random variable with $0 < EN < \infty$ fixed and the X_i are independent and identically distributed (i.i.d) as X with finite mean and nonvanishing characteristic function ϕ and Y_i are i.i.d. as Y with finite mean and nonvanishing characteristic function ψ. Further suppose that if the probability-generating function of N is

$$Q(s) = p_0 + \sum_{n=1}^{\infty} s^n p_n, \quad s \in S \tag{2.151}$$

S has a subset S_0 such that

(i) $\alpha, \beta \in R \Rightarrow \phi(\alpha)\psi(\beta) \in S_0$,

(ii) Q is non-vanishing and one-to-one on S_0, and

(iii) Q can be extended analytically from S_0 to S.

Let

$$U = \begin{cases} 0 & \text{if} \quad N = 0 \\ \sum_{i=1}^{n} X_i & \text{if} \quad N = n > 0 \end{cases} \quad \text{and } V = \begin{cases} 0 & \text{if } N = 0 \\ \sum_{i=1}^{n} Y_i & \text{if } N = n > 0 \end{cases} \tag{2.152}$$

Then the joint distribution of (U, V) uniquely determines the distributions of X, Y and N.

Proof : The characteristic function $\chi(r,t)$ of (U,V) is given by

$$\begin{aligned}
\chi(r,t) &= E[e^{irU+itV}] \\
&= E\{E[e^{irU+itV}|N]\} \\
&= E[e^{irU+itV}|N=0]P(N=0) \\
&\quad +\sum_{n=1}^{\infty} E[\exp\{ir(X_1+\cdots+X_N) \\
&\qquad\qquad +it(Y_1+\cdots+Y_N)\}|_{N=n}]P(N=n) \\
&= P(N=0) \\
&\quad +\sum_{n=1}^{\infty} E[\exp\{ir(X_1+\cdots+X_n)+it(Y_1+\cdots+Y_n)\}]P(N=n) \\
&\quad \text{(by the independence of } N \text{ and } X_i, Y_i, i\geq 1) \\
&= P(N=0)+\sum_{n=1}^{\infty}[\phi(r)]^n[\psi(t)]^n P(N=n) \\
&= Q(\phi(r)\psi(t)), \quad -\infty<r,t<\infty. \qquad (2.153)
\end{aligned}$$

Suppose there is another collection of random variables $\{N^*, X_i^*, Y_i^*, i \geq 1\}$ satisfying the conditions stated in the theorem and define U^*, V^* as before. Suppose further that the joint distribution of (U,V) is the same as that of (U^*, V^*). Then, it follows that

$$\chi(r,t) = Q^*(\phi^*(r)\psi^*(t)), \quad -\infty < r, t < \infty. \qquad (2.154)$$

Relations (2.152) and (2.153) imply that

$$Q^*(\phi^*(r)\psi^*(t)) = Q(\phi(r)\psi(t)), \quad -\infty < r, t < \infty. \qquad (2.155)$$

Since Q and Q^* are one-to-one on S_0 by hypothesis, define

$$q(s) = Q^{*^{-1}}(Q(s)), \ s \in S_0. \qquad (2.156)$$

Then

$$\phi^*(r)\psi^*(t) = q(\phi(r)\psi(t)), r, t \in R. \qquad (2.157)$$

Let $t = 0$ in (2.156). Then $\phi^*(r) = q(\phi(r))$. Similarly $\psi^*(t) = q(\psi(t))$. Hence

$$q(\phi(r))q(\psi(t)) = q(\phi(r)\psi(t)), \ r, t \in R. \qquad (2.158)$$

In view of the properties (i) and (ii) of Q and Q^*, it follows that

$$q(u)q(v) = q(uv), \ u, v \in S_0. \tag{2.159}$$

By property (iii) of Q and Q^*, this relation can be extended to all of S by analyticity and we have

$$q(u)q(v) = q(uv), \ u, v \in S \ . \tag{2.160}$$

Since $q(\cdot)$ is a continuous function, it follows that

$$q(s) = s^c, \ \ s \in S \tag{2.161}$$

for some constant c. In particular, we have

$$Q^{*^{-1}}(Q(s)) = q(s) = s^c, \ \ s \in S$$

or equivalently

$$Q(s) = Q^*(s^c), \ \ s \in S. \tag{2.162}$$

Suppose $\phi(s) = \sum_{n=0}^{\infty} p_n s^n$ and $Q^*(s) = \sum_{n=0}^{\infty} p_n^* s^n$. Since

$$\frac{dQ}{ds} = \frac{dQ^*}{ds} \cdot cs^{c-1}$$

from (2.161), it follows that $E(N) = EN^* \cdot c$. Since EN is given to be a fixed positive constant, it follows that $c = 1$ which in turn proves that

$$Q(s) = Q^*(s), s \in S. \tag{2.163}$$

This relation together with (2.154) proves that

$$Q(\phi^*(r)\psi^*(t)) = Q(\phi(r)\psi(t)), r, t \in R. \tag{2.164}$$

Setting $r = 0$ and $t = 0$ alternately, we have

$$Q(\phi^*(r)) = Q(\phi(r)) \text{ and } Q(\psi^*(t)) = Q(\psi(t)), r, t \in R\ . \tag{2.165}$$

Since $Q(\cdot)$ is one-to-one on S_0, it follows that

$$\phi^*(r) = \phi(r) \text{ and } \psi^*(t) = \psi(t), r, t \in R. \tag{2.166}$$

Relations (2.162) and (2.165) prove that N, X_1, Y_1 have the same distributions as N^*, X_1^*, Y_1^* respectively, completing the proof of the theorem. ∎

Remarks 2.9.1 : It can be shown that, if $0 < EN < \infty$, then the probability generating function Q is one-to-one in a neighbourhood of 1. This implies that there is a neighbourhood of 1 relative to the unit disk such that Q^{-1} exists in this neighbourhood (cf. Choike *et al.* (1980)). The condition that ϕ and ψ are nonvanishing in the theorem can be replaced by that of analyticity. However, the following example shows that, without these assumptions on Q, Q^*, ϕ, ψ, the result may not hold.

Example 2.9.1 : Let N and N^* be nonnegative integer-valued random variables with probability generating function $Q(s) = s^2, |s| \leq 1$. Let X be distributed according to the characteristic function

$$\begin{aligned} \phi(r) &= 1 - \frac{2|r|}{\pi}, \quad -\pi \leq r \leq \pi, \\ \phi(r + 2\pi) &= \phi(r) \quad \text{otherwise} . \end{aligned}$$

Let X^* have the characteristic function $|\phi(r)| \equiv \phi^*(r)$. Suppose Y and Y^* are identically distributed with characteristic function ψ. Then (U, V) and (U^*, V^*) have the same distribution since

$$Q^*(\phi^*(r)\psi^*(t)) = Q(\phi(r)\psi(t)), r, t \in R$$

although $\phi^*(r) \neq \phi(r)$.

Remarks 2.9.2 :If X and Y are symmetric real-valued nondegenerate random variables with characteristic functions ϕ and ψ respectively, in Theorem 2.9.1, then we can conclude that the distribution of (U, V) determines the distribution of X, Y, N uniquely provided $0 < EN < \infty$. No additional conditions on ϕ, ψ or Q are necessary. Note that ϕ and ψ are real-valued functions with $0 \leq \phi(t) \leq 1$ and $0 \leq \psi(t) \leq 1$ for all $t \in R$.

Remarks 2.9.3 (Explicit determination of Q, ϕ and ψ given χ) : Here we consider the problem of explicit determination of the distributions of N, X and Y in terms of the joint distribution of (U, V). It is sufficient to solve the equation (2.152), namely,

$$\chi(r,t) = Q(\phi(r)\psi(t)), \quad -\infty < r, t < \infty \tag{2.167}$$

for Q, ϕ and ψ in terms of χ. Let $q(w) = Q^{-1}(w)$ be the inverse function of $Q(\cdot)$ defined on S_0. Relation (2.166) shows that

$$\phi(r)\psi(t) = q(\chi(r,t)) \ . \tag{2.168}$$

Substituting $r = 0$ and $t = 0$ alternately, we have the equation

$$q(\chi(r,t)) = q(\chi(r,0))q(\chi(0,t)), -\infty < r, t < \infty \tag{2.169}$$

which is a functional equation in the unknown q given known χ. Let

$$q_0 = \log q \tag{2.170}$$

be the continuous branch of the natural logarithm of q satisfying $\log 1 = 0; q_0$ is well defined since $q(\cdot)$ is nonvanishing. Equation (2.168) shows that

$$q_0(\chi(r,t)) = q_0(\chi(r,0)) + q_0(\chi(0,t)), r, t \in R \ . \tag{2.171}$$

Assume that $q_0(\cdot)$ is differentiable twice and that $\chi(r,t)$ has continuous second-order partial derivatives with respect to r and t. Taking partial derivatives with respect to t and then with respect to r we have

$$q_0''(\chi(r,t)) \frac{\partial \chi}{\partial r} \frac{\partial \chi}{\partial t} + q_0'(\chi(r,t)) \frac{\partial^2 \chi}{\partial r \partial t} = 0, r, t \in R$$

or equivalently

$$\frac{q_0''(\chi(r,t))}{q_0'(\chi(r,t))} = -\frac{\frac{\partial^2 \chi}{\partial r \partial t}}{\frac{\partial \chi}{\partial r}\frac{\partial \chi}{\partial t}}, \qquad r, t \in R \tag{2.172}$$

where q_0' and q_0'' denote the first and second derivatives of q_0. The above differentiation can be justified since $E(X)$ and $E(Y)$ are finite. Note that $q_0'(w) \neq 0$ in a neighbourhood of 1 since $q_0'(1) = \frac{1}{EN} > 0$. Since the left side of the equation (2.170) is a function of $w = \chi(r,t)$, we can write the equation (2.171) in the form

$$\frac{q_0''(w)}{q_0'(w)} = -\frac{\frac{\partial^2 \chi}{\partial r \partial t}}{\frac{\partial \chi}{\partial r}\frac{\partial \chi}{\partial t}}\Big|_{w=\chi(r,t)} \tag{2.173}$$

with boundary conditions $q_0(1) = 0, q_0'(1) = \frac{1}{EN} > 0$. Given $\chi(\cdot,\cdot)$, we solve this second order differential equation (2.172) subject to the boundary conditions $q_0(1) = 0, q_0'(1) = \frac{1}{EN} > 0$ to obtain q_0. Having obtained q_0 or equivalently q, the functions ϕ and ψ are determined by

$$\phi(r) = q(\chi(r,0)), \psi(t) = q(\chi(0,t)), r,t \in R.$$

Example 2.9.2 : Suppose (U, V) as defined above has the characteristic function

$$\chi(r,t) = \frac{1 + e^{-(r^2+t^2)}}{2}, r,t \in R$$

and $EN = 1$. Here $\chi(r,t)$ is a real-valued function. It is easy to see that the equation (2.172) reduces to

$$\frac{q_0''(w)}{q_0'(w)} = -\frac{1}{w - \frac{1}{2}}, \frac{1}{2} < w \le 1,$$

where $q_0(1) = 0$ and $q_0'(1) = 1$. The solution of this differential equation is

$$q_0(w) = \log\sqrt{2w-1}, \frac{1}{2} < w \le 1.$$

Hence

$$Q(s) = \frac{1+s^2}{2}, \quad -1 \le s \le 1.$$

recalling that $q_0 = \log\, q$ and q is the inverse of Q. Hence N is an integer-valued random variable with

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

Furthermore

$$\phi(r) = q(\chi(r,0)) = e^{-r^2/2}, r \in R$$

and

$$\psi(t) = q(\chi(0,t)) = e^{-t^2/2}, t \in R$$

which show that X and Y have the standard normal distribution.

The results in this section are due to Choike *et al.* (1980) and Kotlarski (1984).

2.10 Identifiability by the Maximum of a Random Number of Random Variables

We now obtain an analogue of Theorem 2.9.1 given in the previous section for the maximum of a random number of random variables.

Theorem 2.10.1 : Let $N, X_i, Y_i, i \geq 1$ be independent random variables and suppose N is a nonnegative integer-valued random variable with $p_1 = P(N = 1) > 0$ fixed. Further suppose that $X_i, i \geq 1$, are i.i.d. with continuous strictly increasing distribution function $F(\cdot)$, and $Y_i, i \geq 1$, are i.i.d. with continuous strictly increasing distribution function $G(\cdot)$ where

$$F(a) = 0, F(b) = 1, 0 < F(x) < 1 \text{ for } -\infty \leq a < x < b \leq \infty \tag{2.174}$$

and

$$G(c) = 0, G(d) = 1, 0 < G(y) < 1 \text{ for } -\infty \leq c < y < d \leq \infty.$$

Let

$$\begin{aligned} U &= a && \text{for } N = 0 \\ &= \max_{1 \leq i \leq N} X_i && \text{for } N > 0 \end{aligned} \tag{2.175}$$

and

$$\begin{aligned} V &= c && \text{for } N = 0 \\ &= \max_{1 \leq i \leq N} Y_i && \text{for } N > 0\,. \end{aligned} \tag{2.176}$$

Then the joint distribution of (U, V) *uniquely* determines the distributions of N, X_1 and Y_1.

Proof : Let $Q(s)$ be the probability generating function of N. Then

$$Q(s) = \sum_{n=0}^{\infty} s^n P(N = n) = \sum_{n=0}^{\infty} s^n p_n, \; 0 \leq s \leq 1. \tag{2.177}$$

Since $p_1 > 0$, it follows that $p_0 < 1$. Note that the range of $Q(\cdot)$ is $[p_0, 1]$. Let $H(u, v)$ be the joint distribution function of (U, V). Then

$$\begin{aligned} H(u,v) &= P[U \leq u, V \leq v] \\ &= \sum_{n=0}^{\infty} P[U \leq u, V \leq v | N = n] P(N = n) \\ &= p_0 + \sum_{n=1}^{\infty} P(\max_{1\leq i\leq n} X_i \leq u, \max_{1\leq i\leq n} Y_i \leq v | N = n) P(N = n) \\ &= p_0 + \sum_{n=1}^{\infty} P(\max_{1\leq i\leq n} X_i \leq u, \max_{[1\leq i\leq n} Y_i \leq v) p_n \\ &\quad \text{(by the independence of the } X_i\text{'s and } Y_i\text{'s with } N \text{)} \\ &= p_0 + \sum_{n=1}^{\infty} (F(u))^n (G(v))^n p_n \\ &\quad \text{(by the independence of the } X_i\text{'s and } Y_i\text{'s)} \\ &= Q(F(u)G(v)), \;\; -\infty < u, v < \infty. \end{aligned} \tag{2.178}$$

Suppose $N^*, X_i^*, Y_i^*, i \geq 1$ is another collection of random variables having the same properties as N, X_i and Y_i, and define U^*, V^* in analogy with U, V. Suppose (U, V) and (U^*, V^*) have the same distribution. We shall prove that N, X_1, Y_1 have the same distributions as N^*, X_1^*, Y_1^* respectively.

Since (U, V) and (U^*, V^*) have the same distribution, it follows from (2.177) that

$$Q(F(u)G(v)) = Q^*(F^*(u)G^*(v)), a \leq u \leq b, c \leq v \leq d \tag{2.179}$$

where F^* and G^* are the distribution functions of X_1^* and Y_1^* respectively. Let $u = a$ and $v = c$ in (2.178). Then it follows that

$$Q(0) = Q^*(0) \tag{2.180}$$

and hence

$$p_0^* = p_0. \tag{2.181}$$

Let

$$q(s) = Q^{*^{-1}}(Q(s)), 0 \leq s \leq 1. \tag{2.182}$$

Then $q(s)$ is a continuous function from $[0,1]$ onto $[0,1]$. The equation (2.178) can be rewritten in the form

$$F^*(u)G^*(v) = q(F(u)G(v)), a \leq u \leq b, c \leq v \leq d. \tag{2.183}$$

Substituting $v = d$ in (2.182), we get

$$F^*(u) = q(F(u)), a \leq u \leq b \ . \tag{2.184}$$

Similarly, let $u = b$ in (2.182). Then we have

$$G^*(v) = q(G(v)), c \leq v \leq d \ . \tag{2.185}$$

Combining the above relations, we obtain the functional equation

$$q(F(u)G(v)) = q(F(u))q(G(v)), a \leq u \leq b, c \leq v \leq d. \tag{2.186}$$

Let $\alpha = F(u)$ and $\beta = G(v)$. Note that $F(u)$ and $G(v)$ are continuous strictly increasing from 0 to 1 in the intervals $[a,b]$ and $[c,d]$ respectively. This proves that

$$q(\alpha)q(\beta) = q(\alpha\beta), \ \ 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1 \tag{2.187}$$

and $q(s)$ is a continuous function from $[0,1]$ onto $[0,1]$. Hence the only solution of (2.186) is

$$q(s) = s^a, \ 0 \leq s \leq 1 \tag{2.188}$$

for some constant a. In other words

$$Q(s) = Q^*(s^a), \ 0 \leq s \leq 1 \ , \tag{2.189}$$

that is ,

$$p_0 + p_1 s + \sum_{n=1}^{\infty} p_n s^n = p_0^* + p_1^* s^a + \sum_{n=1}^{\infty} p_n^* s^{na}, 0 \leq s \leq 1. \tag{2.190}$$

Note that $p_1 > 0$ and $p_1^* > 0$ under the conditions of the theorem and $p_1 = p_1^*$. Since the equality in (2.189) holds for all s in $[0, 1]$, it follows that $a = 1$ and hence

$$q(s) = s, 0 \leq s \leq 1 \ . \tag{2.191}$$

Therefore

$$Q(s) = Q^*(s), 0 \leq s \leq 1 \ . \tag{2.192}$$

Relations (2.183), (2.184) and (2.190) show that

$$F^*(u) = F(u), a \leq u \leq b \text{ and } G^*(v) = G(v), c \leq v \leq d. \tag{2.193}$$

This completes the proof of the theorem. ■

Remarks 2.10.1 : A result analogous to Theorem 2.10.1 can be proved for minima in both U and V or maximum in one of U and V and minimum in the other. The results given in this section are due to Kotlarski (1979).

Remarks 2.10.2 (Explicit determination of the distributions of X, Y, N given that of (U, V) defined by (2.174) **and** (2.175)): In addition to the assumptions stated in Theorem 2.10.1, suppose that the random variables X and Y have positive densities in the interiors of their supports. Let $H(u, v)$ be the distribution function of $(U, V), F$ and G be the distribution functions of X and Y and Q be the probability generating function of N. Then

$$H(u, v) = Q(F(u)G(v)), a \leq u \leq b, c \leq v \leq d \ . \tag{2.194}$$

Under the assumptions that $p_1 > 0$ (and hence $p_0 < 1$) and that the mapping $Q : [0, 1] \to [p_0, 1]$ is invertible, let

$$q(w) = Q^{-1}(w), w \in [p_0, 1]. \tag{2.195}$$

It is easy to check that the relations (2.193) and (2.194) imply that

$$q(H(u, v)) = q(H(u, d))q(H(b, v)), a < u \leq b, c < v \leq d \tag{2.196}$$

where q is a strictly increasing continuous function mapping $[p_0, 1]$ onto $[0, 1]$. Let

$$q_0(w) = \log\, q(w), \quad p_0 < w \leq 1. \tag{2.197}$$

Taking logarithms on both sides of the equation (2.195), we have

$$q_0(H(u,v)) = q_0(H(u,d)) + q_0(H(b,v)), a < u \leq b, c < v \leq d. \tag{2.198}$$

It is easy to solve this functional equation subject to the condition $q(1) = 1$ and $EN = m$ fixed. This can be done as in Remarks 2.9.3 to obtain $q(\cdot)$ and hence obtain F and G. The details are left to the reader (see Kotlarski (1985)).

Example 2.10.3 : If $H(u,v) = \frac{1+u^2v^2}{2}, 0 \leq u, v \leq 1$ and $EN = 1$, then it can be checked that

$$Q(s) = \frac{1+s^2}{2}, \quad 0 \leq s \leq 1,$$

$$q(w) = \sqrt{2w-1}, \frac{1}{2} \leq w \leq 1,$$

and hence

$$F(x) = q(H(x,1)) = x, 0 \leq x \leq 1,$$

and

$$G(y) = q(H(1,y)) = y, 0 \leq y \leq 1.$$

Remarks 2.10.5 : The results in this section and the previous section can be extended to several other variations of (U, V) under suitable conditions. Some of them are of the following type :

$$\begin{aligned} U &= X + Y_0, \\ V &= Y_0 + Y_1 + \cdots + Y_N, \end{aligned} \tag{2.199}$$

where N is a nonnegative integer-valued random variable, $Y_i, i \geq 0$ are i.i.d. and $X, N, Y_i, i \geq 0$ are independent,

$$\begin{aligned} U &= Z + X_1 + \cdots + X_N, \\ V &= Z + Y_1 + \cdots + Y_M \end{aligned} \tag{2.200}$$

where N, M are nonnegative integer-valued random variables, $N, M, Z, X_i, Y_i, i \geq 1$, are independent, and all of Z, X_i, Y_i are i.i.d., or

$$\begin{aligned} U &= X_1 + \cdots + X_N + Z_1 + \cdots + Z_T, \\ V &= Y_1 + \cdots + Y_M + Z_1 + \cdots + Z_T, \end{aligned} \tag{2.201}$$

where N, M, T are nonnegative integer-valued random variables independent of $Z_i, X_j, Y_k, i \geq 1, j \geq 1, k \geq 1$, which in turn are all independent and identically distributed with a known distribution.

In all the above cases, the joint distribution of (U, V) determines the unknown distributions of the random variables involved in their definition.

The discussion given here is based on Kotlarski (1985).

2.11 Identifiability by Random Linear Forms

Suppose X_1, X_2 and X_3 are three independent real-valued random variables. Let Y_1, Y_2, Y_3 be random variables independent of X_1, X_2, X_3 and independent among themselves with *known* distributions. Let

$$\begin{aligned} W_1 &= Y_1 X_1 + Y_2 X_2, \\ W_2 &= Y_1 X_1 + Y_3 X_3 \ . \end{aligned} \tag{2.202}$$

The question now is to find conditions under which the joint distribution of (W_1, W_2) determines the distributions of X_1, X_2, X_3. This is an extension of the problem discussed in Section 2.1. W_1 and W_2 are called linear forms with random coefficients or *random linear forms* .

Theorem 2.11.1 : If the characteristic function of (W_1, W_2) does not vanish, then the distributions of the products $X_i Y_i, 1 \leq i \leq 3$ are determined up to shift. Furthermore if $E(X_i Y_i)$ is finite and fixed, then the distribution of $X_i Y_i$ is *uniquely* determined for $1 \leq i \leq 3$. In addition, if $X_i Y_i$ has moments of all orders, the characteristic function of X_i is analytic and $E(Y_i^k) \neq 0$ for all $k \geq 2$, then the distribution of X_i is *uniquely* determined.

Proof : Suppose $X_i', Y_i', 1 \leq i \leq 3$, is another set of random variables satisfying the conditions stated in the theorem. The first and second parts follow from Theorem 2.1.1. In other words X_iY_i and $X_i'Y_i'$ will have the same distribution for $1 \leq i \leq 3$. Let $\eta_i(t)$ and $\zeta_i(t)$ be the characteristic functions of X_i and X_i' respectively and μ_i be the distribution function of Y_i (or equivalently Y_i'). Then the characteristic function of X_iY_i and $X_i'Y_i'$ are the same and hence

$$\int_{-\infty}^{\infty} n_i(ty)d\mu_i(y) = \int_{-\infty}^{\infty} \zeta_i(ty)d\mu_i(y) \ . \tag{2.203}$$

Differentiating under the integral sign with respect to t, it follows that

$$\int_{-\infty}^{\infty} y^k\eta_i^{(k)}(ty)d\mu_i(y) = \int_{-\infty}^{\infty} y^k\zeta_i^{(k)}(ty)d\mu_i(y), k \geq 1. \tag{2.204}$$

In particular, let $t = 0$ in (2.203). Then, we have

$$[\eta_i^{(k)}(0) - \zeta_i^{(k)}(0)] \int_{-\infty}^{\infty} y^k d\mu_i(y) = 0, k \geq 1. \tag{2.205}$$

Since

$$\int_{-\infty}^{\infty} y^k d\mu_i(y) \neq 0, k \geq 2, \tag{2.206}$$

by hypothesis, it follows that

$$\eta_i^{(k)}(0) = \zeta_i^{(k)}(0), k \geq 2 \ . \tag{2.207}$$

Since the characteristic functions of X_i and X_i' are analytic with $E(X_i) = E(X_i')$, it follows that

$$\eta_i(t) = \zeta_i(t), -\infty < t < \infty \tag{2.208}$$

which shows that X_i and X_i' have the same distribution. ∎

Remarks 2.11.1 : It seems to be impossible to avoid a condition of the type (2.205) or some other condition on Y_i equivalent to (2.205). For, in general, it is not true that if X_iY_i and $X_i'Y_i'$ have the same distribution and Y_i and Y_i' are identically distributed, then X_i and X_i' have the same distribution even when Y_i and Y_i' are independent of X_i and X_i' respectively.

For instance *different* combinations of distributions with a given mixing distribution might lead to the same mixture (see Chapter 8 on identifiability for mixtures) . The condition on the analyticity of the characteristic function of X_i in Theorem 2.11.1 can be weakened to the condition that the distribution of X_i be determined by its moments.

In analogy with (2.16) and (2.17), let us now consider random linear forms

$$\begin{aligned} W_1 &= Y_1X_1 + Y_2X_2 + Y_3X_3, \\ W_2 &= T_1X_1 + T_2X_2 + T_3X_3 \end{aligned} \tag{2.209}$$

where X_1, X_2, X_3 are independent, identically distributed random variables, (T_1, T_2, T_3) and (Y_1, Y_2, Y_3) are random vectors independent of (X_1, X_2, X_3) and the distributions of (T_1, T_2, T_3) and (Y_1, Y_2, Y_3) are specified. Let $\phi(t_1, t_2)$ be the characteristic function of (W_1, W_2). Then

$$\begin{aligned} \phi(t_1, t_2) &= E[\exp(it_1W_1 + it_2W_2)] \\ &= E[\exp\{it_1(Y_1X_1 + Y_2X_2 + Y_3X_3) \\ &\qquad +it_2(T_1X_1 + T_2X_2 + T_3X_3)\}] \\ &= E_{\boldsymbol{Y},\boldsymbol{T}}[E\exp(i(t_1Y_1 + t_2T_1)X_1 + i(t_1Y_2 + t_2T_2)X_2 \\ &\qquad +i(t_1Y_3 + t_2T_3)X_3)|Y_1, Y_2, Y_3; T_1, T_2, T_3\}] \\ &= E_{\boldsymbol{Y},\boldsymbol{T}}[\eta(t_1Y_1 + t_2T_1)\eta(t_1Y_2 + t_2T_2)\eta(t_1Y_3 + t_2T_3)] \end{aligned} \tag{2.210}$$

by the independence of the X_i's with the Y_i's and T_i's and by the independence of the X_i's among themselves. Let $\mu(\boldsymbol{u}, \boldsymbol{v})$ denote the joint distribution of $(\boldsymbol{Y}, \boldsymbol{T})$. Suppose $\boldsymbol{X}^*, \boldsymbol{Y}^*$ and $\boldsymbol{T}^*$ satisfy the conditions stated above for $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{T}$, and let $\zeta(t)$ denote the characteristic function of X_1^*. Note that $(\boldsymbol{Y}^*, \boldsymbol{T}^*)$ has the same distribution $(\boldsymbol{Y}, \boldsymbol{T})$. Define W_1^* and W_2^* in analogy with W_1 and W_2. Suppose the distribution of (W_1, W_2) is the same as that of (W_1^*, W_2^*). Relation (2.209) implies that

$$\int_{R^6} \Pi_{j=1}^3 \eta(t_1 u_j + t_2 v_j) d\mu(\boldsymbol{u}, \boldsymbol{v})$$
$$= \int_{R^6} \Pi_{j=1}^3 \zeta(t_1 u_j + t_2 v_j) d\mu(\boldsymbol{u}, \boldsymbol{v}). \tag{2.211}$$

Suppose that, in the above functional equation, differentiation with respect to t_1, t_2 under the integral sign is permissible any number of times. Differentiate twice with respect to t_1 and substitute $t_1 = 0$. Differentiate the equation so obtained with respect to t_2 and then substitute $t_2 = 0$. After some easy though tedious computations, it can be shown that

$$c_1\eta^{(3)}(0) + c_2\eta^{(2)}(0)\eta^{(1)}(0) + c_3[\eta^{(1)}(0)]^3$$
$$= c_1\zeta^{(3)}(0) + c_2\zeta^{(2)}(0)\zeta^{(1)}(0) + c_3[\zeta^{(1)}(0)]^3 \tag{2.212}$$

where c_1, c_2 and c_3 are known constants depending on μ but not η or ζ and $\eta^{(k)}(0)$ denotes the kth derivative of $\eta(\cdot)$ evaluated at zero. This equation can be written in the form

$$c_1\eta^{(3)}(0) + Q(\eta^{(j)}(0), 1 \le j \le 2)$$
$$= c_1\zeta^{(3)}(0) + Q(\zeta^{(j)}(0), 1 \le j \le 2) \tag{2.213}$$

where Q is a known function depending on the derivatives of order *less than* three evaluated at zero. Relation (2.212) has the property that substitution of ζ_i for η_i in Q for $1 \le i \le 2$ on the left side of (2.212) leads to the equation

$$c_1\eta^{(3)}(0) = c_1\zeta^{(3)}(0). \tag{2.214}$$

This method allows us to use induction and establish that for suitable constants c_k depending on μ,

$$c_k\eta^{(k)}(0) = c_k\zeta^{(k)}(0), \quad k \ge 3. \tag{2.215}$$

If

$$c_k \ne 0 \text{ for } k \ge 3\ , \tag{2.216}$$

then we can conclude that

$$\eta^{(k)}(0) = \zeta^{(k)}(0), k \ge 3. \tag{2.217}$$

If $E(X_1) = E(X_1^*)$ and if $\eta(t)$ and $\zeta(t)$ are analytic charcteristic functions or the distribution of X_1 is determined by its moments assuming that they exist, then the equation (2.216) implies that

$$\eta(t) = \zeta(t) \text{ for all } t \tag{2.218}$$

and hence X_1 and X_1^* have the same distribution. We have the following theorem.

Theorem 2.11.2: Consider random linear forms defined by (2.208) where X_1, X_2, X_3 are independent and identically distributed, $(Y_1, Y_2, Y_3; T_1, T_2, T_3)$ is independent of (X_1, X_2, X_3), and the condition (2.215) holds. Suppose the characteristic function of X_1 is either analytic or the distribution of X_1 is determined by its moments assuming that they exist.. Then the distribution of (W_1, W_2) determines the distribution of X_1 up to location. If further $E(X_1)$ is fixed, then the distribution of X_1 is completely determined.

The results in this section are due to Prakasa Rao (1990).

2.12 Stability of Identifiability

In all the discussions so far, we have considered the question of finding conditions under which the distribution of a statistic defined in terms of a sequence of random variables determines the distributions of the individual random variables up to a change in location or scale. We now consider stability of this property. Suppose the density is given by

$$p(x, \theta) = p(x - \theta), -\infty < \theta < \infty \tag{2.219}$$

or

$$p(x, \theta) = \frac{1}{\sigma} p(\frac{x - \mu}{\sigma}), \theta = (\mu, \sigma), -\infty < \mu < \infty, 0 < \sigma < \infty\,. \tag{2.220}$$

The former class is called a *location parameter family* and the latter class a *location–scale parameter family*. Let $X_1, X_2, \ldots, X_N$ be i.i.d. random variables and define

$$Y = (X_1 - X_N, X_2 - X_N, \ldots, X_{N-1} - X_N) \tag{2.221}$$

in the case of a location parameter family and

$$Y^* = \left(\frac{X_1 - \overline{X}}{s}, \frac{X_2 - \overline{X}}{s}, \ldots, \frac{X_{N-1} - \overline{X}}{s}\right) \tag{2.222}$$

in the case of a location–scale parameter family where $\bar{X}$ is the sample mean and s is the sample standard deviation. Denote the analogues of $\boldsymbol{Y}$ and $\boldsymbol{Y}^*$ by $\boldsymbol{Y}_n$ and $\boldsymbol{Y}_n^*$ when the density is p_n instead of p. Let F_n and F be the distribution functions corresponding to p_n and p. It is known from the theory of weak convergence that if F_n converges weakly to F, then $\boldsymbol{Y}_n \xrightarrow{\ell} \boldsymbol{Y}$ (or $\boldsymbol{Y}_n^* \xrightarrow{\ell} \boldsymbol{Y}^*$). The problem is that, if $\boldsymbol{Y}_n \xrightarrow{\ell} \boldsymbol{Y}$ (or $\boldsymbol{Y}_n^* \xrightarrow{\ell} \boldsymbol{Y}^*$), can we conclude that $F_n \xrightarrow{w} F$ or $F_n \xrightarrow{w} F(\cdot - \theta)$ for some θ in the location case and $F_n \xrightarrow{w} F(\frac{\cdot - \mu}{\sigma})$ for some $\theta = (\mu, \sigma)$ in the location–scale case ?

Theorem 2.12.1 : Suppose the distribution of $\boldsymbol{Y}$ determines the distribution F up to shift in the location case and the distribution of $\boldsymbol{Y}^*$ determines F up to location–scale in the location–scale parameter case. Then

$$\boldsymbol{Y}_n \xrightarrow{\ell} \boldsymbol{Y} \Rightarrow F_n \xrightarrow{w} F \tag{2.223}$$

with possibly a shift in the case of location parameter families and

$$\boldsymbol{Y}_n^* \xrightarrow{\ell} \boldsymbol{Y}^* \Rightarrow F_n \xrightarrow{w} F \tag{2.224}$$

with possibly changes in location and scale in the case of location–scale parameter families.

Proof (Location parameter case) : Since $\boldsymbol{Y}_n \xrightarrow{\ell} \boldsymbol{Y}$, it follows that the distribution of $X_1 - X_2$ under p_n converges weakly to the distribution of $X_1 - X_2$ under p. Let $\phi_n(t)$ be the characteristic function of X_1 under p_n and $\phi(t)$ be that under p. Then

$$|\phi_n(t)|^2 \to |\phi(t)|^2 \text{ as } n \to \infty, -\infty < t < \infty \tag{2.225}$$

since $|\phi(t)|^2$ is the characteristic function of $X_1 - X_2$ under p. It is known that (2.224) implies that $\{p_n\}$ is "shift compact" in the sense of Parthasarathy (1968), that is, there exists a suitable sequence of constants θ_n such that the sequence of distributions with densities $p_n(x - \theta_n)$ is weakly compact. Let $\{n_k\}$ be a subsequence such that the sequence of distributions with densities $p_{n_k}(x - \theta_{n_k}), k \geq 1$ converges weakly to a limiting distribution with density p'. But

$$\boldsymbol{Y}_{n_k} \xrightarrow{\ell} \boldsymbol{Y}$$

when p is the density of X_1 by hypothesis. From earlier remarks it follows that

$$\boldsymbol{Y}_{n_k} \xrightarrow{\ell} \boldsymbol{Z}$$

where $\boldsymbol{Z}$ corresponds to $\boldsymbol{Y}$ when p' is the density of X_1. Hence the distribution of $\boldsymbol{Y}$ when p is the density of X_1 and the distribution of $\boldsymbol{Z}$ when p' is the density of X_1 are the same. But the distribution of $\boldsymbol{Y}$ determines the density $p(\cdot)$ up to shift by hypothesis. Hence, for some $\theta \in R$,

$$p'(x) = p(x - \theta), -\infty < x < \infty.$$

A similar argument proves the result in the location–scale parameters case.

■

The results of this section are due to Klebanov (1973b) .

Chapter 3

Identifiability of Probability Measures on Abstract Spaces

We will now discuss generalizations of some of the results obtained in Chapter 2 to probability measures on abstract spaces. For the general theory of probability measures on metric spaces, see Parthasarathy (1968).

3.1 Hilbert Spaces

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and H be a real separable Hilbert space. Let $\mathcal{B}$ be the σ-algebra of Borel susets of H generated by the norm topology. X is said to be a *random element* defined on Ω and taking values in H if $X : \Omega \to H$ is such that $X^{-1}B \in \mathcal{F}$ for every $B \in \mathcal{B}$. Define

$$\mu_X(B) = \mu(X^{-1}B), \quad B \in \mathcal{B}. \tag{3.1}$$

μ_X is called the *probability measure* induced by X on $\mathcal{B}$. Let (x, y) denote the inner product defined on H for $x, y \in H$. For any probability measure ν on $(H, \mathcal{B})$, the *characteristic functional* $\hat{\nu}(\cdot)$ is a functional defined on H

by the relation

$$\hat{\nu}(y) = \int_H e^{i(x,y)} d\nu(x), \quad y \in H. \tag{3.2}$$

The characteristic functional $\phi_X(\cdot)$ of X is given by

$$\begin{aligned} \phi_X(y) &= E[e^{i(y,X)}] \\ &= \int_H e^{i(x,y)} d\mu_X(x), \quad y \in H \\ &= \int_\Omega e^{i(X(\omega),y)} d\mu(\omega), \quad y \in H. \end{aligned} \tag{3.3}$$

It is known that there is a one-to-one correspondence between the characteristic functionals and the probability measures on H and the characteristic functional $\phi_X(\cdot)$ of a random element X satisfies the conditions

$$\phi_X(0) = 1, \quad |\phi_X(y)| \leq 1, \phi_X(y) = \overline{\phi_X(-y)}, y \in H, \tag{3.4}$$

where 0 denotes the null element in H. Moreover $\phi_X(\cdot)$ is continuous in the norm topology and positive definite. Further, if X and Y are independent random elements taking values in H, then $X + Y$ is a random element taking values in H and

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

For proofs of these results, see Parthasarathy (1968) or Grenander (1963).

We now prove an analogue of Theorem 2.1.1 for random elements taking values in a Hilbert space.

Theorem 3.1.1 : Let X_1, X_2 and X_3 be independent random elements taking values in a real separable Hilbert space H. Define

$$Z_1 = X_1 - X_3, \; Z_2 = X_2 - X_3. \tag{3.5}$$

Suppose the characteristic functional of (Z_1, Z_2) does not vanish. Then the probability measure of (Z_1, Z_2) determines the probability measures of X_1, X_2, X_3 up to change in location.

Proof : The characteristic functional of (Z_1, Z_2) is given by

$$\begin{aligned}\psi(y_1, y_2) &\equiv E[e^{i(Z_1,y_1)+i(Z_2,y_2)}] \\ &= E[e^{i(X_1-X_3,y_1)+i(X_2-X_3,y_2)}] \\ &= E[e^{i(X_1,y_1)}e^{i(X_2,y_2)}e^{i(X_3,-y_1-y_2)}] \\ &= \phi_1(y_1)\phi_2(y_2)\phi_3(-y_1-y_2), \quad y_1, y_2 \in H \end{aligned} \tag{3.6}$$

where $\phi_i(y)$ denotes the characteristic functional of X_i. Since $\psi(y_1, y_2) \neq 0$ for all y_1, y_2 in H by hypothesis, it follows that $\phi_i(y) \neq 0$ for $y \in H$ for $i = 1, 2, 3$. Suppose $\eta_i(y)$ is another possible characteristic functional of $X_i, i = 1, 2, 3$. Then

$$\phi_1(y_1)\phi_2(y_2)\phi_3(-y_1-y_2) = \eta_1(y_1)\eta_2(y_2)\eta_3(-y_1-y_2) \tag{3.7}$$

for all y_1, y_2 in H. Note that $\eta_i(y) \neq 0$ for $y \in H, 1 \leq i \leq 3$. Define $\zeta_i(y) = \log \frac{\phi_i(y)}{\eta_i(y)}$ where the logarithm denotes the continuous branch of the logarithm with $\zeta_i(0) = 0$. Note that $\zeta_i(y)$ is a continuous functional on H with $\zeta_i(0) = 0$ and $\zeta_i(y) = \overline{\zeta_i(-y)}$. Relation (3.7) implies that

$$\zeta_1(y_1) + \zeta_2(y_2) + \zeta_3(-y_1 - y_2) = 0\,. \tag{3.8}$$

Substituting $y_1 = 0 \in H$ in (3.8) we have

$$\zeta_2(y_2) = -\zeta_3(-y_2). \tag{3.9}$$

Let $y_2 = 0 \in H$ in (3.8). Then it follows that

$$\zeta_1(y_1) = -\zeta_3(-y_1)\ . \tag{3.10}$$

The above relations imply that

$$\zeta_3(-y_2) + \zeta_3(-y_1) = \zeta_3(-y_1 - y_2) \tag{3.11}$$

for all $y_1, y_2 \in H$ or equivalently

$$\zeta_3(y_1) + \zeta_3(y_2) = \zeta_3(y_1 + y_2), y_1, y_2 \in H. \tag{3.12}$$

Hence $\zeta_3(\cdot)$ is a complex-valued continuous linear functional on H. Since the space H is reflexive, every real-valued continuous linear functional is of the form (γ, y) for some $\gamma \in H$. In particular

$$\zeta_3(y) = (\alpha + i\delta, y), \quad y \in H \tag{3.13}$$

where $\alpha \in H$ and $\delta \in H$. Since $\zeta_3(y) = \overline{\zeta_3(-y)}$, it follows that

$$(\alpha, y) = (-\alpha, y), \qquad y \in H. \tag{3.14}$$

This proves that $\alpha = 0$ and hence

$$\phi_3(y) = n_3(y)e^{i(\delta,y)}, \quad y \in H\ . \tag{3.15}$$

Using the equations (3.9) and (3.10), it is easy to see that

$$\phi_k(y) = \eta_k(y)e^{i(\delta,y)}, \quad y \in H \tag{3.16}$$

for $k = 1, 2$. From the one-to-one correspondence between the characteristic functionals on H and the probability measures on H (cf. Parthasarathy (1968)), it follows that the distributions of the $X_k, 1 \leq k \leq 3$, are determined up to location. This completes the proof of the theorem. ■

One can extend Theorem 3.1.1 in the following way. The proof of the theorem is left as an exercise for the reader.

Theorem 3.1.2 : Let $X_1, X_2, \ldots, X_n$, be n independent random elements taking values in a real separable Hilbert space H. Define

$$Y_j = X_j + X_n, \quad 1 \leq j \leq n-1. \tag{3.17}$$

If the characteristic functional of $(Y_1, Y_2, \ldots, Y_{n-1})$ does not vanish, then the probability measure of $(Y_1, Y_2, \ldots, Y_{n-1})$ determines the probability measures of $X_1, X_2, \ldots, X_n$ up to change of location.

Remarks 3.1.1 : The results of this section are due to Kotlarski (1966c). As a special case of Theorem 3.1.1, we get an extension of Theorem 2.1.1 for random vectors.

3.2 Locally Convex Topological Vector Spaces

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and $\mathcal{X}$ be a real locally convex separable topological vector space with dual space $\mathcal{X}^*$. A mapping $X : \Omega \to \mathcal{X}$ is said to be a *random element* taking values in $\mathcal{X}$ if $X^{-1}(G) \in \mathcal{F}$ for G open in $\mathcal{X}$. The probability measure induced by X on $(\mathcal{X}, \mathcal{B})$ is defined by

$$\mu_X(B) = \mu[X^{-1}(B)), B \in \mathcal{B} \tag{3.18}$$

where $\mathcal{B}$ is the σ-algebra generated by the topology on $\mathcal{X}$. It is known that the *characteristic functional* of X, namely,

$$\begin{aligned} \phi_X(x^*) &= Ee^{i<x^*,X>} \\ &= \int_{\mathcal{X}} e^{i<x^*,x)} \mu_X(dx) \\ &= \int_{\Omega} e^{i<x^*,X(\omega))} \mu(d\omega))\ , x^* \in \mathcal{X}^* \end{aligned} \tag{3.19}$$

uniquely determines μ_X and it has properties similar to those of the characteristic function of a real-valued random variable (cf. Prohorov (1961), Vakhania (1981), Grenander (1963)). Here $< x^*, x >$ denotes the value of the linear functional $x^* \in \mathcal{X}^*$ at $x \in \mathcal{X}$.

Theorem 3.2.1 : Let $X_k, 1 \leq k \leq 3$, be independent random elements taking values in $\mathcal{X}$ and define

$$Z_1 = X_1 - X_3, Z_2 = X_2 - X_3. \tag{3.20}$$

If the characteristic functional of (Z_1, Z_2) does not vanish, then it determines the distributions of $X_k, 1 \leq k \leq 3$, up to a change of location.

Proof : Let $\phi(x_1^*, x_2^*)$ be the characteristic functional of (Z_1, Z_2). Observe that for x_1^*, x_2^* in $\mathcal{X}^*$,

$$\begin{aligned} \phi(x_1^*, x_2^*) &= E[\exp\{i < x_1^*, Z_1 > +i < x_2^*, Z_2 >\}] \\ &= E[\exp\{i < x_1^*, X_1 - X_3 > +i < x_2^*, X_2 - X_3 >\}] \\ &= E[\exp\{i < x_1^*, X_1 > +i < x_2^*, X_2 > +i < -x_1^* - x_2^*, X_3 >\}] \\ &= \phi_1(x_1^*)\phi_2(x_2^*)\phi_3(-x_1^* - x_2^*) \end{aligned} \tag{3.21}$$

where $\phi_i(x^*)$ is the characteristic functional of X_i. If $\psi_i(x^*)$ is an alternative possible characteristic functional of X_i for $1 \leq i \leq 3$ giving rise to the same distribution for (Z_1, Z_2), then it follows that

$$\phi_1(x_1^*)\phi_2(x_2^*)\phi_3(-x_1^* - x_2^*) = \psi_1(x_1^*)\psi_2(x_2^*)\psi_3(-x_1^* - x_2^*) \tag{3.22}$$

for all x_1^*, x_2^* in $\mathcal{X}^*$. Since $\phi(x_1^*, x_2^*) \neq 0$ for all $x_1^*, x_2^* \in \mathcal{X}^*$ by hypothesis, it follows that none of the ϕ_i and ψ_i vanish. Let

$$g_k(x^*) = \frac{\phi_k(x^*)}{\psi_k(x^*)}, \quad 1 \leq k \leq 3, x^* \in \mathcal{X}^*. \tag{3.23}$$

Then relation (3.22) reduces to

$$g_1(x_1^*)g_2(x_2^*)g_3(-x_1^* - x_2^*) = 1 \tag{3.24}$$

for all x_1^* and x_2^* in $\mathcal{X}^*$. Substituting $x_2^* = 0$ and $x_1^* = 0$ alternately, it is easy to see that

$$g_3(x_1^* + x_2^*) = g_3(x_1^*)g_3(x_2^*) \tag{3.25}$$

for all x_1^*, x_2^* in $\mathcal{X}^*$ using the fact $g_3(0) = 1, g_3(-x^*) = \overline{g_3(x^*)}$ where 0 is the null element in $\mathcal{X}^*$. Let $h(x^*) = \log g_3(x^*)$ where the logarithm is the continuous branch satisfying the condition log $g_3(0) = 0$. Since $g_3(x^*)$ is continuous in the weak* topology, it follows that $h(x^*)$ is also continuous in the weak * topology on $\mathcal{X}^*$. Furthermore

$$h(x_1^*) + h(x_2^*) = h(x_1^* + x_2^*), x_1^*, x_2^* \in \mathcal{X}^* \quad . \tag{3.26}$$

Hence h is a complex-valued linear functional continuous in the weak* topology on $\mathcal{X}^*$. By Banach's theorem (cf. Yosida (1965)), it follows that there exist x_0 and y_0 in $\mathcal{X}$ such that

$$h(x^*) =< y_0, x^* > +i < x_0, x^* >, x^* \in \mathcal{X}^*. \tag{3.27}$$

Note that $h(-x^*) = \overline{h(x^*)}, x^* \in \mathcal{X}$. Hence $y_0 = 0$. This in turn implies that

$$h(x^*) = i < x_0, x^* >, x^* \in \mathcal{X}^* \ . \tag{3.28}$$

Hence

$$g_3(x^*) = e^{i<x_0,x^*>}, x^* \in H \tag{3.29}$$

or equivalently

$$\phi_3(x^*) = \psi_3(x^*)e^{i<x_0,x^*>}, \quad x^* \in H. \tag{3.30}$$

It is easy then to see that

$$\phi_j(x^*) = \psi_j(x^*)e^{i<x_0,x^*>}, \quad x^* \in H, j = 1, 2 \tag{3.31}$$

using (3.24). These relations prove that the distributions of X_j for $1 \le j \le 3$ are determined up to change of location. This completes the proof of theorem. ■

Remarks 3.2.1 : The above theorem can be extended to weak-measurable random elements taking values in $\mathcal{X}$ in the following sense. Suppose that $\mathcal{Y}$ is a subspace of the dual space $\mathcal{X}^*$ of $\mathcal{X}$ and $\mathcal{Y}$ is *total* over $\mathcal{X}$ (cf. Wilansky (1978, p. 95)). A function $X : \Omega \to \mathcal{X}$ is said to be $\mathcal{Y}$-measurable if $< z, X >$ is measurable for all $z \in \mathcal{Y}$. X_1, X_2, X_3 are said to be $\mathcal{Y}$-independent if, for any $\gamma_1, \gamma_2, \gamma_3$ in $\mathcal{Y}$, the elements of the set $< \gamma_i, X_i >, 1 \le i \le 3$, are independent random variables. Alspach and Kotlarski (1986a) obtained a generalization of Theorem 3.2.1 to $\mathcal{Y}$-independent random elements and gave explicit formulae for the characteristic functionals of X_1, X_2, X_3 in terms of the characteristic functional of (Z_1, Z_2) under some additional conditions, where $Z_1 = X_1 - X_3$ and $Z_2 = X_2 - X_3$.

3.3 Locally Compact Abelian Groups

Let $\mathcal{X}$ denote a locally compact abelian separable metric group. Suppose $\mathcal{X}$ is a multiplicative group and $\mathcal{Y}$ its character group. For $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, let (x, y) denote the value of the character y at x. By Pontryagin's duality theory, the relation between $\mathcal{X}$ and $\mathcal{Y}$ is symmetric, that is $\mathcal{X}$ is the character group of $\mathcal{Y}$. Further the character group of the direct product $\mathcal{X} \times \mathcal{X}$ is isomorphic and homeomorphic to $\mathcal{Y} \times \mathcal{Y}$. For more information on such groups, see Loomis (1953) or Hewitt and Ross (1963).

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. A mapping $X : \Omega \to \mathcal{X}$ is said to be a *random element* taking values in $\mathcal{X}$ if $X^{-1}G \in \mathcal{F}$ for every G open in $\mathcal{X}$. The distribution of X is given by the measure

$$\mu_X(B) = \mu\{[\omega : X(\omega) \in B]\} \tag{3.32}$$

for all $B \in \mathcal{B}$ where $\mathcal{B}$ is the σ-algebra generated by the open sets in $\mathcal{X}$. Random elements X_1, X_2 are said to be *independent* if

$$\begin{aligned} &\mu\{[\omega : (X_1(\omega), X_2(\omega)) \in B_1 \times B_2]\} \\ &\qquad = \mu\{[\omega : X_1(\omega) \in B_1]\}\mu\{[\omega : X_2(\omega) \in B_2]\} \end{aligned} \tag{3.33}$$

for all B_1 and B_2 in $\mathcal{B}$. Let ν be a probability measure on $\mathcal{X}$. The *characteristic functional* $\hat{\nu}$ of ν is a complex-valued function on the character group $\mathcal{Y}$ defined by

$$\begin{aligned} \hat{\nu}(y) &= \int_{\mathcal{X}} (x, y) d\nu(x), \quad y \in \mathcal{Y} \\ &= \int_{\Omega} (X(\omega), y) d\mu(\omega)\ , y \in \mathcal{Y} \end{aligned} \tag{3.34}$$

if X is distributed with probability measure ν.

It is known that $\hat{\nu}$ determines ν uniquely, $\hat{\nu}(y)$ is a uniformly continuous functional of y, $\hat{\nu}(e) = 1$ where e is the identity character in $\mathcal{Y}$ and $|\hat{\nu}(y)| \leq 1$ for all $y \in \mathcal{Y}$. For details, see Grenander (1963) or Parthasarathy (1968).

Theorem 3.3.1 : Let X_1, X_2, X_3 be three independent random elements taking values in a locally compact abelian separable metric group $\mathcal{X}$. Let

$$Z_1 = X_1X_2 \text{ and } Z_2 = X_1X_3. \tag{3.35}$$

If the characteristic functional of (Z_1, Z_2) does not vanish, then the joint distribution of (Z_1, Z_2) determines the distributions of X_1, X_2, X_3 up to a change of scale.

Proof : Let λ denote the joint distribution of $\boldsymbol{Z} = (Z_1, Z_2)$. Since the character group of the product $\mathcal{X} \times \mathcal{X}$ is isomorphic and homeomorphic to

$\mathcal{Y} \times \mathcal{Y}$, we can identify the elements of the character group of $\mathcal{X} \times \mathcal{X}$ by $y_1 y_2$ where $y_1 \in \mathcal{Y}$ and $y_2 \in \mathcal{Y}$. By the definition of the characteristic functional $\hat{\lambda}$ of λ, it follows that

$$\begin{aligned}
\hat{\lambda}(y_1 y_2) &= \int_\Omega (Z, y_1 y_2)\mu(d\omega) \\
&= \int_\Omega (Z_1(\omega), y_1)(Z_2(\omega), y_2)\mu(d\omega) \\
&= \int_\Omega (X_1(\omega)X_2(\omega), y_1)(X_1(\omega)X_3(\omega), y_2)\mu(d\omega) \\
&= \int_\Omega (X_1(\omega), y_1 y_2)(X_2(\omega), y_1)(X_3(\omega), y_2)\mu(d\omega) \\
&= \int_\Omega (X_1(\omega), y_1 y_2)\mu(d\omega) \int_\Omega (X_2(\omega), y_1)\mu(d\omega) \int_\Omega (X_3(\omega), y_2)\mu(d\omega) \\
&= \hat{\nu}_1(y_1 y_2)\hat{\nu}_2(y_1)\hat{\nu}_3(y_2), \quad y_1, y_2 \in \mathcal{Y}, \qquad (3.36)
\end{aligned}$$

where $\hat{\nu}_i$ is the characteristic functional of X_i for $1 \leq i \leq 3$. Since $\hat{\lambda}(y_1 y_2) \neq 0$ for all $y_1, y_2 \in \mathcal{Y}$, it follows that

$$\hat{\nu}_i(y) \neq 0 \text{ for } y \in \mathcal{Y}, i = 1, 2, 3. \qquad (3.37)$$

Suppose $\hat{\eta}_i$ is another possible characteristic functional for $X_i, 1 \leq i \leq 3$, such that

$$\hat{\lambda}(y_1 y_2) = \hat{\eta}_1(y_1 y_2)\hat{\eta}_2(y_1)\hat{\eta}_3(y_2), y_1, y_2 \in \mathcal{Y}. \qquad (3.38)$$

Note that $\hat{\eta}_i(y) \neq 0$ for $y \in \mathcal{Y}, 1 \leq i \leq 3$. Let

$$\hat{\psi}(y) = \hat{\nu}_i(y)/\hat{\eta}_i(y), y \in \mathcal{Y}, 1 \leq i \leq 3. \qquad (3.39)$$

$\hat{\psi}_i(y)$ is well defined and the relations (3.37) and (3.38) prove that

$$\hat{\psi}_1(y_1 y_2)\hat{\psi}_2(y_1)\hat{\psi}_3(y_2) = 1, y_1, y_2 \in \mathcal{Y}. \qquad (3.40)$$

Since $\hat{\psi}_2(e) = 1, \hat{\psi}_3(e) = 1$ and $\hat{\psi}(y) = \overline{\hat{\psi}(y^{-1})}$, it is easy to see that

$$\hat{\psi}_1(y_1 y_2) = \hat{\psi}_1(y_1)\hat{\psi}_1(y_2), y_1, y_2 \in \mathcal{Y}. \qquad (3.41)$$

Furthermore $\hat{\psi}_1(y)$ is continuous. Hence $\hat{\psi}_1$ is a continuous homomorphism on the locally compact abelian group $\mathcal{Y}$ into the multiplicative group of complex numbers of absolute value one. Therefore $\hat{\psi}_1$ is a character on $\mathcal{Y}$. Since the character group of $\mathcal{Y}$ is $\mathcal{X}$ by Pontryagin's duality theory (cf. Hewitt and Ross (1963)), it follows that

$$\hat{\psi}_1(y) = (x_0, y) \tag{3.42}$$

for some $x_0 \in \mathcal{X}$. This relation proves that

$$\hat{\nu}_1(y) = \hat{\eta}_1(y)(x_0, y), y \in \mathcal{Y}. \tag{3.43}$$

Similarly, it can be shown, using (3.40), that

$$\hat{\nu}_i(y) = \hat{\eta}_i(y)(x_0^{-1}, y), y \in \mathcal{Y}, i = 1, 2. \tag{3.44}$$

Hence the distributions of $X_i, 1 \leq i \leq 3$, are determined up to a change of scale. This completes the proof of the theorem. ■

Remarks 3.3.1 :The results given above are due to Prakasa Rao (1968). Flusser (1972) extended Theorem 3.3.1 characterizing the marginal distributions of a random vector $\boldsymbol{X} = (X_0, X_1, X_2)$ with X_0, X_1 and X_2 independent and with values in a locally compact abelian group $\mathcal{X}$ in terms of the joint probability measure of $\boldsymbol{Z}$ where $\boldsymbol{Z} = T(\boldsymbol{X})$ and T is a homomorphism on $\mathcal{X}$ satisfying certain conditions. We now state his result. For the proof, see Flusser (1972).

Theorem 3.3.2 : Let $\mathcal{X}$ be a locally compact abelian separable metric group and suppose $\mathcal{X}$ is the direct sum of three of its subgroups $\mathcal{X}_0, \mathcal{X}_1$ and $\mathcal{X}_2$. For $k = 0, 1, 2$, let π_k be the projection of $\mathcal{X}$ onto its kth direct summand. Let $\boldsymbol{X}$ be a random element with values in $\mathcal{X}$ and define $X_k = \pi_k \boldsymbol{X}, k = 0, 1, 2$. Suppose $X_k, k = 0, 1, 2$, are independent random elements with values in $\mathcal{X}_0, \mathcal{X}_1$ and $\mathcal{X}_2$ respectively and that the characteristic functionals of X_0, X_1 and X_2 do not vanish. Let τ be another locally compact abelian separable metric group and let $T : \mathcal{X} \to \tau$ be a continuous homomorphism from $\mathcal{X}$ onto τ. Let $T_k = T\pi_k, k = 0, 1, 2$. Further assume that

(i) $T_0|_{\mathcal{X}_0}$ is injective ,

(ii) $(T_1 + T_2)|_{\mathcal{X}_1 \oplus \mathcal{X}_2}$ is bijective and

(iii) $T(\mathcal{X}_0) \cap T(\mathcal{X}_1) = \{0\}$ and $T(\mathcal{X}_0) \cap T(\mathcal{X}_2) = \{0\}$ where 0 is the identity element in τ.

Let $\boldsymbol{Z} = T(\boldsymbol{X})$. Then the distribution of $\boldsymbol{Z}$ determines the distributions of X_0, X_1 and X_2 up to shifts. The shift for X_0 is given by an element $x_0 \in \mathcal{X}_0$ and those for X_1 and X_2 are determined by x_0 (Here $T_0|_{\mathcal{X}_0}$ denotes restriction of T_0 to the set $\mathcal{X}_0$) .

Remarks 3.3.2 : The relation $\mathcal{X} = \mathcal{X}_0 \oplus \mathcal{X}_1 \oplus \mathcal{X}_2$ and the condition (ii) in Theorem 3.3.2 imply that τ is isomorphic and homomorphic to $\mathcal{X}_1 \oplus \mathcal{X}_2$. In fact $\tau = T(\mathcal{X}_1) \oplus T(\mathcal{X}_2)$. If we define $Y = Y_1 + Y_2$ where $Y_k = \pi'_k Y$ where π'_k is the projection of τ onto $T(\mathcal{X}_k)$, then the joint distribution of (Y_1, Y_2) determines the distributions of X_0, X_1 and X_2 up to shifts.

Remarks 3.3.3 : Rao (1971) proved that if $X_i, 0 \leq i \leq 3$, are four independent real-valued random variables and if Y_1, Y_2 are two linear functions of $X_i, 0 \leq i \leq 3$, then the joint distribution of (Y_1, Y_2) determines the distributions of $X_i, 0 \leq i \leq 3$ up to a normal factor, possibly degenerate under some conditions. Prakasa Rao (1975a) generalized this result to locally compact abelian separable metric groups extending the result of Flusser (1972) and Prakasa Rao (1968).

3.4 Abelian Semigroups

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Let $(\mathcal{X}, \mathcal{B})$ be a measurable space where $\mathcal{X}$ is a separable Hausdorff topological space and $\mathcal{B}$ is a σ-algebra of subsets of $\mathcal{X}$ generated by the open sets of $\mathcal{X}$. X is said to be a *random element* defined on Ω taking values in $\mathcal{X}$ if $X : \Omega \to \mathcal{X}$ is such that $X^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$. Define

$$\mu_X(A) = \mu(X^{-1}A), A \in \mathcal{B}. \tag{3.45}$$

μ_X is called the probability measure induced by X on $(\mathcal{X}, \mathcal{B})$. Let X and

Y be random elements taking values in $\mathcal{X}$. Random elements X and Y are said to be *independent* if

$$\mu\{[\omega : X(\omega) \in A_1, Y(\omega) \in A_2]\}$$

$$= \mu\{[\omega : X(\omega) \in A_1]\}\mu\{[\omega : Y(\omega) \in A_2]\} \tag{3.46}$$

for all A_1, A_2 in $\mathcal{B}$.

Let 0 and ∇ be two abelian semigroup operations on $\mathcal{X}$, i.e.,

(i) $x_1, x_2 \in \mathcal{X} \Rightarrow x_1 o x_2 \in \mathcal{X}$ and $x_1 \nabla x_2 \in \mathcal{X}$;

(ii) $x_1, x_2 \in \mathcal{X} \Rightarrow x_1 o x_2 = x_2 o x_1$ and $x_1 \nabla x_2 = x_2 \nabla x_1$;

(iii) if $x_1, x_2, x_3 \in \mathcal{X}$, then

$$(x_1 o x_2) o x_3 = x_1 o (x_2 o x_3) \tag{3.47}$$

and

$$(x_1 \nabla x_2) \nabla x_3 = x_1 \nabla (x_2 \nabla x_3);$$

(iv) both $x_1 o x_2$ and $x_1 \nabla x_2$ are continuous on $\mathcal{X} \times \mathcal{X}$; and

(v) there exist two identity elements $e^{(1)}$ and $e^{(2)}$ in $\mathcal{X}$ such that $e^{(1)} o x = x = e^{(2)} \nabla x$ for all $x \in \mathcal{X}$.

Let X_0, X_1, X_2 be three independent random elements on $(\Omega, \mathcal{F}, \mu)$ with values in $(\mathcal{X}, \mathcal{B})$. Let

$$Z = (Z_1, Z_2) \tag{3.48}$$

where

$$Z_1 = X_0 o X_1 \text{ and } Z_2 = X_0 \nabla X_2. \tag{3.49}$$

Then $\boldsymbol{Z}$ is a random element on $(\Omega, \mathcal{F}, \mu)$ with values in $\mathcal{X} \times \mathcal{X}$. For any $B_1, B_2 \in \mathcal{B}$,

$$\begin{aligned}
\mu_{(Z_1,Z_2)}&(B_1 \times B_2) \\
&= E(\chi_{B_1}(Z_1)\chi_{B_2}(Z_2)) \\
&= E(\chi_{B_1}(X_0 o X_1)\chi_{B_2}(X_0 \nabla X_2)) \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{X}} \chi_{B_1}(x_0 o x_1)\chi_{B_2}(x_0 \nabla x_2)\mu_{X_0}(dx_0)\mu_{X_1}(dx_1)\mu_{X_2}(dx_2)
\end{aligned} \tag{3.50}$$

where χ_B denotes the indicator function of the set B.

Suppose $(\mathcal{X}, \mathcal{B}, 0, \nabla)$ is a measurable space with a double abelian semigroup operation structure; that is $(\mathcal{X}, \mathcal{B})$ is a measurable space as described above where 0 and ∇ are two (identical or distinct) abelian semigroup operations.

Let the kernels

$$K(x,u), L(x,v), x \in \mathcal{X}, u \in \mathcal{U}, v \in \mathcal{V} \tag{3.51}$$

be two complex-valued functions such that

(i) K and L are both continuous in x on $\mathcal{X}$;

(ii) $|K(x,u)| \leq 1, |L(x,v)| \leq 1$ for all $x \in \mathcal{X}, u \in \mathcal{U}, v \in \mathcal{V}$;

(iii) $K(x_1 o x_2, u) = K(x_1, u)K(x_2, u)$ for all $x_1, x_2 \in \mathcal{X}$ and $u \in \mathcal{U}$, $L(x_1 \nabla x_2, v) = L(x_1, v)L(x_2, v)$ for all $x_1, x_2 \in \mathcal{X}$ and $v \in \mathcal{V}$;

(iv) $$K(e^{(1)}, u) = 1 = L(e^{(2)}, v) \text{ for all } u \in \mathcal{U} \text{ and } v \in \mathcal{V}; \tag{3.52}$$

and

(v) there exist $u_0 \in \mathcal{U}$ and $v_0 \in \mathcal{V}$ such that

$$K(x, u_0) = 1 = L(x, v_0), x \in \mathcal{X}. \tag{3.53}$$

The function $K(x,u), x \in \mathcal{X}, u \in \mathcal{U}$ is called a *kernel* on $\mathcal{X}$ with the *characteristic set* $\mathcal{U}$ for the abelian semigroup operation o. Similarly, the function $L(x,v), x \in \mathcal{X}, v \in \mathcal{V}$ is a *kernel* on $\mathcal{X}$ with the *characteristic set* $\mathcal{V}$ for the semigroup operation ∇. The function

$$K(x,u)L(x,v), x \in \mathcal{X}, u \in \mathcal{U}, v \in \mathcal{V} \tag{3.54}$$

is called a *double kernel* on $\mathcal{X}$ with the characteristic set $\mathcal{U} \times \mathcal{V}$ for the pair of semigroup operations (o, ∇). For any random element X with values in $\mathcal{X}$, define

$$\begin{aligned} \Phi_X^K(u) &= EK(X,u), & u \in \mathcal{U} \\ \Phi_X^L(v) &= EL(X,v), & v \in \mathcal{V} \\ \Phi_X^{KL}(u,v) &= E[K(X,u)L(X,v)], & u \in \mathcal{U}, v \in \mathcal{V}. \end{aligned} \tag{3.55}$$

Note that

$$\Phi_X^{KL}(u, v_0) = \Phi_X^K(u); \Phi_X^{KL}(u_0, v) = \Phi_X^L(v) \tag{3.56}$$

for $u \in \mathcal{U}, v \in \mathcal{V}$ where u_0 and v_0 are as defined by (3.53).

The function Φ_X^K is called the *characteristic functional* of the random element X corresponding to the kernel K if it determines the probability measure of X on $(\mathcal{X}, \mathcal{B})$ *uniquely* .

If X_1 and X_2 are random elements taking values in $(\mathcal{X}, \mathcal{B}, o, \nabla)$, then

$$\Phi_{X_1,X_2}^{KL}(u, v) = E[K(X_1, u)L(X_2, v)], u \in \mathcal{U}, v \in \mathcal{V} \tag{3.57}$$

is the characteristic functional of (X_1, X_2) corresponding to the double kernel KL if it determines the probability measure of (X_1, X_2) *uniquely.*

Define $\boldsymbol{Z} = (Z_1, Z_2)$ by (3.48). Then

$$\begin{aligned}
\Phi_Z^{KL}(u, v) &= \Phi_{(Z_1,Z_2)}^{KL}(u, v) \\
&= E[K(Z_1, u)L(Z_2, v)] \\
&= E[K(X_0 o X_1, u)L(X_0 \nabla X_2, v)] \\
&= E[K(X_0, u)K(X_1, u)L(X_0, v)L(X_2, v)] \\
&= E[K(X_0, u)L(X_0, v)]E[K(X_1, u)]E[L(X_2, v)] \\
&= \Phi_{X_0}^{KL}(u, v)\Phi_{X_1}^K(u)\Phi_{X_2}^L(v), u \in \mathcal{U}, v \in \mathcal{V} \,. \qquad (3.58)
\end{aligned}$$

Let $\Psi_{X_0}^{KL}, \Psi_{X_1}^K$ and $\Psi_{X_2}^L$ be another alternative triple of possible characteristic functionals of X_0, X_1 and X_2 as defined above. Further suppose that

$$\Phi_Z^{KL}(u, v) = \Psi_Z^{KL}(u, v) \neq 0, u \in \mathcal{U}, v \in \mathcal{V}. \tag{3.59}$$

We have the relation

$$\Phi_{X_0}^{KL}(u, v)\Phi_{X_1}^K(u)\Phi_{X_2}^L(v) = \Psi_{X_0}^{KL}(u, v)\Psi_{X_1}^K(u)\Psi_{X_2}^L(v). \tag{3.60}$$

Substituting $u = u_0$ and $v = v_0$ alternately, it can be checked that

$$\Phi_{X_0}^K(u)\Phi_{X_1}^K(u) = \Psi_{X_0}^K(u)\Psi_{X_1}^K(u), u \in \mathcal{U} \tag{3.61}$$

and

$$\Phi_{X_0}^L(v)\Phi_{X_2}^L(v) = \Psi_{X_0}^L(v)\Psi_{X_2}^L(v), v \in \mathcal{V} \,. \tag{3.62}$$

Relations (3.60) to (3.62) show that

$$\frac{\Phi_{X_0}^{KL}(u,v)}{\Phi_{X_0}^{K}(u)\Phi_{X_0}^{L}(v)} = \frac{\Psi_{X_0}^{KL}(u,v)}{\Psi_{X_0}^{K}(u)\Psi_{X_0}^{L}(v)}, u \in \mathcal{U}, v \in \mathcal{V} \ . \tag{3.63}$$

If the solution $(\Phi_{X_0}^K, \Phi_{X_0}^L, \Phi_{X_0}^{KL})$ of this equation is unique, then we obtain that $\Phi_{X_0}^{KL} = \Psi_{X_0}^{KL}$ and the relations (3.61) and (3.62) show that $\Phi_{X_1}^K = \Psi_{X_1}^K$ and $\Phi_{X_2}^L = \Psi_{X_2}^L$. This in turn shows that the distributions of X_0, X_1 and X_2 are determined uniquely.

Remarks 3.4.1 : Examples of the results obtained above are discussed in Chapter 2, for instance, characterizing the probability distributions of components by the joint distribution of their product and their sum. The discussion in this section is from Alspach and Kotlarski (1986b).

3.5 Homogeneous Spaces

Let P and Q be probability measures defined on the σ-algebra $\mathcal{B}$ of Borel subsets of a homogeneous space (cf. Kelley (1953,p. 107)) $\mathcal{X} = G/H$ where G is a locally compact separable group of transformations and H a subgroup of G. Suppose that $X_i, 1 \leq i \leq n$, are independent identically distributed random elements taking values in $\mathcal{X}$ distributed according to P or Q. A function f defined on $\mathcal{X}^n$ is said to be *invariant* with respect to G if

$$f(gx_1, \ldots, gx_n) = f(x_1, \ldots, x_n) \tag{3.64}$$

for all $(x_1, \ldots, x_n) \in \mathcal{X}^n$ and $g \in G$. Suppose that the distribution of any invariant function computed with respect to P is the same as that computed with respect to Q. The problem is to find conditions under which P and Q agree to within a shift by an element of G, that is, $P(E) = Q(g_0^{-1}E)$ for all $E \in \mathcal{B}$ for some $g_0 \in G$. This problem was discussed in Chapter 2 in the case of the real line and in earlier sections of this chapter for the case of Hilbert spaces and locally compact abelian groups.

Let us consider the special case when G is a compact group. Denote by

$u(G)$ the set of all unitary irreducible (finite-dimensional) representations of the group G (cf. Vilenkin (1968)). Let $\mathcal{A}$ be the equivalence classes of sets of representations under the usual definition. Let U_α be a member from the equivalence class for each $\alpha \in \mathcal{A}$. Define $\tilde{P}$, the characteristic functional of P, by the relation

$$\hat{P}(\alpha) = \int_G U_\alpha(g)dP(g), \alpha \in \mathcal{A}. \tag{3.65}$$

Here we have assumed that P is defined on the group G by extending P on G/H using the relation $P(Eh) = P(E)$ for $h \in H$. It is known that the characteristic functional $\hat{P}(\cdot)$ uniquely determines the probability measure P on G (cf. Grenander (1963)).

Rukhin (1975) proved the following theorem. We omit the proof.

Theorem 3.5.1 : Suppose P and Q are probability measures defined on the σ-algebra of Borel sets $\mathcal{B}$ of a compact group G. Further suppose that the characteristic functionals $\hat{P}(\alpha)$ and $\hat{Q}(\alpha)$ are nonsingular for all $\alpha \in \mathcal{A}$. If

$$E_P[f(X_1, \ldots, X_n)] = E_Q[f(X_1, \ldots, X_n)] \tag{3.66}$$

for all invariant functions f and some $n \geq 3$, then

$$Q(E) = P(g_0^{-1}E), E \in \mathcal{B} \tag{3.67}$$

for some $g_0 \in G$.

The result has been extended in the following form to random elements X_i which are independent but not necessarily identically distributed.

Theorem 3.5.2 : Let $X_i, 1 \leq i \leq n, n \geq 3$ be independent random elements with values in a compact group G. Suppose the distribution of each invariant function $f(X_1, X_2, \ldots, X_n)$ when X_i is distributed with probability measure P_i on G for $1 \leq i \leq n$ is the same as its distribution when X_i

is distributed with probability measure Q_i on G for $1 \le i \le n$. Further suppose that

$$\det[\hat{P}_i(\alpha)] \neq 0, \alpha \in \mathcal{A}, 1 \le i \le n \ . \tag{3.68}$$

Then there exists $g_0 \in G$ such that

$$Q_j(E) = P_j(g_0^{-1}E), E \in \mathcal{B}, 1 \le j \le n. \tag{3.69}$$

For proofs of above results and for further remarks, see Rukhin (1975, 1977).

3.6 Generalized Random Fields

Let $\mathcal{X}$ be the space of all real-valued functions $\phi(\boldsymbol{x}) = \phi(x_1, \ldots, x_n)$ of n real variables which are infinitely differentiable and have bounded supports. A sequence $\{\phi_m\}$ of functions in $\mathcal{X}$ is said to converge to zero if there exists a constant a such that ϕ_m vanishes for $||\boldsymbol{x}|| \ge a$ for all m and, if for every q, the sequence $\{\phi_m^{(q)}\}$ converges uniformly to zero. Here $||\boldsymbol{x}||$ is the Euclidean norm on R^n and $\phi^{(q)}$ denotes any qth-order partial derivative of ϕ. Any continuous linear functional on $\mathcal{X}$ is called a *generalized function.*

A *random functional* Φ is defined on $\mathcal{X}$ if for every $\phi \in \mathcal{X}$ there is associated a real-valued random variable $\Phi(\phi)$. In other words, for every k elements $\phi_i, 1 \le i \le k$, in $\mathcal{X}$, the joint distribution of $(\Phi(\phi_1), \ldots, \Phi(\phi_k))$ is specified and these probability distributions form a consistent family in the sense of Kolmogorov. The random functional $\Phi(\cdot)$ is said to be linear if

$$\Phi(\alpha\phi + \beta\psi) = \alpha\Phi(\phi) + \beta\Phi(\psi) \text{ a.s.} \tag{3.70}$$

for $\phi, \psi \in \mathcal{X}$ and α, β real. $\Phi(\cdot)$ is said to be continuous if $\phi_{k_j} \to \phi_j$, $1 \le j \le m$, with $\phi_{k_j}, \phi_j \in \mathcal{X}$, imply that $P_m \Rightarrow P$ where P_m is the probability measure of $(\Phi(\phi_{k_1}), \ldots, \Phi(\phi_{k_m}))$ and P is the probability measure of $(\Phi(\phi_1), \ldots, \Phi(\phi_m))$ on R^m. Here "$\Rightarrow$" denotes the weak convergence of probability measures (cf. Billingsley (1968)).

Any continuous linear random functional on $\mathcal{X}$ is called a *generalized random function* . If $\mathcal{X}$ consists of functions of one variable, then the corresponding random function Φ is called a *generalized random process* . If $\mathcal{X}$ consists of functions of several variables, then the functional Φ is called a *generalized random field.*

Let Φ and Ψ be two generalized random fields on $\mathcal{X}$. Φ and Ψ are said to be *independent* if the set of random variabes $\{\Phi(\phi) : \phi \in \mathcal{X}\}$ is independent of the set $\{\Psi(\phi) : \phi \in \mathcal{X}\}$. This notion can be extended to any finite number of generalized random fields.

For any generalized random field Φ, define

$$L(\phi) = E[e^{i\Phi(\phi)}], \phi \in \mathcal{X}. \tag{3.71}$$

$L(\cdot)$ is called the *characteristic functional* of the generalized random field Φ. It can be shown that

$$L(0) = 1, L(-\phi) = \overline{L(\phi)}, |L(\phi)| \leq 1, \tag{3.72}$$

and $L(\cdot)$ is the continuous functional on $\mathcal{X}$. In fact, there exists a one-to-one correspondence between the characteristic functionals L and generalized random fields Φ on $\mathcal{X}$.

For any two generalized random fields Φ and Ψ, the joint characteristic functional of the two-dimensional generalized random field (Φ, Ψ) is defined by

$$L(\phi, \psi) = E[e^{i\Phi(\phi)+i\Psi(\psi)}], \phi \in \mathcal{X}, \psi \in \mathcal{X}. \tag{3.73}$$

Let Φ_1 and Φ_2 be two generalized random fields on $\mathcal{X}$ and f and g be any two infinitely differentiable functions. The generalized random field $f\Phi_1 + g\Phi_2$ is defined by the relation

$$(f\Phi_1 + g\Phi_2)(\phi) = \Phi_1(f\phi) + \Phi_2(g\phi), \phi \in \mathcal{X}. \tag{3.74}$$

Two generalized random fields Φ_1 and Φ_2 are said to be *determined up to shift* if there exists a generalized function m such that $\Phi_1 = \Phi_2 + m$.

We refer the reader to Gelfand and Vilenkin (1964) for further results on generalized random fields.

Theorem 3.6.1 : Let $\Phi_i, 0 \leq i \leq 2$, be three independent generalized random fields on $\mathcal{X}$ and define

$$\begin{aligned} \Psi_1 &= \Phi_0 + \Phi_1 + \Phi_2, \\ \Psi_2 &= \beta_0\Phi_0 + \beta_1\Phi_1 + \beta_2\Phi_2 \end{aligned} \tag{3.75}$$

where $\beta_i, 0 \leq i \leq 2$ are infinitely differentiable functions such that $\beta_i(\boldsymbol{x}) \neq \beta_j(\boldsymbol{x})$ for $i \neq j$ and all $\boldsymbol{x}$. Suppose the joint characteristic functional of (Ψ_1, Ψ_2) does not vanish. Then the two-dimensional generalized random field (Ψ_1, Ψ_2) determines the generalized random fields Φ_0, Φ_1, Φ_2 up to shift.

Proof: Let $\Gamma_i, 0 \leq i \leq 2$ be three independent generalized random fields on $\mathcal{X}$ such that the two-dimensional generalized random field (Σ_1, Σ_2) where

$$\begin{aligned} \Sigma_1 &= \Gamma_0 + \Gamma_1 + \Gamma_2, \\ \Sigma_2 &= \beta_0\Gamma_0 + \beta_1\Gamma_1 + \beta_2\Gamma_2 \end{aligned} \tag{3.76}$$

has the same joint characteristic functional $H(\phi, \psi)$ as (Ψ_1, Ψ_2). Let $L_i(\cdot)$ and $M_i(\cdot), 0 \leq i \leq 2$ be the characteristic functionals of Φ_i and $\Gamma_i, 0 \leq i \leq 2$, respectively. It is easy to see that

$$H(\phi, \psi) = \Pi_{i=0}^2 M_i(\phi + \beta_i\psi) = \Pi_{i=0}^2 L_i(\phi + \beta_i\psi) \tag{3.77}$$

for ϕ, ψ in $\mathcal{X}$. Since $H(\phi, \psi) \neq 0$ for all ϕ, ψ in $\mathcal{X}$, $J_i(\phi) = \log\{L_i(\phi)/M_i(\phi)\}$ is well defined where the logarithm is taken to be the continuous branch with $J_i(0) = 0$. Then it follows that

$$\sum_{i=0}^{2} J_i(\phi + \beta_i\psi) = 0, \phi, \psi \in \mathcal{X}. \tag{3.78}$$

Let ϕ, ψ and λ be fixed in $\mathcal{X}$ and let $\phi' = \phi - \beta_2\lambda$ and $\psi' = \psi + \lambda$. Then ϕ' and ψ' belong to $\mathcal{X}$ and the equation (3.78) implies that

$$\sum_{i=0}^{2} J_i(\phi' + \beta_i\psi') = 0 \ . \tag{3.79}$$

Substracting (3.78) from (3.79), we obtain the equation

$$\sum_{i=0}^{1}[J_i(\phi' + \beta_i\psi') - J_i(\phi + \beta_i\psi)] = 0 \tag{3.80}$$

for all ϕ, ψ and λ in $\mathcal{X}$ since $\phi' + \beta_2\psi' = \phi + \beta_2\psi$. Let

$$W_i(\phi) = J_i(\phi + \lambda(\beta_i - \beta_2)) - J_i(\phi), i = 0, 1 \tag{3.81}$$

for any fixed λ in $\mathcal{X}$. Relation (3.80) implies that

$$W_0(\phi + \beta_0\psi) + W_1(\phi + \beta_1\psi) = 0 \tag{3.82}$$

for any ϕ, ψ in $\mathcal{X}$. Let ϕ, ψ and ν be fixed in $\mathcal{X}$ and let $\phi' = \phi - \beta_1\nu$ and $\psi' = \psi + \nu$. By arguments similar to those given above, we obtain the relation

$$Y_0(\phi + \beta_0\psi) = 0 \tag{3.83}$$

where

$$Y_0(\phi) = W_0(\phi + \nu(\beta_0 - \beta_1)) - W_0(\phi), \phi \in \mathcal{X}. \tag{3.84}$$

Relation (3.83) implies that

$$Y_0(\phi) = 0, \quad \phi \in \mathcal{X} \tag{3.85}$$

which in turn shows that

$$W_0(\phi + \nu(\beta_0 - \beta_1)) = W_0(\phi), \phi \in \mathcal{X}. \tag{3.86}$$

from (3.84). Using the definition $W_0(\phi)$ in (3.81), we have

$$J_0(\phi + \nu(\beta_0 - \beta_1) + \lambda(\beta_0 - \beta_2)) - J_0(\phi + \nu(\beta_0 - \beta_1))$$

$$= J_0(\phi + \lambda(\beta_0 - \beta_2)) - J_0(\phi) \tag{3.87}$$

for all $\phi \in \mathcal{X}$. Since ν and λ are arbitrary and $\beta_i(\boldsymbol{x}) \neq \beta_j(\boldsymbol{x})$ for all $\boldsymbol{x}$ with $i \neq j$ and infinitely differentiable, it follows that

$$J_0(\phi + \nu + \lambda) - J_0(\phi + \nu) = J_0(\phi + \lambda) - J_0(\phi) \tag{3.88}$$

for all ϕ, ψ, ν and λ in $\mathcal{X}$ or equivalently

$$J_0(\phi + \nu + \lambda) + J_0(\phi) = J_0(\phi + \nu) + J_0(\phi + \lambda) \tag{3.89}$$

for all ϕ, ψ, ν and λ in $\mathcal{X}$. This can be written also in the form

$$J_0(\phi + \psi) = J_0(\phi) + J_0(\psi) \tag{3.90}$$

by choosing $\phi = 0$ in $\mathcal{X}$. Hence $J_0(\cdot)$ is a linear functional on $\mathcal{X}$. By the properties of characteristic functionals, $J_0(\cdot)$ is a complex-valued continuous linear functional on $\mathcal{X}$ with $J_0(-\phi) = \overline{J_0(\phi)}$. In other words

$$L_0(\phi) = M_0(\phi)e^{r_0(\phi)+im_0(\phi)}, \phi \in \mathcal{X} .$$

This proves that the random fields Φ_0 and Γ_0 differ by m_0 with probability one, that is $\Phi_0 = \Gamma_0 + m_0$ a.s. Similar analysis proves that Φ_i and Γ_i differ by m_i for some generalized functions m_i almost surely. This completes the proof of the theorem. ∎

Remarks 3.6.1 : The results in this section are from Prakasa Rao (1976). The theorem holds if β_i are constants all different from zero and different from each other. The results can be extended to multidimensional generalized random fields. If the two-dimensional generalized random field in Theorem 3.6.1 is infinitely divisible, then it is known that its characteristic functional does not vanish and the conclusion in Theorem 3.6.1 holds. Finally these results are not trivial consequences of earlier results for real-valued random variables since for any fixed $\phi \in \mathcal{X}, \Psi_i(\phi)$ is *not* a linear combination of $\Phi_j(\phi), 0 \leq j \leq 2$, since β_i are *not* necessarily constants. A more general result on characterization of generalized random fields up to Gaussian factors is discussed in Prakasa Rao (1976).

Chapter 4

Identifiability for Some Types of Stochastic Processes

We now consider an extension of the results in Chapter 2 to the framework of stochastic processes. Some of these results can be derived as special cases of results in Chapter 3 but direct derivations are of independent interest.

4.1 Point Processes

It is known that every point process $N(\cdot)$ on $[0, \infty)$ corresponds to a triple $(\Omega, \mathcal{F}, P_N)$ where Ω is the set of all countable sequences of real numbers $\{t_i\}$ without limit points and $\mathcal{F}$ is the σ-algebra generated by cylinder sets and P_N is a probability measure (cf. Harris (1963)). The point process $N(\cdot)$ is said to be degenerate if P_N is concentrated at a single point $(r_1, r_2, \ldots,)$ in Ω. Let $\mathcal{V}$ denote the class of measurable functions ξ such that $0 \leq \xi(t) \leq 1$ for all real t and $\xi(t) = 1$ outside a bounded interval. The ***probability generating functional*** of a point process $N(\cdot)$ is defined by

$$G(\xi) = E\{\exp(\int_0^\infty \log\ \xi(t) dN(t))\}, \xi \in \mathcal{V}. \tag{4.1}$$

(If $\xi(t) \equiv 0$ over some set A in $[0, \infty)$, the exponent is defined to be zero unless $N(A) = 0$ when it is defined to be equal to one). The probability-generating functional of a bivariate point process $(N_1(\cdot), N_2(\cdot))$ is defined by

$$H(\xi_1, \xi_2) = E\{\exp[\int \log\, \xi_1(t)dN_1(t) + \int \log\, \xi_2(t)dN_2(t)]\} \tag{4.2}$$

for $\xi_1 \in \mathcal{V}$ and $\xi_2 \in \mathcal{V}$.

Theorem 4.1.1 : Let N_0, N_1 and N_2 be three independent point processes and define

$$M_1 = N_1 + N_0 \text{ and } M_2 = N_2 + N_0. \tag{4.3}$$

Then the bivariate point process (M_1, M_2) *uniquely* determines the point processes N_0, N_1 and N_2.

Proof : Let $G_i(\xi)$ denote the probability generating functional of $N_i, i = 0, 1, 2,$ and $H(\xi_1, \xi_2)$ denote the probability generating functional of (M_1, M_2). It is easy to see that

$$\begin{aligned} H(\xi_1, \xi_2) &= E\{\exp[\int \log\, \xi_1(t)dM_1(t) + \int \log\, \xi_2(t)dM_2(t)]\} \\ &= E\{\exp[\int \log \xi_1(t)dN_1(t) + \int \log\, \xi_2(t)dN_2(t) \\ &\qquad + \int \log(\xi_1(t)\xi_2(t))dN_0(t)]\} \\ &= G_1(\xi_1)G_2(\xi_2)G_0(\xi_1\xi_2) \end{aligned} \tag{4.4}$$

for $\xi_1 \in \mathcal{V}, \xi_2 \in \mathcal{V}$ since N_0, N_1 and N_2 are independent point processes. Suppose that $R_i, i = 0, 1, 2$ are independent point processes such that the bivariate point process (S_1, S_2) has the same probability structure as (M_1, M_2) where

$$S_1 = R_1 + R_0, \quad S_2 = R_2 + R_0. \tag{4.5}$$

Let $K_i(\xi), i = 0, 1, 2,$ be the probability generating functionals of $R_i, i = 0, 1, 2$ respectively. It is easy to see as before that

$$H(\xi_1, \xi_0) = K_1(\xi_1)K_2(\xi_2)K_0(\xi_1\xi_2), \xi_1, \xi_2 \in \mathcal{V}. \tag{4.6}$$

Let $A_j, 1 \leq j \leq m$ be disjoint Borel sets in $[0, \infty)$ and $G_i(\boldsymbol{z})$ and $K_i(\boldsymbol{z})$ denote the probability generating functionals of $(N_i(A_1), \ldots, N_i(A_m))$ and $(R_i(A_1), \ldots, R_i(A_m))$ respectively. Relations (4.4) and (4.6) imply that

$$G_1(\boldsymbol{z}_1)G_2(\boldsymbol{z}_2)G_0(\boldsymbol{z}_1\boldsymbol{z}_2) = K_1(\boldsymbol{z}_1)K_2(\boldsymbol{z}_2)K_0(\boldsymbol{z}_1\boldsymbol{z}_2) \tag{4.7}$$

for all $\boldsymbol{z} \in [0,1]^m$ where $\boldsymbol{z}_1\boldsymbol{z}_2$ denotes the vector obtained by multiplying $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ componentwise. $G_i(\boldsymbol{z})$ and $K_i(\boldsymbol{z}), 0 \leq i \leq 2$ are nonzero in the set $D = \{0 < z_j \leq 1, 1 \leq j \leq m\}$ where $\boldsymbol{z} = (z_1, \ldots, z_m)$. Let

$$J_i(\boldsymbol{z}) = G_i(\boldsymbol{z})/K_i(\boldsymbol{z}), 0 \leq i \leq 2, \boldsymbol{z} \in D. \tag{4.8}$$

Then $J_i(\boldsymbol{z})$ is nonzero in D and

$$J_1(\boldsymbol{z}_1)J_2(\boldsymbol{z}_2)J_0(\boldsymbol{z}_1\boldsymbol{z}_2) = 1, \boldsymbol{z}_1, \boldsymbol{z}_2 \in D. \tag{4.9}$$

Substituting $\boldsymbol{z}_2 = 1$, it follows that

$$J_1(\boldsymbol{z}_1)J_0(\boldsymbol{z}_1) = 1, \boldsymbol{z}_1 \in D \ . \tag{4.10}$$

Similarly, we have

$$J_2(\boldsymbol{z}_2)J_0(\boldsymbol{z}_2) = 1, \boldsymbol{z}_2 \in D \ . \tag{4.11}$$

Hence

$$J_0(\boldsymbol{z}_1)J_0(\boldsymbol{z}_2) = J_0(\boldsymbol{z}_1\boldsymbol{z}_2), \boldsymbol{z}_1, \boldsymbol{z}_2 \in D \tag{4.12}$$

and J_0 is continuous on D. The only continuous solutions of this functional equation are functions of the type

$$\Pi_{j=1}^m z_j^{c_j} \tag{4.13}$$

where $c_j, 1 \leq j \leq m$, are constants by results in Aczel (1966, p. 215). Hence

$$G_0(\boldsymbol{z}) = K_0(\boldsymbol{z})\Pi_{j=1}^m z_j^{c_j}, \boldsymbol{z} \in D. \tag{4.14}$$

In other words

$$G_0(\xi) = K_0(\xi)J_0(\xi) \tag{4.15}$$

for $\xi \in \mathcal{V}$ of the form

$$\xi(t) = 1 - \sum_{j=1}^{m}(1 - z_j)\chi_{A_j}(t), 0 < z_j \leq 1, 1 \leq j \leq m \tag{4.16}$$

where

$$J_0(\xi) = \Pi_{j=1}^{m} z_j^{c_j}. \tag{4.17}$$

Here χ_A is the indicator function of the set A and $A_1, ..., A_m$ are disjoint bounded Borel subsets of the real line. Every $\xi \in \mathcal{V}$ can be uniformly approximated by an increasing sequence of simple functions of the above type. Define

$$J_0(\xi) = \lim_n J_0(\xi_n) \tag{4.18}$$

for any $\xi \in \mathcal{V}$ where $\{\xi_n\}$ is an approximating sequence in $\mathcal{V}$ for ξ of the type (4.16). Therefore

$$G_0(\xi) = K_0(\xi)J_0(\xi), \quad \xi \in \mathcal{V} \tag{4.19}$$

and $J_0(\xi)$ is the probability generating functional of a degenerate point process (Westcott (1972)). But the probability-generating functional *uniquely* determines the point process, by a result of Vere-Jones (1968) (cf. Daley and Vere-Jones (1988, p. 221)). Hence N_0 and R_0 differ by a degenerate point process. A similar argument shows that N_1, R_1 and N_2, R_2 differ by a degenerate point processes. But the structure of the bivariate point process shows that we cannot add a degenerate point process to one without subtracting from the other. Hence N_0, N_1 and N_2 are unique to the process (M_1, M_2). ■

The results in this section are due to Prakasa Rao (1975b).

4.2 Homogeneous Markov Chains

Suppose that

$$\theta_{hj}^{(k)} : 1 \leq h, j \leq p, 1 \leq k \leq n \tag{4.20}$$

is a collection of independent real-valued random variables. Let $\{\eta_j : j \geq 0\}$ be a homogeneous Markov chain with state space $\{1, \ldots, p\}$ and with a nonsingular transition matrix $A = ((a_{hj}))$. We denote this Markov chain by $\{A\}$.

A collection of random variabes $\{\xi_k, 1 \leq k \leq n\}$ is said to be *defined on the homogeneous Markov chain* $\{A\}$ if

$$\xi_k = \theta^{(k)}_{\eta_{k-1}\eta_k}, 1 \leq k \leq n, \tag{4.21}$$

that is,

$$\xi_k = \theta^{(k)}_{hj} \text{ if } \eta_{k-1} = h, \eta_k = j, 1 \leq k \leq n. \tag{4.22}$$

Let

$$a^{(k)}_{hj}(x) = P[\xi_k \leq x, \eta_k = j | \eta_{k-1} = h], \tag{4.23}$$

$$A_k(x) = ((a^{(k)}_{hj}(x))), \tag{4.24}$$

and

$$\phi_k(t) = \int_{-\infty}^{\infty} e^{itx} dA_k(x), 1 \leq k \leq n, t \in R \ . \tag{4.25}$$

Observe that $\phi_k(0) = A$ and $\phi_k(t)$ is continuous in $t \in R$.

$A_k(x)$ is called the *matrix-valued distribution function* of ξ_k and $\phi_k(t)$ is called the *matrix-valued characteristic functional* of ξ_k defined on the homogeneous Markov chain $\{A\}$. It is easy to see that

$$a^{(k)}_{hj}(x) = a_{hj} F^{(k)}_{hj}(x) \tag{4.26}$$

where $F^{(k)}_{hj}(x)$ is the distribution function of $\theta^{(k)}_{hj}$. Further the matrix-valued characteristic functional of the linear form

$$a_1\xi_1 + a_2\xi_2 + a_3\xi_3 \tag{4.27}$$

is

$$\phi_1(a_1 t)\phi_2(a_2 t)\phi_3(a_3 t) \tag{4.28}$$

(cf. Gyires (1981a,b)).

Given a nonsingular matrix M, there always exists a matrix L such that

$$M = \sum_{\nu=0}^{\infty} \frac{L^{\nu}}{\nu!} \tag{4.29}$$

(Hille (1948, p. 125)). The matrix L is called the *logarithm of the matrix* M and is denoted by $\log M$. Since A is nonsingular, it can be seen that the matrix-valued characteristic functional ϕ_k of ξ_k given by (4.25) is nonsinguar in a neighbourhood of zero and $\Phi_k(t) = \log \ \phi_k(t)$ exists in this neighbourhood. We choose that continuous version of $\log \phi_k(t)$ for which $\Phi_k(0) = \log A$. Note that, if two nonsingular matrices M and N commute, then

$$\log MN = \log M + \log N. \tag{4.30}$$

For any $1 \leq i_1 < i_2 < \cdots < i_j \leq n$, the matrix-valued characteristic functional of

$$Z = a_1 \xi_{i_1} + \cdots + a_j \xi_{i_j} \tag{4.31}$$

is

$$A^{i_1-1} \phi_{i_1}(a_1 t) A^{i_2-i_1-1} \phi_{i_2}(a_2 t) \cdots A^{i_j-i_{j-1}-1} \phi_{i_j}(a_j t) A^{n-i_j} \tag{4.32}$$

(cf. Gyires (1981a,b)). In particular, if $\phi_{i_r}(t), 1 \leq r \leq j$, commute with A for every t, then the matrix-valued characteristic functional of Z can be written in the form

$$A^{n-j} \phi_{i_1}(a_1 t) \cdots \phi_{i_j}(a_j t). \tag{4.33}$$

We now have the following analogue of Theorem 2.1.1 for random variables defined on a homogeneous Markov chain.

Theorem 4.2.1 : Let ξ_1, ξ_2, ξ_3 be random variables defined on a homogeneous Markov chain $\{A\}$. Define

$$Z_1 = \xi_1 - \xi_2, Z_2 = \xi_2 - \xi_3. \tag{4.34}$$

If the matrix-valued characteristic functional of (Z_1, Z_2) is nonsingular, then the matrix-valued distribution function of (Z_1, Z_2) determines the matrix-valued distribution functions of ξ_1, ξ_2, ξ_3 up to change in location.

Proof : For any real t and u,

$E[\exp\{it(\xi_1-\xi_2)+iu(\xi_2-\xi_3)\}\chi([\eta_3=j])|\eta_0=h]$

$$= E[\exp\{it\xi_1+i(-t+u)\xi_2-iu\xi_3\}\chi([\eta_3=j])|\eta_0=h] \tag{4.35}$$

where $\chi(A)$ denotes the indicator function of the set A. Hence the matrix-valued characteristic functional of (Z_1, Z_2) is

$$\phi_1(t)\phi_2(u-t)\phi_3(-u) \tag{4.36}$$

from (4.28). Suppose that $\{\gamma_1, \gamma_2, \gamma_3\}$ is another set of random variables defined on the homogeneous Markov chain $\{A\}$ such that the matrix-valued characteristic functional of $(\gamma_1-\gamma_2, \gamma_2-\gamma_3)$ is the same as that of $(\xi_1-\xi_2, \xi_2-\xi_3)$. Let $\psi_i, 1 \le i \le 3$, be the matrix-valued characteristic functionals of $\gamma_i, 1 \le i \le 3$, respectively. It is obvious that

$$\phi_1(t)\phi_2(u-t)\phi_3(-u) = \psi_1(t)\psi_2(u-t)\psi_3(-u) \tag{4.37}$$

for all t, u real. Observe that the ϕ_i's and ψ_i's are nonsingular matrices for all t and u since the joint matrix-valued characteristic functional of $(\xi_1-\xi_2, \xi_2-\xi_3)$ is nonsingular by hypothesis. Substituting $t=0$ in (4.37) we have

$$A\phi_2(u)\phi_3(-u) = A\psi_2(u)\psi_3(-u), -\infty < u < \infty \tag{4.38}$$

or equivalently

$$\psi_2^{-1}(u)\phi_2(u) = \psi_3(-u)\phi_3^{-1}(-u), -\infty < u < \infty \tag{4.39}$$

since A is nonsinguar by hypothesis. Similarly, substituting $u=0$ in (4.37), we have

$$\phi_1(t)\phi_2(-t)A = \psi_1(t)\psi_2(-t)A, -\infty < t < \infty \tag{4.40}$$

or equivalently

$$\psi_1^{-1}(t)\phi_1(t) = \psi_2(-t)\phi_2^{-1}(-t), -\infty < t < \infty. \tag{4.41}$$

But

$$\psi_1^{-1}(t)\phi_1(t)\phi_2(u-t)\phi_3(-u)\psi_3^{-1}(-u) = \psi_2(u-t), -\infty < u, t < \infty \tag{4.42}$$

from (4.37). Using the relations (4.39) and (4.41), it follows that

$$\psi_2(-t)\phi_2^{-1}(-t)\phi_2(u-t)\phi_2^{-1}(u)\psi_2(u) = \psi_2(u-t), -\infty < u,t < \infty. \quad (4.43)$$

Therefore

$$\psi_2(-t)\phi_2^{-1}(-t)\phi_2(u-t) = \psi_2(u-t)\psi_2^{-1}(u)\phi_2(u), -\infty < u,t < \infty. \quad (4.44)$$

Let $\zeta_2 = \psi_2\phi_2^{-1}$. It follows from (4.44) that

$$\zeta_2(-t)\phi_2(u-t) = \psi_2(u-t)\psi_2^{-1}(u)\phi_2(u), -\infty < u,t < \infty \ . \quad (4.45)$$

Substituting $t = u$ in (4.45), we have

$$\zeta_2(-u)A = A\psi_2^{-1}(u)\phi_2(u), -\infty < u < \infty. \quad (4.46)$$

Hence, from (4.45) again, it follows that

$$\zeta_2(-t)\phi_2(u-t) = \psi_2(u-t)A^{-1}\zeta_2(-u)A, -\infty < u,t < \infty \quad (4.47)$$

or equivalently

$$A^{-1}\zeta_2^{-1}(-u)A\zeta_2(-t) = \zeta_2(u-t), -\infty < u,t < \infty. \quad (4.48)$$

The last equation can be written in the form

$$A\zeta_2(-t) = \zeta_2(-u)A\zeta_2(u-t), -\infty < u,t < \infty. \quad (4.49)$$

Hence

$$A\zeta_2(x+y) = \zeta_2(x)A\zeta_2(y), -\infty < x,y < \infty. \quad (4.50)$$

Let $y = 0$ in (4.50). Then it follows that

$$A\zeta_2(x) = \zeta_2(x)A, -\infty < x < \infty. \quad (4.51)$$

Hence A commutes with $\zeta_2(x)$ for all x and we have

$$A\zeta_2(x+y) = A\zeta_2(x)\zeta_2(y), -\infty < x,y < \infty. \quad (4.52)$$

Since A is nonsingular, relation (4.52) implies that

$$\zeta_2(x+y) = \zeta_2(x)\zeta_2(y), -\infty < x,y < \infty \ . \quad (4.53)$$

Note that ζ_2 is continuous with $\zeta_2(0) = I$. It follows from results in Hille and Phillips (1957, Theorem 9.6.1, p. 287) that there exists a matrix D_2 such that

$$\zeta_2(x) = e^{xD_2}, -\infty < x < \infty \tag{4.54}$$

and hence

$$\psi_2(u)e^{-uD_2} = \phi_2(u), ; -\infty < u < \infty\,. \tag{4.55}$$

Similar relations hold for ψ_1, ϕ_1 and ψ_3, ϕ_3. By the uniqueness theorem for characteristic functionals (cf. Gyires 1981a,b)), the above relation implies that the matrix-valued distribution functions of ξ_1, ξ_2, ξ_3 are determined up to changes in location. This completes the proof of Theorem 4.2.1. ■

We now extend Theorem 4.2.1 to more general linear functions of random variabes defined on a homogeneous Markov chain.

Theorem 4.2.2 : Let $\{\xi_k, 1 \le k \le n\}$ be random variables defined on a homogeneous Markov chain $\{A\}$. Suppose $1 \le i_1 < i_2 < i_3 \le n$. Define

$$Z_1 = a_1\xi_{i_1} + a_2\xi_{i_2} + a_3\xi_{i_3}\ ,$$

$$Z_2 = b_1\xi_{i_1} + b_2\xi_{i_2} + b_3\xi_{i_3}\ \ . \tag{4.56}$$

Further suppose that the matrix-valued characteristic functionals $\phi_{i_j}(t), 1 \le j \le 3$ of $\xi_{i_j}, 1 \le j \le 3$, commute with each other and with A. Let $\{\zeta_k, 1 \le k \le n\}$ be another set of random variables defined on the homogeneous Markov chain $\{A\}$ such that the matrix-valued characteristic fucntionals $\psi_{i_j}(t), 1 \le j \le 3$, of $\zeta_{i_j}, 1 \le j \le 3$, commute with each other and with A. Define

$$W_1 = a_1\zeta_{i_1} + a_2\zeta_{i_2} + a_3\zeta_{i_3},$$

$$W_2 = b_1\zeta_{i_1} + b_2\zeta_{i_2} + b_3\zeta_{i_3}\ \ . \tag{4.57}$$

Assume that the joint matrix-valued characteristic functional of (Z_1, Z_2) is the same as that of (W_1, W_2) and is nonsingular. Suppose that $a_i : b_i \neq$

$a_j : b_j$ for $i \neq j, 1 \leq i, j \leq 3$. Then the matrix-valued distribution functions of $\xi_{i_j}, 1 \leq j \leq 3$ are determined up to change of location.

Remarks 4.2.1 : The proof of Theorem 4.2.2 depends on extensions of Lemmas 2.1.1 to 2.1.3 and Corollary 2.1.1 to matrix-valued functions. We omit the proofs. For details, see Prakasa Rao (1987).

4.3 Homogeneous Processes with Independent Increments

Suppose $\{X(t), t \geq 0\}$ is a homogenous stochastic process with independent increments in the sense that the distribution for $X(t_2) - X(t_1)$ for $0 \leq t_1 < t_2 < \infty$ depends on $t_2 - t_1$ and, for $0 \leq t_1 < t_2 < t_3 < \infty, X(t_3) - X(t_2)$ is independent of $X(t_2) - X(t_1)$. Further suppose that the process $\{X(t), t \geq 0\}$ is continuous in the sense that it has no fixed points of discontinuity.

Let $\phi(u; h)$ denote the characteristic function of $X(t+h) - X(t)$ for $h > 0$ and $0 < t < \infty$. It is well known that $\phi(u; h)$ is infinitely divisible and $\phi(u; h) = [\phi(u; 1)]^h$ for all $h > 0$. For simplicity, let $\phi(\cdot)$ denote the function $\phi(\cdot; 1)$. The process $\{X(t), t \geq 0\}$ is uniquely determined by the characteristic function of $X(0)$ and by the function $\phi(\cdot)$. Hereafter *we assume that* $X(0) = 0$.

4.3.1 Stochastic integrals: Let $g(\cdot)$ be a real-valued function defined over an interval $[A, B] \subset [0, \infty)$ and let $w(\cdot)$ be a nonnegative function defined over $[A, B]$. Consider a sequence of subdivisions

$$D_n : A = t_{n,0} < t_{n,1} < \cdots < t_{n,k_n} = B \tag{4.58}$$

of the interval $[A, B]$ such that

$$\lim_{n \to \infty} \max_{1 \leq k \leq k_n} (t_{n,k} - t_{n,k-1}) = 0 \ . \tag{4.59}$$

Let $t^*_{n,k} \in [t_{n,k-1}, t_{n,k}], 1 \leq k \leq k_n$ for all $n \geq 1$. Construct the sequence of partial sums

$$S_n = \sum_{k=1}^{k_n} g(t^*_{n,k})[X(w(t_{n,k})) - X(w(t_{n,k-1}))] \ . \tag{4.60}$$

If the sequence $\{S_n\}$ converges in probability to a random variable S and if this limit does not depend on the choice of $t^*_{n,k}$ and the sequence of subdivisions $\{D_n\}$ satisfying (4.59), then we say that the *stochastic integral* S exists in probability and write

$$S = \int_A^B g(t)dX(w(t)). \tag{4.61}$$

If the sequence $\{S_n\}$ converges in quadratic mean to a random variable S, then we say that the stochastic integral S given by (4.61) exists in quadratic mean.

The following results are known about the existence of such stochastic integrals. We omit the proofs.

Theorem 4.3.1 : Let $\{X(t), t \geq 0\}$ be a continuous homogeneous process with independent increments. Suppose the process $\{X(t), t \geq 0\}$ has finite mean function and finite covariance function both of bounded variation on a finite closed interval $[A, B]$. Further suppose that $g(t)$ is real-valued and continuous on $[A, B]$. Then the stochastic integral

$$\int_A^B g(t)dX(t) \tag{4.62}$$

exists in quadratic mean.

Theorem 4.3.2 : Suppose $\{X(t), t \geq 0\}$ is a continuous homogeneous process with independent increments and $g(t)$ is real-valued and continuous on $[A, B]$. Then the stochastic integral

$$\int_A^B g(t)dX(t) \tag{4.63}$$

exists in probability.

For proofs of Theorems 4.3.1 and 4.3.2, see Lukacs (1968). Suppose $w(\cdot)$

is a nonnegative, nondecreasing and right-continuous function. Define

$$V\{(-\infty,t]\} = \begin{cases} 0 & \text{if } t \leq A \\ w(t)-w(A) & \text{if } A < t \leq B \\ w(B)-w(A) & \text{if } t > B \end{cases} \tag{4.64}$$

where $-\infty < A < B < \infty$. Then V gives rise to a finite measure with support contained in $[A.B]$. Denote this measure also by V. Suppose $g(\cdot)$ is continuous on $[A,B]$. Define

$$w_g(t) = V([x : g(x) \leq t]) \; . \tag{4.65}$$

Then $w_g(\cdot)$ is nondecreasing, nonnegative and right-continuous on $[A,B]$. The following result is due to Riedel (1980a).

Theorem 4.3.3 : Suppose $\{X(t), t \geq 0\}$ is a continuous homogeneous process with independent increments and $g(\cdot)$ is a continuous real-valued function on $[A,B]$. Suppose $w(\cdot)$ is a nondecreasing, nonnegative right-continuous function on $[A,B]$. Define

$$C = \min_{A\leq t\leq B} g(t) \text{ and } D = \max_{A\leq t\leq B} g(t). \tag{4.66}$$

Then the integrals

$$\int_A^B g(t)dX(w(t)) \text{ and } \int_C^D t\, dX(w_g(t)) \tag{4.67}$$

exist in probability and are identically distributed.

The next result gives a representation for the characteristic functions of the stochastic integrals defined above.

Theorem 4.3.4 : Let $\{X(t), t \geq 0\}$ be a continuous homogeneous process with independent increments. Further suppose that the process has finite mean function and finite covariance function which are of bounded variation on any finite closed interval $[A,B]$. Let $g(\cdot)$ and $h(\cdot)$ be continuous in $[A,B]$. Define

$$Y = \int_A^B g(t)dX(t) \text{ and } Z = \int_A^B h(t)dX(t) \tag{4.68}$$

and denote by $\phi(u;h)$ and $\theta(u,v)$ the characteristic functions of $X(t+h)-X(t)$ and (Y,Z) respectively. Then $\theta(u,v)$ is different from zero for all u and v and

$$\log\ \theta(u,v) = \int_A^B \psi[ug(t)+vh(t)]dt \tag{4.69}$$

where $\psi(u) = \log\ \phi(u,1)$ and the logarithm taken here is the continuous branch of the logarithm of $\phi(\cdot;1)$ with $\log\ \phi(0;1) = 0$.

Remarks 4.3.1 : For a proof of Theorem 4.3.4, see Lukacs (1968, pp. 107–108). This theorem continues to hold if the integrals Y and Z exist in probability.

Since $\{X(t), t \geq 0\}, X(0) = 0$ is a homogeneous process with independent increments, the characteristic function $\phi(u) \equiv \phi(u;1)$ of $X(t+1)-X(t)$ is infinitely divisible and the Lévy canonical representation for the characteristic function of $X(1)$ holds as given in Lukacs (1970, Theorem 5.5.2). Riedel (1980a) proved the following theorem. We omit the proof.

Theorem 4.3.5 : Let $\{X(t), t \geq o\}, X(0) = 0$ be a continuous homogeneous process with independent increments. Suppose $w(\cdot)$ is nondecreasing nonnegative and right-continuous on $[0,\infty)$. Let the Lévy canonical representation for the characteristic function of $X(1)$ be given by a, σ, M and N. Then the Lévy–Khintchin canonical representation for the characteristic function of the stochastic integral

$$\int_A^B t dX(w(t)) \tag{4.70}$$

is given by the formulae

$$a_w = \int_A^B (ta + t(1-t^2)\int_{0+}^{\infty} \frac{x^3}{(1+(tx)^2)(1+x^2)}\, d(M(-x)+N(x)))dw(t), \tag{4.71}$$

$$\sigma_w^2 = \sigma^2 \int_A^B t^2 dw(t), \tag{4.72}$$

$$M_w(x) = -\int_{\min(A,0)}^{\min(B,0)} N(\frac{x}{t})dw(t) + \int_{\max(A,0)}^{\max(B,0)} M(\frac{x}{t})dw(t), x < 0 \quad (4.73)$$

and

$$N_w(x) = -\int_{\min(A,0)}^{\min(B,0)} M(\frac{x}{t})dw(t) + \int_{\max(A,0)}^{\max(B,0)} N(\frac{x}{t})dw(t), x > 0. \quad (4.74)$$

4.3.2 Identifiability : We say that the stochastic integral given by (4.61) *determines* the homogeneous process with independent increments $\{X(t), t \geq 0\}$ if the characteristic function of S determines the characteristic function of $X(1)$.

We first present a couple of results identifying such a stochastic process up to shift via stochastic integrals.

Theorem 4.3.6 : Let $\{X(t), t \geq 0\}$ be a continuous homogeneous process with independent increments. Suppose the process has moments of all orders and its mean function and covariance function are of bounded variation in any finite closed interval. Suppose $g(t)$ and $h(t)$ are continuous functions on $[A, B]$ and $[C, D]$ respectively such that $A < C < B < D$. Further suppose that either

$$\int_A^B [g(t)]^k dt \neq 0, k \geq 2 \quad (4.75)$$

or

$$\int_C^D [h(t)]^k dt \neq 0, k \geq 2. \quad (4.76)$$

Let

$$Y = \int_A^B g(t)dX(t), Z = \int_C^D h(t)dX(t). \quad (4.77)$$

Then the joint distribution of (Y, Z) completely determines the process X except possibly for change of location provided the characteristic function of $X(1)$ is entire. In such an event either

$$\int_A^B g(t)dt = \int_C^D h(t)dt = 0 \quad (4.78)$$

or there is no change in location.

Proof : Let $\theta(u,v)$ denote the characteristic function of (Y,Z) and $\psi(u)$ denote the continuous branch of the logarithm of the characteristic function of $X(1)$ with $\psi(0)=0$. It is easy to check that

$$\log\,\theta(u,v) = \int_A^C \psi(ug(t))dt + \int_C^B \psi(ug(t)+vh(t))dt + \int_B^D \psi(vh(t))dt. \quad (4.79)$$

Suppose that $\{W(t), t\geq 0\}$ is another stochastic process with the same properties as $\{X(t), t\geq 0\}$. Let $\eta(u)$ denote the continuous branch of the logarithm of the characteristic function of $W(1)$ with $\eta(0)=0$. Suppose that the random vector (S,R) has the same joint distribution as (Y,Z) where

$$S = \int_A^B g(t)dW(t) \text{ and } R = \int_C^D h(t)dW(t). \quad (4.80)$$

It follows from (4.79) that

$$\int_A^C \psi(ug(t))dt + \int_C^B \psi(ug(t)+vh(t))dt + \int_B^D \psi(vh(t))dt$$

$$= \int_A^C \eta(ug(t))dt + \int_C^B \eta(ug(t)+vh(t))dt + \int_B^D \eta(vh(t))dt \quad (4.81)$$

for all u, v real. Suppose that (4.75) holds. Let $v=0$ in (4.81). Then

$$\int_A^B \psi(ug(t))dt = \int_A^B \eta(ug(t))dt, -\infty < u < \infty. \quad (4.82)$$

Since the processes X and W have moments of all orders, the integrals on both sides can be differentiated with respect to u repeatedly under the integral sign and we have

$$\int_A^B [g(t)]^k \psi^{(k)}(ug(t))dt = \int_A^B [g(t)]^k \eta^{(k)}(ug(t))dt \quad (4.83)$$

where $\psi^{(k)}(\cdot)$ denotes the kth derivative of ψ. Let $u=0$ in (4.83). Then it follows that

$$\psi^{(k)}(0)\int_A^B [g(t)]^k dt = \eta^{(k)}(0)\int_A^B [g(t)]^k dt \quad (4.84)$$

which proves that $\psi^{(k)}(0)=\eta^{(k)}(0)$ for $k\geq 2$ in view of (4.75). Since ψ and η are entire functions with $\psi(0)=\eta(0)=0, \psi(t)=\overline{\psi(-t)}$ and $\eta(t)=\overline{\eta(-t)}$, it follows that

$$\psi(u) = \eta(u) + icu, -\infty < u < \infty \tag{4.85}$$

for some real constant c. This proves that $X(1)$ and $W(1) + c$ have the same distribution. From the fact that $\{X(t), t \geq 0\}$ and $\{W(t), t \geq 0\}$ are homogeneous processes with independent increments, it can be seen that

$$X(t+h) - X(t) \text{ and } W(t+h) - W(t) + ch$$

are identically distributed for all $t \geq 0$ and $h \geq 0$. If $c = 0$, then the processes $\{X(t), t \geq 0\}$ and $\{W(t), t \geq 0\}$ are the same. If $c \neq 0$, then it is easy to check that

$$\int_A^B g(t)dt = 0 = \int_C^D h(t)dt. \tag{4.86}$$

A similar argument proves the result in case (4.76) holds. This completes the proof. ■

As a special case of Theorem 4.3.6, we have the following result by choosing $h(t) = 0$ for all t.

Theorem 4.3.7: Suppose a process $\{X(t), t \geq 0\}$ satisfies the conditions stated in the above theorem. Suppose $g(t)$ is real-valued and continuous on $[A, B]$ and

$$\int_A^B [g(t)]^k dt \neq 0, k \geq 2. \tag{4.87}$$

Let

$$Y = \int_A^B g(t)dX(t). \tag{4.88}$$

Then the distribution of Y completely determines the process $\{X(t), t \geq 0\}$ except for a change of location, provided the characteristic function of $X(1)$ is entire. In such an event either there is no change of location or

$$\int_A^B g(t)dt = 0\ . \tag{4.89}$$

Remarks 4.3.2 : The results obtained above are due to Prakasa Rao (1975c). The conditions that the process $\{X(t), t \geq 0\}$ has moments of all orders and the characteristic function of $X(1)$ is entire are too strong. Riedel (1980b) has weakened these conditions and derived results determining the stochastic processes $\{X(t), t \geq 0\}$ of the above type by means of stochastic integerals. His analysis involves some results on Wiener–Hopf factorization and a modern extension of the Phragmén–Lindelof theory (cf. Rossberg (1975)). We will state the results without proofs.

Let $g(t)$ be real-valued and continuous on $[A, B]$ and $w(t)$ be a nonnegative, nondecreasing and right-continuous function on $[A, B]$. For $\mathrm{Re}(z) \geq 0$, define

$$S(z) = \int_A^B |g(t)|^z dw(t), \tag{4.90}$$

and

$$\hat{S}(z) = \int_A^B |g(t)|^{z-1} g(t) dw(t). \tag{4.91}$$

Theorem 4.3.8 : Suppose $\{X(t), t \geq 0\}$ is a continuous homogeneous process with independent increments and $E|X(1)|^\lambda < \infty$ for some $0 < \lambda < 2$. Then the stochastic integral

$$Y = \int_A^B g(t) dX(w(t)) \tag{4.92}$$

defined in the sense of convergence in probability determines the process $\{X(t), t \geq 0\}$ iff the following conditions are satisfied:

$$\begin{array}{lll} \text{(i)} & S(z) \neq 0, & \lambda \leq \mathrm{Re}(z) < 2, \\ \text{(ii)} & \hat{S}(z) \neq 0, & \lambda \leq \mathrm{Re}(z) < 2 \text{ and} \\ \text{(iii)} & \hat{S}(1) \neq 0. & \end{array} \tag{4.93}$$

Theorem 4.3.9 : Suppose $\{X(t), t \geq 0\}$ is a continuous homogeneous process with independent increments and $E|X(1)|^2 < \infty$. Then the stochastic integral Y defined by (4.94) in probability determines the process $\{X(t), t \geq 0\}$ iff

$$\hat{S}(1) = \int_A^B g(t) dw(t) \neq 0. \tag{4.94}$$

Remarks 4.3.3 : For proofs of Theorems 4.3.8 and 4.3.9 and related results, see Riedel (1980b). These results make use of Theorem 4.3.5 on the representation of the characteristic function of a stochastic integral (cf. Riedel (1980a)). For a comprehensive survey on the identification of stochastic processes by stochastic integrals, see Prakasa Rao (1983a).

4.4 Linear Processes

Let $\{X(t), -\infty < t < \infty\}$ be a homogeneous process with independent increments and f be a function such that $|f|$ and f^2 are integrable. It is known that the stochastic integral

$$\Lambda_f(t) = \int_{-\infty}^{\infty} f(t-u)dX(u), -\infty < t < \infty \tag{4.95}$$

exists in the sense of quadratic mean (cf. Doob (1953)) if

$$E(X(t))^2 < \infty, -\infty < t < \infty. \tag{4.96}$$

$\{\Lambda_f(t), -\infty < t < \infty\}$ is called a *linear process.* The process $\{\Lambda_f(t), -\infty < t < \infty\}$ is a stationary process. Since $\{X(t), -\infty < t < \infty\}$ is a homogeneous process with independent increments, it is known that

$$E(\exp\{i\theta[X(t+u) - X(u)]\}) = \exp\{t\psi(\theta)\} \tag{4.97}$$

where

$$\psi(\theta) = i\gamma\theta - \frac{1}{2}\delta^2\theta^2 + \int_{-\infty}^{\infty} \frac{e^{i\theta x} - 1 - i\theta x}{x^2} K(dx), \tag{4.98}$$

γ and δ are real constants and $K(\cdot)$ is a nondecreasing bounded function with $K(-\infty) = 0, K(0+) - K(0-) = 0$ (cf. Lukacs (1968)).

The *characteristic functional* of such a stochastic process $\{\Lambda_f(t), -\infty < t < \infty\}$ is defined by

$$\phi_{\Lambda_f}(\xi) = E\{\exp[i\int_{-\infty}^{\infty} \Lambda_f(t)\xi(dt)]\} \tag{4.99}$$

where $\xi(\cdot)$ runs through real-valued signed totally finite measures on the σ-algebra of Borel subsets of the real line (cf. Bartlett (1966)). From the fact that $\{\Lambda_f(t), -\infty < t < \infty\}$ is a linear process, it can be shown that

$$\phi_{\Lambda_f}(\xi) = \exp\{\int_{-\infty}^{\infty} \psi(\int_{-\infty}^{\infty} f(t-u)\xi(dt))du\}. \tag{4.100}$$

Let

$$C(f,\theta) = \log\ E\{\exp\ i\theta\Lambda_f(t)\}, -\infty < \theta < \infty. \tag{4.101}$$

Note that $\Lambda_f(t)$ is an infinitely divisible random variable and hence $C(f,\theta)$ is well defined. It can be seen from (4.99) or directly that

$$C(f,\theta) = \int_{-\infty}^{\infty} \psi(\theta f(u))du, -\infty < \theta < \infty\ , \tag{4.102}$$

from the definition of the linear process $\{\Lambda_f(t), -\infty < t < \infty\}$. Making use of the canonical representation (4.98) for $\psi(\theta)$, it can be shown that (cf. Weiss and Westcott (1976))

$$C(f,\theta) = i\hat{\gamma}_f\theta - \frac{1}{2}\hat{\delta}_f^2\theta^2 + \int_{-\infty}^{\infty} \frac{e^{i\theta x} - 1 - i\theta x}{x^2}\ \hat{K}_f(dx) \tag{4.103}$$

where

$$\hat{\gamma}_f = \gamma \int_{-\infty}^{\infty} f(t)dt, \hat{\delta}_f^2 = \delta^2 \int_{-\infty}^{\infty} f^2(t)dt, \tag{4.104}$$

and $\hat{K}_f$ is a nondecreasing bounded function with $\hat{K}_f(-\infty) = 0$, $\hat{K}_f(0+) - \hat{K}_f(0-) = 0$ defined by

$$\hat{K}_f(dv) = \int_0^{b+} z^2 K(\frac{dv}{z})|dh^+(z)| + \int_0^{b-} z^2 K(-\frac{dv}{z})|dh^-(z)|\ . \tag{4.105}$$

Here $b_\pm = \sup_u\ f^\pm(u) \leq \infty, h^\pm(y) = \lambda\{x : f^\pm(x) \geq y\}$ where λ is the Lebesgue measure.

Let B_2 denote the class of all real-valued functions f such that $|f|$ and f^2 are integrable. The following results are due to Weiss and Westcott (1976). We omit the proofs.

Theorem 4.4.1 : ψ is uniquely determined given $\Lambda_f(\cdot)$ and f for all $f \in B_2$.

Suppose a process $\{\Lambda(t), -\infty < t < \infty\}$ can be expressed as a linear process in two different ways :

$$\Lambda(t) = \int_{-\infty}^{\infty} f_i(t-u)dX_i(u), -\infty < t < \infty, i = 1,2 \tag{4.106}$$

where $f_i \in B_2, i = 1,2$ and $\{X_i(t), -\infty < t < \infty\}, i = 1,2$ are homogeneous processes with independent increments. In general, two different representations (4.106) for the same linear process $\{\Lambda(t), -\infty < t < \infty\}$ are possible; for instance, $f_2 = cf_1$ and $X_2(t) = \frac{1}{c}X_1(\cdot)$ and $f_2(t) = \pm f_1(t+a), X_2(t) = \pm X_1(t)$ for constants $c \neq 0$ and a. The next theorem states that the representation is *unique* up to a constant factor and up to translations of f. If $f_2(t) = \pm f_1(t+a)$, then $X_2(t) = \pm X_1(t)$ by Theorem 4.4.1.

Theorem 4.4.2 : If a linear process $\{\Lambda(t), -\infty < t < \infty\}$ has two representations with $f_1, f_2 \in B_2$ and $\int_{-\infty}^{\infty} f_i^2(t)dt = 1, i = 1,2$, and if the cases $f_2(t) = \pm f_1(t+a), X_2(t) = \pm X_1(t)$ are excluded, then the processes $X_1(\cdot)$ and $X_2(\cdot)$ are Gaussian.

Theorem 4.4.3 : If two linear processes

$$\Lambda_i(t) = \int_{-\infty}^{\infty} f_i(t-u)dX_i(u), i = 1,2$$

have the same characteristic functional as defined by (4.99), then either $f_2(t) = cf_1(t+a)$ or $X_1(t)$ and $X_2(t)$ are Gaussian.

Remarks 4.4.1 : For the proofs of Theorems 4.4.1 to 4.4.3, see Weiss and Westcott (1976).

Definition : A stationary stochastic process $\{X(t), -\infty < t < \infty\}$ is said to be *time-reversible* if for all n and $t_1, t_2, \ldots, t_n, (X(t_1), \ldots, X(t_n))$ and $(X(-t_1), \ldots, X(-t_n))$ have the same joint distribution.

Remarks 4.4.2 : For example, stationary Gaussian processes are time-reversible. If $\{X(t), -\infty < t < \infty\}$ is a stationary process which is time-reversible, then for every h and every $t_1, t_2, \ldots, t_n$, $(X(h+t_1), \ldots, X(h+t_n))$ and $(X(h-t_1), \ldots, X(h-t_n))$ have the same joint probability distribution.

The following result is an easy consequence of Theorem 4.4.3.

Theorem 4.4.4 : Let $\Lambda(\cdot)$ be a linear process defined by (4.101). Suppose there does not exist a constant a such that $f(t) = f(a-t)$ for all t or $f(t) = -f(a-t)$ for all t and $X(t)$ has a symmetric distribution for all t. If $\Lambda(\cdot)$ is time-reversible, then $X(\cdot)$ is Gaussian.

Proof : Let $\Lambda_1(t) = \Lambda(-t), -\infty < t < \infty$. Then $\Lambda(\cdot)$ and $\Lambda_1(\cdot)$ have the same probability structure due to the time-reversibility of the process $\Lambda(\cdot)$. Let $f_1(t) = f(-t)$ and $X_1(t) = X(-t)$. Then

$$\begin{aligned}\Lambda_1(t) = \Lambda(-t) &= \int_{-\infty}^{\infty} f(-t-u)dX(u) \\ &= \int_{-\infty}^{\infty} f_1(t-u)dX_1(u)\end{aligned}$$

and

$$\Lambda(t) = \int_{-\infty}^{\infty} f(t-u)dX(u).$$

Since $\{\Lambda(t), -\infty < t < \infty\}$ and $\{\Lambda_1(t), -\infty < t < \infty\}$ are two linear processes with the same probability structure and $f_1(t) \neq cf(t+a)$ by hypothesis, it follows that $\{X(t), -\infty < t < \infty\}$ is Gaussian, by Theorem 4.4.2. ■

For detailed proofs, see Westcott (1970), Weiss (1975) and Weiss and Westcott (1976).

Chapter 5

Generalized Convolutions

Some of the identifiability results studied in Chapter 2 have analogues in the theory of Laplace transforms and lead to methods of solving some partial differential equations. We discuss some of these results in this chapter.

5.1 Generalized Convolutions

Let f_1 and f_2 be two real-valued functions such that $f_1(t) = f_2(t) = 0$ for $t < 0$ and f_i not identically zero for $i = 1, 2$. The convolution of two such functions $f_1(t)$ and $f_2(t)$ is defined by the formula

$$(f_1 \star f_2)(t) = \int_0^t f_1(x) f_2(t - x) dt, t \geq 0 \tag{5.1}$$

assuming that this is defined. However, the convolution $f_1 \star f_2$ does not determine the functions f_1 and f_2 uniquely. For instance, let

$$f_1(t) = 1, t \geq 0, f_2(t) = \frac{1}{2} t^2, t \geq 0 \tag{5.2}$$

and

$$g_1(t) = g_2(t) = t, t \geq 0. \tag{5.3}$$

It is easy to check that

$$(f_1 \star f_2)(t) = (g_1 \star g_2)(t), t \geq 0 \tag{5.4}$$

even though the pair (f_1, f_2) and (g_1, g_2) differ. We now define a notion of *generalized convolution of three functions.* If the generalized convolution is known, then the three functions are determined uniquely under some conditions.

Definition : Let $f_k(t), 0 \leq k \leq 2$ be real-valued functions, locally integrable for $t \geq 0$. Further suppose that

$$\int_0^\infty e^{-st}|f_k(t)|dt \tag{5.5}$$

is well defined whenever $\text{Re}(s) > s_k, s_k$ real, for $0 \leq k \leq 2$. Then the *generalized convolution* of $f_k, 0 \leq k \leq 2$ is defined by

$$(f_0, f_1, f_2)(u_1, u_2) = \int_0^{\min(u_1,u_2)} f_0(t)f_1(u_1 - t)f_2(u_2 - t)dt \tag{5.6}$$

for $0 \leq u_1, u_2 < \infty$.

Lemma 5.1.1 : Let $F_k(s)$ be the Laplace transform of $f_k(t)$. Then the two-dimensional Laplace transform $F(s_1, s_2)$ of the generalized convolution (f_0, f_1, f_2) is given by

$$F(s_1, s_2) = F_0(s_1 + s_2)F_1(s_1)F_2(s_2) \tag{5.7}$$

whenever the expression on the right side of the above relation is defined.

Proof : Note that

$$\begin{aligned}
F(s_1, s_2) & \\
&= \int_0^\infty \int_0^\infty e^{-(s_1u_1+s_2u_2)}(f_0, f_1, f_2)(u_1, u_2)du_1du_2 \\
&= \int_0^\infty \int_0^\infty e^{-(s_1u_1+s_2u_2)}\{\int_0^{\min(u_1,u_2)} f_0(t)f_1(u_1 - t)f_2(u_2 - t)dt\}du_1du_2 \\
&= \int_0^\infty f_0(t)\{\int_t^\infty e^{-s_1u_1}f_1(u_1 - t)du_1 \int_t^\infty e^{-s_2u_2}f_2(u_2 - t)du_2\}dt \\
&= \int_0^\infty f_0(t)e^{-s_1t}F_1(s_1)e^{-s_2t}F_2(s_2)dt \\
&= F_0(s_1 + s_2)F_1(s_1)F_2(s_2).
\end{aligned} \tag{5.8}$$

■

Theorem 5.1.1 : Let $f_k(t), 0 \leq k \leq 2$, be real-valued functions defined for $t \geq 0$. Suppose $f_k(t)$ are not equal to zero almost everywhere. Further suppose that the Laplace transform of $|f_k(t)|$ is $F_k(s)$ defined by

$$F_k(s) = \int_0^\infty e^{-st}|f_k(t)|dt, 0 \leq k \leq 2. \tag{5.9}$$

Then the generalized convolution (f_0, f_1, f_2) of $f_k, 0 \leq k \leq 2$, determines the functions $f_k, 0 \leq k \leq 2$, up to a set of Lebesgue measure zero, up to a shift, and up to nonzero constant factors.

Proof : Suppose $g_k, 0 \leq k \leq 2$ is another set of real-valued functions satisfying the conditions stated in the theorem such that

$$(f_0, f_1, f_2)(t) = (g_0, g_1, g_2)(t), t \geq 0. \tag{5.10}$$

Taking the two-dimensional Laplace transform on both sides of the equation (5.10), we have

$$G_0(s_1 + s_2)G_1(s_1)G_2(s_2) = F_0(s_1 + s_2)F_1(s_1)F_2(s_2) \tag{5.11}$$

by Lemma 5.1.1. Let (s_{10}, s_{20}) be a point at which the expression on the right side of (5.11) does not vanish. Such a point exists since $f_k(t)$, $0 \leq k \leq t$ are not equal to zero almost everywhere. From the continuity of the Laplace transforms $F_k(s), 0 \leq k \leq 2$, it follows that there exists some neighbourhood S of (s_{10}, s_{20}) in which the right side of (5.11) does not vanish. Hereafter, let us restrict attention to points $(s_1, s_2) \in S$. Let

$$G_k(s) = F_k(s)b_k H_k(s), 0 \leq k \leq 2 \tag{5.12}$$

where b_k are nonzero complex constants and $H_k(s)$ are complex-valued functions satisfying the conditions

$$H_0(s_{10} + s_{20}) = H_1(s_{10}) = H_2(s_{20}) = 1. \tag{5.13}$$

Relation (5.11) implies that

$$b_0 b_1 b_2 H_0(s_1 + s_2)H_1(s_1)H_2(s_2) = 1 \tag{5.14}$$

for all $(s_1, s_2) \in S$. Equation (5.13) implies that

$$b_0 b_1 b_2 = 1 \tag{5.15}$$

by choosing $s_1 = s_{10}$ and $s_2 = s_{20}$. In particular it follows that $b_k, 0 \leq k \leq 2$ are nonzero and we have

$$H_0(s_1 + s_2) H_1(s_1) H_2(s_2) = 1 \tag{5.16}$$

for all $(s_1, s_2) \in S$. Let $s_1 = s_{10} + w_1$ and $s_2 = s_{20} + w_2$. Define

$$\begin{aligned} h_0(w) &= H_0(s_{10} + s_{20} + w), \\ h_1(w) &= H_1(s_{10} + w) \end{aligned}$$

and

$$h_2(w) = H_2(s_{20} + w). \tag{5.17}$$

Then, it follows that

$$h_0(w_1 + w_2) h_1(w_1) h_2(w_2) = 1 \tag{5.18}$$

for all (w_1, w_2) in a neighbourhood of $(0,0)$. Furthermore

$$h_k(0) = 1, 0 \leq k \leq 2 \tag{5.19}$$

from (5.13). It is now easy to prove that there exists a complex constant c such that

$$h_0(w) = e^{cw}, h_k(w) = e^{-cw}, k = 1, 2 \tag{5.20}$$

in a neighbourhood of 0. Retracing the definition of $h_k, 0 \leq k \leq 2$, it can be checked that

$$\begin{aligned} G_0(s) &= F_0(s) \, \frac{1}{a_1 a_2} e^{cs}, \\ G_1(s) &= F_1(s) \, a_1 e^{-cs}, \\ G_2(s) &= F_2(s) a_2 e^{-cs} \end{aligned} \tag{5.21}$$

for some nonzero complex constants $a_i, i = 1, 2$ and some complex constant c in a neighbourhood of $s_{10} + s_{20}$ for G_0, in a neighbourhood of s_{10} for

G_1, and in a neighbourhood of s_{20} for G_2. From the analyticity of Laplace transforms, it follows that (5.21) holds for all complex s. Again, from the properties of Laplace transforms, it follows that a_1, a_2 and c are real and we have the result. This completes the proof. ■

Remarks 5.1.1 : In analogy with generalized convolution of the functions, we can also define *generalized convolution of three sequences* of real numbers. Suppose

$$\begin{aligned} \boldsymbol{a} &= (a_0, a_1, \ldots), \\ \boldsymbol{b} &= (b_0, b_1, \ldots) \end{aligned}$$

and

$$\boldsymbol{c} = (c_0, c_1, \ldots) \tag{5.22}$$

are three sequences of real numbers. The *generalized convolution* of these sequences is defined by the sequence

$$d_{n,m} = \sum_{k=0}^{\min(n,m)} a_k b_{n-k} c_{m-k}, n, m \geq 0. \tag{5.23}$$

Let $\boldsymbol{d} = (d_{n,m}; n, m \geq 0)$ and

$$\begin{aligned} A(s) &= \sum_{k=0}^{\infty} a_k s^k, \\ B(s) &= \sum_{k=0}^{\infty} b_k s^k, \\ C(s) &= \sum_{k=0}^{\infty} c_k s^k \end{aligned} \tag{5.24}$$

and

$$D(u, v) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} d_{n,m} u^n v^m \tag{5.25}$$

where s, u, v are complex. Then $A(\cdot), B(\cdot), C(\cdot)$ and $D(\cdot)$ are generating functions of the sequences $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and $\boldsymbol{d}$ respectively. If the above series converge in a neighbourhood of the origin, then it is easy to check that

$$D(u, v) = A(uv)B(u)C(v) \tag{5.26}$$

in that neighbourhood of the origin. The following theorem can be proved characterizing the sequences $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ by their generalized convolution $\boldsymbol{d}$. We omit the proof. For details, see Kotlarski (1968a).

Theorem 5.1.2 : Let $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ be sequences of real numbers as defined above. Suppose $a_0 \neq 0, b_0 \neq 0$ and $c_0 \neq 0$ and the three power series given by (5.24) converge in a neighbourhood of the origin. Then the generalized convolution $\boldsymbol{d}$ defined by the sequence $d_{n,m}$ given by (5.23) determines all the three sequences $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ up to nonzero constant factors.

The results of this section are due to Kotlarski (1968a).

5.2 Applications to Solutions of Partial Differential Equations

We now study a special class of partial differential equations which can be solved by methods described in this book.

Let f and g be real-valued functions defined on $(0, \infty)$. Suppose f and g are different from zero almost everywhere and differentiable up to order n. Suppose the derivatives $f^{(i)}$ and $g^{(i)}, 1 \leq i \leq n$, and f and g are all Laplace originals. Consider the differential equation

$$\sum_{k=0}^{n} a_k (D_x + D_y)^k f(x) g(y) = h(x, y), x \geq 0, y \geq 0 \tag{5.27}$$

with the initial conditions

$$f(0) = f^{(i)}(0) = g(0) = g^{(i)}(0) = 0, 1 \leq i \leq n-1 \tag{5.28}$$

where $a_i, 0 \leq k \leq n$ are *unknown* coefficients, the functions f and g are *unknown* but h is *known*. We assume that $a_n \neq 0$. Here $D_x = \frac{\partial}{\partial x}$ and $D_y = \frac{\partial}{\partial y}$.

We are interested in the existence and the uniqueness of the solution of the equation (5.27) and determining the solution explicitly under some conditions.

Let

$$P(s) = a_0 + \sum_{k=1}^{n} a_k s^k, s \in R. \tag{5.29}$$

The function $P(\cdot)$ is the generating function of $\{a_0, a_1, \ldots, a_n\}$. Let F, G, H be the Laplace transforms of f, g, h respectively given by

$$\begin{aligned} F(u) &= \int_0^\infty e^{-ux} f(x)dx, u > u_0 & (5.30)\\ G(v) &= \int_0^\infty e^{-vy} g(y)dy, v > v_0 \end{aligned}$$

and

$$H(u,v) = \int_0^\infty \int_0^\infty e^{-ux-vy} h(x,y)dxdy, u > u_0, v > v_0 \ . \tag{5.31}$$

Lemma 5.2.1 : Suppose the partial differential equation (5.27) has a solution. Then

$$P(u+v)F(u)G(v) = H(u,v), u > u_0, v > v_0 \ . \tag{5.32}$$

Proof : For $u > u_0$ and $v > v_0$,

$$\begin{aligned} H(u,v) &= \int_0^\infty \int_0^\infty e^{-ux-vy} h(x,y)dxdy \\ &= \int_0^\infty \int_0^\infty e^{-ux-vy} [\sum_{k=0}^{n} a_k (D_x + D_y)^k f(x)g(y)]dxdy \\ &= \sum_{k=0}^{n} a_k \int_0^\infty \int_0^\infty e^{-ux-vy} (D_x + D_y)^k f(x)g(y)dxdy \\ &= \sum_{k=0}^{n} a_k \int_0^\infty \int_0^\infty e^{-ux-vy} \sum_{j=0}^{k} \binom{k}{j} f^{(j)}(x) g^{(k-j)}(y)dxdy \\ &= \sum_{k=0}^{n} a_k \sum_{j=0}^{k} \binom{k}{j} \{\int_0^\infty e^{-ux} f^{(j)}(x)dx \int_0^\infty e^{-vy} g^{(k-j)}(y)dy\} \\ &= \sum_{k=0}^{n} a_k \sum_{j=0}^{k} \binom{k}{j} u^j F(u) v^{k-j} G(v) \\ &= \sum_{k=0}^{n} a_k (u+v)^k F(u)G(v) \\ &= P(u+v)F(u)G(v). & (5.33) \end{aligned}$$

■

Theorem 5.2.1 : Suppose $a_k^*, 0 \leq k \leq n$, and the functions f^* and g^* satisfy conditions similar to those on $a_k, 0 \leq k \leq n$, and the functions f and g, and both sets are solutions of the partial differential equation (5.27). Then

$$\begin{aligned} f^*(x) &= \alpha f(x), x \geq 0 \\ g^*(y) &= \beta g(y), y \geq 0 \end{aligned} \tag{5.34}$$

and

$$a_k^* = (\alpha\beta)^{-1} a_k, 0 \leq k \leq n$$

for some nonzero constants α and β.

Proof : Define P^*, F^* and G^* similar to P, F and G for the sequence $a_k^*, 0 \leq k \leq n$, and the functions f^* and g^*. Lemma 5.2.1 shows that

$$P^*(u+v)F^*(u)G^*(v) = H(u,v), u > u_0, v > v_0 \ . \tag{5.35}$$

Relations (5.32) and (5.35) show that

$$P^*(u+v)F^*(u)G^*(v) = P(u+v)F(u)G(v), u > u_0, v > v_0 \ . \tag{5.36}$$

It is sufficient to prove that

$$\begin{aligned} F^*(u) &= \alpha F(u), u > u_0, \\ G^*(v) &= \beta G(v), v > v_0 \end{aligned} \tag{5.37}$$

(5.38)

and

$$P^*(s) = (\alpha\beta)^{-1} P(s), s \in R \ . \tag{5.39}$$

These relations in turn imply (5.34). Relations (5.37) and (5.38) can be proved using methods similar to those discussed earlier in this book and in Section 5.1. We omit the details (cf. Kotlarski (1986)).

Remarks 5.2.1 (Explicit determination of the solution) : Suppose $h(x,y)$ is a known function and there exist constants $a_k, 0 \leq k \leq n$, and functions $f(\cdot)$ and $g(\cdot)$ satisfying the partial differential equation (5.27) subject to the initial condition (5.28). Define H, P, F and G as before. We now give explicit formulae for computation of P, F and G in terms of H. Lemma 5.2.1 implies that

$$H(u,v) = P(u+v)F(u)G(v), u > u_0, v > v_0. \tag{5.40}$$

Let $v = v_1 > v_0$ in (5.39). Then

$$H(u,v_1) = P(u+v_1)F(u)G(v_1), u > u_0 \tag{5.41}$$

and let $u = u_1 > u_0$ in (5.39). Then

$$H(u_1,v) = P(u_1+v)F(u_1)G(v), v > v_0. \tag{5.42}$$

Furthermore

$$H(u_1,v_1) = P(u_1+v_1)F(u_1)G(v_1). \tag{5.43}$$

It is easy to see from these relations that

$$P(u+v)H(u,v_1)H(u_1,v)P(u_1+v_1)$$

$$= H(u,v)P(u+v_1)P(u_1+v)H(u_1,v_1) \tag{5.44}$$

for all $u > u_0, v > v_0, u_1 > u_0$ and $v_1 > v_0$. This is a functional equation only in the unknown P. Taking the continuous branch of the logarithm satisfying $\log 1 = 0$ on both sides of the equation, differentiating with respect to v and substituting $v = v_1$, we have

$$\left[\frac{\partial}{\partial v}(\log P(u+v))\right]_{v=v_1} = \left[\frac{\partial}{\partial v}(\log \frac{H(u,v)}{H(u_1,v)})\right]_{v=v_1} + C \tag{5.45}$$

for some constant C.

Integrating both sides of (5.44) with respect to u in the range $u_1 \leq u \leq s - v_1$, we obtain that

$$P(s) = C\exp\{cs + \int_{u_1}^{s-v_1} [\frac{\partial}{\partial v} \log \frac{H(u,v)}{H(u_1,v)}]_{v=v_1} du\} \tag{5.46}$$

for $s > u_0 + v_0$ where C is a nonzero real constant and c is a real constant. Since $P(\cdot)$ is the generating function of a finite sequence of constants, it has to be a polynomial. It can be shown that there exists a unique constant c_0 for which $P(s)$ is a polynomial in s, namely,

$$c_0 = \frac{d \log P(s)}{ds}|_{s=u_1+v_1} \quad . \tag{5.47}$$

Choose $c = c_0$ as above in (5.45). Then we have $P(\cdot)$ and

$$\begin{aligned} F(u) &= \alpha[P(u+v_1)]^{-1}H(u,v_1), u > u_0, \\ G(v) &= \beta[P(u_1+v)]^{-1}H(u_1,v), v > v_0 \end{aligned} \tag{5.48}$$

from (5.39) and (5.40) where α and β are arbitrary nonzero constants. Substituting the relations (5.47) in the equation (5.41), we obtain that

$$C = [\alpha\beta H(u_1,v_1)]^{-1} \tag{5.49}$$

where C is the constant given in (5.45). This gives us an explicit form for $P(s)$ and hence for $F(u)$ and $G(v)$ where α, β are arbitrary nonzero constants and u_1, v_1 and c_0 are as chosen above.

Example 5.2.1 : Suppose

$$h(x,y) = xy + x + y, x \geq 0, y \geq 0 \ .$$

Then, following (5.31),

$$H(u,v) = (1+u+v)u^{-2}v^{-2}, u > 0, v > 0.$$

Let $u_1 = v_1 = 1$. It can be checked that

$$\begin{aligned} F(u) &= \alpha u^{-2}, u > 0 \, , \\ G(v) &= \beta v^{-2}, v > 0 \end{aligned}$$

and

$$P(s) = (\alpha\beta)^{-1}(1+s), -\infty < s < \infty$$

where α and β are nonzero constants. If $n = 1$, then

$$f(x) = \alpha x, x \geq 0, g(y) = \beta y, y \geq 0,$$

and

$$a_0 = a_1 = (\alpha\beta)^{-1}$$

is the solution for the equation

$$(\alpha\beta)^{-1} \sum_{k=0}^{1} (D_x + D_y)^k f(x)g(y) = h(x, y)$$

with

$$f(0) = g(0) = 0.$$

The results in this section are due to Kotlarski (1986).

Chapter 6

Identifiability in Some Econometric Models

6.1 Introduction

In many fields of biological, physical or social sciences, the main objective of the investigator is not to find the distribution F of an observed random variable X or a random vector $\boldsymbol{X}$ but to identify the probability structure P involved leading to the distribution F. It is theoretically possible, as we will see later in this chapter, that different underlying probability structures P *may* lead to the same probability distribution F. The basic question then is whether a model specified has the property that, given a sample of observations, there could be one and only one probability structure that could have generated this sample. Loosely speaking, we say that a probability structure P is identifiable if there is one and only one probability structure P leading to a given probability distribution F.

Suppose a random variable X is distributed $N(\mu_1 - \mu_2, 1)$. Obviously $\mu_1 - \mu_2$ can be estimated from X. In fact X is the uniformly minimum variance unbiased estimator of $\mu_1 - \mu_2$. However, μ_1 and μ_2 are not individually estimable. There are infinitely many pairs (μ_{1i}, μ_{2i}) such that

$\mu_{1i} - \mu_{2i} = \mu_1 - \mu_2$ for given μ_1 and μ_2. In other words the pair (μ_1, μ_2) is not identifiable.

Let us discuss another example. Consider a pair of random variables X_1 and X_2 distributed as $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ respectively. Suppose $Y = X_1 + X_2$ is observable but not the individual X_1 and X_2. It is obvious that Y is $N(\mu_1 + \mu_2, 2\sigma^2)$. If $Y_i, 1 \leq i \leq n$ is a random sample from this population, then we have information on $\mu_1 + \mu_2$ only and *not* on the individual μ_1 and μ_2. In fact $(\bar{Y}, s_Y^2)$, where $\bar{Y}$ is the sample mean and s_Y^2 is the sample variance, is a sufficient statistic for $(\mu_1 + \mu_2, \sigma^2)$. If μ_1' and μ_2' is another pair of possible values for μ_1 and μ_2 such that $\mu_1' + \mu_2' = \mu_1 + \mu_2$, then the joint density function of $(Y_1, Y_2, ..., Y_n)$ is

$$\prod_{i=1}^{n} \phi(y_i; \mu_1 + \mu_2, 2\sigma^2) ,$$

where $\phi(y; \mu, \sigma^2)$ is the normal density function with mean μ and variance σ^2, either when X_i is $N(\mu_i, \sigma^2), 1 \leq i \leq 2$ or when X_i is $N(\mu_i', \sigma^2), 1 \leq i \leq 2$ as long as $\mu_1 + \mu_2 = \mu_1' + \mu_2'$. In other words the parameters μ_1 and μ_2 are *not identifiable* in this structure. It is easy to see that σ^2 is identifiable .

Suppose that X_1 is $N(\mu, \sigma_1^2)$ and X_2 is $N(\mu, \sigma_2^2)$ and X_1, X_2 independent. Then Y is $N(2\mu, \sigma_1^2 + \sigma_2^2)$. It is easy to see that μ is identifiable but σ_1^2 and σ_2^2 are not.

Let us consider a more general model

$$\begin{aligned} Y_1 &= \eta_1 + \varepsilon_1, \\ Y_2 &= \eta_2 + \varepsilon_2 \end{aligned} \tag{6.1}$$

where (i) $\eta_2 = \alpha + \beta\eta_1$ for some constants α and β, (ii) η_1 is normally distributed independent of $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ and (iii) $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ is bivariate normal with mean (0,0). It is easy to see that the joint distribution of $\boldsymbol{Y} = (Y_1, Y_2)$ is bivariate normal with the covariance matrix

$$\Sigma_{\boldsymbol{Y}} = \begin{pmatrix} 1 & \beta \\ \beta & \beta^2 \end{pmatrix} \text{Var}(\eta_1) + \Sigma_{\boldsymbol{\varepsilon}} \tag{6.2}$$

where $\Sigma_{\boldsymbol{V}}$ denotes the covariance matrix of $\boldsymbol{V}$. The parameter β is not uniquely determined by the above equation. For a fixed $\Sigma_{\boldsymbol{Y}}$, given an arbitrary β, one can always choose $\mathrm{Var}(\eta_1)$ such that $\Sigma_{\boldsymbol{Y}}$ is positive definite and the above equation holds. Since the distribution of $\boldsymbol{Y} = (Y_1, Y_2)$ is uniquely determined by the mean vector and the covariance matrix $\Sigma_{\boldsymbol{Y}}$, it follows that the joint distribution of (Y_1, Y_2) does not identify β. In fact, the parameter β is *uniquely* determined if and only if the joint distribution of (Y_1, Y_2) is *not* bivariate normal. We will give a rigorous proof of this fact later in this chapter.

The problem of identification of the parameters in a statistical model can be referred to as the problem of whether the values of the parameters are uniquely determined by the probability distribution of the model.

Let us consider another example of a regression model. Let

$$Y_1 = \alpha_0 + \alpha_1\eta_1 + \varepsilon_1,$$

and

$$Y_2 = \beta_0 + \beta_1 Y_1 + \varepsilon_2 \ . \tag{6.3}$$

Suppose Y_1 is not observable but Y_2 is. Then

$$Y_2 = \gamma_0 + \gamma_1\eta_1 + \varepsilon_3 \tag{6.4}$$

where

$$\gamma_0 = \beta_0 + \beta_1\alpha_0$$

$$\gamma_1 = \beta_1\alpha_1, \tag{6.5}$$

and

$$\varepsilon_3 = \varepsilon_2 + \beta_1\varepsilon_1 \ .$$

It is clear that γ_0 and γ_1 can be estimated from observations on Y_2 but γ_0 and γ_1 do not determine $\alpha_0, \alpha_1, \beta_0$ and β_1 uniquely. In other words, γ_0 and γ_1 are identifiable but $\alpha_0, \alpha_1, \beta_0$ and β_1 are not.

The identifiability problem is basic to the problem of statistical inference. Unless the parameters in a model are identifiable, there is no meaning of estimability or estimation of such parameters as several combinations of different values for the parameters may lead to the same probability distribution under the given model. However, as Koopmans and Reirsol (1950) point out "...the temptation to specify models in such a way as to produce identifiability of relevant characteristics is (should be) resisted. Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested." For an introduction to problems of identification in economics, see Bartels (1985). There is an extensive literature on identification problems in time series models. We will not discuss it here. For some details, see Deistler and Hannan (1988) and Tigelaar (1982, 1988, 1990). A generalized proportional hazards model is used in econometric models for the study of duration of unemployment. Identifiability problems arising in such models are also of interest and importance.

6.2 Parametric Identification Problem

Following Rothenberg (1971) and Bowden (1973), we now study parametric identification of a probability structure.

Let $\boldsymbol{Y}$ be an m-dimensional random vector representing the outcome of a random experiment. Suppose the probability distribution for $\boldsymbol{Y}$ is known to belong to a family $\mathcal{F}$ of distribution functions on R^m. A *structure* S is a set of hypotheses which implies a unique distribution function $F(S) \in \mathcal{F}$. The set of a priori possible structures is called a *model* denoted by ζ. There is a unique distribution function $F(S) \in \mathcal{F}$ corresponding to each structure $S \in \zeta$. The identification problem is concerned with the existence of a unique inverse for this mapping.

Definition 6.2.1 : Two structures in ζ are said to be *observationally equivalent* if they imply the same probability distribution for the observable random vector $\boldsymbol{Y}$.

Definition 6.2.2 : A structure S in ζ is said to be *identifiable* if there is no other structure in ζ which is observationally equivalent.

Suppose that every structure S is described by a vector $\boldsymbol{\theta} \in R^m$ and the model ζ is described by a set $\Theta \subset R^m$. Further suppose that the distribution of $\boldsymbol{Y}$ under $\boldsymbol{\theta}$ is $F(\boldsymbol{y}, \theta)$. As you might have noticed, by a model here, we mean a probability distribution $F(\boldsymbol{y}, \cdot)$ of known form and, by a structure, we mean a probability distribution function $F(\boldsymbol{y}, \boldsymbol{\theta})$ for a given parameter $\boldsymbol{\theta}$. Thus the problem of differentiating between structures is converted into a problem of differentiating between different values of the parameter θ. Definitions 6.2.1 and 6.2.2 can be recast in the following form.

Definition 6.2.1′ : Suppose the family $\{F(\cdot, \theta), \theta \in \Theta\}$ is dominated by a σ-finite measure μ. Two parameter values θ_0 and θ_1 are said to be *observationally equivalent* if

$$\frac{dF(\boldsymbol{x}, \theta_0)}{d\mu} = \frac{dF(\boldsymbol{x}; \theta_1)}{d\mu} \text{ a.e}[\mu] \ .$$

Definition 6.2.2′ : Suppose the family $\{F(\cdot, \theta), \theta \in \Theta\}$ is dominated by a σ-finite measure μ. A parameter value $\theta_0 \in \Theta$ is said to be *(globally) identifiable* if there exists no other $\theta \in \Theta$ such that

$$\frac{dF(\boldsymbol{x}, \theta_0)}{d\mu} = \frac{dF(\boldsymbol{x}, \theta)}{d\mu} \text{ a.e}[\mu] \ .$$

Definition 6.2.3′ : Suppose the family $\{F(\cdot, \theta), \theta \in \Theta\}$ is dominated by a σ-finite measure μ. A parameter $\theta_0 \in \Theta$ is said to be *locally identifiable* if there exists an open neighbourhood of θ_0 containing no other $\theta \in \Theta$ which is observationally equivalent to θ_0.

Remarks 6.2.1 : It is easy to check that the property of identifiability does not depend on the choice of the dominating measure. Hereafter we assume that the family $\{F(\cdot,\theta),\theta \in \Theta\}$ is dominated by a σ-finite measure μ.

If μ is the Lebesgue measure on R^m, then $F(\cdot,\theta)$ is an absolutely continuous distribution function for every $\theta \in \Theta$. Let $f(\cdot,\theta)$ denote a version of its density function. Definitions 6.2.1′ to 6.2.3′ can be restated in the following form.

Definition 6.2.1″ : Two parameter values θ_1 and θ_2 in Θ are said to be *observationally equivalent* if

$$f(\boldsymbol{x},\theta_1) = f(\boldsymbol{x},\theta_2) \ \text{ a.e}[\lambda]$$

where λ is the Lebesgue measure on R^m.

Definition 6.2.2″ : A parameter value $\theta_0 \in \Theta$ is said to be *(globally) identifiable* if there is no other $\theta \in \Theta$ which is observationally equivalent to θ_0.

Definition 6.2.3″ : A parameter $\theta_0 \in \Theta$ is said to be *locally identifiable* if there exists an open neighbourhood of θ_0 containing no other $\theta \in \Theta$ which is observationally equivalent to θ_0.

The identification problem can be stated as the problem of finding necessary and sufficient conditions for the identifiability of the parameter $\theta \in \Theta$ based on the family of distribution functions $\{F(y,\theta),\theta \in \Theta\}$ (or the family of density functions $\{f(y,\theta),\theta \in \Theta\}$ whenever they exist) and Θ. It is worth noting that the distribution function $F(y,\theta)$ (or the density function $f(y,\theta)$) discussed above could also arise as a mixture of two distribution functions (or two density functions) and the identifiability of the mixing parameter is of interest (cf. Quandt and Ramsey (1978) and Ghosh and Sen (1985)).

We should again caution that identifiability is logically prior to inference and it is connected with proper specification of the theoretical structure that generates the sample observations. It is expected that suitable prior restrictions on Θ and the family of distribution functions $\{F(x,\theta), \theta \in \Theta\}$ or the family of density functions $\{f(x,\theta), \theta \in \Theta\}$ will bring about the identifiability.

6.3 A General Parametric Identification Criterion

Suppose a family of distribution functions $\{F(\boldsymbol{x},\theta), \theta \in \Theta\}$ is dominated by a σ-finite measure μ on R^k and $\{x : \frac{dF(x,\theta)}{d\mu} > 0\}$ does not depend on $\theta \in \Theta$. Define

$$H(\theta,\theta_0) = E_{\theta_0}\left\{\log[\frac{dF(\boldsymbol{X},\theta)/d\mu}{dF(\boldsymbol{X},\theta_0)/d\mu}]\right\}. \tag{6.6}$$

For simplicity, we write

$$H(\theta,\theta_0) = E_{\theta_0}[\log \frac{dF(\boldsymbol{X},\theta)}{dF(\boldsymbol{X},\theta_0)}] \tag{6.7}$$

$$= \int \log[\frac{dF(\boldsymbol{x},\theta)}{dF(\boldsymbol{x},\theta_0)}]dF(\boldsymbol{x},\theta_0) \ . \tag{6.8}$$

If μ is the Lebesgue measure on R, then

$$H(\theta,\theta_0) = \int_{-\infty}^{\infty} \log[\frac{f(\boldsymbol{x},\theta)}{f(x,\theta_0)}]f(x,\theta_0)dx \ . \tag{6.9}$$

$H(\theta,\theta_0)$ is called the *Kullback–Leibler information* (cf. Kullback (1959)). This measure of information can be interpreted in the following manner.

Let H_θ denote the hypothesis that the true density is $f(x,\theta)$ with respect to a σ-finite measure μ. Then the quantity $\log \frac{f(x,\theta)}{f(x,\theta_0)}$ can be taken as the information at x for discriminating between H_0 and H_{θ_0} and the expected information for discrimination between θ and θ_0 is given by

$$H(\theta,\theta_0) = \int_{-\infty}^{\infty} \log[\frac{f(x,\theta)}{f(x,\theta_0)}]f(x,\theta_0)d\mu(x)$$

which is the Kullback–Leibler information described above.

Theorem 6.3.1 : If the distribution function $F(\cdot,\theta)$ is different from the distribution function $F(\cdot,\theta_0)$ and if $H(\theta,\theta_0) < \infty$, then $H(\theta,\theta_0) < 0$.

Proof : Since the distribution functions $F(\cdot,\theta)$ and $F(\cdot,\theta_0)$ are different, it follows that

$$\frac{dF(\boldsymbol{x},\theta)}{dF(\boldsymbol{x},\theta_0)} \neq 1$$

on a set with positive probability under θ_0. By Jensen's inequality, strict concavity of the function $\log x$ implies that

$$\begin{aligned} H(\theta,\theta_0) &= E_{\theta_0} \log \frac{dF(\boldsymbol{X},\theta)}{dF(\boldsymbol{X},\theta_0)} \\ &< \log E_{\theta_0} \frac{dF(\boldsymbol{X},\theta)}{dF(\boldsymbol{X},\theta_0} = 0\,. \end{aligned} \tag{6.10}$$

■

Remarks 6.3.1 : It is easy to see that, if $\theta = \theta_0$, then $H(\theta,\theta_0) = 0$. Hence the parameter θ is globally identifiable iff the equation $H(\theta,\theta_0) = 0$ has a unique solution $\theta = \theta_0$. Observe that $H(\theta,\theta_0)$ attains its maximum at θ_0. Hence a sufficient condition that θ is globally identifiable is that $H(\theta,\theta_0)$ is strictly concave on Θ and Θ is convex.

Let us now discuss the relation between the Kullback–Leibler information and Fisher information.

Case of scalar parameter : Assume that θ is a scalar parameter, that is, $\Theta \subset R$.

Suppose the function $H(\theta,\theta_0)$ is differentiable twice with respect to θ under the integral sign. Note that

$$H(\theta,\theta_0) = \int \log[\frac{f(\boldsymbol{x},\theta)}{f(\boldsymbol{x},\theta_0}]f(\boldsymbol{x},\theta_0)d\mu(\boldsymbol{x}) \tag{6.11}$$

and hence

$$\begin{aligned} H'(\theta,\theta_0) &= \frac{d}{d\theta}\{\int \log[\frac{f(\boldsymbol{x},\theta)}{f(\boldsymbol{x},\theta_0)}]f(\boldsymbol{x},\theta_0)d\mu(\boldsymbol{x})\} \\ &= \int_R \frac{f'(\boldsymbol{x},\theta)}{f(\boldsymbol{x},\theta)}f(\boldsymbol{x},\theta_0)d\mu(\boldsymbol{x}) \\ &= E_{\theta_0}[\frac{\partial}{\partial\theta} \log f(\boldsymbol{X},\theta)]. \end{aligned} \tag{6.12}$$

Furthermore

$$
\begin{aligned}
H''(\theta,\theta_0) &= \frac{d}{d\theta}\{\int \frac{f'(\boldsymbol{x},\theta)}{f(\boldsymbol{x},\theta)} f(\boldsymbol{x},\theta_0)d\mu(\boldsymbol{x})\} \\
&= \int \frac{d}{d\theta}\{\frac{f'(\boldsymbol{x},\theta)}{f(\boldsymbol{x},\theta)}\}f(\boldsymbol{x},\theta_0)d\mu(\boldsymbol{x}) \\
&= \int \frac{f(\boldsymbol{x},\theta)f''(\boldsymbol{x},\theta)-(f'(\boldsymbol{x},\theta))^2}{(f(\boldsymbol{x},\theta))^2} f(\boldsymbol{x},\theta_0)d\mu(\boldsymbol{x}) \\
&= E_{\theta_0}[\frac{f''(\boldsymbol{X},\theta)}{f(\boldsymbol{X},\theta)}] - E_{\theta_0}[\frac{\partial \log f(\boldsymbol{X},\theta)}{\partial\theta}]^2 . \qquad (6.13)
\end{aligned}
$$

Here g' and g'' denote the first and second derivatives of g respectively. Since

$$\int f(\boldsymbol{x},\theta)d\mu(\boldsymbol{x}) = 1, \qquad (6.14)$$

it follows that

$$\int f'(\boldsymbol{x},\theta)d\mu(\boldsymbol{x}) = \int f''(\boldsymbol{x},\theta)d\mu(\boldsymbol{x}) = 0 \qquad (6.15)$$

under the assumption of differentiability twice under the integral sign with respect to θ. It is now easy to check that

$$H'(\theta_0,\theta_0) = 0 \qquad (6.16)$$

and

$$
\begin{aligned}
H''(\theta_0,\theta_0) &= -E_{\theta_0}[\frac{\partial \log f(\boldsymbol{X},\theta)}{\partial\theta}|_{\theta_0}]^2 \\
&= -I(\theta_0) \qquad (6.17)
\end{aligned}
$$

where $I(\theta_0)$ is the Fisher information. Hence, if $0 < I(\theta_0) < \infty$, then $H''(\theta_0,\theta_0) < 0$. Since $H'(\theta,\theta_0) = 0$ at $\theta = \theta_0$, the function $H(\theta,\theta_0)$ has a local maximum at θ_0 and the parameter θ_0 is locally identifiable.

Case of vector parameter : If θ is a vector parameter, i.e., $\theta = (\theta_1, ..., \theta_k)$ say, then it can be checked, under the classical regularity conditions for the validity of Cramér–Rao inequality, that

$$H''(\theta_0,\theta_0) = -I(\theta_0) \qquad (6.18)$$

where $I(\theta_0)$ is the Fisher information matrix with (i, j)th element

$$I_{ij}(\theta_0) \equiv E_{\theta_0}[\frac{\partial \log f(\boldsymbol{X}, \theta)}{\partial \theta_i}|_{\theta=\theta_0} \frac{\partial \log f(\boldsymbol{X}, \theta)}{\partial \theta_j}|_{\theta=\theta_0}] . \tag{6.19}$$

If $I(\theta_0)$ is of full rank and hence positive definite, then $H''(\theta_0, \theta_0)$ is negative definite and it follows that $H(\theta, \theta_0)$ has a local maximum at $\theta = \theta_0$ since $H'(\theta_0, \theta_0) = 0$. Hence θ_0 is locally identifiable. Here H'' is the Hessian and H' is the gradient of H.

Exponential families : For most of the problems encountered in practice, the interest is in global identifiability of the parameter rather than local identifiability. In general, conditions implying global identifiability are not easy to obtain for the class of densities $\{f(\boldsymbol{y}, \theta), \theta \in \Theta\}$. However, for exponential families, this can be done as will be shown below. Suppose

$$f(\boldsymbol{y}, \theta) = \exp[A(\boldsymbol{y}) + B(\theta) + \sum_{i=1}^{k} \theta_i T_i(\boldsymbol{y})] \tag{6.20}$$

for all $\boldsymbol{y}$ and $\theta \in \Theta$ with respect to a σ-finite measure μ and further assume that, for some $\theta_1 \neq \theta_2$ in Θ,

$$f(\boldsymbol{y}, \theta_1) = f(\boldsymbol{y}, \theta_2) \text{ a.e}[\mu]. \tag{6.21}$$

Then it follows that

$$A(\boldsymbol{y}) + B(\theta_1) + \sum_{i=1}^{k} \theta_{i1} T_i(y) = A(\boldsymbol{y}) + B(\theta_2) + \sum_{i=1}^{k} \theta_{i2} T_i(\boldsymbol{y}) \text{ a.e } [\mu] \tag{6.22}$$

where $\theta_1 = (\theta_{11}, ..., \theta_{k1})$ and $\theta_2 = (\theta_{12}, ..., \theta_{k2})$. Hence

$$B(\theta_1) - B(\theta_2) = -\sum_{i=1}^{k} (\theta_{i1} - \theta_{i2}) T_i(\boldsymbol{y}) \text{ a.e}[\mu] . \tag{6.23}$$

Assume that $B(\theta)$ is continuously differentiable with respect to θ on Θ. Then it follows that there exists $\theta^* \in \Theta$ such that

$$(\theta_1 - \theta_2)^t \nabla B(\theta^*) = -(\theta_1 - \theta_2)^t T(\boldsymbol{y}) \text{ a.e } [\mu] \tag{6.24}$$

where $\nabla B(\theta) = (\frac{\partial B(\theta)}{\partial \theta_1}, \ldots, \frac{\partial B(\theta)}{\partial \theta_k})^t, T(y) = (T_1(y), \ldots, T_k(y))^t$ and α^t denotes the transpose of row vector α. Observe that θ^* does not depend on y. In other words

$$(\theta_1 - \theta_2)^t[\nabla B(\theta^*) + T(y)] = 0 \ \text{a.e}[\mu]. \tag{6.25}$$

Note that

$$\nabla \log f(y, \theta) = \nabla B(\theta) + T(y). \tag{6.26}$$

Hence

$$(\theta_1 - \theta_2)^t \nabla \log f(y, \theta^*) = 0 \ \text{a.e}[\mu] \tag{6.27}$$

where θ^* *does not depend on* y or equivalently

$$(\theta_1 - \theta_2)^t \ \nabla \ \log f(y, \theta^*) \ \nabla \ \log f(y, \theta^*)^t (\theta_1 - \theta_2) = 0 \ \text{a.e} \ [\mu]. \tag{6.28}$$

Taking expectation with respect to θ^*, it follows that

$$(\theta_1 - \theta_2)^t I(\theta^*)(\theta_1 - \theta_2) = 0. \tag{6.29}$$

Since $\theta_1 \neq \theta_2$, it follows that $I(\theta^*)$ is a singular matrix. Hence we have the following theorem.

Theorem 6.3.2 : Suppose the family of density functions $\{f(y, \theta), \theta \in \Theta\}$ is a multivariate exponential family given by

$$\log f(y, \theta) = A(y) + B(\theta) + \sum_{i=1}^{k} \theta_i T_i(y) \tag{6.30}$$

with respect to a σ-finite measure μ. Further suppose that $B(\theta)$ is continuously differentiable in $\theta \in \Theta$. Then every θ in Θ is *globally identifiable* if the Fisher information matrix (assumed to be finite) is nonsingular equivalently of full rank for every $\theta \in \Theta$.

Another situation where global identification is possible is given by the following theorem.

Theorem 6.3.3 : Suppose there exist k known functions $\phi_i(\boldsymbol{y}), 1 \leq i \leq k$ such that

$$\theta_i = E_\theta[\phi_i(\boldsymbol{Y})], 1 \leq i \leq k, \theta \in \Theta \tag{6.31}$$

when $\boldsymbol{Y}$ has the distribution $F(\boldsymbol{y}, \theta)$ under the parameter θ. Then every $\theta \in \Theta$ is *globally identifiable* .

Proof : This result is an easy consequence of the fact that if $F(\boldsymbol{y}, \theta_1) = F(\boldsymbol{y}, \theta_2)$ for all $\boldsymbol{y}$, then

$$\int \phi_i(\boldsymbol{y}) dF(\boldsymbol{y}, \theta_1) = \int \phi_i(\boldsymbol{y}) dF(\boldsymbol{y}, \theta_2), 1 \leq i \leq k$$

and hence $\theta_{i1} = \theta_{i2}, 1 \leq i \leq k$ where $\theta_j = (\theta_{1j}, ..., \theta_{kj}), j = 1, 2.$ ∎

The results in this section are due to Rothenberg (1971) and Bowden (1973).

6.4 Identifiability for Some Structural Models

The identification problem for structural models in econometrics is extensively discussed (cf. Fisher (1966)). We will not discuss all the results in this area but concentrate on some special models.

Example 6.4.1 (Reiersol (1950)). Let us consider the following model :

Model (A)

$$Y_1 = \eta_1 + \varepsilon_1,$$

$$Y_2 = \eta_2 + \varepsilon_2 \tag{6.32}$$

where

$$\begin{array}{ll} \text{(i)} & \eta_2 = \alpha + \beta\eta_1, \\ \text{(ii)} & \eta_1 \text{ independent of } (\varepsilon_1, \varepsilon_2) \text{and} \\ \text{(iii)} & (\varepsilon_1, \varepsilon_2) \text{ is bivariate normal with mean} \\ & \text{(0,0) and covariance matrix } \Sigma\ . \end{array} \tag{6.33}$$

Suppose Y_2 is observable but not Y_1. The problem is to find conditions under which the parameters α, β and other unknown parameters and distributions are identifiable; i.e., the model is identifiable. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \text{ and } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}. \tag{6.34}$$

Let $\phi_{Z_1,Z_2}(t_1, t_2)$ denote the characteristic function of a bivariate random vector (Z_1, Z_2) and $\phi_Z(t)$ denote the characteristic function of a random variable Z. Observe that

$$\begin{aligned} \phi_{\boldsymbol{Y}}(t_1, t_2) &= E[\exp\{iY_1t_1 + iY_2t_2\}] \\ &= E[\exp\{i(\eta_1 + \varepsilon_1)t_1 + i(\alpha + \beta\eta_1 + \varepsilon_2)t_2\}] \\ &= E[\exp\{i\alpha t_2 + i(t_1 + \beta t_2)\eta_1 + it_1\varepsilon_1 + it_2\varepsilon_2\}] \\ &= e^{i\alpha t_2}\phi_{\eta_1}(t_1 + \beta t_2)\phi_{\boldsymbol{\varepsilon}}(t_1, t_2) \end{aligned} \tag{6.35}$$

since η_1 is independent of $\boldsymbol{\varepsilon}^T = (\varepsilon_1, \varepsilon_2)$. But

$$\phi_{\boldsymbol{\varepsilon}}(t_1, t_2) = \exp\{-\frac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t}\} \tag{6.36}$$

where $\boldsymbol{t}^T = (t_1, t_2)$ and $\boldsymbol{t} = \binom{t_1}{t_2}$.

Suppose there exist two different structures

$$S = (\beta, \alpha, \Sigma, \phi_{\eta_1}(t))$$

and

$$S^* = (\beta^*, \alpha^*, \Sigma^*, \phi^*_{\eta_1}(t)) \tag{6.37}$$

generating the *same* joint distribution for $\boldsymbol{Y}$. Then

$$\begin{aligned} \phi_{\boldsymbol{Y}}(\boldsymbol{t}) &= e^{i\alpha t_2}\phi_{\eta_1}(t_1 + \beta t_2)\exp\{-\frac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t}\} \\ &= e^{i\alpha^* t_2}\phi^*_{\eta_1}(t_1 + \beta^* t_2)\exp\{-\frac{1}{2}\boldsymbol{t}^T\Sigma^*\boldsymbol{t}\}. \end{aligned} \tag{6.38}$$

Suppose $\beta \neq \beta^*$. Given an arbitrary u, let us determine t_1 and t_2 such that

$$t_1 + \beta t_2 = u \text{ and } t_1 + \beta^* t_2 = 0\ . \tag{6.39}$$

This can be done by choosing

$$t_1 = \frac{-\beta^* u}{\beta - \beta^*}, t_2 = \frac{u}{\beta - \beta^*} \quad . \tag{6.40}$$

Then, the equation (6.38) implies that

$$\phi_{\eta_1}(u) = \exp\{i\, \frac{\alpha^* - \alpha}{\beta - \beta^*} u - \frac{u^2}{2(\beta - \beta^*)^2}\gamma^T(\Sigma^* - \Sigma)\gamma\} \tag{6.41}$$

where $\gamma^T = (-\beta^*, 1)$. Since $\phi_{\eta_1}(\cdot)$ is the characteristic function of a random variable, it follows that η_1 is either normally distributed or η_1 is a constant with probability one. Since $\eta_2 = \alpha + \beta\eta_1$, it is obvious that η_2 is also normally distributed or η_2 is a constant. In fact (η_1, η_2) has either a nondegenerate bivariate normal distribution or it is a constant with probability one. This proves the following result.

Proposition 6.4.1 : A sufficient condition that the parameter β is identifiable in the Model (A) is that (η_1, η_2) neither is degenerate nor does it have a bivariate normal distribution, or equivalently if (η_1, η_2) has a bivariate normal distribution or it is a constant, then the parameter β is not identifiable.

Let us now suppose that the parameter β is identifiable in the Model (A). Then, for any two equivalent structures S and S^* given by (6.37), we have $\beta = \beta^*$ and hence

$$\phi_{\eta_1}(u)[\phi^*_{\eta_1}(u)]^{-1}$$

$$= \exp\{i(\alpha^* - \alpha)(u - \beta t_2) - \frac{1}{2}(u - \beta t_2, t_2)(\Sigma^* - \Sigma)\begin{pmatrix} u - \beta t_2 \\ t_2 \end{pmatrix}^T\} \tag{6.42}$$

where $u = t_1 + \beta t_2$ from (6.39). Let $\psi_{\eta_1}(u)$ be the principal branch of log $\phi_{\eta_1}(u)$ with $\psi_{\eta_1}(0) = 0$. Since the expression on the left side of equation (6.42) does not depend on t_2, it follows that the coefficients of t_2, t_2^2 and ut_2 must be zero on the right side. Hence

$$\begin{aligned} \alpha^* &= \alpha, \\ \lambda_{12} - \lambda^*_{12} &= \beta(\lambda_{11} - \lambda^*_{11}) \end{aligned} \tag{6.43}$$

and

$$\lambda_{22} - \lambda^*_{22} = \beta(\lambda_{12} - \lambda^*_{12})$$

where λ_{ij} and λ^*_{ij} are (i,j)th elements of Σ and Σ^* respectively. This proves that α is identifiable if β is identifiable in Model (A). Relation (6.43) proves that

$$\phi_{\eta_1}(u)[\phi^*_{\eta_1}(u)]^{-1} = \exp\{-\frac{1}{2}(\lambda_{11} - \lambda^*_{11})u^2\} \ . \tag{6.44}$$

Hence the distributions of η_1 differ by a normal factor under both the structures S and S^* provided $\lambda_{11} - \lambda^*_{11} > 0$.

Example 6.4.2 (Willassen (1979)): Let us now consider a generalization of the Model (A) discussed in Example 6.4.1.

Model (A*)

$$\begin{aligned} X_i &= Y_i + \varepsilon_i, 0 \le i \le k \ , \\ Y_0 &= \gamma_0 + \gamma_1 Y_1 + \ldots + \gamma_k Y_k. \end{aligned} \tag{6.45}$$

Suppose the random variables $\{X_i, 0 \le i \le k\}$ are *observable* where as $\{Y_i, 0 \le i \le k\}$ are *not observable.* $\{Y_i, 0 \le i \le k\}$ are called *latent variables.* Here $\{\varepsilon_i, 0 \le i \le k\}$ are the unobserved errors. Assume that

(i) the vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_1, \ldots, \varepsilon_k)$ is independent of $\boldsymbol{Y} = (Y_0, Y_1, ..., Y_k)$,

(ii) $\boldsymbol{\varepsilon}$ is multivariate normal with mean zero and covariance matrix Σ and

(iii) $\{Y_i, 1 \le i \le k\}$ are independent.

Let $\boldsymbol{X} = (X_0, X_1, \ldots, X_k)$ and $\boldsymbol{t}^T = (t_0, ..., t_k)$. Let $\phi_{\boldsymbol{X}}$ denote the characteristic function of $\boldsymbol{X}$. It is easy to check that

$$\phi_{\boldsymbol{X}}(t) = \exp(i\gamma_0 t_0 - \frac{1}{2}\boldsymbol{t}^T \Sigma \boldsymbol{t}) \prod_{j=1}^{k} \phi_{Y_j}(\gamma_j t_0 + t_j) \ . \tag{6.46}$$

Let us call

$$H = \{\gamma_0, \gamma_1, ..., \gamma_k, \Sigma, \phi_{Y_1}, \ldots, \phi_{Y_k}\}$$

the latent structure of the model. There may exist several different structures H^* which generate the same joint distribution F for $\boldsymbol{X}$. The problem of identification is to find conditions under which the correspondence between the family of latent structures H and the family of distributions F for $\boldsymbol{X}$ is one-to-one.

Suppose H and H^* generate the *same* probability distribution F for $\boldsymbol{X}$. Let

$$H^* = \{\gamma_0^*, \gamma_1^*, ..., \gamma_k^*, \Sigma^*, \phi_{Y_1}^*, ..., \phi_{Y_k}^*\} \ .$$

Equating the characteristic functions of $\boldsymbol{X}$ under both the structures H and H^*, we have

$$\exp(i\gamma_0 t_0 - \tfrac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t})\prod_{j=1}^{k}\phi_{Y_j}(\gamma_j t_0 + t_j)$$

$$= \exp(i\gamma_0^* t_0 - \frac{1}{2}\boldsymbol{t}^T\Sigma^*\boldsymbol{t})\prod_{j=1}^{k}\phi_{Y_j}^*(\gamma_j^* t_0 + t_j) \ . \tag{6.47}$$

Suppose that the parameters $\gamma_1, \gamma_2, ..., \gamma_k$ are not identifiable. Then $\gamma_j \neq \gamma_j^*, 1 \leq j \leq k$ and yet the distribution F of $\boldsymbol{X}$ is the same under both H and H^* . Suppose $\gamma_1^* \neq 0$. Let us choose $(t_0, t_1, ..., t_k)$ such that

$$\begin{aligned} \gamma_1^* t_0 + t_1 &= 0, \\ \gamma_2^* t_0 + t_2 &= 0, \end{aligned} \tag{6.48}$$

and

$$\gamma_k^* t_0 + t_k = 0.$$

Then

$$t_0 = -\frac{t_1}{\gamma_1^*}, t_i = -\gamma_i^* t_0 = \frac{\gamma_i^* t_1}{\gamma_1^*}, 2 \leq i \leq k \ . \tag{6.49}$$

Substituting these values of $t_0, t_1, ..., t_k$ in the equation (6.47) we have

$$\begin{aligned} \prod_{j=1}^{k}\phi_{Y_j}(\frac{(\gamma_j^* - \gamma_j)}{\gamma_1^*}t_1) &= \exp\{i(\gamma_0^* - \gamma_0)t_0\}\exp(-\mu\,\frac{t_1^2}{2}) \\ &= \exp(i\{\frac{\gamma_0 - \gamma_0^*}{\gamma_1^*}\}t_1 - \mu\,\frac{t_1^2}{2}) \end{aligned} \tag{6.50}$$

where μ is a positive constant. In other words, the sum of independent rescaled independent random variables $Y_i, 1 \leq i \leq k$ is normally distributed.

By the decomposition theorem of Cramér (cf. Lukacs (1970, Sec. 8.2)), it follows that $Y_i, 1 \leq i \leq k$ are normally distributed.

Conversely, suppose that $Y_i, 1 \leq i \leq k$ are independent normally distributed random variables under Model (A*) satisfying the assumptions (i) to (iii) . Then $\boldsymbol{X} = (X_0, X_1, ..., X_k)$ has a multivariate normal distribution and the distribution of $\boldsymbol{X}$ is determined by its mean vector and its covariance matrix. Note that

$$E(X_j) = E(Y_j), \quad 0 \leq j \leq k, \tag{6.51A}$$

$$\mathrm{Var}(X_j) = \mathrm{Var}(Y_j) + \mathrm{Var}(\varepsilon_j), \quad 1 \leq j \leq k, \tag{6.51B}$$

$$\mathrm{Var}(X_0) = \sum_{j=1}^{k} \gamma_j^2 \mathrm{Var}(Y_j) + \mathrm{Var}(\varepsilon_0), \tag{6.51C}$$

$$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(Y_i, Y_j) + \mathrm{Cov}(\varepsilon_i, \varepsilon_j), \quad 1 \leq i, j \leq k; \tag{6.51D}$$

and

$$\mathrm{Cov}(X_0, X_j) = \gamma_j \mathrm{Var}(Y_j) + \mathrm{Cov}(\varepsilon_0, \varepsilon_j), 1 \leq j \leq k\ . \tag{6.51E}$$

Apart from the means of $Y_j, 0 \leq j \leq k$ which are identifiable from the means of $X_j, 0 \leq j \leq k$ from (6.51A), the system of independent equations in (6.51B) to (6.51E) is

$$\frac{k(k-1)}{2} + 2k + 1 = \frac{(k+2)(k+1)}{2}$$

in number. However, the number of unknown parameters is

$$2k + \frac{(k+2)(k+1)}{2}$$

since $\mathrm{Var}(Y_j), 1 \leq j \leq k; \gamma_j, 1 \leq j \leq k$ and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j), 0 \leq i, j \leq k$ are unknown. Hence there is no unique solution for the system for a given set of $\mathrm{Cov}(X_i, X_j), 1 \leq i, j \leq k$. In other words $\gamma_1, \gamma_2, ..., \gamma_k$ are not identifiable. This result together with the one obtained above proves the following proposition.

Proposition 6.4.2 : Suppose the assumptions (i) to (iii) hold in Model (A*). A necessary and sufficient condition for the set $\{\gamma_1, \gamma_2, ..., \gamma_k\}$ to be identifiable in the Model (A*) is that the set of $(Y_i, 1 \leq i \leq k\}$ is *not* normally distributed.

Remarks 6.4.1 : (i) If it is assumed that the random vector $\boldsymbol{\varepsilon}$ in Model (A*) has *independent* components, then Willassen (1979) has proved that nonidentifiability of $\gamma_1, \gamma_2, ..., \gamma_k$ implies that $Y_i, 1 \leq i \leq k$ are normally distributed. We omit the proof.

(ii) Results obtained here continue to hold if the Model (A*) contains random vectors $\boldsymbol{X}_i, \boldsymbol{Y}_i, \varepsilon_i$ and the coefficient γ_i are matrices of suitable dimensions. One has to use the Cramér decomposition theorem in the multivariate setup (cf. Cuppens (1975, p. 109)) in this case.

(iii) An alternate approach to obtain these types of results is due to Linnik (1964) and Rao (1966, 1971) via functional equations as discussed in Chapter 2 .

Example 6.4.3 (Rothenberg (1971)): Consider the nonlinear regression model

$$Y_i = h_i(\theta, x_i) + \varepsilon_i, 1 \leq i \leq n, n \geq k$$

where $\theta \in R^k, h_i$ is twice differentiable in θ and $\{\varepsilon_i\}$ are i.i.d. $N(0,1)$ random variables and x_i are known constants. It is easy to check that the Fisher information matrix is

$$I(\theta) = H(\theta)H(\theta)^T$$

where

$$H(\theta) = ((h_i(\theta, x_j)))_{k \times k}$$

and θ is locally identifiable if $H(\theta)$ has full rank.

Example 6.4.4 (Bowden (1973)): Suppose $\boldsymbol{Y} = (Y_1, \ldots, Y_k)^T$ is multivariate normal with mean $\boldsymbol{X}\beta_0$ and nonsingular covariance matrix Σ_0. Let $\theta = (\beta, \Sigma)$. Then the Kullback–Leibler information number is

$$\begin{aligned}
H(\theta, \theta_0) &= \frac{1}{2}\log(\frac{\det \Sigma_0}{\det \Sigma}) \\
&+\frac{1}{2}E_{\beta_0,\Sigma_0}[(\boldsymbol{Y} - \boldsymbol{X}\beta_0)^T\Sigma_0^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta_0)] \\
&-\frac{1}{2}E_{\beta_0,\Sigma_0}[(\boldsymbol{Y} - \boldsymbol{X}\beta)^T\Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta)] \\
&= \frac{1}{2}\log(\frac{\det \Sigma_0}{\det \Sigma}) \\
&+\frac{1}{2}(k - \operatorname{tr}\Sigma^{-1}\Sigma_0 - (\beta - \beta_0)^T\boldsymbol{X}^T\,\Sigma^{-1}\boldsymbol{X}(\beta - \beta_0))
\end{aligned}$$

and

$$\frac{\partial H}{\partial \beta} = -(\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})(\beta - \beta_0).$$

It is clear that $\frac{\partial H}{\partial \beta} = 0$ for $\beta \neq \beta_0$ only if $\boldsymbol{X}$ does not have full rank. In fact if $\boldsymbol{X}$ has full rank, then β_0 is identifiable and if $\theta = (\beta, \Sigma)$ is observationally equivalent to $\theta_0 = (\beta_0, \Sigma_0)$, then $\beta = \beta_0$. In this case

$$H(\theta, \theta_0) = \frac{1}{2}\log\left(\frac{\det\Sigma_0}{\det\Sigma}\right) + \frac{1}{2}(k - \operatorname{tr}\,\Sigma^{-1}\Sigma_0).$$

If this equation $H(\theta, \theta_0) = 0$ in Σ has only one solution $\Sigma = \Sigma_0$, then θ_0 is identifiable. If $\boldsymbol{X}$ does not have full rank, then β_0 is not identifiable and hence θ_0 is not identifiable.

Example 6.4.5 : Consider the linear model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$ where $\boldsymbol{X}$ is the design matrix, $E(\boldsymbol{\varepsilon}) = 0$ and the covariance matrix of $\boldsymbol{\varepsilon}$ is $\sigma^2 I$. Then β is identifiable if $\boldsymbol{X}$ has full rank but σ^2 is always identifiable. This can be seen from the following remarks. Note that

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta \text{ and } \operatorname{Cov}(\boldsymbol{Y}) = \sigma^2 I.$$

It is obvious that two different values of σ^2 in the model cannot give the same covariance matrix for $\boldsymbol{Y}$. Hence σ^2 is identifiable always. Suppose there are two values of β (say) β_0, β_1 for which $E(\boldsymbol{Y})$ is the same. Then

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta_0 = \boldsymbol{X}\beta_1$$

and hence

$$\boldsymbol{X}(\beta_0 - \beta_1) = 0.$$

If $\boldsymbol{X}$ has full rank, then $\beta_0 = \beta_1$ and hence β is identifiable. If $\boldsymbol{X}$ does not have full rank, then there exist β_0 and $\beta_1, \beta_0 \neq \beta_1$, such that

$$\boldsymbol{X}(\beta_0 - \beta_1) = 0.$$

Further information about the distribution or parameter restrictions on β are needed to identify the parameter β.

6.5 Further Remarks on Identifiability

(i) It is useful to note that if a vector parameter θ is identifiable in a model and $g(\cdot)$ is a single–valued function of θ, then $\phi = g(\theta)$ is identifiable. Here, by a single-valued function, we mean that if $\phi = g(\theta)$ and $\phi^* = g(\theta^*)$,then $\phi \neq \phi^*$ implies $\theta \neq \theta^*$. It is possible that the parameter θ itself may not be identifiable but there might be a function $\gamma(\theta)$ (nonconstant) which is identifiable. Then θ is said to be partially identifiable and γ is said to be identifiable.

(ii) It is possible that two structures are not strictly observationally equivalent but nearly identifiable or there might be situations where the problem of near unidentifiability might occur as for instance in the model discussed in Kumar and Gapinski (1974) and Kumar and Asher (1974). Here the question of degree of identifiability is also relevant. We do not go into the discussion here and the problem does not seem to have received attention. The problem is akin to the discussion on stability of characterization of probability distribution.

(iii) It is interesting to observe that if there exists a consistent estimator for a parameter θ, then the parameter is identifiable. This can be seen from the following arguments. Suppose θ is not identifiable. Then there exist at least two different values of the parameter (say) θ_1 and θ_2 leading to the same distribution for the observations. If $\hat{\theta}_n$ is a consistent estimator of θ based on the observation $(X_1, X_2, ..., X_n)$, then $\hat{\theta}_n$ should converge to both θ_1 and θ_2 in probability as $n \to \infty$. This is impossible since $\theta_1 \neq \theta_2$. This proves that the existence of a consistent estimator for θ implies its identifiability . However, the converse is not true in general (Gabrielsen

(1978)). This can be seen by the following example. Consider the stochastic model

$$Y_i = \beta\rho^i + \varepsilon_i, 1 \leq i \leq n \tag{6.52}$$

where ρ is known with $|\rho| < 1, \varepsilon_i$ i.i.d. $N(0,1)$ and $\beta > 0$ but unknown. Since

$$E(Y_i) = \beta\rho^i$$

and ρ is known, it immediately follows that β is identifiable. However, there exists no consistent estimator for β. This can be seen from the following analysis. It is easy to see that

$$\hat{\beta}_n = (\sum_{i=1}^{n} \rho^i Y_i)/(\sum_{i=1}^{n} \rho^{2i}) \tag{6.53}$$

is the maximum likelihood estimator of β based on $(Y_1, ..., Y_n)$. It can be checked that

$$E(\hat{\beta}_n) = \beta \tag{6.54}$$

and

$$\sigma_n^2 \equiv \text{Var}(\hat{\beta}_n) = 1/(\sum_{i=1}^{n} \rho^{2i}) = \frac{1-\rho^2}{\rho^2(1-\rho^{2n})}. \tag{6.54A}$$

Hence $\sigma_n^2 \to \sigma^2 = \frac{1-\rho^2}{\rho^2}$ as $n \to \infty$. Observe that $\hat{\beta}_n$ is $N(\beta, \sigma_n^2)$. Hence $\hat{\beta}_n \xrightarrow{\mathcal{L}} N(\beta, \sigma^2)$. If $\hat{\beta}_n$ were consistent for β, then $\hat{\beta}_n \xrightarrow{p} \beta$ by definition and hence $\hat{\beta} \xrightarrow{\mathcal{L}} \beta$ which contradicts the fact that $\hat{\beta}_n \xrightarrow{\mathcal{L}} N(\beta, \sigma^2)$. Hence $\hat{\beta}_n$ is inconsistent for β.

Let us consider a test of the hypothesis $H_0 : \beta = 0$ against the alternative $H_1 : \beta > 0$. The uniformly most powerful (UMP) level α test for testing H_0 against H_1 has the critical region

$$[\hat{\beta}_n > \sigma_n z_{1-\alpha}]$$

where $z_{1-\alpha}$ is such that $\Pr[Z > z_{1-\alpha}] = \alpha$ and the random variable Z has the standard normal distribution. It is easy to check that the power function $\hat{\gamma}_n(\beta)$ of this UMP test is given by

$$\hat{\gamma}_n(\beta) = \Phi(\frac{\beta}{\sigma_n} - z_{1-\alpha}) \tag{6.55}$$

where Φ is the standard normal distribution function. Observe that no other test depending on $Y_1, ..., Y_n$ has more power than $\hat{\gamma}_n(\beta)$.

Suppose β_n^* is a consistent estimator of β. Let us consider the test which rejects H_0 if $\beta_n^* > 1$. For any $\beta > 0$, power of this test is

$$\gamma_n^*(\beta) = \Pr_\beta[\beta_n^* > 1] .$$

Since the test based on $\hat{\beta}_n$ is the UMP test, it follows that

$$\hat{\gamma}_n(\beta) \geq \gamma_n^*(\beta) .$$

Hence

$$\overline{\lim_{n\to\infty}} \hat{\gamma}_n(\beta) \geq \overline{\lim_{n\to\infty}} \gamma_n^*(\beta) .$$

But

$$\overline{\lim_{n\to\infty}} \hat{\gamma}_n(\beta) \leq \Phi(\frac{\beta}{\sigma} - z_{1-\alpha}) .$$

Hence

$$\overline{\lim_{n\to\infty}} \gamma_n^*(\beta) \leq \Phi(\frac{\beta}{\sigma} - z_{1-\alpha}) . \tag{6.56}$$

Since β_n^* is consistent for β, it follows that

$$\gamma_n^*(\beta) \to \Pr_\beta(\beta > 1) \text{ as } n \to \infty \tag{6.57}$$

which is equal to 1 for $\beta > 1$. This contradicts the inequality (6.56) since $\Phi(\frac{\beta}{\sigma} - z_{1-\alpha}) < 1$. Hence there exists no consistent estimator for β.

(iv) Example 6.4.5 gives the impression that both estimability as discussed in the statistical literature and identifiability are one and the same. Indeed, they are equivalent in the context of linear models or when the distribution of the observation vector Y has a multivariate normal distribution. See the discussion in Mitra (1980) or Bunke and Bunke (1974).

(v) Extensive discussions on identification of structural economic models are given in books on econometrics. For instance, see F.M. Fisher (1966). Moran (1971) surveys the problem of estimating a linear relationship between variables which are observed with errors known as "errors-in-variables model." The variables could be either fixed variables (functional relationship) or random variables (structural relationship). Several results were discussed on identifiability for such models in Moran (1971). For related results on identifiability problems in time series models, see Tigelaar (1982, 1988, 1990) and Deistler and Hannan (1988). We have discussed sufficient conditions for identifiability of a model. The basic problem is to obtain necessary and sufficient conditions. As we have noted already, the model specification for identifiability requires restrictions to be imposed on the family of distribution functions $\{F(x,\theta),\theta \in \Theta\}$ or the family of density functions $\{f(x,\theta),\theta \in \Theta\}$ and Θ so that identifiability is achieved. However, there may exist different sets of restrictions that might achieve the same goal, namely identifiability. The question is how to choose among such sets of conditions. Is it possible to arrive at a minimal set of sufficient conditions for identifiability?

6.6 Identifiability for a Generalized Proportional Hazard Model

Econometricians studying labour market phenomena have developed methods for the analysis of duration of unemployment (cf. Lancaster (1979)). One of the methods that was proposed in Lancaster (1979) is a generalization of the proportional hazard model developed by Cox (1972). The model tries to explain the length of an individual spell of unemployment or equivalently the probability of leaving the state of unemployment. Let the probability that an individual leaves unemployment in the interval $[t, t+\Delta t)$ be $\theta(t,\boldsymbol{x},\boldsymbol{\beta})\Delta t$ where t is the time elapsed since the beginning of the spell of unemployment, $\boldsymbol{x}$ is a vector of covariates and $\boldsymbol{\beta}$ is a parameter vector . Suppose

$$\theta(t,\boldsymbol{x},\boldsymbol{\beta}) = \phi(\boldsymbol{x},\boldsymbol{\beta})\psi(t)V \tag{6.58}$$

which generalizes the proportional hazard model introduced by Cox (1972). The reasoning behind the model (6.58) is that the function $\phi(\boldsymbol{x},\boldsymbol{\beta})$ is possibly subject to a specification error since there might be some covariates which have been ignored either due to unobservability or due to the ignorance of the underlying mechanism and this specification error may be measured by a positive multiplicative disturbance V. The function $\phi(\boldsymbol{x},\beta)$

is interpreted as the observed and V as the unobserved heterogeneity . The function ψ specifies the time dependence on the probability. If $\psi \equiv 1$, there is no time dependence on the probability. It is easy to see that the duration distribution is given by

$$G(t, \boldsymbol{x}, \boldsymbol{\beta}) = 1 - \int_0^\infty \exp[-\phi(\boldsymbol{x}, \boldsymbol{\beta}) Z(t) v] F(dv) \tag{6.59}$$

where $Z(t) = \int_0^t \psi(s) ds$ and F is the distribution function of the random variable V. Lancaster (1979) has given methods of estimation for $\boldsymbol{\beta}$ for given functional forms of ϕ and Z and a given distribution F of V. These methods of estimation are meaningful only when G identifies ϕ, Z and the distribution F of Y. Identifiability problems of this nature were investigated by Elbers and Ridder (1982) and Heckman and Singer (1984) under different conditions mainly on the random variable V and covariates $\boldsymbol{x}$. We now discuss their results briefly.

Identifiability when $E(V) < \infty$: Let $\{G(t, \boldsymbol{x}), \boldsymbol{x} \in S\}, S \subset R^k$ be a family of strictly increasing distribution functions represented by the relation

$$G(t, \boldsymbol{x}) = 1 - E\{e^{-\phi(\boldsymbol{x}) Z(t) V}\}, t \geq 0, \boldsymbol{x} \in S. \tag{6.60}$$

Here and in the following discussion, we suppress the parameter vector $\boldsymbol{\beta}$ in ϕ. Let us assume that the following conditions hold :

(A i) $V \geq 0$ and $E(V) < \infty$.

Without loss of generality, assume that

$$E(V) = 1 ; \tag{6.61}$$

$$\text{(A ii)} \qquad Z(t) = \int_0^t \psi(s) ds, \qquad t \geq 0 \tag{6.62}$$

where $\psi > 0$ and ψ is locally integrable .

(A iii) The function ϕ is positive, differentiable and nonconstant on R^k.

(A iv) S is open in R^k .

Since $G(t, \boldsymbol{x}) \to 1$ as $t \to +\infty$, it follows that $Z(t) \to \infty$ as $t \to \infty$ from (6.60). Furthermore $G(0, \boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in S$ since $Z(0) = 0$ from (6.62).

Let

$$M_V(s) = \int_0^\infty e^{sv} F(dv), s \leq 0. \tag{6.63}$$

Since the support of F is contained in $[0, \infty)$, it follows that the moment generating function $M_V(s)$ exists and is bounded between 0 and 1 for all $s \leq 0$. Since $E(V)$ is finite, it follows that M_V is differentiable on $(-\infty, 0)$ and infact

$$M_V^{(1)}(0) = E(V)$$

(cf. Feller (1966, p. 412)). Let us note that F is uniquely determined by M_V . Observe that

$$G(t, \boldsymbol{x}) = 1 - M_V(-Z(t)\phi(\boldsymbol{x})), t \geq 0. \tag{6.64}$$

Theorem 6.6.1 (Elbers and Ridder (1982)): Suppose differention under the integral sign with respect to t is permissible in (6.60) and the assumptions (A i) to (A iv) hold. Then G identifies (ϕ, Z, F) with the proviso that ϕ is identified up to translation by a constant.

Proof : It is easy to see that G is differentiable with respect to t under the hypothesis. Let $g(t, x)$ denote the derivative of G with respect to t. Then

$$g(t, \boldsymbol{x}) = \phi(\boldsymbol{x})\psi(t) \int_0^\infty v e^{-v\phi(\boldsymbol{x})Z(t)} F(dv), \boldsymbol{x} \in S, t \geq 0. \tag{6.65}$$

It is easy to check that $g(t, \boldsymbol{x}) > 0$ for all t and $\boldsymbol{x} \in S$. In particular, for a given $\boldsymbol{x}_0 \in S$,

$$\frac{g(t, \boldsymbol{x})}{g(t, \boldsymbol{x}_0)} = \frac{\phi(x)}{\phi(\boldsymbol{x}_0)} \frac{\int_0^\infty v e^{-v\phi(\boldsymbol{x})Z(t)} F(dv)}{\int_0^\infty v e^{-v\phi(\boldsymbol{x}_0)Z(t)} F(dv)} \tag{6.66}$$

for all $t \geq 0$ and $\boldsymbol{x} \in S$. Let $t \to 0$. By the bounded convergence theorem, it follows that

$$\int_0^\infty v e^{-v\phi(\boldsymbol{x})Z(t)} F(dv) \to \int_0^\infty vF(dv) < \infty \tag{6.67}$$

for every $\boldsymbol{x} \in S$. Hence

$$\lim_{t \to 0} \frac{g(t, x)}{g(t, \boldsymbol{x}_0)} = \frac{\phi(\boldsymbol{x})}{\phi(\boldsymbol{x}_0)}$$

or equivalently

$$\phi(\boldsymbol{x}) = \phi(\boldsymbol{x}_0) \lim_{t\to 0} \frac{g(t,x)}{g(t,\boldsymbol{x}_0)} \ . \tag{6.68}$$

This relation shows that the family of densities $\{g(t,\boldsymbol{x}), t \geq 0\}$ determine $\phi(\boldsymbol{x})$ up to a constant factor. In view of the relation (6.64), it follows that

$$1 - G(t,\boldsymbol{x}) = M_V(-Z(t)\phi(\boldsymbol{x})) \tag{6.69}$$

and hence

$$-Z(t)\phi(\boldsymbol{x}) = M_V^{-1}(1 - G(t,\boldsymbol{x}))$$

or equivalently

$$Z(t) = \frac{-M_V^{-1}(1 - G(t,\boldsymbol{x}))}{\phi(\boldsymbol{x})}, t \geq 0, \boldsymbol{x} \in S \ . \tag{6.70}$$

This relation defines the function $Z(t)$ provided $M_V^{-1}(\cdot)$ is well defined. Let

$$T(t,\boldsymbol{x}) = 1 - G(t,\boldsymbol{x}) \ . \tag{6.71}$$

Then

$$-Z(t) = \frac{M_V^{-1}(T(t,\boldsymbol{x}))}{\phi(\boldsymbol{x})} \ . \tag{6.72}$$

Note that the left side of the equation does not depend on $\boldsymbol{x}$. In particular, the partial derivatives with respect to $\boldsymbol{x}$ of the function on the right side are equal to zero for all $\boldsymbol{x}$ and

$$\phi(\boldsymbol{x}) \frac{\partial}{\partial x_i}[M_V^{-1}(T(t,\boldsymbol{x}))] - M_V^{-1}(T(t,\boldsymbol{x}))\frac{\partial \phi(\boldsymbol{x})}{\partial x_i} = 0, \boldsymbol{x} \in S, t \geq 0$$

or equivalently

$$\phi(\boldsymbol{x})\frac{\partial}{\partial s}[M_V^{-1}(T(t,\boldsymbol{x}))]\frac{\partial T(t,\boldsymbol{x})}{\partial x_i} - M_v^{-1}(T(t,\boldsymbol{x})) \frac{\partial \phi(\boldsymbol{x})}{\partial x_i} = 0, \boldsymbol{x} \in S, t \geq 0 \tag{6.73}$$

where $s = T(t,\boldsymbol{x}) = 1 - G(t,\boldsymbol{x})$ or equivalently $t = K(s,\boldsymbol{x})$. Note that $t \geq 0$ and $0 \leq s < 1$. Such an inverse map K exists since G is strictly increasing and differentiable in t. Then it follows that, for any fixed $i, 1 \leq i \leq k$,

$$\phi(\boldsymbol{x})\frac{\partial}{\partial s}[M_V^{-1}(s)] \ \frac{\partial T(t,x)}{\partial x_i}|_{(K(s,\boldsymbol{x}),\boldsymbol{x})} - M_V^{-1}(s) \ \frac{\partial \phi(\boldsymbol{x})}{\partial x_i} = 0, \boldsymbol{x} \in S, 0 \leq s < 1. \tag{6.74}$$

Solving this differential equation for M_V^{-1}, it follows that

$$M_V^{-1}(s) = C \exp\{\frac{\partial \log \phi(\boldsymbol{x})}{\partial x_i} \int_{1/2}^{s} \frac{1}{\frac{\partial T}{\partial x_i}|_{(K(u,\boldsymbol{x}),\boldsymbol{x})}} du\} \tag{6.75}$$

for some constant C. This proves that $M_V^{-1}(\cdot)$ is well defined by ϕ and G. We have already seen that ϕ is determined by G up to a constant factor. Suppose there exists another random variable W with a distribution F^* and another function $r(t)$ but with the same regression function $\phi(\boldsymbol{x})$ satisfying (A i) – (A iv) such that

$$1 - G(t, \boldsymbol{x}) = M_W(-r(t)\phi(\boldsymbol{x})). \tag{6.76}$$

Then, relation (6.74) implies that

$$M_W^{-1}(s) = C^* \exp\{\frac{\partial \log \phi(\boldsymbol{x})}{\partial x_i} \int_{1/2}^{s} \frac{1}{\frac{\partial T}{\partial x_i}|_{(K(u,\boldsymbol{x}),\boldsymbol{x})}} du\} \tag{6.77}$$

and hence, from (6.75) and (6.77), it follows that

$$M_W^{-1}(s) = \frac{C^*}{C} M_V^{-1}(s), 0 \leq s < 1 .$$

But

$$\frac{dM_V(u)}{du}|_{u=0} = E(V) = 1$$

and

$$\frac{dM_W(u)}{du}|_{u=0} = E(W) = 1 .$$

Hence

$$\frac{dM_V^{-1}(s)}{ds}|_{s=1} = 1 = \frac{dM_W^{-1}(s)}{ds}|_{s=1} .$$

This proves that $C^* = C$ and therefore

$$M_V^{-1}(s) = M_W^{-1}(s), \quad 0 \leq s < 1$$

or equivalently

$$M_V(u) = M_W(u), u \leq 0 . \tag{6.78}$$

Since V and W have supports on $[0, \infty)$, it follows that the distributions of V and W are the same from Feller (1966, p. 230). ■

Theorem 6.6.1 makes use of the fact that the distribution function $G(t, \boldsymbol{x})$ is absolutely continuous. However, Theorem 6.6.1 continues to hold even for discrete distributions $G(t, \boldsymbol{x})$ as the following arguments will show.

Theorem 6.6.2 (Elbers and Ridder (1982)): Suppose $G_1(t)$ and $G_2(t)$ are distribution functions such that

$$1 - G_i(t) = M_V(-\phi_i Z(t)), t \geq 0, i = 1, 2 \tag{6.79}$$

where $\phi_i > 0, i = 1, 2, \phi_1 \neq \phi_2, Z(t)$ nondecreasing continuous with $Z(0) = 0$ and $M_V(\cdot)$ is the moment generating function of a nonnegative random variable V. Further suppose that $E(V) = 1$ and M_V^{-1} is well defined. Then the numbers $\phi_i, i = 1, 2$, the function $Z(t)$ and the distribution of V are *uniquely* determined by $G_i, i = 1, 2$.

Proof : Without loss of generality, assume that $\phi_1 = 1$ and $\phi_2 < \phi_1$. Suppose that both the triples $(Z(t), V, \phi_2)$ and $(R(t), W, \psi_2)$ satisfy the relations (6.79). Let

$$L_V(s) = M_V(-s). \tag{6.80}$$

Then

$$L_V(Z(t)) = L_W(R(t)), t \geq 0 \tag{6.81}$$

and

$$L_V(\phi_2 Z(t)) = L_W(\psi_2 R(t)), t \geq 0. \tag{6.81A}$$

Therefore

$$\psi_2 R(t) = L_W^{-1}(L_V(\phi_2 Z(t))), t \geq 0 \tag{6.82}$$

and

$$\psi_2 R(t) = \psi_2 L_W^{-1}(L_V(Z(t)), t \geq 0. \tag{6.82A}$$

Hence

$$\psi_2 R(t) = L_W^{-1}(L_V(\phi_2 Z(t))) = \psi_2 L_W^{-1} L_V(Z(t)), t \geq 0 \ . \tag{6.83}$$

Let

$$f = L_W^{-1} o L_V \ . \tag{6.84}$$

Note that $R(t) \to \infty$ as $t \to \infty$ and $Z(t) \to \infty$ as $t \to \infty$ since G_1 and G_2 are distribution functions. Hence

$$f(t) \to \infty \text{ as } t \to \infty \tag{6.85}$$

and

$$f(\phi_2 s) = \psi_2 f(s), s \geq 0 . \tag{6.86}$$

Since $EV = EW = 1$, it follows that f is differentiable with respect to s and $f^{(1)}(0+) = 1$ where $f^{(1)}(s)$ denotes the derivative of $f(s)$ for $s > 0$ and

$$f^{(1)}(0+) = \lim_{s \downarrow 0} f^{(1)}(s) \tag{6.87}$$

(cf. Feller (1966, p. 412)). Let $s == \phi_2 s'$. Then

$$f(\phi_2^2 s') = \psi_2^2 f(s') \tag{6.88}$$

from (6.86). In general

$$f(\phi_2^n s) = \psi_2^n f(s), s \geq 0, n \geq 1 . \tag{6.89}$$

Differentiating with respect to s on both sides, it follows that

$$f^{(1)}(s) = \left(\frac{\phi_2}{\psi_2}\right)^n f^{(1)}(\phi_2^n s), \quad s \geq 0, n \geq 1 \ . \tag{6.90}$$

Since $0 < \phi_2 < 1$, taking limit as $n \to \infty$, we obtain that

$$\begin{aligned} f^{(1)}(s) &= f^{(1)}(0+) \lim_{n \to \infty} (\frac{\phi_2}{\psi_2})^n \\ &= \lim_{n \to \infty} (\frac{\phi_2}{\psi_2})^n, s \geq 0 . \end{aligned} \tag{6.91}$$

Let $s \downarrow 0$. Then it follows that

$$1 = f^{(1)}(0+) = \lim_{n \to \infty} (\frac{\phi_2}{\psi_2})^n . \tag{6.92}$$

The last relation holds iff $\phi_2 = \psi_2$ and hence $f^{(1)}(s) = 1, s \geq 0$ from (6.91). Since $f(0) = 1$, it follows that $f(s) = s$ for all s. In other words

$$(L_W^{-1} o L_V)(s) = s, s \geq 0 \ . \tag{6.93}$$

Therefore

$$L_V(s) = L_W(s), s \geq 0 \,. \tag{6.94}$$

Since L_V and L_W uniquely determine the distributions of V and W, it follows that the distributions of V and W are the same. Since $\phi_2 = \psi_2$ and $(L_W^{-1} o L_V)(s) = s$, it follows that $R(t) = Z(t)$ from equation (6.83). This completes the proof of Theorem 6.6.2. ■

Identifiability when $E(V) = \infty$: One of the major assumptions in Theorem 6.6.1 is that $E(V) < \infty$ where V is the positive multiplicative disturbance. There are examples of positive random variabes for which $E(V) = \infty$, for instance, if the density of V is given by

$$f(v) = \frac{2}{\pi(1+x^2)}, 0 < x < \infty \,. \tag{6.95}$$

Heckman and Singer (1984) have given alternate sufficient conditions for identifiability to take care of this situation. The condition $E(V) < \infty$ is replaced by a condition on the tail behaviour of the distribution of V. They assume that V has an absolutely continuous distribution with density f such that

$$f(v) \simeq \frac{c}{(\log v)^\delta v^{1+\varepsilon} L(v)} \text{ as } v \to \infty \tag{6.96}$$

where $c > 0, 0 < \varepsilon < 1, \delta \geq 0$ and $L(\cdot)$ is a slowly varying function in the sense that

$$\frac{L(vu)}{L(v)} \to 1 \text{ as } v \to \infty \text{ for } u > 0 \,. \tag{6.97}$$

Here ε is specified number in (0,1). If V is a discrete random variable having masses at $0 < v_0 < v_1 \cdots$ with jumps p_k at v_k, then it is assumed that

$$v_k \simeq ck \text{ and } p_k \simeq \frac{c}{(\log k)^\delta k^{1+\varepsilon} L(k)} \tag{6.98}$$

where $c > 0, \delta \geq 0, 0 < \varepsilon < 1$ (specified) and L is slowly varying. In addition to these conditions on the distribution of V, Heckman and Singer (1984) prove identifiability under additional conditions on $Z(t)$ and $\phi(\boldsymbol{x})$. We omit the details.

Identifiability in some parametric models when covariates are not present : Suppose a distribution function G satisfies the relation

$$G(t) = 1 - M_V[-Z(t)], t \geq 0 \tag{6.99}$$

or equivalently

$$g(t) = Z^{(1)}(t) \int_0^\infty ve^{-vZ(t)} F(dv), t \geq 0 \tag{6.100}$$

where $g(t)$ denotes the derivative of $G(t)$ and $Z^{(1)}(t)$ denotes the derivative of $Z(t)$. Let us suppose that $Z(t)$ belongs to a parametric family $Z(\alpha, t)$. Observe that $\phi(\boldsymbol{x}) \equiv 1$ in (6.99) when compared with relation (6.64). The presence of at least one covariates in the model is *essential* for the validity of results in Theorem 6.6.1. We prove now identifiability in a parametric model even when no covariate is present in the model. For a general discussion of such results, see Heckman and Singer (1984). The identifiability problem can be stated as follows : suppose

$$g_i(t) = Z^{(1)}(\alpha_i, t) \int_0^\infty ve^{-vZ(\alpha_i,t)} F_i(dv) \tag{6.101}$$

for $i = 1, 2$. If $g_0(t) = g_1(t)$ for all $t \geq 0$, can we conclude that $\alpha_0 = \alpha_1$ and $F_0 = F_1$?

We now discuss one such example due to Heckman and Singer (1984). Suppose

$$Z^{(1)}(\alpha, t) = \exp(\gamma(\frac{t^\lambda - 1}{\lambda})) \tag{6.102}$$

where $\alpha = (\gamma, \lambda), \lambda \neq 0$. This class of models is called the Box-Cox hazard models introduced by Flinn and Heckman (1982). If $\lambda = 1$, then the model reduces to the Gompertz hazard model. If $\gamma = 0$, then the model is exponential and, as $\lambda \to 0$, the model approaches the Weibull hazard model.

Proposition 6.6.1 (Heckman and Singer (1984)) : Suppose $E(V)$ is finite and $\lambda < 0$. Then $\alpha^* = (\gamma, \lambda, F)$ is uniquely determined by g defined by (6.101) whenever $\gamma \neq 0$. If $\gamma = 0$, then F is uniquely determined by g.

Proof : Suppose there exists $\alpha_i^* = (\gamma_i, \lambda_i, F_i), i = 0, 1$ such that

$$g_0(t) = g_1(t), t \geq 0$$

where g_i is as defined by the relations (6.101) and (6.102). Then

$$1 = \frac{g_1(t)}{g_0(t)} = \frac{\exp[\gamma_1(\frac{t^{\lambda_1}-1}{\lambda_1})] \int_0^\infty ve^{-Z(\alpha_1,t)v}F_1(dv)}{\exp[\gamma_0(\frac{t^{\lambda_0}-1}{\lambda_0})] \int_0^\infty ve^{-Z(\alpha_0,t)}F_0(dv)}, t \geq 0 . \tag{6.103}$$

Suppose $\gamma_0 \neq 0$ and $\lambda_0 < 0$. It can be checked that

$$\lim_{t\to 0} \frac{g_1(t)}{g_0(t)} = 0 \text{ or } \infty$$

whenever $\lambda_1 \neq \lambda_0$. This contradicts the relation (6.103). Hence $\lambda_1 = \lambda_0$. If $\gamma_1 \neq \gamma_0$, then

$$\lim_{t\to 0} \exp(\frac{t^{\lambda_0} - 1}{\lambda_0}(\gamma_1 - \gamma_0)) = \begin{cases} 0 & \text{if } \gamma_1 > \gamma_0 \\ \infty & \text{if } \gamma_1 < \gamma_0 \end{cases} \tag{6.104}$$

again contradicting (6.103) since

$$Z(\alpha_i, t) \to 0 \text{ as } t \to 0 \tag{6.105}$$

and

$$\int_0^\infty ve^{-Z(\alpha_i,t)v}F_i(dv) \to E_{F_i}(V) < \infty \text{ as } t \to 0 . \tag{6.106}$$

Hence $\gamma_1 = \gamma_0$. This proves that

$$\int_0^\infty ve^{-Z(\alpha,t)v}F_0(dv) = \int_0^\infty ve^{-Z(\alpha,t)v}F_1(dv), t \geq 0 \tag{6.107}$$

where $\alpha_0 = \alpha_1 = \alpha$. Since $Z(\alpha, t)$ is a continuous function taking all values between $[0, \infty)$, it follows that the Laplace of transform of F_i^* defined by

$$F_i^*(dv) = vF_i(dv), i = 0, 1 \tag{6.108}$$

is identical. Since $F_i^*, i = 0, 1$ with supports on $[0, \infty)$ are uniquely determined by their Laplace transforms, it follows that F_0 and F_1 are identical being distribution functions.

If $\gamma_0 = 0$, then $Z^{(1)}(\alpha, t) \equiv 1$ for all t and the result is a consequence of uniqueness of Laplace of transforms. This completes the proof. ■

For more examples, see Heckman and Singer (1984).

Chapter 7

Identifiability in Reliability and Survival Analysis

7.1 Introduction

In the previous chapter, we have seen problems of identifiability in many stochastic models encountered in econometric modeling. As we have pointed out earlier, the notion of estimability of a parameter in a model is meaningful only when the parameter is identifiable in the model. Recall that a parameter $\theta \in \Theta$ is nonidentifiable by a random vector $\boldsymbol{Y}$ if there is at least one pair $(\theta, \theta'), \theta \neq \theta'$ in Θ such that the distributions of $\boldsymbol{Y}$ are the same under both θ and θ'. This type of identifiability may be termed as parametric identifiability. Suppose the class of distribution functions under consideration are *not* indexed by a parameter. Then we have the problem of identifiability in a nonparametric framework. Problems of this type occur in reliability as well as survival analysis. Let us discuss such problems.

An individual may be subject to two causes of death (or two types of terminal illness). Let X_i represent the lifetime of the individual exposed to cause i (or disease i) *alone* . $X_i, i = 1, 2$ are not observable in practice and $Y = \min(X_1, X_2)$ is observable. Does the distribution of Y identify the distributions of X_1 and X_2? Consider a 2-components system when the components $i = 1, 2$ are connected in series. Let X_i be the lifetime of the ith component. Suppose the system fails if at least one of the components fails and one can observe only $Y = \min(X_1, X_2)$ the lifetime of the system. Does the distribution of Y identify the distributions of X_1 and X_2 ? Let X_1 and X_2 be the demand and supply for a commodity at a given price p. Then the amount that is transacted in the market is $Y = \min(X_1, X_2)$. Does the distribution of Y determine the distributions of X_1 and X_2 ? Such problems are termed the problems of *competing risks* in the literature on reliability and survivial analysis. Associated with the problems of competing risks is the dual problem of *complementary risks* (Basu and Ghosh (1980)). Let us again consider a 2-component system connected in parallel. Let X_1 and X_2 be the lifetimes of the two components. The system life $Z = \max(X_1, X_2)$ is observable. There are examples where X_1 and X_2 are not individually observable but Z is, for instance, the flight of a twin engine aircraft or a satellite etc. Another example is the failure of internal body organs like kidneys : exact time of failure of each kidney may not be known but when both kidneys fail to function, the time to death can be recorded. The problem again is to find whether the distributions of the components X_1 and X_2 are identifiable when the distribution of Z is observable.

For a survey of identifiability results in problem of this nature, see Basu (1981), Puri (1979) and Birnbaum (1979).

Let us consider a specific example. Suppose X_1 and X_2 are independent random variables with distribution functions F_1 and F_2 respectively where

$$\begin{aligned} F_i(x) &= 1 - e^{-\lambda_i x}, \quad x > 0 \\ &= 0, \qquad\qquad x \leq 0 \end{aligned}$$

for $i = 1, 2$ where $\lambda_i > 0, i = 1, 2$. It is easy to see that $Y = \min(X_1, X_2)$ has the exponential distribution F_Y given by

$$\begin{aligned} F_Y(y) &= 1 - e^{-(\lambda_1+\lambda_2)y}, && y > 0 \\ &= 0, && y \leq 0. \end{aligned}$$

Note that the parameters λ_1 and λ_2 in F_1 and F_2 respectively are not identifiable from the distribution F_Y of Y since there are infinite number of pairs (λ_1, λ_2) leading to the same value of $\lambda = \lambda_1 + \lambda_2$.

Even though the problem discussed above leads to nonidentifiability, it is sometimes possible to rectify the problem by observing another random variable. We will discuss this later in this chapter.

7.2 Identifiability

Let us recall the definition of identifiability given in Chapter 6.

Let $\boldsymbol{Y}$ be an observable random vector with distribution function $F_\theta \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$, a family of distribution functions indexed by a parameter $\theta \in \Theta$. θ is said to be *nonindentifiable* if there is at least one pair $(\theta, \theta'), \theta \neq \theta', \theta, \theta' \in \Theta$ such that $F_\theta(\boldsymbol{y}) = F_{\theta'}(\boldsymbol{y})$ for all $\boldsymbol{y}$. Otherwise θ is said to be *identifiable.*

Suppose θ itself is not identifiable but there exists a function $\gamma(\theta)$ (non-constant) which is identifiable, that is , for any θ, θ' in $\Theta, F_\theta(y) = F_{\theta'}(y)$ for all y implies that $\gamma(\theta') = \gamma(\theta)$. Then θ is said to be *partially identifiable* and γ is said be *identifiable.*

Suppose θ is not identifiable but an additional random variable I can be introduced such that the joint distribution of $(\boldsymbol{Y}, I)$ identifies θ. Then the identifiability problem is said to be *rectifiable.*

From the definition identifiability of a parametric function $\gamma(\theta)$ of θ, it follows that $\gamma(\theta)$ is identifiable iff different points in the range of γ correspond to different F in $\mathcal{F}$, or equivalently, iff γ coincides with a function α

on $\mathcal{F}$ such that

$$\gamma(\theta) = \alpha(F_\theta) .$$

It is easy to see that every function $\psi(\cdot)$ of an identifiable parametric function $\gamma(\theta)$ is identifiable and a vectorial function is identifiable iff all its components are identifiable.

For an extensive discussion on identification in statistical inference, see Van der Genugten (1977).

7.3 Identifiability in the Problem of Competing or Complementary Risks (Independent Case)

Suppose $X_1, X_2, ..., X_k$ are independent random variables with continuous distribution functions $F_1, F_2, ..., F_k$ respectively. Let

$$Y = \min(X_1, X_2, ..., X_k) . \tag{7.0}$$

It is clear that the distribution function of Y is given by

$$F_Y(y) = 1 - \prod_{i=1}^{k}(1 - F_i(y)), \quad -\infty < y < \infty. \tag{7.1}$$

If $X_i, 1 \leq i \leq k$ are i.i.d. random variables, then $F_i(y) = F(y)$ for $1 \leq i \leq n$ for some distribution function F and hence

$$F_Y(y) = 1 - (1 - F(y))^k, \; -\infty < y < \infty . \tag{7.2}$$

It is obvious that the distribution function $F_Y(\cdot)$ determines $F(\cdot)$ uniquely. In fact

$$F(y) = 1 - [1 - F_Y(y)]^{1/k}, \quad -\infty < y < \infty . \tag{7.3}$$

If $X_i, 1 \leq i \leq k$ are independent but not identically distributed, then the distribution functions $F_i, 1 \leq i \leq k$ may not be uniquely determined from F_Y using the equation (7.1). In other words, the individual distribution functions $F_i, 1 \leq i \leq k$ may not be identifiable from the distribution function F_Y. However, it is easy to check that identifiability holds for k-out of-p

identical component systems. For a discussion of identifiability for k-out of-p systems, see Section 7.7.

It is clear that $\Pr(X_i = X_j) = 0$ for all $i \neq j$. Let I be the random index j for which $Y = X_j$.

Theorem 7.3.1 (Berman (1963)): The joint distribution of (Y, I) *uniquely* determines the distribution functions $F_i, 1 \leq i \leq k$.

Proof : Let

$$H_j(x) = \Pr[Y \leq x, I = j] \,. \tag{7.4}$$

Then, for x such that $F_j(x) < 1$,

$$\begin{aligned}
H_j(x) &= \text{Probability that } X_j \text{ is the minimum} \\
&\quad \text{among } X_1, \ldots, X_n \text{ and } X_j \text{ is less than or equal to } x \\
&= \int_{-\infty}^{x} \prod_{\substack{i \neq j \\ 1 \leq i \leq k}} (1 - F_i(t)) dF_j(t) \\
&= \int_{-\infty}^{x} \frac{\prod_{i=1}^{k}(1 - F_i(t))}{1 - F_j(t)} \, dF_j(t) \\
&= -\int_{-\infty}^{x} \{1 - \sum_{i=1}^{k} H_i(t)\} d \log[1 - F_j(t)] \qquad (7.5)
\end{aligned}$$

since

$$\begin{aligned}
1 - \sum_{i=1}^{k} H_i(t) &= 1 - \Pr(Y \leq t) \\
&= \Pr(Y > t) \\
&= \prod_{i=1}^{k}(1 - F_i(t)) \,.
\end{aligned}$$

Therefore

$$dH_j(x) = -[1 - \sum_{i=1}^{k} H_i(x)] d \log[1 - F_j(x)] \tag{7.6}$$

or equivalently

$$F_j(x) = 1 - \exp\{-\int_{-\infty}^{x}(1 - \sum_{i=1}^{k} H_i(t))^{-1} dH_j(x)\}, 1 \leq j \leq k \ . \qquad (7.7)$$

This proves that the distribution functions $F_j(x), 1 \leq j \leq k$ are determined uniquely by the class of functions $H_i(x), 1 \leq i \leq k$. In other words, the joint distribution function of (Y, I) identifies the distribution functions F_i, $1 \leq i \leq k$. ■

Remarks 7.3.1 : The random variable $Y = \min(X_1, \ldots X_k)$ is called *nonidentified minimum* and the random vector (Y, I) is said to be *identified minimum.* Since $\max(X_1, \ldots, X_k)$ is the same as $-\min(-X_1, \ldots, -X_k)$, it follows that the distribution of the identified maximum (Z, J), where $Z = \max(X_1, \ldots, X_k)$ and J is the random index j for which $Z = X_j$, *uniquely* determines the distribution functions $F_j, 1 \leq j \leq k$.

Let us now suppose that the extrema Z or Y do not identify the distribution functions $F_i, 1 \leq i \leq k$. We now give some sufficient conditions on the family $\{F_i\}$ for identifiability.

Theorem 7.3.2 (Anderson and Ghurye (1977)) : Let $\mathcal{F}$ be a family of density functions f on the real line which are continuous and positive to the right of some point α and such that if f and g belong to $\mathcal{F}$, then $\lim_{x\to\infty}[f(x)/g(x)]$ exists and is either 0 or ∞. Suppose $X_1, ..., X_k$ are independent random variables with densities $f_1, ..., f_k$ respectively in $\mathcal{F}$ and $W_1, W_2, ..., W_\ell$ are independent random variables with densities $g_1, g_2, ..., g_\ell$ respectively in $\mathcal{F}$. Further suppose that $\max(X_1, ..., X_k)$ and $\max(W_1, ..., W_\ell)$ are identically distributed. Then $k = \ell$ and there exists a permutation $\{i_1, ..., i_k\}$ of $\{1, 2, ..., k\}$ such that the probability density function of W_j is $f_{i_j}, 1 \leq j \leq k$.

Proof : Let $F_i(x)$ be the distribution function of X_i and $G_j(x)$ be the distribution function of W_j. By hypothesis, we have

$$\Pr[\max_{1\le i\le k} X_i \le x] = \Pr[\max_{1\le j\le \ell} W_j \le x],\ -\infty < x < \infty\,.$$

Independence of $\{X_i\}$ and independence of $\{W_j\}$ imply that

$$\prod_{i=1}^{k} F_i(x) = \prod_{j=1}^{\ell} G_j(x),\ -\infty < x < \infty. \tag{7.8}$$

Hence, for all $x > \alpha$,

$$\sum_{i=1}^{k} \log F_i(x) = \sum_{j=1}^{\ell} \log G_j(x)\,.$$

Differentiating with respect to x on both sides, we have

$$\sum_{i=1}^{k} \frac{f_i(x)}{F_i(x)} = \sum_{j=1}^{\ell} \frac{g_j(x)}{G_j(x)}, \alpha < x < \infty\,. \tag{7.9}$$

By a change in the notation, we can rewrite the equation (7.9) in the form

$$\sum_{i=1}^{k+\ell} a_i \frac{f_i(x)}{F_i(x)} = 0, \alpha < x < \infty \tag{7.10}$$

where

$$a_i = +1, 1 \le i \le k;\ a_{k+j} = -1, 1 \le j \le \ell;$$

and

$$f_{k+j} = g_j, F_{k+j} = G_j, 1 \le j \le \ell\,.$$

Suppose there exists a density function f_i (say) f_1 among $f_i, 1 \le i \le k+1$ such that

$$\lim_{x\to\infty} \frac{f_r(x)}{f_1(x)} = 0 \text{ or } 1$$

for $1 \le r \le k+\ell$. Let

$$\mathcal{N} = \{i : \frac{f_i(x)}{f_1(x)} \to 1 \text{ as } x \to \infty\}\,. \tag{7.11}$$

Dividing both sides of the relation (7.10) by $f_1(x)$ and allowing $x \to \infty$, we have

$$\Sigma^* a_i = 0 \tag{7.12}$$

where Σ^* denotes the sum over $i \in \mathcal{N}$. Since $\Sigma^* a_i = 0$ and a_i is either +1 or -1 , it follows that $\mathcal{N}$ contains an even number of elements and half of these are from $\{1, 2, ..., k\}$. Hence a certain number of f_i in (7.9) are identical to one another and to the *same* number of g_i . Observe that if $i \in \mathcal{N}$, then $f_i(x) = f_1(x)$ for all x, for if $f_i(x) \neq f_1(x)$ for some x, then $\lim_{x \to +\infty} \frac{f_i(x)}{f_1(x)} = 0$ or ∞ by hypothesis contradicting the definition of $\mathcal{N}$. Subtracting these identical terms from both sides of (7.9), we have a new equation of the same form but with fewer terms. Repeat the process until each term on one side of (7.9) is matched with a term from the other side of (7.9). If $k = \ell$, then the theorem is proved. If $k \neq \ell$, (say), $k < \ell$, then $\ell - k$ of g_i are such that $g_i(x) = 0$ for $x > \alpha$ contradicting the assumption on $\mathcal{F}$. Hence $k = \ell$ and $\{f_1, ..., f_k\}$ is a permutation of $\{g_1, ..., g_k\}$. ∎

A result analogous to Theorem 7.3.2 can be proved for the case of minima of sets of random variables. We omit the proof but state the result.

Theorem 7.3.3 (Basu and Ghosh (1980)): Let $\mathcal{F}$ be a family of probability density functions on the real line with support $(a, b), -\infty \leq a < b \leq \infty$ which are continuous and are positive to the left of some point α and such that if f and g are any two distinct members of $\mathcal{F}$, then $\lim_{x \to a}(f(x)/g(x))$ exists and is equal to either 0 or ∞. Let $X_1, X_2, ..., X_k$ be independent random variables with density functions $f_1, f_2, ..., f_k$ respectively in $\mathcal{F}$ and $W_1, W_2, ..., W_\ell$ be independent random variables with density functions in $\mathcal{F}$. Suppose that $\min(X_1, ..., X_k)$ and $\min(W_1, ..., W_\ell)$ have identical distributions. Then $k = \ell$ and there exists a permutation $\{i_1, ..., i_k\}$ of $\{1, 2, ..., \ell\}$ such that the density function of W_j is $f_{i_j}, j = 1, 2, ..., \ell$.

Example 7.3.1 (Anderson and Ghurye (1977)): Consider the family $\mathcal{F}$ of normal density functions

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}, -\infty < \mu < \infty, 0 < \sigma < \infty \,.$$

The family $\mathcal{F}$ satisfies the conditions of Theorem 7.3.2 and Theorem 7.3.3. In fact

$$\lim_{x \to \pm\infty} \frac{f(x;\mu_2,\sigma_2^2)}{f(x;\mu_1,\sigma_1^2)} = \begin{cases} 0 & \text{if} \quad \sigma_1^2 > \sigma_2^2 \text{ or if } \sigma_1^2 = \sigma_2^2, \mu_1 > \mu_2 \\ \infty & \text{if} \quad \sigma_1^2 < \sigma_2^2 \text{ or if } \sigma_1^2 = \sigma_2^2, \mu_1 < \mu_2 \\ 1 & \text{if} \quad \sigma_1^2 = \sigma_2^2, \mu_1 = \mu_2 \quad . \end{cases}$$

Hence the conclusions of the Theorems 7.3.2 and 7.3.3 hold and the family $\mathcal{F}$ is identified by the minimum or by the maximum up to a permutation.

Example 7.3.2 : Consider the family $\mathcal{F}$ of exponential densities

$$\begin{aligned} f(x;\lambda) &= \lambda e^{-\lambda x}, \quad x > 0 \\ &= 0, \qquad\quad x \leq 0 \end{aligned}$$

where $\lambda > 0$. This family $\mathcal{F}$ satisfies the conditions stated in Theorem 7.3.2. Hence the conclusion of Theorem 7.3.2 holds and the family $\mathcal{F}$ is identified by the maximum up to a permutation.

Remarks 7.3.2: There are families of densities for which the assumptions in Theorem 7.3.2 do not hold and yet they are identified by the maximum up to a permutation . This can be seen by the following examples.

Example 7.3.3 (Anderson and Ghurye (1977)) : Consider the family $\mathcal{F}$ of exponential densities

$$\begin{aligned} f(x,\theta) &= e^{-(x-\theta)}, \quad x > \theta \\ &= 0, \qquad\qquad x \leq \theta \end{aligned}$$

where $-\infty < \theta < \infty$. It is easy to check that the family $\mathcal{F}$ does not satisfy the conditions stated in Theorem 7.3.2. In fact, for $0 < \theta_1 < \theta_2$,

$$\frac{f(x,\theta_2)}{f(x,\theta_1)} = \frac{e^{-(x-\theta_2)}}{e^{-(x-\theta_1)}} = e^{\theta_2-\theta_1}$$

for $x > \theta_2$ and hence

$$\lim_{x\to\infty} \frac{f(x,\theta_2)}{f(x,\theta_1)} = e^{\theta_2-\theta_1}$$

which is neither zero nor infinity. Let $X_1, ..., X_K$ be independent random variables with exponential densities $f(x,\theta_i), 1 \leq i \leq k$ respectively. Let $W_1, ..., W_\ell$ be independent random variables with exponential densities $f(x,\theta'_j), 1 \leq j \leq \ell$. Suppose that $\max_{1\leq i\leq k} X_i$ and $\max_{1\leq j\leq \ell} W_j$ are identically distributed. It is easy to see from the structure of the distribution functions of $\max_{1\leq i\leq k} X_i$ and $\max_{1\leq j\leq \ell} W_j$ and the fact that they are identically distributed that

$$\max_{1\leq i\leq k} \theta_i = \max_{1\leq j\leq \ell} \theta'_j\ .$$

Since the distribution functions of $\max_{1\leq i\leq k} X_i$ and $\max_{1\leq j\leq \ell} W_j$ are the same for values of x between the largest of θ'_i's and the second largest of distinct $\theta'_i s$, it follows that the number of $\theta_i, 1 \leq i \leq k$ equal to the largest of θ_i's is the same as the number of $\theta'_j = \max_{1\leq i\leq k} \theta_i (= \max_{1\leq j\leq \ell} \theta'_j)$ and the second largest θ_i is equal to the second largest θ'_j. Proceeding this way, we obtain that $k = \ell$ and $\{\theta_1, ..., \theta_k\}$ is a permutation of $\{\theta'_1, ..., \theta'_\ell\}$.

Example 7.3.4 (Anderson and Ghurye (1977)) : Consider the family $\mathcal{F}$ of double exponential densities

$$f(x,\theta) = \frac{1}{2}e^{-|x-\theta|}, -\infty < x < \infty$$

where $-\infty < \theta < \infty$. It is again easy to check that this family does not satisfy the conditions stated in Theorem 7.3.2 and yet it is identifiable by the maximum up to a permutation. This can be seen in the following way. Suppose $X_i, 1 \leq i \leq k$ are independent random variables with densities $f(x,\theta_i), 1 \leq i \leq k$ respectively and $W_j, 1 \leq j \leq \ell$ are independent random variables with densities $f(x,\theta'_j), 1 \leq j \leq \ell$ respectively. Further suppose that

$$\max_{1\leq i\leq k} X_i \text{ and } \max_{1\leq j\leq \ell} W_j$$

are identically distributed. Then, it is easy to check that

$$\prod_{i=1}^{k} |1 - \frac{1}{2}e^{-(x-\theta_i)}| = \prod_{j=1}^{\ell} |1 - \frac{1}{2}e^{-(x-\theta_j')}|$$

for large x. Without loss of generality, assume that $k \leq \ell$. Let $z = e^x$. Multiplying both sides by z^ℓ, we have

$$z^{\ell-k} \prod_{i=1}^{k} (z - \frac{1}{2}e^{\theta_i}) = \prod_{j=1}^{\ell} (z - \frac{1}{2}e^{\theta_j'}) .$$

Since this equality holds for all $z = e^x, x$ large, it follows that the zeroes of polynomials in z on both sides should be the same. Hence $k = \ell$ and $\{\theta_i, 1 \leq i \leq k\} = \{\theta_j', 1 \leq j \leq \ell\}$.

Example 7.3.5 : Suppose a random variable X_i has the gamma density function

$$\begin{aligned} f(x; \alpha_i, \beta_i) &= \frac{e^{-x/\beta_i} x^{\alpha_i - 1}}{\beta^{\alpha_i} \Gamma(\alpha_i)}, x > 0 \\ &= 0, \qquad\qquad x \leq 0 \end{aligned}$$

where $\alpha_i > 0, \beta_i > 0$ and at least one of α_i and β_i is different from unity. Suppose $X_i, 1 \leq i \leq 4$ are independent random variables and the random variables $\min(X_1, X_2)$ and $\min(X_3, X_4)$ are identically distributed. Then, Basu and Ghosh (1980) proved that either

$$(\alpha_1, \alpha_2) = (\alpha_3, \alpha_4) \text{ and } (\beta_1, \beta_2) = (\beta_3, \beta_4)$$

or

$$(\alpha_1, \alpha_2) = (\alpha_4, \alpha_3) \text{ and } (\beta_1, \beta_2) = (\beta_4, \beta_3).$$

However the family $\mathcal{F}$ of density functions $\{f(x; \alpha, \beta)\}$ does not satisfy the conditions of the theorem. We omit the details.

Example 7.3.6 : Suppose X_i has the Weibull density

$$\begin{aligned} f(x; p_i, \theta_i) &= \tfrac{p_i}{\theta_i} x^{p_i - 1} e^{-x^{p_i}/\theta_i}, \quad x > 0 \\ &= 0, \qquad\qquad x \leq 0 \end{aligned}$$

where $\theta_i > 0$ and $p_i > 0$. It is easy to check that the distribution function of X_i is

$$\begin{aligned} F(x; p_i, \theta_i) &= 1 - e^{-x^{p_i}/\theta_i}, \quad x > 0 \\ &= 0, \qquad\qquad\qquad x \leq 0 . \end{aligned}$$

Suppose $X_i, 1 \leq i \leq 4$ are independent. We leave it to the reader to check that the above family of densities does not satisfy the conditions stated in Theorem 7.3.3 (cf. Basu and Ghosh (1980)). Suppose the distribution of $\min(X_1, X_2)$ is the same as that of $\min(X_3, X_4)$. Then

$$(1 - F_1(x))(1 - F_2(x)) = (1 - F_3(x))(1 - F_4(x)), -\infty < x < \infty .$$

Taking logarithms on both sides, it follows that

$$\frac{x^{p_1}}{\theta_1} + \frac{x^{p_2}}{\theta_2} = \frac{x^{p_3}}{\theta_3} + \frac{x^{p_4}}{\theta_4}, x > 0 .$$

Suppose $p_1 \neq p_2$. Without loss of generality, assume that $p_1 < p_2$. Taking limits as $x \to 0$ and $x \to \infty$ in the above relation, it can be shown that $p_1 = \min(p_3, p_4)$ and $p_2 = \max(p_3, p_4)$. Since $p_1 \neq p_2$, it follows that $p_3 \neq p_4$. Suppose $p_3 < p_4$. Then it can be checked that $p_1 = p_3$ and $p_2 = p_4$. It is easy to see that $\theta_1 = \theta_3$ and $\theta_2 = \theta_4$ by the linear independence of the family $\{x^p, p > 0\}$. Hence

$$(p_1, \theta_1) = (p_3, \theta_3) \text{ and } (p_2, \theta_2) = (p_4, \theta_4).$$

If $p_1 > p_2$, then it can be shown by similar arguments that

$$(p_1, \theta_1) = (p_4, \theta_4) \text{ and } (p_2, \theta_2) = (p_3, \theta_3).$$

This shows that the Weibull family is identifiable up to a permutation.

Remarks 7.3.3 : We remark that even though several examples given above illustrate families of densities identified by the maximum or minimum up to permutation, there exist families which are not identifiable by the maximum as shown by the following example .

Example 7.3.7 : Suppose a random variable X_i has the exponential density

$$\begin{aligned} f(x, \lambda_i) &= \lambda_i e^{-\lambda_i x}, x > 0 \\ &= 0, \qquad x \leq 0 \end{aligned}$$

for $1 \leq i \leq n$. Suppose $X_i, 1 \leq i \leq n$ are independent. $Y = \max_{1 \leq i \leq n} X_i$. Then Y has the exponential density

$$\begin{aligned} f(y, \lambda) &= \lambda e^{-\lambda y}, y > 0 \\ &= 0, \qquad y \leq 0 \end{aligned}$$

where $\lambda = \sum_{i=1}^{n} \lambda_i$. Hence the distribution of Y specifies $\sum_{i=1}^{n} \lambda_i$ but not the individual $\lambda_i, 1 \leq i \leq n$. This does not contradict the conclusion in Example 7.3.2 where two independent samples were considered.

Let us again consider the problem studied in Theorem 7.3.2. This can be stated as follows: If $F_1 F_2 \cdots F_k = G_1 G_2 \cdots G_\ell$ where F_i and G_j are univariate distribution functions, then, is $k = \ell$ and is $\{F_1, F_2, ..., F_k\}$ a permutation of $\{G_1, G_2, ..., G_\ell\}$ when $k = \ell$? Let us consider a special case of this problem again when

$$F_i(x) = F(a_i x) \text{ and } G_j(x) = F(b_j x)$$

where F is a distribution function and $a_i, b_j, 1 \leq i \leq k, 1 \leq j \leq \ell$ are real numbers. Define

$$\prod_{i=1}^{k} F(a_i x) = \prod_{j=1}^{\ell} F(b_j x), -\infty < x < \infty . \tag{7.13}$$

Note that a_i and b_j are necessarily positive since $F_i(x) = F(a_i x)$ and $G_j(x) = G(b_j x)$ are distribution functions by assumption. The question is to find out whether $k = \ell$ and $\{a_1, a_2, ..., a_k\}$ is a permutation of $\{b_1, b_2, ..., b_\ell\}$ under some conditions.

(A1) Suppose the function

$$g(x) = \frac{F'(x)}{F(x)} \tag{7.14}$$

where F' is the derivative of F can be expanded in an *infinite* power series about zero so that

$$g(x) = \sum_{n=0}^{\infty} x^n \frac{g^{(n)}(0)}{n!}, -\alpha < x < \alpha \tag{7.15}$$

where $g^{(n)}(0)$ is the nth derivative of $g(\cdot)$ evaluated at 0 and $0 < \alpha \leq \infty$. Taking logarithms on both sides of (7.13) and differentiating with respect to x, we have

$$\sum_{i=1}^{k} a_i \frac{F'(a_i x)}{F(a_i x)} = \sum_{j=1}^{\ell} b_j \frac{F'(b_j x)}{F(b_j x)}, -\alpha < x < \alpha \,. \tag{7.16}$$

Under the assumption (A1) on $g(x)$ stated above, it follows that

$$\sum_{i=1}^{k} (\sum_{s=0}^{\infty} a_i^{s+1} x^s \frac{g^{(s)}(0)}{s!}) = \sum_{j=1}^{\ell} (\sum_{s=0}^{\infty} b_j^{s+1} x^s \frac{g^{(s)}(0)}{s!}), -\alpha < x < \alpha \tag{7.17}$$

or equivalently, for $-\alpha < x < \alpha$,

$$\sum_{s=0}^{\infty} (\sum_{i=1}^{k} a_i^{s+1}) x^s \frac{g^{(s)}(0)}{s!} = \sum_{s=0}^{\infty} (\sum_{j=1}^{\ell} b_j^{s+1}) x^s \frac{g^{(s)}(0)}{s!} \tag{7.18}$$

by Fubini's theorem under the additonal assumption that the series
(A2)

$$\sum_{s=0}^{\infty} \sum_{i=1}^{k} a_i^{s+1} x^s \frac{g^{(s)}(0)}{s!}$$

and

$$\sum_{s=0}^{\infty} \sum_{j=1}^{\ell} b_j^{s+1} x^s \frac{g^{(s)}(0)}{s!}$$

are absolutely convergent for every x in $-\alpha < x < \alpha$. Equation (7.18) implies that

$$(\sum_{i=1}^{k} a_i^{s+1}) \frac{g^{(s)}(0)}{s!} = (\sum_{j=1}^{\ell} b_j^{s+1}) \frac{g^{(s)}(0)}{s!} \tag{7.19}$$

for all integers $s \geq 0$ since the power series are identical in $-\alpha < x < \alpha$, where $\alpha \geq 0$. In particular, it follows that

$$\sum_{i=1}^{k} a_i^{s+1} = \sum_{j=1}^{\ell} b_j^{s+1}, s \geq 0 \tag{7.20}$$

since $g^{(s)}(a) \neq 0$ for sufficiently large $s \geq 0$ which in turn follows from the fact that the function $g(x)$ has *infinite* power series expansion by **(A1)**. Note that $a_i > 0, b_j > 0$ for $1 \leq i \leq k, 1 \leq j \leq \ell$. Let $s \to \infty$. Then

$$(\sum_{i=1}^{k} a_i^s)^{1/s} \to \max_{1 \leq i \leq k} a_i \text{ as } s \to \infty \tag{7.21}$$

and

$$(\sum_{j=1}^{\ell} b_j^s)^{1/s} \to \max_{1 \leq j \leq \ell} b_j \text{ as } s \to \infty .$$

This proves that

$$\max_{1 \leq i \leq k} a_i = \max_{1 \leq j \leq \ell} b_j . \tag{7.22}$$

Delete the maximum terms on both sides of (7.20) and then let $s \to \infty$. Repeat the procedure. Then it follows that $k = \ell$ and $\{a_i, 1 \leq i \leq k\}$ is a permutation of $\{b_j, 1 \leq j \leq \ell\}$ in view of the fact that $a_i > 0$ and $b_j > 0, 1 \leq i \leq k, 1 \leq j \leq \ell$. The above discussion leads to the following theorem due to Mukherjea *et al.* (1986).

Theorem 7.3.4: (Mukherjea *et al* . (1986)) : Let $F(x)$ be a distribution function and a_i and b_j be positive numbers such that

$$\prod_{i=1}^{k} F(a_i x) = \prod_{j=1}^{\ell} F(b_j x), -\alpha < x < \alpha$$

for some $\alpha > 0$. Suppose that function $g(x) = \frac{F'(x)}{F(x)}$ satisfies the assumption **(A1)** and the assumption **(A2)** holds for $g(\cdot), \{a_i\}$ and $\{b_j\}$. Then $k = \ell$ and $\{a_1, ..., a_k\}$ is a permutation of $\{b_1, b_2, ..., b_\ell\}$.

Remarks 7.3.4 (Mukherjea *et al.* (1986)): gives another set of sufficient conditions for the conclusion in Theorem 7.3.4 to hold. For general parametric families, the following result due to Basu and Ghosh (1983) holds. We omit the proof.

Theorem 7.3.5 (Basu and Ghosh (1983)): Let $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$ be a family of distribution functions with failure rate functions $\lambda(x,\theta)$, that is

$$\lambda(x,\theta) = \frac{f(x,\theta)}{1 - F(x,\theta)}$$

for all x such that $F(x,\theta) < 1$ where $f(x,\theta)$ is the density function of $F(x,\theta)$. Suppose $X_i, 1 \leq i \leq k$ are k independent random variables with distribution functions $F(x,\theta_i)$ and failure rate functions $\lambda(x,\theta_i)$ for $1 \leq i \leq k$. Then $Z = \min(X_1, ..., X_k)$ identifies $\theta_i, 1 \leq i \leq k$ up to a permutation iff $\lambda(x,\theta_i), 1 \leq i \leq k$ are linearly independent .

Remarks 7.3.5 : An example of a family of distributions satisfying the conditions stated in Theorem 7.3.5 is the Weibull family discussed in Example 7.3.6.

7.4 Identifiability in the Dependent Case

Let $X_1, X_2, ..., X_k$ be k random variables with joint distribution function $F(x_1, x_2, ..., x_k)$. Let $Z = \min(X_1, ..., X_k)$ and $I = i$ if $Z = X_i$. In the previous section, we have discussed the identifiability problem, that is, identifying the distribution of $X_i, 1 \leq i \leq k$ given the distribution of Z or that of the identified minimum (Z, I) when $X_1, X_2, ..., X_k$ are independent random variables. There are physical situations where $X_i, 1 \leq i \leq k$ might not be independent. The problem of interest is to know whether Z or (Z, I) still identifies the joint distribution function of $(X_1, ..., X_k)$.

Let us first consider the case $k = 2$. Suppose (X_1, X_2) has the joint distribution $F(x_1, x_2)$ and the joint density $f(x_1, x_2) > 0$ for all x_1 and x_2. Let

$$\bar{F}(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2) \tag{7.23}$$

and

$$\bar{F}_i(x_1, x_2) = \frac{\partial \bar{F}(x_1, x_2)}{\partial x_i}, i = 1, 2 \, . \tag{7.24}$$

Define

$$\bar{G}_i(x) = \exp\{-\int_{-\infty}^{x} \frac{-\bar{F}_i(z,z)}{\bar{F}(z,z)} dz\}, -\infty < x < \infty \tag{7.25}$$

and suppose that

$$\int_{-\infty}^{\infty} \frac{-\bar{F}_i(z,z)}{F(z,z)} dz = \infty, i = 1,2 . \tag{7.26}$$

Then $G_i(x), i = 1,2$ will be distribution functions. We leave it to the reader to check that the random vector (Z, I) has the same distribution function whether (X_1, X_2) is distributed with joint distribution function $F(x_1, x_2)$ or with joint distribution function

$$F^*(x_1, x_2) = G_1(x_1)G_2(x_2) . \tag{7.27}$$

In other words, independent random variables X_1^* and X_2^* with distribution functions $G_1(x_1)$ and $G_2(x_2)$ respectively and random vector (X_1, X_2) with joint distribution function $F(x_1, x_2)$ give rise to the same distribution function for the identified minimum (Z, I). Hence (Z, I) does not identify the joint distribution $F(x_1, x_2)$.

The nonidentifiability aspect of the problem was noted at first by Cox (1959) and further investigations were made by Tsiatis (1975). The above discussion due to Basu and Ghosh (1978) shows that the problem of identifiability cannot be resolved in a nonparametric framework when the components are dependent. This leads us to the question of identifiability in parametric families.

We will discuss identifiability for families of bivariate normal distributions later in this chapter. Let us consider some other families of bivariate distributions.

Example 7.4.1 : The tail probability of the bivariate exponential distribution introduced by Marshall and Olkin (1967) is given by

$$\begin{aligned} \bar{F}(x_1, x_2) &= \Pr(X_1 > x_1, X_2 > x_2) \\ &= \exp[-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} \max(x_1, x_2)] \qquad \text{if } x_1 > 0, x_2 > 0 \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{7.28}$$

where $\lambda_1 > 0, \lambda_2 > 0$ and $\lambda_{12} \geq 0$. Let $f_i(z)$ be the conditional density of Z given $I = i$. Observe that the joint density of (Z, I) is

$$\begin{aligned} p_i f_i(z) &= \lambda_{12} e^{-\lambda z} && \text{if } i = 0, z > 0 \\ &= \lambda_1 e^{-\lambda z} && \text{if } i = 1, z > 0 \\ &= \lambda_2 e^{-\lambda z} && \text{if } i = 2, z > 0 \\ &= 0 && \text{otherwise} \end{aligned} \tag{7.29}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$ and $i = 0$ if $X_1 = X_2, i = 1$ if $Z = X_1$ and $i = 2$ if $Z = X_2$. Here $p_i = \Pr(I = i)$. It is clear that all the parameters are identifiable from the distribution of (Z, I). However, if Z is only observable, then the parameters are *not* identifiable since the density of Z is

$$\begin{aligned} f(z) &= \lambda e^{-\lambda z} && \text{if } z > 0 \\ &= 0 && \text{otherwise} \end{aligned}$$

and $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$.

Example 7.4.2 : Consider the absolutely continuous bivariate exponential distribution with density given by

$$\begin{aligned} f(x_1, x_2) &= \{\frac{\lambda\lambda_1(\lambda_2 + \lambda_{12})}{\lambda_1 + \lambda_2}\} \exp\{-\lambda_1 x_1 - (\lambda_2 + \lambda_{12}) x_2\} \\ &\qquad\qquad \text{if } x_1 < x_2 \\ &= \{\frac{\lambda\lambda_2(\lambda_1 + \lambda_{12})}{\lambda_1 + \lambda_2}\} \exp\{-(\lambda_1 + \lambda_{12}) x_1 - \lambda_2 x_2\} \\ &\qquad\qquad \text{if } x_1 > x_2 \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned} \tag{7.30}$$

This distribution was introduced by Block and Basu (1974). Here the joint density of (Z, I) is given by

$$\begin{aligned} p_i f_i(z) &= \tfrac{\lambda\lambda_i}{\lambda_1\lambda_2} e^{-\lambda z}, && z > 0 \\ &= 0, && z \leq 0 \end{aligned}$$

for $i = 1, 2$ where $\lambda > 0$ and the parameter set $(\lambda, \lambda_1, \lambda_2, \lambda_{12})$ is not identifiable but the set $(\lambda, \frac{\lambda_1}{\lambda_1+\lambda_2}, \frac{\lambda_2}{\lambda_1+\lambda_2})$ is identifiable. If Z is only observable, then λ is the only parameter which is identifiable.

Similar results can be obtained for the bivariate exponential distribution introduced by Gumbel (1960). For further discussion, see Basu and Ghosh (1978, 1980).

7.5 Identifiability for Families of Bivariate Normal Distributions (The Case of Identified Minimum)

Suppose a random vector (X_1, X_2) has the bivariate normal distribution with the mean vector (μ_1, μ_2) and the covariance matrix $\sum = ((\sigma_{ij}))$. Let $\sigma_{ii} \equiv \sigma_i^2$ and $\sigma_{ij} \equiv \rho\sigma_i\sigma_j, i \neq j$. For simplicity, we write (X_1, X_2) is $\mathrm{BVN}(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho)$. Assume that $|\rho| < 1$.

Theorem 7.5.1 (Basu and Ghosh (1978), Nadas (1971)): Suppose (X_1, X_2) is BVN $(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho)$. Let $Z = \min(X_1, X_2)$ and $I = i$ if $Z = X_i, i = 1, 2$. Further assume that (X_3, X_4) is $\mathrm{BVN}(\mu_3, \mu_4; \sigma_3, \sigma_4; \rho')$. Define $Z' = \min(X_3, X_4)$ and $I' = i$ if $Z' = X_i, i = 3, 4$. Let

$$\alpha_1 = 1 - \rho\,\frac{\sigma_1}{\sigma_2}, \alpha_2 = 1 - \rho\,\frac{\sigma_2}{\sigma_1}\,, \tag{7.31}$$

and

$$\alpha_3 = 1 - \rho'\,\frac{\sigma_3}{\sigma_4}, \alpha_4 = 1 - \rho'\,\frac{\sigma_4}{\sigma_3}\,. \tag{7.32}$$

If (Z, I) and (Z', I') have the same distribution, then

$$(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho) = (\mu_3, \mu_4; \sigma_3, \sigma_4; \rho')\,. \tag{7.33}$$

Proof : Recall that (Z, I) is called the identified minimum of (X_1, X_2). Note that

$$(\rho\,\frac{\sigma_2}{\sigma_1})(\rho\,\frac{\sigma_1}{\sigma_2}) = \rho^2 < 1 \tag{7.34}$$

and hence at least one of $\rho\,\frac{\sigma_2}{\sigma_1}$ and $\rho\,\frac{\sigma_1}{\sigma_2}$ is less than one. In other words, either α_1 or α_2 is positive. Similarly either α_3 or α_4 is positive. Let $f_i(z)$ be the conditional density of Z given $I = i$ and $p_i = \Pr(I = i)$. Let H be the distribution function of (Z, I). Observe that

$$
\begin{aligned}
H(z,I) &= \Pr(Z \le z, I=1) \\
&= \Pr(I=1) - \Pr(z < X_1 < X_2) \\
&= p_1 - \int_z^\infty \int_{x_1}^\infty n(x_2|m(x_1), \sigma_2^2(1-\rho^2)) n(x_1|\mu_1,\sigma_1^2) dx_2 dx_1 \\
&= p_1 - \int_z^\infty \{1 - N(\frac{x - m(x)}{\sigma_2(1-\rho^2)^{1/2}}|0,1)\} n(x|\mu_1,\sigma_1^2) dx \qquad (7.35)
\end{aligned}
$$

where

$$m(x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad (7.36)$$

$$n(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\}, \qquad (7.37)$$

and

$$N(x|\mu,\sigma^2) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\} dy. \qquad (7.38)$$

This identity implies that

$$p_1 f_1(z) = n(z|\mu_1,\sigma_1^2)\{1 - N(\frac{z - m(z)}{\sigma_2(1-\rho^2)^{1/2}}|0,1)\}\ . \qquad (7.39)$$

Observe that

$$\frac{z - m(z)}{\sigma_2(1-\rho^2)^{1/2}} = \frac{(1 - \rho\frac{\sigma_2}{\sigma_1}) - \{\mu_2 - (\rho\frac{\sigma_2}{\sigma_2})\mu_1\}}{\sigma_2(1-\rho^2)^{1/2}}\ . \qquad (7.40)$$

If $\alpha_2 = 1 - \frac{\rho\sigma_2}{\sigma_1} > 0$, then

$$p_1 f_1(z) = \phi_{11}(z)[1 - \Phi_{2*2*}(z)] \qquad (7.41)$$

where

$$\phi_{11}(z) \text{ is the density function of } N(\mu_1,\sigma_1^2), \qquad (7.42)$$

$$\phi_{2*2*}(z) \text{ is the density function of } N(\mu_2^*,\sigma_2^{*2}), \qquad (7.42A)$$

$$\mu_2^* = (\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1)/\alpha_2 \qquad (7.42B)$$

and

$$\sigma_2^* = \sigma_2(1-\rho^2)^{1/2}/\alpha_2\,. \qquad (7.42C)$$

Here $\Phi(z)$ denotes the distribution function corresponding to $\phi(z)$. If $\alpha_2 = 0$, then

$$p_1 f_1(z) = \phi_{11}(z)(1 - \Phi_{2*2*}(0)) \qquad (7.43)$$

where

$$\mu_2^* = \mu_2 - \mu_1 \quad (\text{for } \alpha_2 = 0) \tag{7.44}$$

and

$$\sigma_2^* = \sigma_2(1-\rho^2)^{1/2}(\text{for } \alpha_2 = 0)\,. \tag{7.45}$$

If $\alpha_2 = 1 - \frac{\rho\sigma_2}{\sigma_1} < 0$, then

$$p_1 f_1(z) = \phi_{11}(z)\Phi_{2*2*}(z) \tag{7.46}$$

where ϕ and Φ are as defined above. Similar relations hold for $p_2 f_2(z)$. In fact

$$\begin{aligned} p_2 f_2(z) &= \phi_{22}(z)[1-\Phi_{1*1*}(z)] \quad \text{if } \alpha_1 > 0 && (7.47)\\ &= \phi_{22}(z)\Phi_{1*1*}(z) \quad \text{if } \alpha_1 < 0 && (7.47A)\\ &= \phi_{22}(z)[1-\Phi_{1*1*}(0)] \quad \text{if } \alpha_1 = 0\,. && (7.47B)\end{aligned}$$

Analogous relations hold for the BVN$(\mu_3, \mu_4; \sigma_3, \sigma_4; \rho')$. We have already noted that at least one of α_1, α_2 and at least one of α_3, α_4 are positive.

Case (1) Suppose $\alpha_i > 0, 1 \le i \le 4$. Then

$$p_1 f_1(z) = \phi_{11}(z)[1-\Phi_{2*2*}(z)], \tag{7.48}$$

$$p_2 f_2(z) = \phi_{22}(z)[1-\Phi_{1*1*}(z)], \tag{7.48A}$$

$$p_3 f_3(z') = \phi_{33}(z')[1-\Phi_{4*4*}(z'))], \tag{7.48B}$$

and

$$p_4 f_4(z') = \phi_{44}(z')[1-\Phi_{3*3*}(z')]\,. \tag{7.48C}$$

Note that

$$\frac{p_1 f_1(z)}{\phi_{11}(z)} \to 1 \text{ as } z \to -\infty\,. \tag{7.49}$$

Similarly

$$\frac{p_3 f_3(z')}{\phi_{33}(z')} \to 1 \text{ as } z' \to -\infty\,. \tag{7.50}$$

Since the distributions of (Z, I) and (Z', I') are identical by hypothesis, it follows that $p_1 f_1(z) = p_3 f_3(z)$ and hence

$$\frac{\phi_{11}(z)}{\phi_{33}(z)} \to 1 \text{ as } z \to -\infty\,. \tag{7.51}$$

In other words

$$\frac{\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\{-\frac{1}{2}(\frac{z-\mu_1}{\sigma_1})^2\}}{\frac{1}{\sqrt{2\pi\sigma_3^2}}\exp\{-\frac{1}{2}(\frac{z-\mu_3}{\sigma_3})^2\}} \to 1 \text{ as } z \to -\infty \,. \tag{7.52}$$

Hence

$$\frac{\sigma_3}{\sigma_1}\exp\{-\frac{1}{2}(\frac{z-\mu_1}{\sigma_1})^2 + \frac{1}{2}(\frac{z-\mu_3}{\sigma_3})^2\} \to 1 \text{ as } z \to -\infty \,.$$

It is easy to check that this limit holds iff $\sigma_1 = \sigma_3$ and $\mu_1 = \mu_3$. Similarly we obtain that $\sigma_2 = \sigma_4$ and $\mu_2 = \mu_4, p_1 = \Pr(I = 1) = \Pr(X_1 < X_2)$ and $p_3 = \Pr(I = 3) = \Pr(X_3 < X_4)$. Since $p_1 = p_3$, it follows that

$$\Phi(\frac{\mu_2 - \mu_1}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}) = \Phi(\frac{\mu_4 - \mu_3}{\sigma_3^2 + \sigma_4^2 - 2\rho'\sigma_3\sigma_4})$$

which implies that $\rho = \rho'$ as $\sigma_1 = \sigma_3, \sigma_2 = \sigma_4, \mu_1 = \mu_3$ and $\mu_2 = \mu_4$. This proves that

$$(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho) = (\mu_3, \mu_4; \sigma_3, \sigma_4; \rho') \tag{7.53}$$

and hence (Z, I) identifies the BVN $(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho)$.

Case (2) Suppose that exactly one of (α_1, α_2) and one of (α_3, α_4) is positive. Without loss of generality, assume that $\alpha_3 > 0, \alpha_2 < 0$. Then either $\alpha_3 > 0$ and $\alpha_4 < 0$ or $\alpha_3 < 0$ and $\alpha_4 > 0$. Assume that $\alpha_3 > 0$ and $\alpha_4 < 0$. Then we have $\alpha_1 > 0, \alpha_2 < 0, \alpha_3 > 0$ and $\alpha_4 < 0$. Since $\alpha_2 < 0$ and $\alpha_4 < 0$, it follows that

$$p_1 f_1(z) = \phi_{11}(z)\Phi_{2*2*}(z) = \phi_{33}(z)\Phi_{4*4*}(z) = p_3 f_3(z) \tag{7.54}$$

from (7.47). Hence

$$\Phi_{2*2*}(z) = (\phi_{11}(z))^{-1}\phi_{33}(z)\Phi_{4*4*}(z), -\infty < z < \infty \,. \tag{7.55}$$

Let $z \to \infty$. Then it follows that

$$(\phi_{11}(z))^{-1}\phi_{33}(z) \to 1 \text{ as } z \to \infty \,. \tag{7.56}$$

It is easy to check that this relation holds iff $\mu_1 = \mu_3$ and $\sigma_1 = \sigma_3$. Since $\alpha_1 > 0$ and $\alpha_3 > 0$, it can be shown that $\mu_2 = \mu_4$ and $\sigma_2 = \sigma_4$ by the arguments given in Case (1). These two relation show that $\rho = \rho'$ again by the arguments developed in Case (1). Hence

$$(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho) = (\mu_3, \mu_4; \sigma_3, \sigma_4; \rho') . \tag{7.57}$$

Suppose that $\alpha_1 > 0, \alpha_2 < 0$ but $\alpha_3 < 0$ and $\alpha_4 > 0$. Then

$$p_1 f_1(z) = \phi_{11}(z)\Phi_{2*2*}(z) = \phi_{33}(z)(1 - \Phi_{4*4*}(z)] = p_3 f_3(z) \tag{7.58}$$

from relations of the type given in (7.48) and the fact that (Z, I) and (Z', I') are identically distributed. Hence

$$[\phi_{33}(z)]^{-1}\phi_{11}(z)\Phi_{2*2*}(z) = 1 - \Phi_{4*4*}(z), -\infty < z < \infty . \tag{7.59}$$

Let $z \to -\infty$ in (7.59) on both sides. The term on the right side tends to 1 and hence

$$[\phi_{33}(z)]^{-1}\phi_{11}(z)\Phi_{2*2*}(z) \to 1 \text{ as } z \to -\infty . \tag{7.60}$$

It is easy to show that, for any X_i distributed $N(\mu_i, \sigma_i^2)$ and X_j distributed $N(\mu_j, \sigma_j^2)$,

$$\frac{\phi_{jj}(z)}{\phi_{ii}(z)} \to 1, 0 \text{ or } \infty \text{ as } z \to\to \pm\infty . \tag{7.61}$$

Furthermore

$$\Phi_{2*2*}(z) \to 0 \text{ as } z \to -\infty . \tag{7.62}$$

Hence the equation (7.60) does not hold and the condition

$$\alpha_1 > 0, \alpha_2 < 0, \alpha_3 < 0, \alpha_4 > 0 \tag{7.63}$$

is impossible whenever $(Z, I), (Z', I')$ are identically distributed. Similarly it can be shown that in all other cases on $\alpha_i, 1 \leq i \leq 4$ either there is identifiability given (Z, I) or the conditions on $\alpha_i, 1 \leq i \leq 4$ will not hold. For details, see Basu and Ghosh (1978). ■

7.6 Identifiability for Families of Bivariate Normal Distributions (The Case of Nonidentified Minimum)

Let us assume again that (X_1, X_2) has $\text{BVN}(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho)$ and (X_3, X_4) has $\text{BVN}(\mu_3, \mu_4; \sigma_3, \sigma_4; \rho^1)$. Let $Z = \min(X_1, X_2)$ and $Z' = \min(X_3, X_4)$. Z is called the *nonidentified minimum* of X_1 and X_2. The problem of interest is to find whether the distribution of Z identifies the distribution of (X_1, X_2). The following result is due to Basu and Ghosh (1978).

Theorem 7.6.1 (Basu and Ghosh (1978)): If Z and Z' have the same distribution, then either

$$\mu_1 = \mu_3, \sigma_1 = \sigma_3, \mu_2 = \mu_4, \sigma_2 = \sigma_4 \text{ and } \rho = \rho'$$

or

$$\mu_1 = \mu_4, \sigma_1 = \sigma_4, \mu_2 = \mu_3, \sigma_2 = \sigma_3 \text{ and } \rho = \rho' .$$

In other words, either the distributions of (X_3, X_4) and (X_1, X_2) are the same bivariate normal distribution or the distributions of (X_3, X_4) and (X_2, X_1) are the same bivariate normal distribution.

Remarks 7.6.1 : Proof of Theorem 7.6.1 in Basu and Ghosh (1978) essentially uses the methods developed in the proof of Theorem 7.5.1 and the fact that the normal distribution function is not an elementary function. Alternate proofs of this result are given in Anderson and Ghurye (1979), Mukherjea *et al.* (1986) and Gilliland and Hannan (1980). We have already seen that the identifiability question is an important problem in econometrics. In the Fair–Jaffee model with μ_1 and μ_2, regression of supply and demand on some regression variables and full rank covariance matrix, Hartley and Mallela (1977) consider the problem of estimation of σ_1^2, σ_2^2 and ρ based on the observed minima of supply and demand. Identifiability was implicitly assumed by them. The general problem of identification of parameters by the distribution of the maximum random variable in the trivariate normal case and the general multivariate normal case is studied in Basu and Ghosh (1978) and Mukherjea and Stephens (1990a,b). The result in the multivariate normal case can be stated in the following manner. We omit the proofs.

Theorem 7.6.2 (Mukherjea and Stephens (1990b)): Let $\boldsymbol{X}_i, 1 \le i \le k$ be k independent n-dimensional random vectors each with a nonsingular multivariate normal distribution with zero mean vector and *positive* partial correlations. Suppose that $\boldsymbol{X}_i = (X_{i1}, ..., X_{in})$ and $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$ where $Y_j = \max(X_{ij}, 1 \le i \le k)$. Let $\boldsymbol{W}$ be another n-dimensional random vector which is the vector of maxima componentwise of another such family of independent n-dimensional random vectors $\boldsymbol{Z}_j, 1 \le j \le \ell$. Then the distributions of $\boldsymbol{X}_i$'s , $1 \le i \le k$ are a rearrangement of the distribution of $\boldsymbol{Z}_j$'s ,$1 \le j \le \ell$ (and hence necessarily $k = \ell$) whenever $\boldsymbol{Y}$ and $\boldsymbol{W}$ have the same distribution.

7.7 Identifiability for a k-out of -p System

We now consider a generalization of the problem discussed in Section 7.3. Let us consider a generalization of the concepts of competing and complementary risks. The problem can be paraphrased as the problem of identifying the distributions of component lifetimes from that of system lifetime where the system is a k-out of - p system; that is, the system with p components works if and only if k or more of p components of the system function or equivalently the system fails when the first $r = p - k + 1$ components fail. It can be checked that identifiability holds for a k-out of-p *identical* component system following arguments similar to those given at the beginning of Section 7.3.

Let $X_i, 1 \le i \le p$ be the component lifetimes and $X_{(r)}$ denote the rth-order statistic. Suppose the random variable $X_{(r)}$ is the only observable. Given the distribution function of $X_{(r)}$, is it possible to determine the joint distribution of $(X_1, ..., X_p)$? If $r = 1$, the problem reduces to the problem of competing risks and, if $r = p$, then it reduces to the problem of complementary risks. Note that if $r = 1$, then the system is in series and if $r = p$, then the system is in parallel.

Let $I = i$ when $X_{(r)} = X_i, 1 \leq i \leq p$. The pair $(X_{(r)}, I)$ is called the *identified rth order statistic.* Observe $X_{(r)}$ is termed *nonidentified rth-order statistic.*

Basu and Ghosh (1983) proved that, for $1 \leq r \leq p, (X_{(r)}, I)$ identifies the distributions of $X_i, 1 \leq i \leq p$ when X_i follows the exponential distribution and $X_i, 1 \leq i \leq p$ are independent. If $X_{(r)}$ is nonidentified, then, for $2 \leq r \leq p$, the distribution of $X_{(r)}$ determines the distributions of $X_i, 1 \leq i \leq p$ up to a permutation whenever X_i follows the exponential distribution and $X_i, 1 \leq i \leq p$ are independent. Results for the case of general distributions are unknown as far as the author is aware. We now discuss these results from Basu and Ghosh (1983).

Theorem 7.7.1 (Basu and Ghosh (1983)): Suppose $X_i, 1 \leq i \leq p$ are independent random variables and X_i follows the exponential distribution with parameter $\lambda_i > 0$, that is, the density function of X_i is

$$\begin{aligned} f(x_i, \lambda_i) &= \lambda_i e^{-\lambda_i x}, x_i > 0 \\ &= 0, \qquad x_i \leq 0 \end{aligned}$$

for $1 \leq i \leq p$. Then the distribution of the identified rth order statistic $(X_{(r)}, I)$ *uniquely* determines $\lambda_1, \lambda_2, ..., \lambda_p$ whenever $1 \leq r \leq p$.

Proof : Let p_j be the probability that $I = j$ and $f_j(y)$ be the conditional density of $Y = X_{(r)}$ gives $I = j$. It is easy to check that

$$p_j f_j(y) = \lambda_j e^{-\lambda_j y} \sum [\prod_{i=1}^{r-1} (1 - e^{-\lambda_{\alpha_i} y}) \prod_{s=r+1}^{p} e^{-\lambda_{\beta_s} y}] \tag{7.64}$$

where $(\alpha_1, \alpha_2, ..., \alpha_{r-1}; \beta_{r+1}, ..., \beta_p)$ is a permutation of the integers $(1, 2, ..., j - 1, j + 1, ..., p)$ partitioned into two sets $(\alpha_1, ..., \alpha_{r-1})$ and $(\beta_{r+1}, ..., \beta_p)$ and $X_{\alpha_i} < X_{(r)}$ and $X_{\beta_s} > X_{(r)}, 1 \leq i \leq r-1$ and $r+1 \leq s \leq p$. The summation Σ runs over all such sets $(\alpha_1, ..., \alpha_{r-1}; \beta_{r+1}, ..., \beta_p)$. Note that the term with the highest power of e^{-y} on the left side of (7.64) is

$$(-1)^{r-1} \lambda_j e^{-(\lambda_1 + ... + \lambda_p) y} .$$

This identifies λ_j. By repeating the procedure for $1 \le j \le p$, all the parameters $\lambda_1, \lambda_2, ..., \lambda_p$ can be identified. ■

Theorem 7.7.2 (Basu and Ghosh (1983)): Let $X_1, X_2, ..., X_p$ be independent random variables and suppose X_i has the exponential distribution with parameter $\lambda_i > 0$. Then the distribution of the nonidentified $X_{(r)}$ determines the values of $\lambda_1, \lambda_2, \cdots, \lambda_p$ up to a permutation whenever $2 \le r \le p$.

Proof : The density of $X_{(r)}$ is

$$f_{(r)}(y) = \sum_{j=1}^{p}\{\lambda_j e^{-\lambda_j y}[\prod_{i=1}^{r-1}(1 - e^{-\lambda_{\alpha_i} y}) \prod_{s=r+1}^{p} e^{-\lambda_{\beta_s} y}]\} \tag{7.65}$$

where the expression in $\{\cdots\}$ on the right side of (7.65) is obtained from the expression on the right side in (7.64). A typical term on the right side of (7.65) (after collecting together different expressions involving the same power of e^{-y}) is of the form

$$(\lambda - \lambda_{i_1} - \lambda_{i_2} - \cdots - \lambda_{i_\ell})(-1)^{r-\ell-1}\binom{p-\ell-1}{r-\ell-1}e^{-(\lambda-\lambda_{i_1}\cdots-\lambda_{i_\ell})y}$$

where $1 \le i_1, i_2, ..., i_\ell \le p, i_s \ne i_t$ for $s \ne t, 0 \le \ell \le p-1$ and $\lambda = \lambda_1 + \cdots + \lambda_p$. Thus we can identify λ and $\lambda - \lambda_{i_1} - \lambda_{i_2} - \cdots - \lambda_{i_\ell}$, $1 \le i_1, ..., i_\ell \le p, i_s \ne i_t$ for $s \ne t, 1 \le \ell \le p-1$. It can be checked that these values uniquely determine $\lambda_1, \lambda_2, ..., \lambda_p$ up to a permutation (cf. Basu and Ghosh (1983)). ■

Remarks 7.7.1 : It is interesting to observe that in the case of exponential distribution, nonidentifiability occurs if and only if $r = 1$, that is, the minimum does not identify the component exponential distributions.

Remarks 7.7.2 (Identifiability in coherent systems): Competing risks deal with a system failing as a consequence of the failure of one of its components. It was shown in Theorem 7.3.1 that if the components have independent lifetimes, then the joint distribution of the sytems failure time and the identity of the failed component uniquely determine the lifetime distribution of each of the components.

In reliability theory, coherent systems are also used to model the systems (cf. Barlow and Proschan (1975)). Coherent systems extend the theory of competing risks to systems failing as a consequence of the failure of some of its components rather than just one.

Given a coherent system with n components having independent lifetimes X_i let Z be the age of the system at breakdown and I be the set of componets failed by time Z. I is called the *diagnostic set.* Then $I = \{i : X_i \leq Z\}$. A set of components E is called a *cut set* if, when all components in E have failed, the system fails. E is a minimal cut set if it is a cut set which does not contain a proper subset which is itself a cut set. Let $\{I_1, I_2, ..., I_m\}$ be the collection of all minimal cut sets. Let $\mathbf{M}$ be the $m \times n$ incidence matrix of $\{I_1, I_2, ..., I_m\}$. In other words $M_{ij} = 1$ if $j \in I_i$ and $M_{ij} = 0$ otherwise. Meilijson (1985) showed that if the component lifetimes $X_1, X_2, ..., X_n$ are nonatomic, independent and possess the same essential extrema and if the rank of the matrix $\mathbf{M}$ is n, then the joint distribution of Z and I determine the distribution of each $X_i, 1 \leq i \leq n$ *uniquely.* In other words, the system is identifiable. Meilijson (1985) also proved that a necessary condition for identifiability is that no two components be in parallel, that is, belong to the same minimal cut sets. Suppose that the independent lifetimes $X_i, 1 \leq i \leq n$ have mutually absolutely continuous distributions and that each component lifetime possesses a single positive atom at the common essential infimum. Nowik (1990) proved that the joint distribution of (Z, I) identifies the life time distribution of each component if and only if there is at most one component belonging to all cut sets or equivalently no two components are in parallel. For further details, see Nowik (1990).

7.8 Identifiability from Survival Functions

Let (T_1, T_2) be a bivariate nonnegative random vector with the joint survival function

$$S_{T_1,T_2}(t_1, t_2) = \Pr(T_1 > t_1, T_2 > t_2)$$

with $S(0,0) = 1$. Suppose the variables T_1 and T_2 are subject to censoring by random intervals $[X_1, Y_1]$ and $[X_2, Y_2]$ respectively. In other words, T_1 and T_2 are observable iff $X_1 \leq T_1 \leq Y_1$ and $X_2 \leq Y_2 \leq T_2$. The information on T_1 and T_2 can be expressed by the random vectors (W_1, W_2) and (δ_1, δ_2) where

$$W_i = \max(\min(Y_i, T_i), X_i)), i = 1, 2$$

and

$$\delta_i = \begin{cases} 1 & \text{if} \quad X_i \leq T_i \leq Y_i \\ 2 & \text{if} \quad T_i > Y_i \\ 3 & \text{if} \quad T_i < X_i \end{cases}$$

for $i = 1, 2$.

An example of this type of double censoring can be illustrated by the following scenario. Suppose we have a follow-up study for determining the ages T_1 and T_2 respectively at which a male-child and a female-child of the same family developed a particular skill for the first time. T_1 and T_2 are observable if the skills were developed after admitting into a program. It is possible that for some females or males in the program, the individual might have developed the skill prior to joining the program resulting in left censoring of T_1 or T_2. On the other hand, a right censoring may occur due to withdrawal of the child either due to withdrawal from the study or by not attaining the skill before the program is terminated. Here the random vector (T_1, T_2) is subject to double censoring. The joint survival function of (T_1, T_2) is unobservable but is of importance. The problem is to determine sufficient conditions under which the distribution of $(\boldsymbol{W}, \delta)$ determines the joint survival function $S_{T_1,T_2}(t_1, t_2)$ *uniquely.*

Another example where right censoring is only present can be described as follows. Assume that a pair of individuals, a wife and a husband for instance, are under study. The observation on each of the individuals is terminated in the event of death or in the case of withdrawal from the study. The joint life length of the two individuals is of importance but is unobservable. Again the problem is to determine sufficient conditions under which the observed distribution *uniquely* determines the unobservable distributions.

By observing a series system of d components, we can only determine its life length and the components that cause the system to fail. In particular, the life length of a series subsystem consisting of k components where $(0 < k < d)$ is unobservable. Some k of the d components may be in the system to support operation of the remaining $d-k$ main components. The distribution function of the life length of the series subsystem consisting of the k main components is unobservable. The problem is again to determine sufficient conditions under which the observed distribution *uniquely* determines the unobservable distributions.

Langberg and Shaked (1982) discussed the identifiability problem from multivariate survival functions under right censoring. Chang (1984) discussed the univariate case under double censoring. We briefly discuss results due to Ebrahimi (1988).

Let

$$\begin{aligned} S_{Y_1,Y_2}(t_1,t_2) &= \Pr(Y_1 > t_1, Y_2 > t_2), \\ S_{X_1,X_2}(t_1,t_2) &= \Pr(X_1 > t_1, X_2 > t_2), \\ S_{X_1,Y_2}(t_1,t_2) &= \Pr(X_1 > t_1, Y_2 > t_2), \\ S_{Y_1,X_2}(t_1,t_2) &= \Pr(Y_1 > t_1, X_2 > t_2) \end{aligned}$$

and

$$S_{X_1,X_2}(t_1,t_2) = \Pr(X_1 > t_1, X_2 > t_2)\ .$$

We assume that $\Pr(X_1 \le Y_1, X_2 \le Y_2) = 1$ and the above survival functions are continuously differentiable for $t_1 > 0, t_2 > 0$. We further assume that (T_1, T_2) and $\{(X_1, Y_1), (X_2, Y_2)\}$ are *independent* random vectors. Let

$$Q_{ij}(t_1,t_2) = \Pr(W_1 > t_1, W_2 > t_2, \delta_1 = i, \delta_2 = j)$$

for $1 \leq i \leq 3, 1 \leq j \leq 3$. Ebrahimi (1988) proved the following result generalizing results of Langberg and Shaked (1982) in the bivariate case for random right censoring and Chang (1984) in the univariate case for random double censoring. We omit the proof.

Theorem 7.8.1 (Ebrahimi (1988)) : In addition to the conditions stated earlier, suppose that for all $(t_1, t_2), t_1 > 0, t_2 > 0$,

$$\begin{aligned} S_{Y_1,Y_2}(t_1,t_2) - S_{X_1,Y_2}(t_1,t_2) - S_{Y_1,X_2}(t_1,t_2) + S_{X_1,X_2}(t_1,t_2) &> 0, \\ \frac{\partial}{\partial t_2}[S_{Y_1,X_2}(t_1,t_2) - S_{X_1,X_2}(t_1,t_2)] &< 0, \\ \frac{\partial}{\partial t_1}[S_{X_1,Y_2}(t_1,t_2) - S_{X_1,X_2}(t_1,t_2)] &< 0, \\ \frac{\partial}{\partial t_2}[S_{Y_1,Y_2}(t_1,t_2) - S_{X_1,Y_2}(t_1,t_2)] &< 0, \\ \frac{\partial}{\partial t_1}[S_{Y_1,Y_2}(t_1,t_2) - S_{X_1,X_2}(t_1,t_2)] &< 0 \end{aligned}$$

and

$$S_{T_1,T_2}(t_1,t_2) > 0 \,.$$

Then the unobservable survival functions $S_{Y_1,X_2}(t_1,t_2)$, $S_{X_1,X_2}(t_1,t_2)$, $S_{Y_1,Y_2}(t_1,t_2)$, $S_{X_1,Y_2}(t_1,t_2)$ and $S_{T_1,T_2}(t_1,t_2)$ are *uniquely* determined by the observable survival functions

$$Q_{ij}(t_1,t_2), 1 \leq i,j \leq 3 \,.$$

Remarks 7.8.1 : The result can be extended to the multivariate case.

7.9 Nonidentifiability in Some Stochastic Models

7.9.1 (Accident models): Occurrence of nonidentifiability in some stochastic models fitted to accident data was pointed out by Cane (1972, 1977). Negative binomial distribution is often used as a model for fitting for accident data and it was found to be a good fit most often. It is known that two explanations, one in terms of the accident proneness and the other involving contagion, can be given for fitting a negative binomial distribution as a model. It is generally assumed that one can decide the underlying model if complete information, that is, the time of each accident for every individual in the sample is known. Cane (1972) indicated that, even with such complete information, it is not possible to pick the underlying mechanism from the two described above. In fact there are infinite number of mechanisms each of which gives rise to the same type of data and hence the presence of nonidentifiability in modeling. We now discuss some of these results due to Cane (1972, 1977).

Model 1 : The Poisson process is generally used for modeling the occurrence of accidents. Here it is assumed that the accidents occur at a rate λu where λ refers to the accident proneness of any individual at risk, u refers to the danger of the situation in which accidents occur and the distribution of the number of the accidents in a time T has Poisson distribution with mean λuT with the probability generality function (p.g.f.) $\phi_1(s) = \exp\{\lambda uT(s-1)\}$.

It was found that the accident data in factories do not conform to the Model 1. An alternate model was proposed by Greenwood and Yule (1920).

Model 2 (Accident proneness model) : Here it is assumed that Model 1 holds for any given individual but that the individuals may have different λ values and that the variation in λ can be described by a gamma density

$$\begin{aligned} f(\lambda) &= \frac{c^k \lambda^{k-1} e^{-\lambda c}}{\Gamma(k)}, \lambda > 0 \\ &= 0, \qquad\qquad \lambda \leq 0 . \end{aligned}$$

The p.g.f. of the distribution of accidents in time T is

$$\phi_2(s) = E_\lambda[\phi_1(s)] = c^k(c - uT(s-1))^{-k} .$$

It is convenient to absorb c into u and replace c by 1.

A third model was suggested by McKendrick (1926).

Model 3 (Contagion model) : Here it is assumed that a person who has had n accidents in time $(0,t)$ has a conditional probability $\frac{k+n}{1+ut}udt$ of having another accident in $(t, t+dt)$ independent of the times of the preceding accidents. All individuals of the population have the same probability $ku\, dt$ of an accident in $(0, dt)$.

Nonidentifiability : Let us now show that the Model 2 and Model 3 are equivalent.

Suppose that an individual has n accidents at times $t_i, 1 \leq i \leq n$,

$$0 = t_0 < t_1 < \cdots < t_n < T\,.$$

The conditional probability for such an event under Model 2 given λ is

$$\prod_{i=1}^{n} e^{-\lambda u(t_i - t_{i-1})} \lambda u\, dt_i e^{-\lambda u(T-t_n)} \tag{7.66}$$

and the probability under Model 3 is

$$\prod_{i=1}^{n}\{(\frac{1+ut_{i-1}}{1+ut_i})^{k+i-1}\frac{k+i-1}{1+ut_i}udt_i\}(\frac{1+ut_n}{1+uT})^{k+n}\,. \tag{7.67}$$

These expressions can be rewritten in the form

$$(n!dt_1...dt_nT^{-n})(\lambda uT)^n\frac{e^{-\lambda uT}}{n!} \tag{7.68}$$

and

$$(n!dt_1...dt_nT^{-n})(uT)^n(1+uT)^{-n-k}\binom{k+n-1}{n} \tag{7.69}$$

respectively. Observe that the term

$$(n!dt_1\cdots dt_nT^{-n})$$

in (7.68) and (7.69) gives the conditional probability that accidents occur at the *specified* times in $[0,T)$ given that there are n accidents in all in $[0,T)$. Thus the distribution of $(t_1, t_2, ..., t_n)$, conditional on n accidents in time T, is the same under both the Models 2 and 3. In fact, it is the joint distribution of the order statistics for a random sample of size n from the uniform distribution on $[0, T)$. Furthermore, it can be checked that each model gives the same value for the (unconditional) probability $p_n(t)$ of n accidents in time t (choosing $c = 1 = u$ in Model 2 and $u = 1$ in Model 3). Thus the Models 2 and 3 are indistinguishable. In other words, if the accident records of a large number of people are such that the distribution of accidents in time $[0, t)$ (possibly rescaled) fits the negative binomial distribution with probability generating function $\psi(s) = (1+t-ts)^{-k}$, then no additional data on individual records will provide information in distinguishing between the two models and there is no mathematical difference between the Models 2 and 3.

For general discussion on this problem, see Cane (1972, 1977) and Puri (1979). For earlier remarks on this problem, see Feller (1966, p. 57).

7.9.2 (A threshold-type shock model): Consider a system involving a single component. Suppose the system is subject to "shocks" at random times. Assume that the system fails as soon as the threshold K for the number of shocks is reached. Suppose the shocks are governed by a time-homogeneous Poisson process with parameter λ. Hence the lifetime L of the system has the distribution

$$H_1(t) = \Pr(L \leq t) = 1 - \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \bar{p}_k, t > 0 \tag{7.70}$$

where

$$\bar{p}_k = \Pr(K > k), k = 0, 1, 2, ..., \text{and } \bar{p}_0 = 1. \tag{7.71}$$

This model is called a "threshold-type" model. In practice, it is not possible to observe the occurrence of shocks and L is the only observable quantity. Let $\mathcal{F}$ be the family of distributions of L generated by varying λ and $\{\bar{p}_k\}$. This family $\mathcal{F}$ is not identifiable. This can be seen as follows. Let

$$\begin{aligned} \bar{H}_1(t) &= 1 - H_1(t) \\ &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \bar{p}_k, t > 0 . \end{aligned} \tag{7.72}$$

However, $\bar{H}_1(t)$ can also be written in the form

$$\bar{H}_1(t) = \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!} e^{-\nu t} \bar{q}_k, t > 0 \tag{7.73}$$

where

$$\bar{q}_k = \nu^{-k} \sum_{j=0}^{k} \binom{k}{j} (\nu - \lambda)^{k-j} \bar{p}_k, k = 0, 1, 2, ... \tag{7.74}$$

with $\bar{q}_0 = 1$. Hence the family $\mathcal{F}$ is not identifiable.

7.9.3 (A nonthreshold-type shock model) : Here we assume the existence of a nonnegative risk function $\beta(N(t), t)$, where $N(t)$ denotes the number of shocks received up to time t, such that

Pr [Failure of the system occurs in time $(t, t + \Delta t)$ given that no failure of the system occurred until time t and $N(t) = n$]

$$= \beta(n, t)\Delta t + 0(\Delta t) . \tag{7.75}$$

Suppose that $\beta(n, t) = n\ \alpha(t)$ where $\alpha(\cdot)$ is a nonnegative locally integrable function on $[0, \infty)$ such that

$$\int_0^\infty [1 - \exp\{-\int_\tau^\infty \alpha(u)du\}]d\tau = \infty . \tag{7.76}$$

Then the lifetime L has the distribution function

$$\begin{aligned} H_2(t) &= 1 - \Pr(L > t) \\ &= 1 - E\{\exp[-\int_0^t N(u)\alpha(u)du]\} . \end{aligned} \tag{7.77}$$

Hence

$$\begin{aligned} \bar{H}_2(t) &= 1 - H_2(t) \\ &= \exp\{-\lambda \int_0^t [1 - \exp(-\int_\tau^t \alpha(u)du)]d\tau\} \end{aligned} \tag{7.78}$$

when $N(t)$ is a Poisson process with parameter λ. Let $\mathcal{F}^*$ be the family of distributions of L generated by varying λ and α subject to the conditions stated above. $\mathcal{F}^*$ is not identifiable. In fact, given $\theta = (\lambda, \alpha(\cdot))$ generating $\bar{H}_2(\cdot)$ defined by (7.78), it can be checked that the same $\bar{H}_2(\cdot)$ is also generated by $\theta' = (\lambda', \alpha'(\cdot))$ where $\lambda' > \lambda$,

$$\alpha'(t) = \lambda\, \alpha(t)h(t)[(\lambda' - \lambda)t + \lambda h(t)]^{-1} \tag{7.79}$$

and

$$h(t) = \int_0^t \exp[-\int_\tau^t \alpha(u)du]d\tau\ .$$

Remarks 7.9.1 : Discussion in subsections 7.9.2 and 7.9.3 is based on Puri (1979).

Chapter 8

Identifiability for Mixtures of Distributions

8.1 Introduction

Mixtures of distributions are used in building probability models quite frequently in biological and physical sciences. For instance, in order to study certain characteristics in natural populations of fish, a random sample might be taken and the characteristic measured for each member of the sample; since the characteristic varies with the age of the fish, the distribution of the characteristic in the total population will be a mixture of the distributions at different ages. In order to analyze the qualitative character of inheritance, a geneticist might observe a phenotypic value that has a mixture distribution because each genotype might produce phenotypic values over an interval. For applications where mixtures of distributions arise, see Bruni *et al.* (1983), Merz (1980) and Christensen *et al.* (1980). Other applications are in the area of pattern recognition, for instance, in image reconstruction and statistical model building for positron emission tomography (Vardi *et al.* (1975)).

In order to devise statistical procedures for inferential purposes, an important problem is the identifiability of the mixing distribution. Unless the mixing distribution is identifiable in the model, it is not meaningful to estimate the distribution either nonparametrically or in a parametric framework. Some discussion on identifiability in the problem is given in Everitt and Hand (1981), Prakasa Rao (1983b), Titterington *et al.* (1985) and Maritz and Levin (1989). In this chapter we discuss the identifiability aspect of the problem more extensively.

Let $(\mathcal{X}, \mathcal{F})$ and $(\Theta, \mathcal{B})$ be measurable spaces such that $\mathcal{B}$ contains all singletons of Θ. Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a family of probability measures on $(\mathcal{X}, \mathcal{F})$ such that the mapping $\theta \to P_\theta(A)$ is $\mathcal{B}$ - measurable for each $A \in \mathcal{F}$. Let G be a probability measure on $(\Theta, \mathcal{B})$ and define

$$H(A) = \int_\Theta P_\theta(A) dG(\theta), A \in \mathcal{F} . \tag{8.1}$$

Then H is a probability measure on $(\mathcal{X}, \mathcal{F})$. H is called a *mixture* of the family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. G is called a *mixing distribution* . Let Λ be the class of all mixing distributions G on $(\Theta, \mathcal{B})$ and ζ be the corresponding class of mixtures. Define $Q : \Lambda \to \zeta$ by $Q(G) = H$. The class Λ and equivalently the family ζ is said to be *identifiable* with respect to $\mathcal{P}$ if the mapping Q is a *one-to-one mapping* between Λ and ζ.

As was pointed out earlier, the problem of estimation of G is meaningful only when the family Λ is identifiable. It is easy to see that if T is a measurable mapping from $(\mathcal{X}, \mathcal{F})$ to $(\mathcal{Y}, \mathcal{T})$ and if the family Λ is identifiable with respect to the family $\mathcal{P}T^{-1} = \{P_\theta T^{-1}, \theta \in \Theta\}$ on $(\mathcal{Y}, \mathcal{T})$, then Λ is identifiable with respect to $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ on $(\mathcal{X}, \mathcal{F})$.

The distribution H defined by (8.1) is called a *finite mixture* if the mixing distribution G is a discrete distribution with finite number of mass points. H is said to be a *countable mixture* if the mixing distribution G is a discrete distribution possibly with countable number of mass points. H is said to be an *arbitrary mixture* if G is any general mixing probability distribution.

In order to indicate that the problem of nonidentifiability does arise in these problems, we now present some examples.

Example 8.1.1 : Let P_θ be the binomial distribution $B(2,\theta)$ with two trials and θ as the probability of success, $0 < \theta < 1$. Let $G_{\theta_1,\theta_2,\alpha}$ be a mixing distribution given by

$$\Pr(\theta = \theta_1) = \alpha, Pr(\theta = \theta_2) = 1 - \alpha \tag{8.2}$$

where $\theta_1 \neq \theta_2, 0 < \alpha < 1$. Let X denote a random variable with the distribution which is a mixture of $\{P_\theta, 0 < \theta < 1\}$ with respect to the mixing distribution $G_{\theta_1,\theta_2,\alpha}$. Then

$$\Pr(X = 0) = \alpha(1-\theta_1)^2 + (1-\alpha)(1-\theta_2)^2, \tag{8.3}$$

$$\Pr(X = 1) = 2\alpha\theta_1(1-\theta_1) + 2\alpha\theta_2(1-\theta_2), \tag{8.3A}$$

and

$$\Pr(X = 2) = \alpha\theta_1^2 + (1-\alpha)\theta_2^2 \, . \tag{8.3B}$$

Since $\sum_{i=0}^{2}\Pr(X = i) = 1$, two of the above equations (8.3) to (8.3B) determine $\Pr(X = i)$ for $i = 0, 1, 2$. Let us consider the equations (8.3) and (8.3A). These are two equations containing three parameters α, θ_1 and θ_2 . Obviously there are infinitely many solutions $(\alpha, \theta_1, \theta_2)$ for a given pair of values for $\Pr(X = 0)$ and $\Pr(X = 1)$. Hence the family

$$\Lambda \equiv \{G_{\theta_1\theta_2,\alpha},: 0 < \theta_1, \theta_2 < 1, \theta, \neq \theta_2, 0 < \alpha < 1\}$$

is not identifiable with respect to $\mathcal{P} = \{B(2,\theta), 0 < \theta < 1\}$. In other words, the family of convex mixtures of two binomials $B(2,\theta_1)$ and $B(2,\theta_2)$ is not identifiable.

Example 8.1.2 : Let

$$p(k|\lambda) = \binom{n}{k}\lambda^k(1-\lambda)^{n-k}, k = 0, 1, 2, ..., n \tag{8.4}$$

and $G(\lambda)$ be an arbitrary mixing distribution on [0,1]. Let X be a random variable with the distribution which is a mixture of $p(k|\lambda)$ with respect to the mixing distribution G. Then, for $0 \leq k \leq n$,

$$\begin{aligned} \Pr(X = k) &= \int_0^1 p(k|\lambda)G(d\lambda) \\ &= \int_0^1 \binom{n}{k}\lambda^k(1-\lambda)^{n-k}G(d\lambda) \end{aligned} \tag{8.5}$$

and $\Pr(X = k)$ is a linear function of the first n moments of G, namely,

$$\mu_r = \int_0^1 \lambda^r dG(\lambda), 0 \leq r \leq n\,. \tag{8.6}$$

Hence any other distribution G^*, with the same first n moments as those of G, will yield the same value for $\Pr(X = k)$ as that given by G for $0 \leq k \leq n$. This shows the lack of identifiability of G with respect to the family $\{B(n,\lambda), 0 \leq \lambda \leq 1\}$ where n is known.

Example 8.1.3 : Let $U_{\alpha,\beta}(x)$ denote the uniform distribution function on the interval (α, β). It is easy to check that

$$U_{0,1}(x) = \alpha U_{0,\alpha}(x) + (1-\alpha)U_{\alpha,1}(x), -\infty < x < \infty \tag{8.7}$$

for any $0 < \alpha < 1$. In other words, the standard uniform distribution on (0,1) is a convex mixture of the uniform distributions on $(0, \alpha)$ and $(\alpha, 1)$ for every $\alpha, 0 \leq \alpha \leq 1$. This proves that the family of discrete distributions $\{G_\alpha, 0 < \alpha < 1\}$ with

$$\begin{aligned} G_\alpha(\beta) &= \alpha \quad \text{for } \beta = 0 \\ &= 1-\alpha \ \text{ for } \beta = \alpha \end{aligned} \tag{8.8}$$

is not identifiable with respect to the family $\{U(\alpha, \beta), 0 \leq \alpha, \beta \leq 1\}$. Hence the family of mixtures of uniform distributions is not identifiable.

Examples given above illustrate the fact that the problem of identifiability for mixtures is not artificial. Hence we would like to obtain sufficient conditions for identifiability in the later sections of this chapter. We point out that if a family Λ is identifiable with respect to a family $\mathcal{P}$, then any subfamily of Λ is also identifiable with respect to $\mathcal{P}$. This follows from the fact that the mapping $Q : \Lambda \to \zeta$ is one-to-one where ζ is the family of mixtures.

Remarks 8.1.1 : It is trivial to check that if Λ contains all degenerate distributions over Θ and if

$$P_{\theta'}(A) = \int_A P_\theta(A) dG(\theta), A \in \mathcal{F} \tag{8.9}$$

for some $\theta' \in \Theta$ and some nondegenerate distribution $G \in \Lambda$, then Λ is not identifiable.

According to our discussion, Λ or equivalently ζ is identifiable if the mapping $Q : \Lambda \to \zeta$ is one-to-one where ζ is the class of mixtures. In some of the literature, ζ is said to be identifiable (cf. Teicher (1954), Yakowitz and Spragins (1968)) in such an event. This need not create confusion among the readers in the light of explanation given earlier. In view of this duality, we interchangeably use the notion of identifiability either for ζ or for Λ depending on the context, convenience in interpretation and applicability.

8.2 Identifiability for Finite Mixtures

The following result gives a necessary and sufficient condition for the identifiability of a finite mixing distribution. A discrete mixing distribution with finite number of mass points is called a *finite mixing distribution.*

Theorem 8.2.1 (Yakowitz and Spragins (1968)) : A necessary and sufficient condition, on a family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ of probability measures so that the class Λ of all finite mixing distributions is identifiable relative to $\mathcal{P}$, is that the family $\mathcal{P} \equiv \{P_\theta, \theta \in \Theta \in \Theta\}$ is *linearly independent* as functions on $\mathcal{F}$.

Proof : Suppose the family $\mathcal{P}$ is not linearly independent as functions on

$\mathcal{F}$. Then there exists constants C_i not zero and $\theta_i \in \Theta, 1 \leq i \leq N$ such that

$$\sum_{i=1}^{N} C_i P_{\theta_i}(A) = 0, A \in \mathcal{F} . \tag{8.10}$$

Without loss of generality, assume that

$$C_1 \leq C_2 \leq \cdots \leq C_M < 0 < C_{M+1} < \cdots < C_N . \tag{8.11}$$

Then

$$\sum_{i=1}^{M} |C_i| P_{\theta_i}(A) = \sum_{i=M+1}^{N} |C_i| P_{\theta_i}(A) , \tag{8.12}$$

and hence

$$\sum_{i=1}^{M} |C_i| P_{\theta_i}(\mathcal{X}) = \sum_{M+1}^{N} |C_i| P_{\theta_i}(\mathcal{X}) \tag{8.13}$$

which proves that

$$\sum_{i=1}^{N} |C_i| = \sum_{i=M+1}^{N} |C_i| = b \text{ (say)} . \tag{8.14}$$

It is obvious that $b > 0$. Let $a_i = |C_i|/b, 1 \leq i \leq N$. Then, it follows that

$$\sum_{i=1}^{M} a_i P_{\theta_i}(A) = \sum_{i=M+1}^{N} a_i P_{\theta_i}(A), A \in \mathcal{F} \tag{8.15}$$

are two distinct representations of the same finite mixture. Hence Λ is not identifiable and equivalently $\mathcal{P}^*$, the family of convex mixtures of elements of $\mathcal{P}$, is not identifiable.

Conversely, suppose the family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ are linearly independent. Then they form a basis for the linear space $< \mathcal{P} >$ spanned by $\mathcal{P}$. Since $\zeta \subset < \mathcal{P} >$, the identifiability of Λ is a consequence of the uniqueness of the representation of elements in ζ with respect to the basis $\mathcal{P}$. ■

As a corollary to Theorem 8.2.1, the following result holds.

Corollary 8.2.1 : A necessary and sufficient condition on the family $\mathcal{P}$, so that the class Λ of all finite mixing distributions is identifiable with respect to the family $\mathcal{P}$, is that the image of $\mathcal{P}$ under any isomorphism on $< \mathcal{P} >$

consists of linearly independent elements in the image space. Here $< \mathcal{P} >$ is the linear space spanned by $\mathcal{P}$.

Proof : This result is a consequence of Theorem 8.2.1 by observing that the set $\mathcal{P}$ is linearly independent iff its image is linearly independent in the image space. ■

Remarks 8.2.1 : Corollary 8.2.1 is quite useful in checking identifiability. For instance, it is often convenient to check the linear independence of the family of Fourier transforms of distribution functions (characteristic functions) rather than the linear independence of the family of distribution functions themselves.

Example 8.2.1 : Let $\mathcal{P}$ be the family of distribution functions $\{F(x+\theta), -\infty < \theta < \infty\}$ where F is a given distribution function. We claim that the family $\mathcal{P}$ is linearly independent and hence the corresponding Λ of finite mixing distributions is identifiable. Let $\phi(t,\theta)$ denote the characteristic function of the distribution function $F(x+\theta)$. Then

$$\phi(t,\theta) = e^{it\theta}\phi(t,0), -\infty < t < \infty\,. \tag{8.16}$$

Since the correspondence between the characteristic functions and the distribution functions on the real line is one-to-one and linear, it is sufficient to prove that

$$\sum_{j=1}^{k} a_j\phi(t,\theta_j) = 0 \Rightarrow a_j = 0, 1 \le j \le k\ . \tag{8.17}$$

Note that

$$\sum_{j=1}^{k} a_j\phi(t,\theta_j) = \sum_{j=1}^{k} a_j e^{it\theta_j}\phi(t,0),\ -\infty < t < \infty\,. \tag{8.18}$$

Hence

$$\sum_{j=1}^{k} a_j\phi(t,\theta_j) = 0, -\infty < t < \infty \tag{8.19}$$

implies that

$$\sum_{j=1}^{k} a_j e^{it\theta_j} = 0 \tag{8.20}$$

in a neighbourhood $\{t : |t| < \delta\}$ of zero for some $\delta > 0$ since $\phi(t, 0) \neq 0$ in a neighbourhood of zero. Suppose $a_j \neq 0$ for $j = i_1, i_2, ..., i_\ell$. Without loss of generality, assume that $\ell = k$ and

$$a_1 < a_2 < \cdots < a_m < 0 < a_{m+1} < \cdots < a_k \ .$$

Then, it follows that

$$-\sum_{j=1}^{m} a_j e^{it\theta_j} = \sum_{j=m+1}^{k} a_j e^{it\theta_j}, -\delta < t < \delta \ .$$

Let $t = 0$. Then it follows that

$$-\sum_{j=1}^{m} a_j = \sum_{j=m+1}^{k} a_j = b(\text{say}).$$

Note that $b > 0$. We have

$$-\sum_{j=1}^{m} \frac{a_j}{b} e^{it\theta_j} = \sum_{j=m+1}^{k} \frac{a_j}{b} e^{it\theta_j}, -\delta < t < \delta \ . \tag{8.21}$$

The function on the left side of (8.21) can be interpreted as the characteristic function of a random variable X taking values $\theta_1, \theta_2, ..., \theta_m$ with probabilities $-\frac{a_1}{b}, \ldots, -\frac{a_m}{b}$ respectively. Similarly the right side of the equation (8.21) is the characteristic function of another random variable Y taking values $\theta_{m+1}, ..., \theta_k$ with probabilities $\frac{a_{m+1}}{b}, ..., \frac{a_{m+k}}{b}$. Relation (8.21) implies that

$$\phi_X(t) = \phi_Y(t) \text{ for } |t| < \delta \ . \tag{8.22}$$

Since $\phi_X(t)$ and $\phi_Y(t)$ are entire characteristic functions being linear functions of exponentials, it follows that

$$\phi_X(t) = \phi_Y(t) \text{ for all } t \tag{8.23}$$

which proves that X and Y should have the same distribution. However, the distributions of X and Y are different by earlier remarks. This is a contradiction. Hence $a_j = 0$ for $1 \leq j \leq k$.

Example 8.2.2 : Let $\mathcal{P}$ be the family of univariate normal distributions $N(\mu, \sigma^2), -\infty < \mu < \infty, 0 < \sigma^2 < \infty$. We claim that the corresponding family Λ of finite mixing distributions is identifiable. Corollary 8.2.1 implies that the identifiability will hold provided

$$\sum_{i=1}^{k} a_i e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2} = 0, -\infty < t < \infty \tag{8.24}$$

implies that $a_i = 0, 1 \le i \le k$. Observe that the function $\exp(\mu_i t + \frac{1}{2}\sigma_i^2 t^2)$ is the moment generating function of $N(\mu_i, \sigma_i^2)$. Let us choose $t_j, 1 \le j \le k$ such that the matrix $((\gamma_{ij}))$ is nonsingular where

$$\gamma_{ij} = \exp\{\mu_i t_j + \frac{1}{2}\sigma_i^2 t_j^2\}, 1 \le i, j \le k\ . \tag{8.25}$$

Since

$$\sum_{i=1}^{k} a_i \gamma_{ij} = 0, 1 \le j \le k\ ,$$

it follows that $a_1 = \cdots = a_k = 0$ as the matrix $((\gamma_{ij}))$ is nonsingular. Hence the family of finite mixtures of univariate normal distributions or equivalently the family Λ of finite mixing distributions is identifiable. This result does not hold if Λ is the class of all arbitrary mixing distributions (Teicher (1960)). See Example 8.4.2.

Remarks 8.2.2 : Suppose the family $\mathcal{P}$ consists of distributions with the property

$$\sum_{j=1}^{m} \theta_j \phi_j(t) = 0 \text{ for } |t| < \delta, \delta > 0$$

implies that $\theta_j = 0, 1 \le j \le m$ whenever $\phi_j(t), 1 \le j \le m$ are the characteristic functions of distributions $F_1, ..., F_m$ in $\mathcal{P}$. Then it follows that the family $\mathcal{P}$ is linearly independent and the class Λ of finite mixing distributions is identifiable.

Remarks 8.2.3 : Suppose the family $\mathcal{P}$ consists of a finite number of distribution functions $\{F_i(x), 1 \le i \le k\}$. Theorem 8.2.1 implies that the family of finite mixtures of $\mathcal{P}$ or equivalently the family Λ of finite mixing distributions on $\mathcal{P}$ is identifiable iff there exist k distinct values $x_1, x_2, ..., x_k$ such that

$$\begin{vmatrix} F_1(x_1) & \cdots & F_k(x_1) \\ F_1(x_2) & \cdots & F_k(x_2) \\ \ldots & \ldots & \ldots \\ F_1(x_k) & \cdots & F_k(x_k) \end{vmatrix} \neq 0 \,. \tag{8.26}$$

Similar results can be given in case the family $\mathcal{P}$ is defined either through density functions or through probability mass functions.

Example 8.2.3 : The family of finite mixtures of geometric distributions $P_\lambda, 0 < \lambda < 1$ defined by $P_\lambda(X = i) = \lambda^{i-1}(1 - \lambda), i \geq 1$ is identifiable. This can be shown by choosing $x_i = i, 1 \leq i \leq k$ and checking that

$$\begin{vmatrix} P_{\lambda_1}(X=1) & P_{\lambda_2}(X=1) & \ldots & P_{\lambda_k}(X=1) \\ P_{\lambda_1}(X=2) & P_{\lambda_2}(X=2) & \ldots & P_{\lambda_k}(X=2) \\ \ldots & \ldots & \ldots & \ldots \\ P_{\lambda_1}(X=k) & P_{\lambda_2}(X=k) & & P_{\lambda_k}(X=k) \end{vmatrix}$$

$$= \{\prod_{i=1}^{k}(1-\lambda_i)\} \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_k \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \ldots. & \lambda_k^{k-1} \end{vmatrix} \,. \tag{8.27}$$

Let $\mathcal{F}^k(\mu_1, \mu_2, ..., \mu_k)$ denote the class of all univariate distributions F such that the first $(k + 1)$ central moments of F are $\mu_0 = 1, \mu_1, ..., \mu_k$ respectively. Let

$$\alpha_i = E_{F_i}(X) = \int x F_i(dx) \tag{8.28}$$

and

$$\alpha_{i,r} = E_{F_i}[X^r] = \int x^r F_i(dx), r \geq 2 \,. \tag{8.29}$$

Then

$$\mu_{i,r} = \int (x - \alpha_i)^r F_i(dx) \,. \tag{8.30}$$

We are assuming that $\mu_{ir} < \infty, 1 \leq r \leq k$. For any *fixed* $\mu_i, 1 \leq i \leq k$, denote $\mathcal{F}^k(\mu_1, \mu_2, ..., \mu_k)$ by $\mathcal{F}^k$. An example of such a family is a translation parameter family as discussed in Example 8.2.1.

Theorem 8.2.2 (Rennie (1974)): The class Λ of finite mixing distributions is identifiable with respect to $\mathcal{P} = \{F_i, 1 \leq i \leq k\}$ where $F_i \in \mathcal{F}^{k-1}$ with unequal means.

Proof : It is sufficient to prove that $\mathcal{P}$ is linearly independent. Suppose

$$\sum_{i=1}^{k} C_i F_i(x) = 0, \ -\infty < x < \infty \ . \tag{8.31}$$

Then we claim that

$$\sum_{i=1}^{k} C_i \alpha_i^r = 0, 0 \leq r \leq k \ . \tag{8.32}$$

This can be seen by induction argument on r. Suppose $r = 0$. It is obvious that (8.32) holds for $r = 0$ by letting $x \to \infty$ in (8.31). Suppose the relation (8.32) holds for some $0 \leq m \leq r \leq k$. We will show that (8.32) holds for $m = r + 1$. Note that

$$\alpha_{i,r} = \sum_{m=0}^{r} \binom{r}{m} \mu_{r-m} \alpha^m \ . \tag{8.33}$$

Furthermore

$$\sum_{i=1}^{k} C_i \alpha_{i,r+1} = 0 \tag{8.34}$$

since $\sum_{i=1}^{k} C_i F_i(x) = 0, -\infty < x < \infty$ and

$$\begin{aligned}
\sum_{i=1}^{k} C_i \alpha_{i,r+1} &= \sum_{i=1}^{k} C_i [\sum_{m=0}^{r+1} \binom{r+1}{m} \mu_{r+1-m} \alpha_i^m] \\
&= \sum_{m=0}^{r+1} \binom{r+1}{m} \mu_{r+1-m} (\sum_{i=1}^{k} C_i \alpha_i^m)
\end{aligned}$$

$$= \sum_{m=0}^{r}\binom{r+1}{m}\mu_{r+1-m}(\sum_{i=1}^{k}C_i\alpha_i^m) + \binom{r+1}{r+1}\mu_0\sum_{i=1}^{k}C_i\alpha_i^{r+1}$$

$$= \sum_{i=1}^{k}C_i\alpha_i^{r+1} \tag{8.35}$$

by the induction hypotheses. Relations (8.34) and (8.35) prove that

$$\sum_{i=1}^{k}C_i\alpha_i^{r+1} = 0 \tag{8.36}$$

which completes the induction argument. Hence

$$\sum_{i=1}^{k}C_i\alpha_i^{r} = 0, 0 \leq r \leq k\ . \tag{8.37}$$

Writing the above set of equations in matrix from, we have

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_k \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \cdots & \alpha_k^{k-1} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \tag{8.38}$$

The matrix $((\alpha_i^j))_{k\times k}$ is the Vandermonde matrix with determinant $\prod_{1\leq i<j\leq k}(\alpha_j - \alpha_i)$ nonzero since $\alpha_i \neq \alpha_j$ for $i \neq j$ by assumption. Hence the matrix $((\alpha_i^j))_{k\times k}$ is nonsingular and it follows that $C_1 = C_2 = \cdots = C_k = 0$ which in turn implies the identifiability of finite mixtures of $\{F_1, F_2, ..., F_k\}$ whenever $E_{F_i}(X) \neq E_{F_j}(X)$ for $i \neq j, 1 \leq i, j \leq k$. ■

Remarks 8.2.4 : As a consequence of the above theorem, we obtain that the family of finite mixtures generated by two distributions with different means is identifiable. Similarly the family generated by three distributions with different means but common variances is identifiable and the family generated by four distributions with different means but common variance and common third absolute central moments (for example symmetric distributions) is identifiable.

For other applications of Theorem 8.2.1, see Yakowtiz and Spragins (1968).

Let us now consider another result due to Teicher (1963) which gives a sufficient condition for identifiability in the case of finite mixtures.

Theorem 8.2.3 : Suppose that to each $P \in \mathcal{P}$ is associated a transform ϕ with domain of definition D_ϕ and the mapping $M : \mathcal{P} \to \phi$ is linear. Further suppose that there is a total ordering $\leq$ of $\mathcal{P}$ such that $P_1 \leq P_2 \Rightarrow D_{\phi_1} \subset D_{\phi_2}$ and for *each* $P_1 \in \mathcal{P}$ there exists $t_1 \in \overline{T}_1 = \overline{\{t : \phi_1(t) \neq 0\}}$ such that

$$\lim_{\substack{t \to t_1 \\ t \in T_1}} \frac{\phi_2(t)}{\phi_1(t)} = 0 \tag{8.39}$$

whenever $P_1 < P_2, P_1, P_2 \in \mathcal{P}$. Then the class Λ of all finite mixing distributions is identifiable.

Proof: Suppose

$$\sum_{i=1}^{N} C_i P_i = 0, P_i \in \mathcal{P}, 1 \leq i \leq N\,. \tag{8.40}$$

Without loss of generality, assume that $P_i < P_j$ if $i < j$. By hypothesis,

$$\sum_{i=1}^{N} C_i \phi_i(t) = 0, -\infty < t < \infty\,. \tag{8.41}$$

Let $T_1 = \{t \in D_{\phi_1} : \phi_1(t) \neq 0\}$. For $t \in T_1$,

$$C_1 + \sum_{i=2}^{N} C_i \frac{\phi_i(t)}{\phi_1(t)} = 0 \tag{8.42}$$

and hence as $t \to t_1 \in \overline{T}_1$ through values of T_1, we get that $C_1 = 0$ by (8.39) . Hence

$$\sum_{i=2}^{N} C_i P_i = 0\,. \tag{8.43}$$

Repeating the process, we get that $C_i = 0, 1 \leq i \leq N$. Hence we have the identifiability of Λ. ■

Example 8.2.4 (Teicher (1963)): An application of Theorem 8.2.3 shows that the finite mixtures of gamma densities are identifiable. This can be checked in the following way. Consider the gamma density

$$\begin{aligned} f(x;\theta,\alpha) &= \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}, \quad 0 < x < \infty \\ &= 0 \qquad\qquad\qquad \text{otherwise} \end{aligned}$$

where $\theta > 0$ and $\alpha > 0$. The moment generating function of this density is given by

$$\psi(t;\theta,\alpha) = (\frac{\theta}{\theta - t})^\alpha = (1 - \frac{t}{\theta})^{-\alpha} \text{ for } -\infty < t < \theta\,. \tag{8.44}$$

Let us order the family of distributions $F(x;\theta,\alpha)$ corresponding to the densities $f(x;\theta,\alpha)$ by the ordering

$$F(x,\theta_1,\alpha_1) \leq F(x,\theta_2,\alpha_2) \tag{8.45}$$

if $\theta_1 < \theta_2$ or $\theta_1 = \theta_2$ but $\alpha_1 > \alpha_2$. Note that if $F_1 \equiv F(\cdot,\theta_1,\alpha_1) \leq F(\cdot,\theta_2,\alpha_2) \equiv F_2$, then $D_{\psi_1} = (-\infty,\theta_1)$ is contained in $D_{\psi_2} = (-\infty,\theta_2)$ and we can take $t_1 = \theta_1$ in Theorem 8.2.3. Furthermore

$$\lim_{t\to t_1} \frac{\psi(t,\theta_2,\alpha_2)}{\psi(t,\theta_1,\alpha_1)} = \lim_{t\to t_1} \frac{(1-\frac{t}{\theta_2})^{-\alpha_2}}{(1-\frac{t}{\theta_1})^{-\alpha_1}} = \frac{(1-\frac{t}{\theta_1})^{\alpha_1}}{(1-\frac{t}{\theta_2})^{\alpha_2}} = 0 \tag{8.46}$$

since $t_1 = \theta_1$. Hence the class of finite mixtures of gamma distributions is identifiable by Theorem 8.2.3. Choosing $\alpha = 1$, we note that the finite mixtures of exponential distributions are identifiable.

8.3 Identifiability of Finite Mixtures for Directional Data

One of the distributions that is widely used for modeling directional data (circular data) is the Von-Mises distribution with density given by

$$\begin{aligned} f(\theta;\alpha,k) &= (2\pi I_0(k))^{-1} \exp[k\cos(\theta-\alpha)], \quad 0 \leq \theta < 2\pi \\ &= 0 \qquad\qquad\qquad \text{otherwise} \end{aligned}$$

where $0 \leq \alpha < 2\pi, k > 0$ and $I_0(k)$ is the modified Bessel function of the first kind and order zero (cf. Mardia (1972)). However, when modeling multimodal directional data, finite mixtures of these distributions or finite mixtures of other circular distributions are used. Hence the question of identifiability of these mixtures is of importance prior to statistical inference aspects for directional data. Results in this direction are given in Fraser *et al.* (1981) and Kent (1983). Fraser *et al.* (1981) proved that the finite mixtures of Von-Mises distributions are identifiable using Theorem 8.2.1 due to Yakowitz and Spragins (1968). More general resuts are obtained in Kent (1983). We now discuss results from Kent (1983).

Let M be a connected manifold which can be embedded in the Euclidean space R^k. Let $E(M)$ denote the family of functions on M of the form $\exp\{P(x)\}$ where $P(x)$ is a polynomial on R^k of arbitrary but finite degree. We are interested in the identifiability of finite mixtures of probability densities on M with respect to some underlying σ-finite measure μ on M when the density is proportional to an element in $E(M)$. It is easy to see that the identifiability holds iff the collection $E(M)$ is identifiable as a collection of functions on M provided the support of μ contains an open subset of M. The last condition is natural since we are dealing with probability measures on M. In other words, for the study of identifiability, the form of μ is irrelevant and it is sufficient to discuss identifiability of $E(M)$ in the following sense following Theorem 8.2.1 due to Yakowitz and Spragins (1968).

A family τ of functions on M is called *identifiable* if all finite sets of essentially distinct functions in τ are *linearly independent.* That is, if $f_i(x), 1 \leq i \leq n$ are essentially distinct functions on M such that

$$\sum_{i=1}^{n} \lambda_i f_i(x) = 0, x \in M, \tag{8.47}$$

then $\lambda_1 = \cdots = \lambda_n = 0$. Here f_1 and f_2 are said to be *essentially distinct* if f_1 and f_2 are not proportional to each other.

Note that two distinct polynomials on R^k need not define two distinct polynomials on M. If $P_1(x) - P_2(x)$ is a constant for $x \in M$ where $P_i(x), 1 \leq i \leq 2$ are polynomials on R^k, then $\exp\{P_1(x)\}$ and $\exp\{P_2(x)\}$ define essentially the same function in $E(M)$ but not on R^k. For example, the polynomials $P(\boldsymbol{x}) \equiv 1$ and $P(\boldsymbol{x}) = (x_1^2 + x_2^2)^3, \boldsymbol{x} = (x_1, x_2) \in R^2$ are same on the unit circle $x_1^2 + x_2^2 = 1$ but are not the same on all of R^2.

It is clear from the definition of identifiability given above that if Γ_1 and Γ_2 are identifiable families of functions on manifolds M_1 and M_2, then the class $\{f(x)g(y) : f \in \Gamma_1, g \in \Gamma_2\}$ is identifiable on the product manifold $M_1 \times M_2$.

Suppose the manifold M is a direct product of Stiefel manifolds and copies of the real line R. A Stiefel manifold $0(p,k)$ can be embedded in R^{pk} as the set of $p \times k$ matrices $\boldsymbol{X}$ such that $\boldsymbol{X}^T\boldsymbol{X} = I_k$, the $k \times k$ identity matrix. If $k = p$, then we add the additional condition $\det(\boldsymbol{X}) = 1$. If $k = 1$, then we obtain the unit sphere in p-dimension as an example of a Stiefel manifold. Manifolds of this type occur in modeling directional data (Beran (1979), Johnson and Wehrly (1978), Mardia and Sutton (1978)).

Theorem 8.3.1 : Let M be a finite direct product of Stiefel manifolds and copies of the real line. Then the family $E(M)$ is identifiable.

Proof : In view of earlier remarks on identifiability on products of manifolds, it is sufficient to study identifiability on the real line and on all Stiefel manifolds.

Case (1) (M is a circle 0(2,1)) : Every point (x_1, x_2) on the unit circle can be represented in the form $x_1 = \cos\,\theta, x_2 = \sin\,\theta$ and every element in $E(0(2,1))$ can be represented uniquely in the form

$$g(\theta) = c\ \exp[\sum_{j=1}^{m} k_j\ \cos(j\theta - \alpha_j)] \tag{8.48}$$

for some $m \geq 0$. The parameters $k_j \geq 0$ are uniquely determined and, if $k_j > 0$, then $\alpha_j \in [0, 2\pi)$ is uniquely determined. Let

$$v_j(\sigma) = k_j\ \cos(j\sigma - \alpha_j), 1 \leq j \leq m \tag{8.49}$$

and

$$\boldsymbol{v}(\sigma)^T = (v_1(\sigma), ..., v_m(\sigma))\ . \tag{8.50}$$

Define $\boldsymbol{v}^{(1)}(\sigma) > \boldsymbol{v}^{(2)}(\sigma)$ if for some j with $1 \leq j \leq m$, $v_j^{(1)}(\sigma) > v_j^{(2)}(\sigma)$ and $v_{j'}^{(1)}(\sigma) = v_{j'}^{(2)}(\sigma)$ for $j < j' \leq m$ whenever $\boldsymbol{v}^{(1)}(\sigma)$ and $\boldsymbol{v}^{(2)}(\sigma)$ are of the same length. Any two vectors $\boldsymbol{v}(\sigma)$ of possibly different lengths can be compared by appending zeroes to the end of the shorter vector. This will give a total ordering on the collection $\{\boldsymbol{v}(\sigma)\}$.

In order to prove the identifiability for $E(M)$, it is sufficient to show that

$$\sum_{r=1}^{N} \lambda_r g^{(r)}(\theta) = 0, 0 \leq \theta < 2\pi \Rightarrow \lambda_1 = \cdots = \lambda_N = 0 \tag{8.51}$$

where

$$g^{(r)}(\theta) = \exp\{\sum_{j=1}^{m^{(r)}} k_j^{(r)} \cos(j\theta - \alpha_j^{(r)})\} . \tag{8.52}$$

Since $g^{(r)}(\theta), 1 \leq r \leq N$ are entire functions of θ, it follows that

$$\sum_{r=1}^{N} \lambda_r g^{(r)}(\theta) = 0, \theta = \sigma + i\tau, -\infty < \sigma < \infty, -\infty < \tau < \infty \tag{8.53}$$

from (8.53). Note that

$$\begin{aligned} \cos(j\sigma + ji\tau - \alpha_j^{(r)}) &= \cos(j\sigma - \alpha_j^{(r)})\cos(ij\tau) \\ &\quad - \sin(j\sigma - \alpha_j^{(r)})\sin(ij\tau) \\ &= \cos(j\sigma - \alpha_j^{(r)})\cosh(j\tau) \\ &\quad -i\,\sin(j\sigma - \alpha_j^{(r)})\sinh(j\tau) . \end{aligned} \tag{8.54}$$

Hence

$$\begin{aligned} |g^{(r)}(\sigma + i\tau)| &= \exp\{\sum_{j=1}^{m^{(r)}} k_j^{(r)} \cos(j\sigma - \alpha_j^{(r)})\cosh(j\tau)\} \\ &= \exp\{\sum_{j=1}^{m^{r)}} v_j^{(r)}(\sigma)\cosh(j\tau)\} \end{aligned} \tag{8.55}$$

and

$$\left| \frac{g^{(r)}(\sigma + i\tau)}{g^{(s)}(\sigma + i\tau)} \right| \to 0 \text{ as } \tau \to \infty \tag{8.56}$$

provided

$$\boldsymbol{v}^{(r)}(\sigma) < \boldsymbol{v}^{(s)}(\sigma) . \tag{8.57}$$

If $g^{(r)}$ and $g^{(s)}$ are two different components in the mixture, there exists at least one j such that

$$(k_j^{(r)}, \alpha_j^{(r)}) \neq (k_j^{(s)}, \alpha_j^{(s)}) . \tag{8.58}$$

Hence, for all but finitely many $\sigma \in [0, 2\pi]), \boldsymbol{v}_j^{(r)}(\sigma) \neq \boldsymbol{v}_j^{(s)}(\sigma)$. Therefore $\boldsymbol{v}^{(r)}(\sigma) \neq \boldsymbol{v}^{(s)}(\sigma)$ for all but finitely many $\sigma \in [0, 2\pi)$. Hence there exists at least one σ for which $\boldsymbol{v}^{(r)}(\sigma), 1 \leq r \leq N$ are all distinct. Choose such a σ and order the functions $g^{(r)}(\theta), 1 \leq r \leq n$ so that $\boldsymbol{v}^{(1)}(\sigma) > \cdots > \boldsymbol{v}^{(N)}(\sigma)$. Dividing (8.55) by $g^{(1)}(\theta)$ with $\theta = \sigma + i\tau$ and allowing $\tau \to \infty$, we get that $\lambda_1 = 0$. Proceeding in a similar way with the remaining terms, we obtain that $\lambda_1 = \cdots = \lambda_N = 0$. This proves the identifiability of $E(M)$ when M is a circle .

Case (2) (M is a Stiefel manifold) : We reduce the problem of identifiability to that of a circle discussed above in Case (1) and apply the result obtained therein.

Let us denote an element of the Stiefel manifold by a matrix $\boldsymbol{X}$ of order $p \times k$. Without loss of generality, we assume that $k = p$, for, if $k < p$, then any polynomial $P_1(\boldsymbol{X}_1)$ defined for $\boldsymbol{X}_1 \in 0(p, k)$ can be extended to $0(p, p)$ by the relation

$$P(\boldsymbol{X}) = P_1(\boldsymbol{X}_1), \boldsymbol{X} \in 0(p, p) \tag{8.59}$$

where $\boldsymbol{X}_1$ contains the first k columns of $\boldsymbol{X}$. Any linear relation between essentially distinct functions on $0(p, k)$ leads to another linear relation between essentially distinct functions on $0(p, p)$.

Let us assume $k = p$. Suppose

$$\sum_{i=1}^{N} \lambda_i f^{(i)}(\boldsymbol{X}) = 0 \tag{8.60}$$

for $\boldsymbol{X} \in M = 0(p, p)$ where $f^{(i)}(\boldsymbol{X}) = \exp\{P^{(i)}(\boldsymbol{X})\}$ and $f^{(i)}(\boldsymbol{X})$ are essentially distinct functions.

Note that $\boldsymbol{I}_p \in 0(p, p)$ where $\boldsymbol{I}_p$ is the identity matrix of order p. Without loss of generality, let us choose the constants in the polynomials $P^{(r)}(\boldsymbol{X}), r = 1, ..., N$ so that $P^{(r)}(\boldsymbol{I}_p) = 0$. Since $P^{(r)}(\boldsymbol{X})$ is an analytic function on the analytic manifold $0(p,p)$, it is determined by its values on any open subset in $0(p,p)$. Therefore, given any two distinct polynomials on $0(p,p)$, the points at which they differ must be dense in $0(p,p)$. Hence the points at which $P^{(r)}(\boldsymbol{X}), 1 \leq r \leq N$ take N distinct values are dense in $0(p,p)$. Let $\boldsymbol{X}^*$ be such a point in $0(p,p)$. Let

$$\boldsymbol{J}(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, 0 \leq \theta < 2\pi\,. \tag{8.61}$$

Define

$$\boldsymbol{B}(\theta_1, \theta_2, ..., \theta_q) = \begin{bmatrix} \boldsymbol{J}(\theta_1) & & & \\ & \boldsymbol{J}(\theta_2) & & \\ \cdots & \cdots & \cdots & \cdots \\ & & & \boldsymbol{J}(\theta_q) \end{bmatrix} \tag{8.62}$$

where $q = \frac{p}{2}$ when p is even and

$$\boldsymbol{B}(\theta_1, \theta_2, ..., \theta_q) = \begin{bmatrix} \boldsymbol{J}(\theta_1) & & & & \\ & \boldsymbol{J}(\theta_2) & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & \boldsymbol{J}(\theta_q) & \cdots \\ & & & & 1 \end{bmatrix} \tag{8.63}$$

for $q = [\frac{p}{2}]$ when p is odd. Here $\boldsymbol{B}(\theta_1, \theta_2, ..., \theta_q)$ is a block diagonal orthogonal matrix. By the decomposition theorem for orthogonal matrices (cf. Herstein (1964, p. 306)), it follows that there exists an orthogonal matrix $\boldsymbol{H}$ such that

$$\boldsymbol{X}^* = \boldsymbol{H}\boldsymbol{B}(\theta_1^*, ..., \theta_q^*)\boldsymbol{H}^T \tag{8.64}$$

where $0 \leq \theta_i^* < 2\pi, 1 \leq i \leq q$. Consider the submanifold

$$M_0 = \{\boldsymbol{H}\boldsymbol{B}(\theta_1, \ldots, \theta_q)\boldsymbol{H}^T : 0 \leq \theta_i < 2\pi, 1 \leq i \leq q\}. \tag{8.65}$$

Then $M_0 \subset 0(p,p)$ and M_0 is a multidimensional torus containing both $\boldsymbol{I}_p$ and $\boldsymbol{X}^*$. Any polynomial in $\boldsymbol{X}$ can be regarded as a polynomial in $(\cos\theta_i, \sin\theta_i), 1 \leq i \leq q$ on M_0 and the functions $f^{(r)}(\boldsymbol{X}), 1 \leq r \leq N$ can be considered as essentially distinct functions in $E(M_0)$.

In view of the fact that the property of identifiability is closed under direct products and the result for a circle holds as proved in Case (1), it follows that the result as stated in the theorem holds for $E(M_0)$. In other words $f^{(r)}(\boldsymbol{X}), 1 \leq r \leq N$ are linearly independent and $\lambda_1 = \cdots = \lambda_N = 0$.

Case (3) (M **is the real line**): Here the proof follows along the same method as that given in Case (1). Observe that the ratio for any two distinct functions in $E(R)$ tends to zero or infinity as $X \to \infty$.

This completes the proof of Theorem 8.3.1. ∎

Remarks 8.3.1 : As a consequence of Theorem 8.3.1, it follows that the finite mixtures of Von-Mises densities

$$g_1(\theta) = \sum_{i=1}^{m} \lambda_i \exp(K_i \cos(\theta - \alpha_i)) \tag{8.66}$$

are identifiable. Similarly finite mixtures of densities of the form

$$g_2(\theta) = c \exp\{\sum_{j=1}^{m} \gamma_j \cos(j\theta - \beta_j)\} \tag{8.67}$$

are identifiable. In addition, it follows that the finite mixtures of multivariate normal distributions on R^n are identifiable (see Remarks 8.6.2).

8.4 Identifiability for Countable Mixtures

Let $\{F_i, i \geq 1\}$ be a sequence of distribution functions and

$$F(x) = \sum_{i=1}^{\infty} \beta_i F_i(x) \tag{8.68}$$

where $\sum_i |\beta_i| < \infty, \sum \beta_i = 1$. F is called a *countable mixture* of $\{F_i\}$. Note that β_i could be negative. If β_i are all nonnegative, then F will be a distribution function. The mixture F or equivalently the sequence $\{\beta_i\}$ is said to be *identifiable* if

$$F(x) = \sum_{i=1}^{\infty} \beta_i F_i(x), \sum_{i=1}^{\infty} |\beta_i| < \infty, \sum_{i=1}^{\infty} \beta_i = 1 \tag{8.69}$$

and

$$F(x) = \sum_{i=1}^{\infty} \gamma_i F_i(x), \sum_{i=1}^{\infty} |\gamma_i| < \infty, \sum_{i=1}^{\infty} \gamma_i = 1 \tag{8.69A}$$

imply that

$$\beta_i = \gamma_i, i \geq 1 .$$

In other words, the representation (8.68) is unique. The problem is to find conditions on the family $\{F_i\}$ for the identifiability of the mixture F.

The infinite set $\{F_i, i \geq 1\}$ is said to be *linearly independent* if *every* finite subset is linearly independent. It is said to be *strongly linearly independent* if

$$\sum_{i=1}^{\infty} a_i F_i(x) = 0, \sum_{i=1}^{\infty} |a_i| < \infty \Rightarrow a_i = 0 \text{ for all } i \geq 1. \tag{8.70}$$

Theorem 8.4.1: A necessary condition that the mixture F defined by (8.68) is identifiable is that the set $\{F_i, i \geq 1\}$ is linearly independent.

Proof : Suppose the set $\{F_i, i \geq 1\}$ is not linearly independent. Then there exists a finite subset which is linearly dependent. By renumbering if necessary, we can assume without loss of generality that

$$F_k(x) = \sum_{i=1}^{k-1} a_i F_i(x). \tag{8.71}$$

Hence

$$F(x) = \sum_{i=1}^{k-1} (\beta_i + a_i \beta_k) F_i(x) + \sum_{j=k+1}^{\infty} \beta_j F_j(x). \tag{8.72}$$

An alternate representation for $F(x)$ is

$$F(x) = \sum_{i=1}^{\infty} \beta_i' F_i(x) \tag{8.73}$$

where

$$\beta_i' = \beta_i + \varepsilon a_i, 1 \leq i \leq k-1, \tag{8.74}$$

$$\beta_k' = \beta_k - \varepsilon \tag{8.74A}$$

and

$$\beta_i' = \beta_i, i > k. \qquad (8.74B).$$

This can be seen from the fact

$$\sum_{i=1}^{k-1} a_i = 1 \qquad (8.75)$$

which follows from the equation (8.71) by letting $x \to +\infty$. It is easy to see that

$$\sum |\beta_i'| < \infty, \sum \beta_i' = 1.$$

The relations (8.74) and (8.75) give two distinct representations for $F(x)$. Hence the mixture F is not identifiable. ■

Remarks 8.4.1 : The condition of linear independence of the set $\{F_i\}$ is a necessary and sufficient condition for the identifiability finite mixtures. It is a necessary condition for the identifiability of countable mixtures. As we will show below, it is *not* a sufficient condition for the identifiability of countable mixtures.

Let $\{F_i^*, i \geq 1\}$ be a strongly linearly independent family. Define

$$F_{i+1} = F_i^*, i \geq 1$$

and

$$F_1 = \sum_{i=1}^{\infty} \beta_i F_i^*$$

where $\beta_i > 0, i \geq 1$ and $\sum_{i=1}^{\infty} \beta_i = 1$. Then the set $\{F_i\}$ is linearly independent but not strongly linearly independent and hence mixtures of $\{F_i\}$ are not identifiable. It is easy to see that the mixture F defined by (8.68) is identifiable iff the set $\{F_i, i \geq 1\}$ is strongly linearly dependent.

In view of Theorem 8.4.1, we will assume that $\{F_i\}$ is linearly independent. Suppose $F_i \in L^2(R)$, the space of square integrable functions with respect to the Lebesgue measure on R. Applying Gram–Schmidt orthogonalization process, we can obtain an associated orthonormal system $\{\phi_j\}$

under the inner product

$$< f, g >= \int_{-\infty}^{\infty} f(x)\bar{g}(x)dx, f, g \in L^2(R). \tag{8.76}$$

Let

$$k_{ij} = \int_{-\infty}^{\infty} \phi_i(x)F_j(x)dx \tag{8.77}$$

and $\boldsymbol{K} = ((k_{ij}))$. $\boldsymbol{K}$ is an infinite (dimensional) matrix. For results on infinite matrices, see Cooke (1950) and Kantorovich and Krylov (1959). Dienes (1932) discusses linear equations in infinite matrices.

Remarks 8.4.2 : Suppose $F_i(x) \leq H(x)$ for all x and $H \in L^2(R)$. Assume that there exists a vector $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \ldots)$ such that

$$F(x) = \sum_{i=1}^{\infty} \beta_i F_i(x) . \tag{8.78}$$

Let

$$\alpha_i = \int_{-\infty}^{\infty} \phi_i(x)F(x)dx . \tag{8.79}$$

Then

$$\alpha_i = \int_{-\infty}^{\infty} \phi_i(x) \lim_{k\to\infty} \{\sum_{j=1}^{k} \beta_j F_j(x)\}dx \tag{8.80}$$

$$= \lim_{k\to\infty} \int_{-\infty}^{\infty} \phi_i(x)\{\sum_{j=1}^{k} \beta_j F_j(x)\}dx \tag{8.81}$$

by the dominated convergence theorem since

$$|\phi_i(x) \sum_{j=1}^{k} \beta_i F_j(x)| \leq |\phi_i(x)|H(x) \sum_{j=1}^{\infty} |\beta_j| \tag{8.82}$$

for all k and the function $\phi_i(x)H(x)$ is integrable. Hence

$$\alpha_i = \lim_{k\to\infty} \sum_{j=1}^{k} \beta_j K_{ij}$$

$$= \sum_{j=1}^{\infty} \beta_j K_{ij} . \tag{8.83}$$

Let

$$\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \ldots). \tag{8.84}$$

Then it follows that

$$\boldsymbol{\alpha} = \boldsymbol{K\beta}. \tag{8.85}$$

On the other hand, suppose there exists a solution $\boldsymbol{\beta}$ to the equation $K\boldsymbol{\beta} = \boldsymbol{\alpha}$. Then

$$\int_{-\infty}^{\infty} \phi_i(x)[F(x) - \sum_{j=1}^{\infty} \beta_j F_j(x)]dx = 0 \tag{8.86}$$

for all $i \geq 1$. Let Γ be the closed subspace spanned by $\{\phi_i, i \geq 1\}$ or equivalently by $\{F_j, j \geq 1\}$. Relation (8.86) shows that

$$F(x) = \sum_{j=1}^{\infty} \beta_j F_j(x) \text{ a.e.}$$

Hence we have the following theorem .

Theorem 8.4.2 : Suppose

$$F_i(x) \leq H(x) \quad \text{where } H \in L^2(R) \tag{8.87}$$

for all $i \geq 1$. If $\boldsymbol{\beta}$ is a solution of the equation

$$F(x) = \sum_{j=1}^{\infty} \beta_j F_j(x) \text{ a.e} \tag{8.88}$$

then $\boldsymbol{\alpha} = \boldsymbol{K\beta}$. Conversely if $\boldsymbol{\alpha} = \boldsymbol{K\beta}$, then $\boldsymbol{\beta}$ is a solution of the equation (8.90).

Remarks 8.4.3 : The above theorem continues to hold without the condition (8.87) if we insist that $\beta_i \geq 0$ for all $i \geq 1$ in (8.68). The result follows from an application of the monotone convergence theorem in equations (8.80) and (8.81).

Remarks 8.4.4 : It is clear that the solution $\boldsymbol{\beta}$ for the equation

$$\boldsymbol{\alpha} = \boldsymbol{K\beta} \tag{8.89}$$

is unique iff $\boldsymbol{K}^{-1}$ exists. In fact, in such an event,

$$\boldsymbol{\beta} = \boldsymbol{K}^{-1}\boldsymbol{\alpha} . \tag{8.90}$$

Let us consider a mixture

$$G(x) = \sum_{j=1}^{\infty} w_j F_j(x), \sum |w_j| < \infty, \sum w_j = 1 . \tag{8.91}$$

Suppose

$$\sum_{j=1}^{\infty} w_j F_j(x) = \sum_{j=1}^{\infty} y_j \phi_j(x) \text{ a.e.} \tag{8.92}$$

where $\{\phi_j\}$ is the orthonormal system for $L^2(R)$ described earlier. Multiplying both sides by ϕ_i and integrating over the real line, we have

$$\sum_{j=1}^{\infty} w_j K_{ij} = y_i, i \geq 1 \tag{8.93}$$

or equivalently

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{w}. \tag{8.94}$$

Let

$$d_{ij} = \int_{-\infty}^{\infty} F_i(x) F_j(x) dx, i \geq 1, j \geq 1. \tag{8.95}$$

Multiplying both sides of (8.95) by F_i and integrating over the real line, we have

$$\sum_{j=1}^{\infty} w_j d_{ij} = \sum_{j=1}^{\infty} y_i K_{ji} \tag{8.96}$$

or equivalently

$$\boldsymbol{D}\boldsymbol{w} = \boldsymbol{K}^T \boldsymbol{y} \tag{8.97}$$

where $\boldsymbol{D} = ((d_{ij}))$. Relations (8.94) and (8.97) prove that

$$\boldsymbol{D}\boldsymbol{w} = \boldsymbol{K}^T \boldsymbol{K}\boldsymbol{w} . \tag{8.98}$$

It can be shown that $\boldsymbol{K}^{-1}$ exists iff $\boldsymbol{D}^{-1}$ exists. Hence the countable mixture F is identifiable iff $\boldsymbol{D}^{-1}$ exists. Recall that we have assumed that the set $\{F_i\}$ is linearly independent and $F_i(x) \leq H(x) \in L^2(R)$ for all $i \geq 1$.

Remarks 8.4.5: It is easy to see that the condition that $\boldsymbol{D}^{-1}$ exists is also necessary and sufficient for identifiability if we consider convex mixtures of $\{F_i\}$, that is, mixtures of the form

$$\sum_{i=1}^{\infty} \beta_i F_i(x), \beta_i > 0, \sum_{i=1}^{\infty} \beta_i = 1. \tag{8.99}$$

Furthermore, the results obtained above continue to hold if we replace F_i by its density f_i or by its characteristic function ϕ_i for every i.

Example 8.4.1 : Suppose

$$f_i(x) = \begin{cases} 1 & \text{if } \frac{i-1}{2} \leq x \leq \frac{i+1}{2} \\ 0 & \text{otherwise} \end{cases}$$

for $i \geq 1$. Let

$$d_{i,j} = \int_{-\infty}^{\infty} f_i(x) f_j(x) dx \,.$$

It is easy to see that d_{ij} is either 0, $\frac{1}{2}$ or 1. In fact, for any $i \geq 1$,

$$d_{i,i-1} = \frac{1}{2}, d_{i,i} = 1, d_{i,i+1} = \frac{1}{2}$$

and $d_{i,j} = 0$ for all other j. The equation

$$\boldsymbol{D}\boldsymbol{x} = 0$$

leads to the set of equations $\frac{1}{2}x_{i-1} + x_i + \frac{1}{2}x_{i+1} = 0, i \geq 1$ where we define $x_0 = 0$ and the condition $\sum_{i=1}^{\infty} |x_i| < \infty$ holds. Let

$$g(s) = \sum_{i=1}^{\infty} x_i s^i \,.$$

Then it follows that

$$g(s)(1 + \frac{s}{2} + \frac{1}{2s}) = \frac{1}{2}x_1 \,.$$

If $x_1 \neq 0$, then

$$g(s) = \frac{\frac{1}{2}x_1}{1 + \frac{s}{2} + \frac{1}{2s}}, 0 < s \leq 1$$

and the power series expansion of g is obviously not of the form $\sum_{i=1}^{\infty} x_i s^i$ with positive powers of s. Hence $x_1 = 0$ which in turn implies that $g(s) = 0$ for $0 < s \leq 1$. Clearly $g(0) = 0$. Hence $g(s) \equiv 0$ for $0 \leq s \leq 1$ which shows that $\boldsymbol{x} = 0$. This proves that $\boldsymbol{D}^{-1}$ exists and the family of convex mixtures of $\{f_i, i \geq 1\}$ is identifiable.

Example 8.4.2 : Let

$$\begin{aligned} f_i(x) &= 2^i \quad \text{if } 1 - \tfrac{1}{2^{i-1}} \leq x \leq 1 - \tfrac{1}{2^i} \\ &= 0 \quad \text{otherwise} \end{aligned}$$

for $i \geq 1$. Here $\boldsymbol{D}$ is a diagonal matrix and $\boldsymbol{Dx} = 0$ iff $\boldsymbol{x} = 0$. Hence the family of convex mixtures of $\{f_i, i \geq 1\}$ is identifiable.

Remarks 8.4.6 : Results in this section are due to Tallis (1969) with slight modification. Patil and Bildikar (1966) discussed identifiability of countable mixtures of discrete probability distributions using methods of infinite matrices. Luexmann(1987) investigated the identifiability of mixtures of infinitely divisible power series distributions.

8.5 Identifiability for Arbitrary Mixtures

Corollary 8.2.1 deals with a necessary and sufficient condition for the identifiability of the class Λ of finite mixing distributions with respect to a family $\mathcal{P}$ of probability measures. In general, this result does not hold for the class Λ of arbitrary mixing distributions. For instance, the class of arbitrary mixtures of normal distributions $\{N(\mu, \sigma^2), -\infty < \mu < \infty, 0 < \sigma^2 < \infty\}$ is *not* identifiable (Teicher (1960)) whereas the class finite mixtures of normal distributions $\{N(\mu, \sigma^2), -\infty < \mu < \infty, 0 < \sigma^2 < \infty\}$ forms an identifiable family as shown in Example 8.2.2.

We shall now obtain some sufficient conditions for identifiability of arbitrary mixtures.

Let $\{f(\cdot, \theta), \theta \in \Theta\}$ be a family of densities on the real line where Θ is an interval on the real line. Let G be a probability distribution on Θ and define

$$f_G(x) = \int_\Theta f(x, \theta) dG(\theta), -\infty < x < \infty. \tag{8.100}$$

Let $\mathcal{P} = \{f(\cdot, \theta), \theta \in \Theta\}$ and $\Gamma = \{f(x, \cdot), -\infty < x < \infty\}$. Let $C_0(\Theta)$ be the Banach space of continuous functions on the interval Θ vanishing at infinity and normed by

$$||g|| = \sup_{y \in \Theta} |g(y)| \tag{8.101}$$

for $g \in C_0(\Theta)$.

Theorem 8.5.1 (Blum and Susarla (1977)): Suppose $\Gamma \subset C_0(\Theta)$. Then the family Λ of arbitrary mixing distributions is identifiable, that is,

$$f_G(x) = f_H(x), -\infty < x < \infty \Rightarrow H(\theta) = G(\theta), \theta \in \Theta \tag{8.102}$$

iff Γ generates $C_0(\Theta)$ under the supremum norm defined by (8.103).

Proof : Suppose the family Λ is identifiable. Let B be the closed subspace of $C_0(\Theta)$ generated by Γ. If possible, suppose there exists $g \in C_0(\Theta) - B, g \neq 0$. By the Hahn–Banach theorem, there exists a bounded linear functional ψ on $C_0(\Theta)$ such that

$$\psi(g) = 1 \text{ and } \psi(h) = 0,\ h \in B. \tag{8.103}$$

But, by the Riesz representation theorem, there exist nondecreasing non-negative functions K_1 and K_2 of bounded variation on Θ such that

$$\psi(f) = \int_\Theta f(\theta) d(K_1 - K_2)(\theta), f \in C_0(\Theta)\,. \tag{8.104}$$

Hence

$$\int_\Theta h(\theta) dK_1(\theta) = \int_\Theta h(\theta) dK_2(\theta), h \in B \tag{8.105}$$

by (8.103). This proves that $K_1(\theta) = K_2(\theta) + C$ for some constant C by the identifiability of Λ and the fact that B is generated by Γ. Hence $\psi(f) = 0, f \in C_0(\theta)$. In particular $\psi(g) = 0$. This contradicts the fact that

$\psi(g) = 1$ given by (8.103). Hence there exists no element $g \in C_0(\Theta) - B, g \neq 0$. In other words, Γ generates $C_0(\Theta)$.

Conversely, assume that Γ generates $C_0(\Theta)$. Suppose

$$f_G(x) = \int_\Theta f(x,\theta)G(d\theta) = \int_\Theta f(x,\theta)H(d\theta) = f_H(x), -\infty < x < \infty \tag{8.106}$$

for some probability distributions G and H on Θ. Since Γ generates $C_0(\Theta)$ under the supremum norm, it is easy to check that

$$\int_\Theta g(\theta)dG(\theta) = \int_\Theta g(\theta)dH(\theta), g \in C_0(\Theta). \tag{8.107}$$

Let

$$\psi(g) = \int_\Theta g(\theta)dG(\theta), g \in C_0(\Theta). \tag{8.108}$$

Then $\psi(\cdot)$ is a bounded linear functional on $C_0(\theta)$ since G is of bounded variation on Θ. From the uniqueness in the Riesz representation theorem, it follows that $G-H$ is a constant. Since G and H are probability distributions on Θ, it follows that $G(\theta) = h(\theta), \theta \in \Theta$ proving the identifiability of the class Λ. ∎

Remarks 8.5.1 : Theorem 8.5.1 essentially generalizes Theorem 8.2.1 to the family Λ of arbitrary mixing distributions. Teicher (1961) extended the result discussed in Example 8.2.1 to the family Λ of arbitrary mixing distributions. His result is as follows: suppose F is a distribution with characteristic function $|\phi(t)| > 0$ in a neighbourhood of zero; then the family Λ is identifiable with respect to the family $\{F(x+\theta); \theta \in \Theta\}$ for any interval Θ contained in R. We omit the proof.

Example 8.5.1: Suppose $f(x, \lambda)$ is normal density with mean λ and variance one. Let us define

$$f_G(x) = \int_{-\infty}^{\infty} f(x,\lambda)dG(\lambda), -\infty < x < \infty\,. \tag{8.109}$$

Then $f_G(x)$ is the density of a mixture of normal densities with mean λ and variance one with mixing distribution $G(\lambda)$. The characteristic function

$\psi(t)$ of the mixture $f_G(\cdot)$ is

$$\begin{aligned}\psi_G(t) \equiv \int_{-\infty}^{\infty} e^{itx} f_G(x)dx &= \int_{-\infty}^{\infty} e^{itx}[\int_{-\infty}^{\infty} f(x,\lambda)dG(\lambda)]dx \\ &= \int_{-\infty}^{\infty}[\int_{-\infty}^{\infty} e^{itx} f(x,\lambda)dx]dG(\lambda) \\ &= e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{i\lambda t} dG(\lambda)\,. \end{aligned} \tag{8.110}$$

All the above equations are justified by Fubini's theorem. Let $\phi_G(t)$ denote the characteristic function of the distribution function G. It follows that

$$\psi_G(t) = e^{-\frac{t^2}{2}} \phi_G(t), -\infty < t < \infty. \tag{8.111}$$

This relation shows that there is a one-to-one correspondence between the characteristic function corresponding to G and the characteristic function corresponding to f_G. Hence, it follows that the distribution function corresponding to f_G is uniquely determined by the distribution function G. Thus G is identifiable and the family of mixtures of $\{N(\lambda, 1), -\infty < \lambda < \infty\}$ is identifiable. It is clear that the result holds for any family of normal distributions with *specified* variance. As we have already mentioned earlier, the result is not true if the variance is not specified (Teicher (1960)). See the next example for details.

Example 8.5.2 (Teicher 1960): Let $\mathcal{P} = \{N(\theta, \sigma^2), -\infty < \mu < \infty,$ $0 < \sigma^2 < \infty\}$ and G be a probability measure on the space $R \times R^+$. Let

$$H(x) = \int_{R\times R^+} \Phi(x; \theta, \sigma^2) dG(\theta, \sigma^2) \tag{8.112}$$

where $\Phi(x; \theta, \sigma^2)$ denotes the normal distribution function with mean θ and variance σ^2. Let $G_{\theta|\sigma^2}(\cdot|\sigma^2)$ be the conditional distribution function of θ given σ^2. Note that

$$\begin{aligned} H(x) &= \int_{R\times R^+} \Phi(\frac{x-\theta}{\sigma}; 0.1) dG(\theta, \sigma^2) \\ &= \int_{R^+}\int_{R} \Phi(\frac{x-\theta}{\sigma}; 0, 1) dG_{\theta|\sigma^2}(\theta|\sigma^2) dG_1(\sigma^2) \end{aligned} \tag{8.113}$$

where $G_1(\sigma^2)$ denotes the marginal distribution of σ^2. Therefore

$$\begin{aligned} H(x) &= \int_0^\infty [\int_{-\infty}^\infty \Phi(\frac{x-\theta}{\sigma};0,1)dG_{\theta|\sigma^2}(\theta|\sigma^2)]dG_1(\sigma^2) \\ &= \int_0^\infty [\Phi(x;0,\sigma^2) * G_{\theta|\sigma^2}(x|\sigma^2)]dG_1(\sigma^2) \end{aligned} \tag{8.114}$$

where $*$ denotes convolution. Let $h(t)$ denote the charcteristic function of the distribution function $H(x)$ and $\psi(t|\sigma^2)$ denote the characteristic function of the conditional distribution function $G_{\theta|\sigma^2}(x|\sigma^2)$. Then

$$h(t) = \int_0^\infty e^{-\sigma^2 t^2/2}\psi(t|\sigma^2)dG_1(\sigma^2). \tag{8.115}$$

Suppose the probability measure G is such that the conditional distribution of θ given σ^2 is symmetric. Then $\psi(t|\sigma^2)$ is a real-valued function and relations (8.114) and (8.115) prove that H is a G_1-mixture of normal distribution function $\Phi(x;\theta,\sigma^2)$. However H is also a G-mixture of normal distribution functions $\Phi(x;\theta,\sigma^2)$ from (8.112). Hence arbitrary mixtures of normal distributions are not identifiable.

Remarks 8.5.2 (Convolution): Suppose

$$H(x) = \int_{-\infty}^\infty K(x-\lambda)G(dx), -\infty < x < \infty \tag{8.116}$$

where K and G are distribution functions. In other words $H = K * G$. Convolution is a special mixture of distributions. We claim that H identifies G if the characteristic function of G is analytic. This can be seen from following observations. Suppose

$$H(x) = \int_{-\infty}^\infty K(x-\lambda)G_1(d\lambda) = \int_{-\infty}^\infty K(x-\lambda)G_2(d\lambda), -\infty < x < \infty. \tag{8.117}$$

Then

$$\phi_K(t)\phi_{G_1}(t) = \phi_H(t) = \phi_K(t)\phi_{G_2}(t) \tag{8.118}$$

for all t where $\phi_F(t)$ denotes the characteristic function of the distribution function F. Since $\phi_K(t)$ does not vanish in a neighbourhood (say) V of zero, it follows that

$$\phi_{G_1}(t) = \phi_{G_2}(t), t \in V\ . \tag{8.119}$$

This will in general not prove that $G_1 = G_2$. However, if the characteristic functions of G_1 and G_2 are analytic, then it follows that

$$\phi_{G_1}(t) = \phi_{G_2}(t), -\infty < t < \infty \tag{8.120}$$

and $G_1 = G_2$. However, if $\phi_K(t) \neq 0$ for all t, then the relation (8.118) implies that $\phi_{G_1}(t) = \phi_{G_2}(t)$ for all t, and hence $G_1 = G_2$. For instance, if K is an infinitely divisible distribution, then $\phi_K(t) \neq 0$ for all t and hence $G_1 = G_2$. In particular, if $K(\cdot)$ is a normal distribution function, then $G_1 = G_2$.

Remarks 8.5.3 (Additively closed families) : Suppose we consider mixtures of the form

$$H(x) = \int_{-\infty}^{\infty} K(x, \lambda)G(d\lambda), -\infty < x < \infty \tag{8.121}$$

where K belongs to an additively closed family of distribution functions in the sense

$$K(x, \lambda_1) * K(x, \lambda_2) = K(x, \lambda_1 + \lambda_2) \tag{8.122}$$

and $*$ denotes convolution. An example of an additively closed family is $P(\lambda), 0 < \lambda < \infty$ where $P(\lambda)$ denotes the Poisson distribution with parameter λ. Let $\phi_k(t, \lambda)$ denote the characteristic function of $K(x, \lambda)$. Then

$$\phi_k(t, \lambda_1)\phi_k(t, \lambda_2) = \phi_k(t, \lambda_1 + \lambda_2), -\infty < t < \infty\ . \tag{8.123}$$

Since $\phi_k(t, 1)$ is a measurable function, the only measurable solution of the above functional equation is

$$\phi_k(t, \lambda) = e^{\lambda c(t)}, -\infty < t < \infty \tag{8.124}$$

for some function $c(t)$. Hence

$$\phi_k(t, \lambda) = [\phi_k(t, 1)]^{\lambda}, -\infty < t < \infty\ . \tag{8.125}$$

Since $\phi_k(t, \lambda)$ is a characteristic function, it follows that $\lambda \geq 0$ and G has to be a measure on $[0, \infty)$. Let $\phi_H(t)$ denote the characteristic function of H.

Then

$$\begin{aligned}\phi_H(t) &= \int_0^\infty \phi_K(t,\lambda)G(d\lambda)\\ &= \int_0^\infty [\phi_K(t,1)]^\lambda G(d\lambda), -\infty < t < \infty. \end{aligned} \tag{8.126}$$

Suppose that G_1 and G_2 are two probability measures with support on $[0, \infty)$ such that

$$H(x) = \int_0^\infty K(x,\lambda)G_1(d\lambda) = \int_0^\infty K(x,\lambda)G_2(d\lambda) \tag{8.127}$$

for all x. Then

$$\phi_H(t) = \int_0^\infty [\phi_K(t,1)]^\lambda G_1(d\lambda) = \int_0^\infty [\phi_K(t,1)]^\lambda G_2(d\lambda) \tag{8.128}$$

for $-\infty < t < \infty$. Let

$$\psi_G(z) = \int_0^\infty z^\lambda G(d\lambda). \tag{8.129}$$

The function $\psi_G(z)$ is analytic in $\{z : 0 < |z| < 1\}$. Since $\psi_{G_1}(z) = \psi_{G_2}(z)$ for $z = \phi_k(t, 1)$ and for all $t \in R$, it follows that $\psi_{G_1}(z) = \psi_{G_2}(z)$ for $0 < |z| < 1$. In particular

$$\psi_{G_1}(\rho e^{it}) = \psi_{G_2}(\rho e^{it}) \tag{8.130}$$

for $0 < \rho < 1$ and $-\infty < t < \infty$. Applying the dominated convergence theorem, it follows that (8.130) holds for $\rho = 1$. Therefore

$$\int_0^\infty e^{it\lambda}G_1(d\lambda) = \int_0^\infty e^{it\lambda}G_2(d\lambda), -\infty < t < \infty\ . \tag{8.131}$$

Since the characteristic functions of G_1 and G_2 are identical, it follows that $G_1 = G_2$ by the inversion theorem.

Results discussed above are due to Teicher (1961). We leave it to the reader to check that the family of mixtures of gamma densities

$$k(x,\lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, 0 < \lambda < \infty \tag{8.132}$$

is identifiable (assuming that α is known) using the above result.

Remarks 8.5.4: It is possible to use the techniques from the theory of integral equations for identifiability. Let us suppose that

$$F(x) = \int_{-1}^{1} F(x,\theta)dG(\theta), -\infty < x < \infty \tag{8.133}$$

where $F(x,\theta)$ is a distribution function for every θ. Here G is a function of bounded variation with $G(1) = 1$ and $G(-1) = 0$. Suppose that

$$T(x,\theta) = \frac{\partial F(x,\theta)}{\partial\theta} \tag{8.134}$$

exists and is continuous in θ. Further assume that T is square integrable on $[-1,1] \times [-1,1]$ with respect to the Lebesgue measure. Then

$$\begin{aligned} F(x) &= \int_{-1}^{1} F(x,\theta)dG(\theta) \\ &= [F(x,\theta)G(\theta)]_{-1}^{1} - \int_{-1}^{1} \frac{\partial F(x,\theta)}{\partial\theta} G(\theta)d\theta \\ &= F(x,1)G(1) - F(x,-1)G(-1) - \int_{-1}^{1} T(x,\theta)G(\theta)d\theta \\ &= F(x,1) - \int_{-1}^{1} T(x,\theta)G(\theta)d\theta\,. \end{aligned} \tag{8.135}$$

Let

$$\begin{aligned} L(x) &= F(x,1) - F(x) \\ &= \int_{-1}^{1} T(x,\theta)G(\theta)d\theta\,, \end{aligned} \tag{8.136}$$

$$K(x,y) = \int_{-1}^{1} T(x,z)T(y,z)dz \tag{8.137}$$

and λ_i and ϕ_i be the eigenvalues and the corresponding eigenfunctions of K, that is,

$$\int_{-1}^{1} \phi_i(z)K(z,y)dz = \lambda_i\phi_i(y)\,.$$

It can be shown that the mixture F defined by (8.133) is identifiable iff $\{\phi_i\}$ is a complete orthonormal system for $L^2([-1,1])$ following Tricomi (1957, p.150). Recall that F is identifiable iff there exists a unique square integrable solution $G(\cdot)$ in $L^2(-1,1)$ for (8.133).

Discussion here is based on Tallis (1969). For more details and further discussion, see Tallis (1969).

8.6 Identifiability for Multivariate Mixtures

The following characterization of identifiability is useful in studying the connection between the identifiability problem in the multivariate case and the identifiability of the marginals.

Theorem 8.6.1 (Chandra (1977)): Let $(\mathcal{X}, \mathcal{F})$ and $(\Theta, \mathcal{B})$ be two measurable spaces and $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a family of probability measures on $(\mathcal{X}, \mathcal{F})$ such that the mapping $\theta \to P_\theta(A)$ is $\mathcal{B}$-measurable for each $A \in \mathcal{F}$. Suppose there exists a measurable mapping T from $(\mathcal{X}, \mathcal{F})$ onto $(\mathcal{Y}, \mathcal{T})$ such that a family Λ of mixing distributions on $(\Theta, \mathcal{B})$ is identifiable with respect to $\mathcal{P}T^{-1} = \{P_\theta T^{-1}, \theta \in \Theta\}$ on $(\mathcal{Y}, \mathcal{T})$. Then the family Λ is identifiable with respect to the family $\mathcal{P}$.

Proof: Suppose

$$\int_\Theta P_\theta(A) G_1(d\theta) = \int_\Theta P_\theta(A) G_2(d\theta), A \in \mathcal{F} \tag{8.138}$$

where G_1 and $G_2 \in \Lambda$. Let $B \in \mathcal{T}$. Then $A = T^{-1}B \in \mathcal{F}$ by the measurability of the mapping T. Relation (8.138) implies that

$$\int_\Theta P_\theta(T^{-1}B) G_1(d\theta) = \int_\Theta P_\theta(T^{-1}B) G_2(d\theta), B \in \mathcal{T}. \tag{8.139}$$

Hence

$$\int_\Theta P_\theta T^{-1}(B) G_1(d\theta) = \int_\Theta P_\theta T^{-1}(B) G_2(d\theta), B \in \mathcal{T}. \tag{8.140}$$

By the identifiablity of Λ relative to the family $\mathcal{P}T^{-1} \equiv \{P_\theta T^{-1}, \theta \in \Theta\}$, it follows that $G_1 = G_2$. This proves that the family Λ is identifiable relative to $\mathcal{P}$. ■

As a consequence of the above theorem, identifiability relative to a family of multivariate distributions can be studied from identifiability relative to the corresponding marginals.

Corollary 8.6.1: Let X_i be a random variable with probability measure $P_{\theta_i}, \theta_i \in \Theta_i, 1 \leq i \leq k$. Let $\mathcal{P}_i = \{P_{\theta_i}, \theta_i \in \Theta_i\}, 1 \leq i \leq k$. Suppose the class Λ_i of arbitrary mixing distributions on Θ_i is identifiable relative to the family $\mathcal{P}_i$. Then the class $\Lambda = \prod_{i=1}^k \Lambda_i$ is identifiable with respect to the family of joint distributions P_θ of $\boldsymbol{X} = (X_1, X_2, \cdots, X_k)$ where $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$.

Proof : Let T be a map from $\boldsymbol{X}$ to $\boldsymbol{X}^*$ where components of $\boldsymbol{X}^*$ are treated as independent components. Let P_θ be the joint distribution of $\boldsymbol{X}$ where $\theta = (\theta_1, \cdots, \theta_k)$. It is easy to see that

$$P_\theta T^{-1} = \prod_{i=1}^{k} P_{\theta_i} \equiv P_\theta^* \text{ say} \tag{8.141}$$

by the construction of the measurable map T.

Suppose G and H are arbitrary mixing distributions on Θ such that

$$\int_\Theta P_\theta^*(A) G(d\theta) = \int_\Theta P_\theta^*(A) H(d\theta) \tag{8.142}$$

for all measurable sets A in $\prod_{i-1}^k \mathcal{X}_i$ where $\mathcal{X}_i$ is the range space of X_i. In particular, it follows that

$$\int_{\Theta_i} P_{\theta_i}(A_i) G_i(d\theta) = \int_{\Theta_i} P_{\theta_i}(A_i) H_i(d\theta), 1 \leq i \leq k \tag{8.143}$$

for all measurable sets A_i in $\mathcal{X}_i$ where G_i is the marginal of G corresponding to θ_i. This can be done by choosing $A = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_{i-1} \times A_i \times \mathcal{X}_{i+1} \times \cdots \times \mathcal{X}_k$. Since the class Λ_i of arbitary mixing distributions on Θ_i is identifiable relative to $\mathcal{P}_i = \{P_{\theta_i}, \theta_i \in \Theta_i\}$, it follows that the probability measures G_i and H_i are identical on Θ_i. Hence

$$G_1 \times G_2 \times \cdots \times G_k = H_1 \times H_2 \times \cdots \times H_k \tag{8.144}$$

on $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_k$. In other words, the family Λ of product measures on Θ is identifiable relative to the family $\mathcal{P}^* = \{P_\theta^*, \theta \in \Theta\}$. An application of Theorem 8.6.1 shows that the family Λ is identifiable relative the family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ of joint distributions of $\boldsymbol{X}$. ■

Remarks 8.6.1 : As a special case of the above result, we obtain that if the class Λ_i of arbitrary mixing distributions is identifiable relative to a family $\mathcal{P}_i$ for $1 \leq i \leq k$, then the class $\Lambda = \prod_{i=1}^{k} \Lambda_i$ of product mixing distributions is identifiable relative to the family of product measures $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2 \times \cdots \times \mathcal{P}_k$. It is easy to check that if the class of arbitrary mixing distributions is identifiable relative to $\mathcal{P} = \prod_{i=1}^{k} \mathcal{P}_i$, then the class Λ_i of mixing distributions is identifiable relative to $\mathcal{P}_i$ for $1 \leq i \leq k$. If the measures in $\mathcal{P}$ are not product measures, then it is not true in general that the identifiability relative to the joint distributions implies the identifiability relative to the corresponding marginals. The following examples due to Rennie (1972) illustrate our remarks.

Example 8.6.1 (Rennie (1972)): Consider the family $\mathcal{P} = \{f_1, f_2, f_3\}$ of bivariate densities where

$$\begin{aligned} f_1(x,y) &= 1 \text{ if } 0 \leq x < 1, 1 \leq y < 2 \\ &= 0 \quad \text{otherwise} , \end{aligned} \tag{8.145}$$

$$\begin{aligned} f_2(x,y) &= 1 \text{ if } \ 1 \leq x < 2, 1 \leq y < 3 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{8.147A}$$

and

$$\begin{aligned} f_3(x,y) &= \tfrac{1}{2} \text{ if } \ 1 \leq x < 3, 3 \leq y < 4 \\ &= 0 \quad \text{otherwise} . \end{aligned} \tag{8.147B}$$

Let Λ be the family of finite mixing distributions on the class $\mathcal{P}$. Any mixture is of the form

$$f(x,y) = p_1 f_1(x,y) + p_2 f_2(x,y) + p_3 f_3(x,y) \tag{8.148}$$

where $0 \leq p_i \leq 1$ and $p_1 + p_2 + p_3 = 1$. Since $f_i, i = 1,2,3$ have disjoint supports, it follows that Λ is identifiable relative to $\mathcal{P}$. This can also be seen as a consequence of Theorem 8.2.1. It is easy to check that the marginals of $\mathcal{X}$ for the family $\mathcal{P}$ are

$$f_{1X}(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{8.149}$$

$$f_{2X}(x) = \begin{cases} 1 & \text{if } 1 \leq x < 2 \\ 0 & \text{otherwise} \end{cases} \tag{8.149A}$$

and

$$f_{3X}(x) = \begin{cases} 1 & \text{if } 2 \leq x < 3 \\ 0 & \text{otherwise} \end{cases} . \tag{8.149B}$$

It is again obvious that Λ is identifiable relative to $\mathcal{P}_X = \{f_{1X}, f_{2X}, f_{3X}\}$ as $f_{iX}, 1 \leq i \leq 3$ have disjoint supports. But the marginals of Y for the family are given by

$$f_{1Y}(y) = \begin{cases} 1 & \text{if } 1 \leq y < 2 \\ 0 & \text{otherwise,} \end{cases} \tag{8.150}$$

$$f_{2Y}(y) = \begin{cases} \frac{1}{2} & \text{if } 1 \leq y < 3 \\ 0 & \text{otherwise} \end{cases} \tag{8.150A}$$

and

$$f_{3Y}(y) = \begin{cases} 1 & \text{if } 2 \leq y < 3 \\ 0 & \text{otherwise.} \end{cases} \tag{8.150B}$$

Note that Λ is not identifiable relative to $\mathcal{P}_Y = \{f_{1Y}, f_{2Y}, f_{3Y}\}$. In fact

$$f_{2Y}(y) = \frac{1}{2} f_{1Y}(y) + \frac{1}{2} f_{3Y}(y), -\infty < y < \infty. \tag{8.151}$$

Here is an example of a family of bivariate mixtures which is identifiable but the mixture of one of its marginals is not identifiable.

It is also possible to give examples when mixtures of marginals for all components fail to be identifiable while the mixture of joint distribution is identifiable as shown below.

Example 8.6.2 (Rennie (1972)): Define $f_i, 1 \leq i \leq 3$ as in Example 8.6.1 and

$$f_4(x, y) = \begin{cases} \frac{1}{2} & \text{if } 1 \leq x < 3, 3 \leq y < 4 \\ 0 & \text{otherwise.} \end{cases} \tag{8.152}$$

Let $\mathcal{P} = \{f_1, f_2, f_3, f_4\}$. It can be checked that the mixtures of $\mathcal{P}$ are identifiable but the mixtures of marginals of either X or Y are not identifiable. In fact

$$f_{4X}(x) = \frac{1}{2}f_{2X}(x) + \frac{1}{2}f_{3X}(x), -\infty < x < \infty \tag{8.153}$$

and

$$f_{2Y}(y) = \frac{1}{2}f_{1Y}(y) + \frac{1}{2}f_{3Y}(y), -\infty < y < \infty\,. \tag{8.154}$$

The next example shows that it is possible that the mixtures of one of the marginals are identifiable while the mixtures of the joint distribution are not identifiable.

Example 8.6.3 (Rennie (1972)): Let $\mathcal{P}_Y = \{f_{1Y}, f_{2Y}, f_{3Y}\}$ be as defined in Example 8.6.1 and $\mathcal{P}_X$ be the family of *all* univariate normal distributions. We have seen that finite mixtures of $\mathcal{P}_Y$ do not form an identifiable family from Example 8.6.1. But finite mixtures of members of $\mathcal{P}_X$ form an identifiable family as shown in Example 8.2.2. Let

$$\mathcal{P} = \{f(x,y) = f_X(x)f_Y(y) : f_X \in \mathcal{P}_X, f_Y \in \mathcal{P}_Y\}. \tag{8.155}$$

Then mixtures of $\mathcal{P}$ are not identifiable. In fact

$$f_X(x)f_{2Y}(y) = \frac{1}{2}f_X(x)f_{1Y}(y) + \frac{1}{2}f_X(x)f_{3Y}(y) \tag{8.156}$$

for all x and y.

Remarks 8.6.2 (Identifiability for mixtures of multivariate normal distributions) : Let us consider the class of bivariate normal distributions BVN$(\mu_1, \mu_2; \sum)$ where $\sum$ is a *known* covariance matrix. Suppose (X_1, X_2) is distributed as BVN$(\mu_1, \mu_2; \sum)$. Let G be a probability measure on R^2 and (X_{1G}, X_{2G}) be a random vector with the joint characteristic function

$$\begin{aligned}\psi_G(\boldsymbol{t}) &\equiv \int_{R^2} e^{i\boldsymbol{t}^T\boldsymbol{x}}dF_G(\boldsymbol{x}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\boldsymbol{t}^T\boldsymbol{x}}dF(\boldsymbol{x}|\mu_1,\mu_2)dG(\mu_1,\mu_2) \end{aligned}\tag{8.157}$$

where $F_G(\cdot,\cdot)$ is the joint distribution of (X_{1G}, X_{2_G}). It is the mixture of the family $\{\text{BVN}(\mu_1,\mu_2,\sum)\}$ with mixing measure G. Here $F(\boldsymbol{x}|\mu_1,\mu_2)$ is the bivariate normal distribution function with mean vector (μ_1,μ_2) and known covariance matix $\sum$. It is easy to see that

$$\begin{aligned}\psi_G(\boldsymbol{t}) &\equiv \int e^{i\boldsymbol{t}^T\boldsymbol{x}} dF_G(\boldsymbol{x}) \\ &= e^{-\frac{1}{2}\boldsymbol{t}^T\sum\boldsymbol{t}} \int_{R^2} e^{i\boldsymbol{t}^T\mu} dG(\mu) = e^{-\frac{1}{2}\boldsymbol{t}^T\sum\boldsymbol{t}}\phi_G(\boldsymbol{t}) \quad (8.158)\end{aligned}$$

where $\boldsymbol{t}^T = (t_1,t_2), \mu^T = (\mu_1,\mu_2)$, $\boldsymbol{x}^T = (x_1,x_2)$ and $\phi_G(\boldsymbol{t})$ denotes the characteristic function of G. This relation proves that ψ_G uniquely determines ϕ_G and hence G is identifiable. In other words arbitrary mixtures of bivariate normal distributions with a specified covariance matrix are identifiable.

Bruni and Koch (1985) considered the following equation :

$$f(\boldsymbol{x}) = \int_D N_p(\boldsymbol{x};\lambda(\boldsymbol{y}))G(d\boldsymbol{y}) \quad (8.159)$$

where $\boldsymbol{x} \in R^p, D$ is a compact subset of R^n, G is a probability measure on D, $\lambda(\boldsymbol{y}) = (m_{\boldsymbol{y}}, \sum_{\boldsymbol{y}})$ denotes the mean vector and covariance matrix $\sum$ defined on D and $N_p(\boldsymbol{x};\lambda(\boldsymbol{y}))$ is the multivariate normal density:

$$N_p(\boldsymbol{x};\lambda(\boldsymbol{y})) = (2\pi)^{-\frac{p}{2}}|\Sigma_{\boldsymbol{y}}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m_y})^t\Sigma_{\boldsymbol{y}}^{-1}(\boldsymbol{x}-\boldsymbol{m_y})\}. \quad (8.160)$$

Without loss of generality, D is assumed to be connected by adding sets of G-measure zero. The problem is to identify λ and G given f. They have also investigated whether f is uniquely and continuously associated to the pair (λ, G). Bruni and Koch (1985) furher considered equations of the type

$$f(\boldsymbol{x}) = \sum_{i=1}^{\nu}\int_D \alpha_i N_p(\boldsymbol{x};\lambda_i(\boldsymbol{y}))G_i(d\boldsymbol{y}) \quad (8.161)$$

where ν is a *known* integer, $\alpha_i \geq 0, \sum_{i=1}^{\nu}\alpha_i = 1$ and $G_i, 1 \leq i \leq \nu$ are probability measures on D. The assumption that D is compact is necessary here for it is known that the family of arbitrary Gaussian mixtures over R^2 is not identifiable from results discussed earlier (cf. Teicher (1961)).

8.7 Identifiability for Mixtures on Abstract Spaces

Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and S be a set of probability measures on $(\mathcal{X}, \mathcal{F})$. Let $\mathcal{S}$ be a σ-algebra of subsets of S and μ be a probability measure defined on $(S, \mathcal{S})$. Define

$$M_\mu(B) = \int_S P(B)d\mu(P), B \in \mathcal{F} \ . \tag{8.162}$$

M_μ is a probability measure on $(\mathcal{X}, \mathcal{F})$. M_μ is called a *mixture* over S with *mixing measure* μ.

Definition 8.7.1: The mapping $M : \mu \to M_\mu$, $\mu \in \tau$ is said to be *identifiable* if the mapping M is one-to-one from τ to the class $\{M_\mu : \mu \in \tau\}$.

In the above definition, μ is considered to be a probability measure. However, if μ is allowed to be any signed measure on $(S, \mathcal{S})$ with $\mu(S) = 1$, then the set of mixtures of S with mixing measures $\mu \in \tau$ is said to be identifiable if

$$M_\mu = \int_S P d\mu(P) \equiv 0 \Rightarrow \mu \equiv 0 \ . \tag{8.163}$$

It is clear that if the set of mixtures of S is identifiable in the sense described earlier, then (8.162) holds. The converse follows from the following proposition.

Proposition 8.7.1. If μ is a nonzero signed measure on $(S, \mathcal{S})$ such that

$$M_\mu = \int_S P d\mu(P) \equiv 0 \tag{8.164}$$

holds, then there are two different probability measures μ_1 and μ_2 on $(S, \mathcal{S})$ such that

$$M_{\mu_1} \equiv \int_S P d\mu_1(P) = \int_S P d\mu_2(P) \equiv M_{\mu_2}$$

and hence the set of mixtures of S is not identifiable.

Proof : Supppose μ is a nonzero signed measure on $(S, \mathcal{S})$ such that (8.164) holds. Let $\mu = \mu_1 - \mu_2$ where μ_1 and μ_2 are measures on $(S, \mathcal{S})$ such that either $\mu_1(S) < \infty$ or $\mu_2(S) < \infty$. Since $P(\mathcal{X}) = 1$ for all $P \in S$, it follows that

$$M_\mu(\mathcal{X}) = \int_S P(\mathcal{X})d\mu(P) = \int_S d\mu(P) = \mu(S). \tag{8.165}$$

Relation (8.164) implies that $\mu_1(S) - \mu_2(S) = 0$ and hence both $\mu_1(S)$ and μ_2 are finite. Rescaling, if necessary, we can choose $\mu_1(S) = \mu_2(S) = 1$ since μ is a nonzero measure. Hence

$$M_{\mu_1} = \int P d\mu_1(P) = \int_S P d\mu_2(P) \equiv M_{\mu_2}$$

for probability measures μ_1 and μ_2 on $(S, \mathcal{S})$. Hence the set of mixtures of $\mathcal{S}$ is not identifiable. ■

Remarks 8.7.1 : We assume that $\mathcal{S}$ contains all singletons $\{P\}, P \in S$. In particular, the set of mixtures of S contains the convex hull of S, that is, the set $\{\sum_{i=1}^n \lambda_i P_i, P_i \in S, \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1, n \geq 1\}$.

Remarks 8.7.2 : Suppose $\mathcal{X}$ is a Polish space (complete separable metric space) and $\mathcal{F}$ is the associated Borel σ-algebra. Any probability measure P on $(\mathcal{X}, \mathcal{F})$ is regular and is determined by its values on open sets (cf. Billingsley (1968)). Since $\mathcal{X}$ is separable, every open set in $\mathcal{X}$ is a union of members of a countable collection of open sets $\{U_i\}$ in $\mathcal{X}$. Without loss of generality, it can be assumed that $\{U_i\}$ are disjoint; if not, let $V_1 = U_1, V_2 = U_2 - U_1$ and in general, let $V_n = U_n - \cup_{i=1}^{n-1} U_i, n \geq 1$. Then $\{V_n\}$ is a countable basis for $\mathcal{X}$. Hence every probability measure P on $(\mathcal{X}, \mathcal{F})$ is determined by its values on the countable collection $\{V_i\}$. Let S be a set of probability measures on $(\mathcal{X}, \mathcal{F})$ and $\mathcal{S}$ be the associated Borel σ-algebra generated by the weak convergence of probability measures on $(\mathcal{X}, \mathcal{F})$. It is easy to see that the singleton $\{P\}$ is a closed subset of S for any probability measure on P on $(\mathcal{X}, \mathcal{F})$ and hence $\{P\} \in \mathcal{S}$ for every $P \in S$. In particular, the set of mixtures over S contains the convex hull described in Remarks 8.7.1.

Let $\mathcal{M}(\mathcal{X})$ denote the space of all probability measures on $(\mathcal{X}, \mathcal{F})$ and the topology on $\mathcal{M}(\mathcal{X})$ be determined by the weak convergence of probability measures on $(\mathcal{X}, \mathcal{F})$. Let $D = \{P_x : x \in \mathcal{X}\}$. It is known that $\mathcal{X}$ is homeomorphic to the subset $D \subset \mathcal{M}(\mathcal{X})$ and D is a sequentially closed subset of $\mathcal{M}(\mathcal{X})$. Furthermore $\mathcal{M}(\mathcal{X})$ is metrizable as a separable metric space since $\mathcal{X}$ is a separable metric space. All these facts follow from results in Parthasarathy (1968). Let ρ_1 denote a metric metrizing $\mathcal{M}(\mathcal{X})$ as a separable metric space.

Note that $S \subset \mathcal{M}(\mathcal{X})$ and (S, ρ_1) is also a separable metric space. Let $\mathcal{M}(S)$ denote the set of all probability measures on $(S, \mathcal{S})$ where $\mathcal{S}$ is the associated Borel σ-algebra generated by the topology on S (which in turn is generated by the weak convergence of probability measures on $(\mathcal{X}, \mathcal{F})$). Since S is a separable metric space, it follows that $\mathcal{M}(S)$ is also separable metric space. Let ρ_2 denote a metric metrizing $\mathcal{M}(S)$ as a separable metric space. Define the function

$$f : \mathcal{M}(S) \to \mathcal{M}(\mathcal{X}) \tag{8.166}$$

by

$$f(\mu) = M_\mu = \int_S P d\mu(P), \mu \in \mathcal{M}(S) \ . \tag{8.167}$$

The problem of identifiability essentially reduces to the existence of an inverse for the mapping f. We first prove a general result regarding existence of a bounded inverse for mapping between two sets.

Suppose $\mathcal{M}_1$ and $\mathcal{M}_2$ are two sets and $\rho_i : \mathcal{M}_i \times \mathcal{M}_i \to R^+$ such that $\rho_i(x, x) = 0$ for $x \in \mathcal{M}_i$ and $\rho_i(x, y) \neq 0$ for $x \neq y \in \mathcal{M}_i, i = 1, 2$. Let $f : \mathcal{M}_1 \to \mathcal{M}_2$ with $\mathcal{D}(f) \subset \mathcal{M}_1$ and $\mathbb{R}(f) \subset \boldsymbol{m}_2$ where $\mathcal{D}(f)$ denotes the domain of f and $\mathbb{R}(f)$ denotes its range. Define the norm of f by

$$||f|| = \inf_{\alpha > 0} \{\rho_2(f(x), f(y)) \leq \alpha\ \rho_1(x, y), x, y \in \mathcal{D}(f)\} \tag{8.168}$$

and if f^{-1} exists, define

$$||f^{-1}|| = \inf_{\alpha > 0} \{\rho_1(f^{-1}(u), f^{-1}(v)) \leq \alpha\ \rho_2(u, v), u, v \in \mathbb{R}(f)\} \ . \tag{8.169}$$

It is easy to see that $||f^{-1}|| > 0$ unless $\mathcal{D}(f)$ is a singleton. Assume that $\mathcal{D}(f)$ is not a singleton.

Lemma 8.7.1 : f^{-1} exists and $0 < ||f^{-1}|| < \infty$ if and only if there exists $\alpha > 0$ such that

$$\rho_2(f(x), f(y)) \geq \alpha \rho_1(x, y), x, y \in \mathcal{D}(f) . \tag{8.170}$$

Proof: Suppose there exists $\alpha > 0$ such that (8.170) holds. Let $x, y \in \mathcal{D}(f)$ such that $f(x) = f(y)$. Then $\rho_2(f(x), f(y)) = 0$ and hence $\rho_1(x, y) = 0$ from (8.170). Therefore $x = y$ from the definition of ρ_1. This proves that there exists one-to-one correspondence between $\mathcal{D}(f)$ and $\mathbb{R}(f)$. In other words, f^{-1} exists. Furthermore, relation (8.170) implies that

$$\rho_1(x, y) \leq \frac{1}{\alpha} \rho_2(f(x), f(y)), x, y \in \mathcal{D}(f) \tag{8.171}$$

or equivalently

$$\rho_1(f^{-1}(u), f^{-1}(v)) \leq \frac{1}{\alpha} \rho_2(u, v), u, v \in \mathbb{R}(f) \ . \tag{8.172}$$

This shows that $||f^{-1}|| < \infty$ from (8.169) and $||f^{-1}|| > 0$ since $\mathcal{D}(f)$ is not a singleton.

Conversely, if f^{-1} exists and $0 < ||f^{-1}|| < \infty$, then

$$\begin{aligned} \rho_1(x, y) &= \rho_1(f^{-1}(f(x)), f^{-1}(f(y))) \\ &\leq ||f^{-1}|| \, \rho_2(f(x), f(y)) \end{aligned} \tag{8.173}$$

from the definition of f^{-1}. Therefore

$$\begin{aligned} \rho_2(f(x), f(y)) &\geq \frac{1}{||f^{-1}||} \rho_1(x, y) \\ &= \alpha \, \rho_1(x, y) \end{aligned} \tag{8.174}$$

with $\alpha^{-1} = ||f^{-1}||$ for all $x, y \in \mathcal{D}(f)$. This shows that (8.170) holds. ■

Let us now apply Lemma 8.7.1 to the separable metric spaces $\mathcal{M}_2 = \mathcal{M}(S)$ and $\mathcal{M}_1 = \mathcal{M}(\mathcal{X})$ and the mapping f defined by (8.166) and (8.167). The following theorem is a consequence of Lemma 8.7.1.

Theorem 8.7.1 : The set of mixtures over S is identifiable and the mapping f has bounded inverse if and only if

$$\begin{aligned}\rho_1(m_1, m_2) &= \rho_1\left(\int_S P d\mu_1(P), \int_S P d\mu_2(P)\right) \\ &\geq \alpha\, \rho_2(\mu_1, \mu_2)\end{aligned} \tag{8.175}$$

for $\mu_1, \mu_2 \in \mathcal{M}_2 = \mathcal{M}(\mathcal{X})$ for some $\alpha > 0$ where ρ_1 and ρ_2 are metrics on $\mathcal{M}_1 = \mathcal{M}(S)$ and $\mathcal{M}_2 = \mathcal{M}(\mathcal{X})$ respectively.

Example 8.7.1 : Let $(\mathcal{X}, \mathcal{F}, P_\theta)$ be a probability space and suppose $\{P_\theta, \theta \in \Omega\}$ is a family of probability measures on $(\mathcal{X}, \mathcal{F})$ dominated by a σ-finite measure ν. Suppose (Ω, τ, λ) is a measure space and C is class of probability measures on (Ω, τ) such that every $\mu \in C$ is dominated by λ. Let

$$p(x, \theta) = \frac{dP_\theta}{d\nu}(x), x \in \mathcal{X}, \theta \in \Omega. \tag{8.176}$$

Assume that

$$\int_{\mathcal{X}} \int_\Omega p^2(x, \theta) d\mu(\theta) d\nu(x) < \infty \tag{8.177}$$

and

$$\int_\Omega \left(\frac{d\mu}{d\lambda}\right)^2 d\lambda < \infty. \tag{8.178}$$

We want to obtain sufficient conditions under which the set of mixtures of $\{P_\theta, \theta \in \Omega\}$ is identifiable with respect to members of C . If Q is a mixture of $\{P_\theta, \theta \in \Omega\}$, then

$$\begin{aligned} Q(A) &= \int_\Omega P(A, \theta) d\mu(\theta), && A \in \mathcal{F} \\ &= \int_\Omega \left[\int_A p(x, \theta) d\nu(x)\right] d\mu(\theta), && A \in \mathcal{F} \\ &= \int_A \left[\int_\Omega p(x, \theta) \tfrac{d\mu}{d\lambda}(\theta) d\lambda(\theta)\right] d\nu(x), && A \in \mathcal{F} \\ &= \int_A q(x) d\nu(x), && A \in \mathcal{F} \end{aligned}$$

where

$$q(x) = \int_\Omega p(x, \theta) \frac{d\mu}{d\lambda}(\theta) d\lambda(\theta). \tag{8.179}$$

Applying Lemma 8.7.1. it follows that the set of mixtures $\mathcal{M}_2$ over $\{P_\theta, \theta \in \Omega\}$ with respect to C is identifiable iff

$$\int_\Omega \left[\frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta)\right]^2 > 0 \Leftrightarrow \int_{\mathcal{X}} [q_1(x) - q_2(x)]^2 d\nu(x) > 0. \tag{8.180}$$

Suppose $p(x, \theta)$ can be expanded as an infinite series given by

$$p(x,\theta) = \sum_n \rho_n \phi_n(x)\psi_n(\theta), x \in \mathcal{X}, \theta \in \Omega \tag{8.181}$$

where $\{\phi_n\}$ and $\{\psi_n\}$ are biorthonormal series and $\rho_n > 0$ on $L^2(\mathcal{X} \times \Omega, \mathcal{F} \times \tau, \nu \times \lambda)$. Note that

$$\begin{aligned}
&\int_{\mathcal{X}} [q_1(x) - q_2(x)]^2 d\nu(x) \\
&= \int_{\mathcal{X}} \{\int_{\Omega} p(x,\theta)[\frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta)]d\lambda(\theta)\}^2 d\nu(x) \\
&= \int_{\mathcal{X}} |\int_{\Omega} \sum_n \rho_n \phi_n(x)\psi_n(\theta)(\frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta))d\lambda(\theta)|^2 d\nu(x) \\
&= \int_{\mathcal{X}} |\sum_n \rho_n \phi_n(x) \int_{\Omega} \psi_n(\theta)(\frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta))d\lambda(\theta)|^2 d\nu(x) \\
&= \int_{x} |\sum_n \alpha_n \rho_n \phi_n(x)|^2 d\nu(x) \\
&= \sum_n \rho_n^2 |\alpha_n|^2
\end{aligned} \tag{8.182}$$

where α_n is as defined by (8.184) given below. All the above statements can be justified by using Fubini's theorem. The statement (8.176) and the relation (8.182) prove that the set of mixtures is identifiable if and only if

$$\int_{\Omega} [\frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta)]^2 d\lambda(\theta) > 0 \Leftrightarrow \sum_n \rho_n^2 |\alpha_n|^2 > 0 \tag{8.183}$$

where

$$\begin{aligned}
\alpha_n &= \int_{\Omega} \psi_n(\theta)[\frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta)]d\lambda(\theta) \\
&= < \psi_n, \frac{d\mu_1}{d\lambda}(\theta) - \frac{d\mu_2}{d\lambda}(\theta) > .
\end{aligned} \tag{8.184}$$

The statement (8.183) holds if $\{\psi_n\}$ forms a complete family for $L^2(\Omega, \tau, \lambda)$. Hence the set of mixtures $\mathcal{M}_2$ over $\{P_\theta, \theta \in \Omega\}$ is identifiable with respect to C iff the family $\{\psi_n\}$ given by (8.181) is complete.

Remarks 8.7.3 : The results in this section are due to Tallis and Chesson (1982). Estimation of mixing measures in metric spaces is investigated in Fisher and Yakowitz (1970).

References

Aczel, J. (1966) *Lectures on Functional Equations and Their Applications.* Academic Press, New York.

Alspach, D.E. and Kotlarski, I.I. (1986a) Some characterization theorems for distributions in stochastic locally convex linear topological spaces (unpublished) (Preprint) Oklahama State University.

Alspach, D.E. and Kotlarski, I.I. (1986b) On a characterization of probability distributions on semigroups (unpublished) (Preprint) Oklahama State University.

Anderson, T.W. and Ghurye, S.G. (1977) Identification of parameters by the distribution of a maximum random variable. *J.Royal Statist. Soc. B* **39** , 337–342 .

Barlow, R.E. and Proschan, F. (1975) *Statistical Theory of Reliability and Life Testing : Probability Models.* Hold, Reinehart and Winston, New York.

Bartels, R. (1985) Identification in econometrics. *Amer. Statistician* **39**, 102–104.

Bartlett, M.S. (1966) *An Introduction to Stochastic Processes.* (2nd Ed.). Cambridge University Press, Cambridge.

Basu, A.P. (1981) Identifiability problems in the theory of competing and complementary risks – A Survey. In *Statistical Distributions* *in Scientific Work*, Vol. **5** (Ed. Taille *et al.*). D. Reidel, Dordrecht, 335–348.

Basu, A.P. (1983) Identifiability. In *Encyclopedia in Statistical Scinces*, Vol. **4** (Ed.S. Kotz and N.L. Johnson) 2–6.

Basu, A.P. and Ghosh, J.K. (1978) Identifiability of the multinomial and other distributions under competing risks model, *J. Multivariate Anal.* **8**, 413–429.

Basu, A.P. and Ghosh, J.K. (1980) Identifiability of distributions under competing risks and complementary risks model. *Comm. Statist. Theory and Methods A* **9**, 1515–1525.

Basu, A.P. and Ghosh, J.K. (1983) Identifiability results for a k-out of-p system, *Comm. Statist. Theory and Methods A* **12**, 199–205.

Beran, R. (1979) Exponential models for directional data. *Ann. Statist.* **7**, 1162–1178.

Berman, S.M. (1963) Notes on extreme values, competing risks and semi–Markov processes. *Ann.Math.Statist.* **34**, 1104–1106.

Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

Birnbaum, Z.W. (1979) On the mathematics of competing risks. DHEW Publication No. (PHS) 79–1351, U.S. Dept. of Health, Education and Welfare.

Block, H.W. and Basu, A.P. (1974) A continuous bivariate exponential distribution. *J. Amer. Statist. Assoc.* **69**, 1031–1037.

Blum, J.R. and Susarla, V. (1977) Estimation of a mixing distribution function. *Ann. Probability* **5**, 200–209.

Bondesson, L.(1974) Characterizations of probability laws through constant regression. *Z. Wahrscheinlichkeitstheorie verw Geb.* **3**, 93–115.

Bowden, R. (1973) The theory of parametric identification. *Econometrica* **41**, 1069–1074.

Bruni, C. and Koch, G. (1985) Identifiability of continuous mixtures of unknown gaussian distributions. *Ann. Probability* **13**, 1341–1357.

Bruni, C., Koch, G. and Rossi, C. (1983) On the inverse problem in cytofluorometry : Recovering DNA distribution from FMF data. *Cell Biophys.* **5**, 5–19.

Bunke, H. and Bunke, O. (1974) Identifiability and estimability. *Math. Operations–forsch. Statist.* **5**, 223–233.

Cacoullos, T. (1965) A relation between t and F distributions, *J. Amer. Statist. Assoc.* **60**, 528–531.

Cane, V.R. (1972) The concept of accident proneness. *Bull. Inst. Math Bulgariau.* **15**, 183–189.

Cane, V.R. (1977) A class of non identifiable stochastic models *J. Appl. Probability* **14**, 475–482.

Chandra, S. (1977) On mixtures of probability distributions. *Scand. J. Statist.* **4**, 105–112.

Chang, M. (1984) *Non–parametric Estimations in Doubly Censored Data.* Ph.D. Dissertation. University of Maryland.

Choike, J. , Kotlarski, I.I. and Smith, M. (1980) Characterization of probability distributions from samples of random size. *Pacific J. Math.* **91**, 71–77.

Christensen, P.R. *et al.* (1980) Gamma–ray multiplicity moments from 86_{Kr} reactions on 144, 154_{Sm} at 490 MeV. *Nuclear Phys A* 349, 217–257.

Cooke, R.G. (1950) *Infinite Matrices and Sequence Spaces.* McMillan, London.

Cox, D.R. (1959) The analysis of exponentially distributed life times with two types of failure. *J. Royal Statist. Soc. B* **21**, 411–421.

Cox, D.R. (1972) Regression models with life tables. *J.Royal Statist. Soc.* B **34**, 187–220.

Cramér, H. (1936) Uber eine eigenschaft der normalen verteilungfunction. *Math. Zeitschift* **41**, 405–414.

Cuppens, R. (1975) *Decomposition of Multivariate Probability Distributions.* Academic Press, New York.

Daley, D.J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes.* Springer-Verlag, New York.

Deistler , M. and Hannan, E.J. (1988) *The Statistical Theory of Linear Systems.* Wiley, New York.

Dienes, R.(1932) On linear equations in infinite matrices. *Quart. J. Math* **3**, 253–268.

Doob, J.L. (1953) *Stochastic Processes.* Wiley, New York.

Ebrahimi, N. (1988) On the non–identifiability of multivariate survival functions. *J. Multivariate Anal.* **25**, 164–173.

Elbers, C. and Ridder, G. (1982) True and spurious durational dependence: The identifiability of the proportional hazard model. *Review of Economic Studies* **49**, 403–409.

Everitt, B.S. and Hand, D.J. (1981) *Finite Mixtures of Distributions.* Chapman and Hall, London.

Feller, W.(1966) *Probability Theory,* Vol. II. Wiley, New York.

Fisher, F.M.(1966) *The Identification Problems in Econometrics.* McGraw Hill, New York.

Fisher, L. and Yakowitz, S.J. (1970) Estimating mixing distributions in metric spaces. *Sankhyā Ser. A* **32**, 411–418.

Flinn, C. and Heckman, J. (1982) Models for the analysis of labor force dynamics. In *Advances in Econometrics* Vol. I (Ed. R. Bassman, and G. Rhodes), 35–95.

Flusser, P. (1972) A characterization of the distribution of three independent random variables with values in a locally compact abelian group. *Sankhyā Ser A* **34**, 99–110.

Fraser, M.D., Hsu Yu-Sheng and Walker, J.J. (1981) Identifiabilities of finite mixtures of Von-Mises distribution. *Ann. Statist.* **9**, 1130–1134.

Gabrielsen, Anne (1978) Consistency and identifiability. *J. Econometrics* **8**, 261–263.

Gelfand, I.M. and Vilenkin, N.Ya (1964) *Generalized Functions*, Vol.4. Academic Press, New York.

Ghosh, J.K. and Sen, P.K. (1985) On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II (Ed. L.M. Lecam and R.A. Olshen). Wadsworth Inc., 789–806.

Gilliland, D. and Hannan, J.(1980) Identification of the ordered bivariate normal distribution by minimum variate. *J. Amer. Statist. Assoc.***75**, 651–654.

Greenwood, M. and Yule, U. (1920) *J. Royal Statist. Soc.* **83**, 255–275.

Grenander, U. (1963) *Probabilities on Algebraic Structures.* Wiley, New York.

Gumbel, E.J. (1960) Bivariate exponential distribution. *J. Amer. Statist. Assoc.* **55**, 698–707.

Gyires, B. (1981a) Linear forms in random variables defined on a homogenous Markov chain. In *The First Pannonian Symposium on Mathematical Statistics* (Ed. P.Revesz *et al.*) Lecture Notes in Statistics, Vol. **9**. Springer–Verlag, New York, 110–121.

Gyires, B. (1981b) Constant regression of quadratic statistics on the sum of random variables defined on a Markov chain. In *Contributions in Probability* (Ed.J. Gani and V.K. Rohatgi). Academic Press, New York, 255–266.

Harris, T.E. (1963) *The Theory of Branching Processes.* Springer–Verlag, Berlin.

Hartley, M.J. and Mallela, P. (1977) The asymptotic properties of a maximum likelihood estimator for a model of markets in disequilibrium. *Econometrica* **45**, 1205–1220.

Heckman, J. and Singer, B. (1984) The identifiability of the proportional hazard model. *Review of Economic Studies* **51**, 231–241

Herstein, I.N. (1964) *Topics in Algebra.* Blaisdell, New York.

Hewitt, E. and Ross, K. (1963) *Abstract Harmonic Analysis.* Springer–Verlag, Berlin.

Hille, E. (1948) *Functional Analysis and Semigroups.* Amer. Math. Soc., Providence, Rhode Island.

Hille, E. and Phillips, R.S. (1957) *Functional Analysis and Semigroups.* Amer.Math.Soc., Providence, Rhode Island.

Johnson, R.A. and Wehrly, T. (1978) Some angular – linear distributions and related regression models. *J.Amer.Statist.Assoc.* **73**, 602–606.

Kagan, A.M., Linnik, Yu.V. and Rao, C.R. (1973) *Characterization Problems in Mathematical Statistics.* Wiley, New York.

Kantorovich, L.V. and Krylov, V.I. (1959) *Approximate Methods of Higher Analysis.* Noordhoff, Groningen.

Kelley, J.L. (1953) *General Topology.* Van Nostrand, Princeton.

Kent, J.T. (1983) Identifiability of finite mixtures for directional data. *Ann. Statist.* **11**, 984–988.

Khatri, C.G. and Rao, C.R. (1972) Functional equations and characterization of probability laws through linear functions of random variables. *J. Multivariate Anal.* **2**, 162–173.

Klebanov, L.B. (1973a) A characterization of the normal distribution by a property of order statistics. *Math. Notes* **13**, 71–72.

Klebanov, L.B. (1973b) Reconstituting the ditribution of the components of a random vector from distributions of certain statistics. *Mathematical Notes* **13**, 531–532.

Koopmans, T.C. (1949) Identification problems in economic model construction. *Econometrica* **17**, 125–144.

Koopmans, T.C. and Reiersol, O. (1950) The identification of structural characteristics. *Ann. Math. Statist.* **21**, 165–187.

Kotlarski, I.I. (1960) On random variables whose quotient follows the Cauchy law, *Colloquium Mathematicum.* **7**, 277–284.

Kotlarski, I.I.(1966a) On characterizing the chi-square distribution by the Student law. *J.Amer. Statist. Assoc.***61**, 971–981.

Kotlarski, I.I. (1966b) On characterizing the normal distribution by Student's law. *Biometrika* **53**, 603–606.

Kotlarski, I.I. (1966c) On some characterizations of probability distributions in Hilbert spaces. *Annal di Matematica Pura et Applicate.* **74**, 129–134.

Kotlarski, I.I.(1967) On characterizing the gamma and the normal distributions. *Pacific J. Math.* **20**, 69–76.

Kotlarski, I.I. (1968a) On generalized convolution of sequences. *Zesyty Naukowe Politickniki. Warszawskiez. Matemetey.***11**, 77–83.

Kotlarski, I.I. (1968b) The generalized Laplace convolution. *Bull. de l'Acad. Pol. Scie. Ser. Math. Astronom et Phys.***14**, 367–370.

Kotlarski, I.I. (1968c) A theorem on joint probability distributions in stochastic locally convex linear topological spaces *Colloquium Mathemticum* **19**, 175–177.

Kotlarski, I.I. (1971) On a characterization of probability distributions by the joint distribution of some of their linear forms. *Sankhyā* **33**, 73–80.

Kotlarski, I.I. (1978) On some characterization in probability by using minima and maxima of random variables. *Acquationes Mathematicae* **17**, 77–82.

Kotlarski, I.I. (1979) On a characterization of probability distributions by using maxima of a random number of random variables. *Sankhyā Ser. A*. **41**, 133–136.

Kotlarski, I.I. (1984) Explicit formulas in characterization of distributions of random variables by using random sums. *Metron* **42**, 131–136.

Kotlarski, I.I. (1985) Explicit formulas for characterization of probability distributions by using maxima of a random number of random variables. *Sankhyā Ser. A***47**, 406–409.

Kotlarski, I.I. (1986) Solution of a differential equation with unknown coefficients. *J. Math. Anal. and Appl.* **114**, 116–222.

Kotlarski, I.I. and Cook, L. (1977) Characterization of normality from samples of random size. *Sankhyā Ser. B.* **39**, 196–200.

Kotlarski, I.I. and Hinds, W. (1980) Characterization using intersection of random events. *Aequationes Mathematicae* **21**, 156–158.

Kovalenko, I.N. (1960) On recovering the additive type of a distribution over a sequence of runs of independent observations. In *Proceedings of the All Union Congress on the Theory of Probability and Mathematical Statistics.* Erevan, 148–159 (In Russian).

Kullback, S. (1959) *Information Theory and Statistics.* Wiley, New York.

Kumar, T.K. and Asher, E. (1974) Soviet postwar economic growth and Capital–Labor substitution : Comment. *Amer. Economic Review* **64**, 240–242.

Kumar, T.K. and Gapinski, J.H. (1974) Nonlinear estimation of the CES production parameters : A Monte Carlo study. *Review of Economics and Statist.* **56**, 563–567.

Laha, R.G. (1958) An example of a nonnormal distribution where the quotient follows the Cauchy law. *Proc. Natl. Acad. Sci.* **44**, 222–223.

Laha, R.G. (1959a) On a class of distribution functions where the quotient follows the Cauchy law. *Trans. Amer. Math. Soc.* **93**, 205–215.

Laha, R.G. (1959b) On the laws of Cauchy and Gauss. *Ann. Math. Statist.* 1165–1174.

Lancaster, T. (1979) Econometric methods for the duration of unemployment. *Econometrica* **47**, 939–956.

Langberg, N. and Shaked, M. (1982) On the identifiability of multivariate life distribution function. *Ann. Probability* **10**, 773–779.

Linnik, Yu.V. (1964) *Decomposition of Probability Distributions.* Oliver and Boyd, London .

Loomis, L.H. (1953) *An Introduction to Abstract Harmonic Analysis* . Van Nostrand, Princeton.

Luexmann, Elling Vam (1987) On the identifiability of mixtures of infinitely divisible power series distribution. *Statist. Probability Letters* **5**, 375–378.

Lukacs, E. (1968) *Stochastic Convergence.* D.C. Heath, Lexington, Massachusetts.

Lukacs, E. (1970) *Characteristic Functions.* Griffin, London.

Mardia, K.V. (1972) *Statistics of Directional Data.* Academic Press, London.

Mardia, K.V. and Sutton, T.W. (1978) A model for cylindrical variables with applications. *J. Royal Statist. Soc. B* **40**, 229–233.

Maritz, J. and Levin, T. (1989) *Emprical Bayes Methods* (2nd Ed.). Methuen, London.

Marshall, A.W. and Olkin, I. (1967) A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62** , 30–44.

Mauldon, J.G. (1956) Characterizing properties of statistical distributions. *Quart. J. Math. Oxford Ser. 2* **7**, 155–160.

McKendrick, A.G. (1926) *Proc. Edin. Math. Soc.* **44**, 98–130.

Meilijson, I. (1985) Competing risks on coherent reliability systems. Technical Report, Tel–Aviv University.

Merz, P.H. (1980) Determination of absorption energy distribution by regularization and a characterization of certain absorption isotherms. *J. Comput. Phys.* **34**, 64–85.

Miller, P.G. (1970) Characterizing the distribution of three independent n-dimensional random variables X_1, X_2, X_3 having analytic characteristic functions by the joint distribution of $(X_1 + X_3, X_2 + X_3)$. *Pacific J. Math.***34**, 487–491.

Mitra, S.K. (1980) Generalized inverses of matrices and applications to linear models. In *Handbook of Statistics*, Vol. I (Ed. P. R. Krishnaiah). North–Holland, 471–512.

Moran, P.A.P. (1971) Estimating structural and functional relations. *J. Multivariate Anal.* **1**, 232–253.

Mukherjea, A., Nakassis. A. and Miyashta, J. (1986) The problems of identification of parameters by the distribution of the maximum random variable. *J. Multivariate Anal.* **18**, 178–186.

Mukherjea, A. and Stephens, R. (1990a) The problem of identification of parameters by the distribution of the maximum random variable: Solution for the trivariate normal case. *J. Multivariate Anal.* **34**, 95–115.

Mukherjea, A. and Stephens, R. (1990b) Identification of parameters by the distribution of the maximum random variable. The general multivariate normal case. *Probability Theory and Related Fields* **84**, 289–296.

Nadas, A. (1971) The distribution of the identified minimum of a normal pair determines the distribution of the pair. *Technometrics* **13**, 201–202.

Nowik, S. (1990) Identifiability problem in coherent sytems. *J. Appl. Probability* **28**, 862–872.

Parthasarathy, K.R. (1968) *Probability Measures on Metric Spaces.* Academic Press, London.

Patil, G.P. and Bildikar, S. (1966) Identifiability of countable mixtures of discrete distributions using methods of infinite matrices. *Math. Proc. Camb. Philos. Soc.* **62**, 485–494.

Prakasa Rao, B.L.S. (1968) On a characterization of probability distributions on locally compact abelian groups. *Z. Wahrscheinlichkeitstheorie verw Geb.* **9**, 98–100.

Prakasa Rao, B.L.S. (1975a) On a characterization of probability distributions on locally compact abelian groups II. In *Statistical Distribution in Scientific Work.* Vol. 3. *Characterization and Applications.* (Ed. G.P. Patil *et al.*) D. Reidel, Dordrecht, 231–236.

Prakasa Rao, B.L.S. (1975b)On a characteristic property of point processes. *J. Australian Math. Soc.Ser A* **21** , 108–111.

Prakasa Rao, B.L.S. (1975c) Characterization of stochastic processes determined up to shift. *Theory of Probability and its Applications* **20**, 623–626.

Prakasa Rao, B.L.S. (1976) On a property of generalized random fields. *Studia Sci. Math. Hung.* **11**, 277–282.

Prakasa Rao, B.L.S. (1977) Characterization of multivariate normal distribution from samples of random size. (Preprint) Indian Statistical Institute, New Delhi.

Prakasa Rao, B.L.S. (1983a) Characterization of stochastic processes by stochastic integrals. *Adv. Appl. Probability* **15**, 81–98.

Prakasa Rao, B.L.S. (1983b) *Nonparametric Functional Estimation.* Academic Press, New York.

Prakasa Rao, B.L.S. (1987) Characterization of probability measures by linear functions defined on a homogenous Markov chain. *Sankhyā Ser. A* **49**, 199–206.

Prakasa Rao, B.L.S. (1990) Characterization of probability distributions by random linear forms (Preprint) Indian Statistical Institute, New Delhi.

Prohorov, Yu.V. (1961) The method of characteristic functionals. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol.II, 403–419.

Prohorov, Yu.V. (1965) On a characterization of class of probability distributions by distributions of some statistics. *Theory of Probability and Its Applications* **10**, 438–445.

Prohorov, Yu.V. (1967) Some characterization problems in statistics. In *Proceedings of the Fifth Berkeley symposium in Mathematical Statistics and Probability*, Vol. I, 341–349.

Puri, P.S. (1979) On certain problems involving nonidentifiability of distributions arising in stochastic modelling. In *Optimizing Methods in Statistics* (Ed. J.S. Rustagi). Academic Press, New York, 403–417.

Quandt, R.E. and Ramsey, J.B. (1978) Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Assoc.* **73**, 730–738.

Rao, C.R. (1966) Characterization of the distribution of random variables in linear structural relations. *Sankhyā* **28**, 251–260.

Rao, C.R. (1967) On some charcterization of the normal law. *Sankhyā* **29**, 1–14.

Rao, C.R. (1969) A decomposition theorem for vector variables with a linear structure. *Ann. Math. Statist.* **40**, 1845–1849.

Rao, C.R. (1971) Characterization of probability laws by linear functions. *Sankhyā Ser. A* **33**, 265–270.

Rao, C.R. (1974) Functional equations and characterization of probability distributions (Invited address, International Congress of Mathematics 1974, Vancouver).

Reiersol, O. (1950) Identifiability of linear relation between variables subject to error. *Econometrica* **18**, 375–389.

Rennie, R. (1972) On the interdependence of the identifiability of finite multivariate mixtures and the identifiability of the marginal mixtures. *Sankhyā Ser. A* **34**, 449–452.

Rennie, R. (1974) An identification algorithm for finite mixtures of distributions with common central moments. *Sankhyā Ser. A* **36**, 315–320.

Riedel, M. (1980a) Representation of the characteristic function of a stochastic integral. *J. Appl. Probability* **17**, 448–455.

Riedel, M. (1980b) Determination of a stochastic process by means of stochastic integrals. *Theory of Probability and Its Applications* 25, 339–349.

Robbins, H. (1948) Mixtures of distributions. *Ann. Math. Statist.* **19**, 360–369.

Robbins, H. (1964) The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1–20.

Robbins, H and Pitman, E.J.G. (1949) Application of the method of mixtures to quadratic forms in normal variates. *Ann. Math. Statist.* **20**, 552–560.

Rossberg, H. (1975) An extension of the Phragmén–Lindelöf theory which is relevent for characterization theory. In *A Modern Course on Statistical Distributions in Scientific Work*, Vol. **3** (Ed. G.P. Patil *et al.*) D. Reidel, Dordrecht.

Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* **39**, 577–591.

Rukhin, A.L. (1975) Invariant statistics and characterization of probability distributions. *Theory of Probability and Its Applications* **20**, 596–609.

Rukhin, A.L. (1977) Charakterisierung der Transformationsparameterfamilie. *Z. Wahrscheinlichkeitstheorie verw Geb.* **38**, 287–291.

Sasvari, Z. (1984) Einigc Bermerkungen über die Fortsetzung positiv definiter Funktionen. *Z. Anal.Anwend.* **3**, 435–440.

Sasvari, Z. (1985) Über die Nullstellenmenge von charakteristichen Funktionen. *Math. Nachrichten* **121**, 33–40.

Sasvari, Z. (1986) Characterizing the distribution of the random variables X_1, X_2, X_3 by the distribution of $(X_1\text{-}X_3, X_2\text{-}X_3)$. *Probability Theory and Related Fields* **73**, 43–49.

Sasvari, Z. and Wolff, W. (1985) Über die Wiederherstellung der Ausgangsverteilung durch die Statistik $(X_1\text{-}X_3, X_2\text{-}X_3)$. *Math. Operationsforsch. Ser. Statist.* **16**, 233–241.

Sasvari, Z. and Wolff, W. (1986) Characterizing the distribution of the random vectors X_1, X_2, X_3 by the distribution of the statistic $(X_1\text{-}X_3, X_2\text{-}X_3)$. In *Probability Theory and Mathematical Statistics* , Vol. **2** (Ed.Yu.V. Prohorov *et al.*). VNU Science Press, Utrecht, The Netherlands, 535–539.

Schmidt, P. (1983) Identification problems. In *Encycolopedia in Statistical Sciences,* Vol. **4** (Ed. S. Kotz and N.L. Johnson), 10–14.

Steck, G. (1958)A uniqueness property not enjoyed by the normal distribution, *Ann. Math. Statist.* **29**, 604–606.

Tallis, G.M. (1969) The identifiability of mixtures of distributions. *J.Appl. Probability* **6**, 389–398.

Tallis, G.M. and Chesson, P. (1982) Identifiability of mixtures. *J. Australian Math. Soc. Ser. A* **32**, 339–348.

Teicher, H. (1954) On convolutions of distributions. *Ann. Math. Statist.* **25**, 775–788.

Teicher, H. (1960) On the mixture of distributions. *Ann. Math. Statist.* **31** , 55–73.

Teicher, H. (1961) Identifiability of mixtures. *Ann. Math. Statist.* **32**, 244–248.

Teicher, H. (1963) Identifiability of finite mixtures. *Ann. Math. Statist.*, **34**, 1265–1269.

Teicher, H. (1967) Identifiability of mixture of product measures. *Ann. Math. Statist.* **38**, 1300–1302.

Tigelaar, H.H. (1982) *Identification and Informative Sample Size.* Mathematical Centre Tracts 147, Amsterdam.

Tigelaar, H.H. (1988) Sample size, informative and predictive. In *Encyclopaedia of Statistical Sciences*, Vol.**8**(Ed. S. Kotz and N.L. Johnson), 238–240.

Tigelaar, H.H. (1990) Informative sampling for multivariate ARMAX–system. *Int. J. of Systems Science* (To appear).

Titterington, D.M., Smith, A.F.M. and Makov, V.E. (1985) *Statistical Analysis of Finite Mixture Distributions.* Wiley, Chichester.

Tricomi, F.G. (1957) *Integral Equations.* Interscience, New York.

Tsiatis, A. (1975) A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. USA* **72**, 20–22.

Vakhania, N.N. (1981) *Probability Distributions on Linear Spaces.* North–Holland, Amsterdam.

Van der Genugten, B.B. (1977) Identification in statistical inference. *Statist. Neerlandica* **31**, 69–89.

Vardi, Y., Shepp, L.A. and Kaufman, L. (1975) A statistical model for position emission tomography. *J. Amer. Statist. Assoc.* **80**, 8–20.

Vere–Jones, D. (1968) Some applications of probability generating functionals to the study of input–output streams. *J.Royal Statist. Soc. B* **30**, 321–333.

Vilenkin , N.Ya. (1968) *Special Functions and the Theory of Group Representations.* Amer. Math. Soc., Providence, Rhode Island.

Weiss, G. (1975) Time reversibility of linear stochastic processes. *J. Appl. Probability* **12**, 831–836.

Weiss, G. and Westcott, M. (1976) A note on identification, characterization of the gaussian distribution and time reversibility in linear stochastic processes. *Z. Wahrscheinlichkeitstheorie verw Geb.* **35**, 151–157.

Westcott, M. (1970) Identifiability in linear processes. *Z. Wahrscheinichkeitstheorie verw. Geb.* **16**, 39–46.

Westcott, M. (1972) The probability generating functional. *J. Australian Math. Soc.* **14**, 448–466.

Wilansky, A. (1978) *Modern Methods in Topological Vector Spaces.* McGraw–Hill, New York.

Willassen, Y. (1979) Extension of some results by Reiersol to multivariate models. *Scand. J. Statist.* **6**, 89–91

Yakowitz, S.J. and Spragins, J.D. (1968) On the identifiability of finite mixtures. *Ann. Math. Statist.* **39**, 209–214.

Yosida, K. (1965) *Functional Analysis* Springer–Verlag, Berlin.

Zinger, A.A. (1956) On a problem of A.N. Kolmogorov. *Vestnik LGU* **1**, 53–56 (In Russian).

Author Index

Numbers in italics refer to the pages on which the complete references are listed.

Subject Index

A

B

C

D

E

F

G

H

I

K

L

M

N

PROBABILITY AND MATHEMATICAL STATISTICS

Thomas Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*
Howard Tucker, *A Graduate Course in Probability*
K. R. Parthasarathy, *Probability Measures on Metric Spaces*
P. Révész, *The Laws of Large Numbers*
H. P. McKean, Jr., *Stochastic Integrals*
B. V. Gnedenko, Yu. K. Belyayev, and A. D. Solovyev, *Mathematical Methods of Reliability Theory*
Demetrios A. Kappos, *Probability Algebras and Stochastic Spaces*
Ivan N. Pesin, *Classical and Modern Integration Theories*
S. Vajda, *Probabilistic Programming*
Robert B. Ash, *Real Analysis and Probability*
V. V. Fedorov, *Theory of Optimal Experiments*
K. V. Mardia, *Statistics of Directional Data*
H. Dym and H. P. McKean, *Fourier Series and Integrals*
Tatsuo Kawata, *Fourier Analysis in Probability Theory*
Fritz Oberhettinger, *Fourier Transforms of Distributions and Their Inverses: A Collection of Tables*
Paul Erdös and Joel Spencer, *Probabilistic Methods of Statistical Quality Control*
K. Sarkadi and I. Vincze, *Mathematical Methods of Statistical Quality Control*
Michael R. Anderberg, *Cluster Analysis for Applications*
W. Hengartner and R. Theodorescu, *Concentration Functions*
Kai Lai Chung, *A Course in Probability Theory, Second Edition*
L. H. Koopmans, *The Spectral Analysis of Time Series*
L. E. Maistrov, *Probability Theory: A Historical Sketch*
William F. Stout, *Almost Sure Convergence*
E. J. McShane, *Stochastic Calculus and Stochastic Models*
Robert B. Ash and Melvin F. Gardner, *Topics in Stochastic Processes*
Avner Friedman, *Stochastic Differential Equations and Applications, Volume 1, Volume 2*
Roger Cuppens, *Decomposition of Multivariate Probabilities*
Eugene Lukacs, *Stochastic Convergence, Second Edition*
H. Dym and H. P. McKean, *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*
N. C. Giri, *Multivariate Statistical Inference*
Lloyd Fisher and John McDonald, *Fixed Effects Analysis of Variance*
Sidney C. Port and Charles J. Stone, *Brownian Motion and Classical Potential Theory*

Konrad Jacobs, *Measure and Integral*
K. V. Mardia, J. T. Kent, and J. M. Biddy, *Multivariate Analysis*
Sri Gopal Mohanty, *Lattice Path Counting and Applications*
Y. L. Tong, *Probability Inequalities in Multivariate Distributions*
Michel Metivier and J. Pellaumail, *Stochastic Integration*
M. B. Priestly, *Spectral Analysis and Time Series*
Ishwar V. Basawa and B. L. S. Prakasa Rao, *Statistical Inference for Stochastic Processes*
M. Csörgö and P. Révész, *Strong Approximations in Probability and Statistics*
Sheldon Ross, *Introduction to Probability Models, Second Edition*
P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Applications*
Imre Csiszár and János Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*
A. Hald, *Statistical Theory of Sampling Inspection by Attributes*
H. Bauer, *Probability Theory and Elements of Measure Theory*
M. M. Rao, *Foundations of Stochastic Analysis*
Jean-Rene Barra, *Mathematical Basis of Statistics*
Harald Bergström, *Weak Convergence of Measures*
Sheldon Ross, *Introduction to Stochastic Dynamic Programming*
B. L. S. Prakasa Rao, *Nonparametric Functional Estimation*
M. M. Rao, *Probability Theory with Applications*
A. T. Bharucha-Reid and M. Sambandham, *Random Polynomials*
Sudhakar Dharmadhikari and Kumar Joag-dev, *Unimodality, Convexity, and Applications*
Stanley P. Gudder, *Quantum Probability*
B. Ramachandran and Ka-Sing Lau, *Functional Equations in Probability Theory*
B. L. S. Prakasa Rao, *Identifiability in Stochastic Models: Characterization of Probability Distributions*