MULTIVARIATE STATISTICAL SIMULATION

Mark E. Johnson

A Volume in the Wiley Series in Probability and Mathematical Statistics: Vic Barnett, Ralph A. Bradley, J. Stuart Hunter, David G. Kendall, Adrian F. M. Smith, Stephen M. Stigler, Geoffrey S. Watson —Advisory Editors

Multivariate Statistical Simulation

Multivariate Statistical Simulation

MARK E. JOHNSON

Statistics and Operations Research
Los Alamos National Laboratory
Los Alamos, New Mexico

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging in Publication Data:

Johnson, Mark E. 1952–
Multivariate statistical simulation.

(Wiley series in probability and mathematical statistics. Applied probability and statistics)
Includes bibliographies and index.
1. Multivariate analysis—Data processing. I. Title. II. Series.

QA278.J62 1987 519.5'35'0724 86-22469
ISBN 0-471-82290-6

Preface

Multivariate Statistical Simulation concerns the computer generation of multivariate probability distributions. Generation is used in a broader context than solely algorithm development. An important aspect of generating multivariate probability distributions is what should be generated, as opposed merely to what could be generated. This viewpoint necessitates an examination of distributional properties and of the potential payoffs of including particular distributions in simulation studies. Since other available books document many of the mathematical properties of distributions (e.g., Johnson and Kotz, *Distributions in Statistics*: *Continuous Multivariate Distributions*), a complementary approach is taken here. Generation algorithms are presented in tandem with many graphic aids (three-dimensional and contour plots) that highlight distributional properties from a unique perspective. These plots reveal features of distributions that rarely emerge from preliminary algebraic manipulations.

The primary beneficiary of this book is the researcher who is confronted with the task of designing and executing a simulation study that will employ continuous multivariate distributions. The prerequisite for the reader is a relentless curiosity as to the behavior of the method, estimator, test, or system under investigation when various multivariate distributions are assumed. The multivariate distributions presented in this text can serve as simulation drones to satiate the researcher's curiosity.

For the past ten years or so, my research efforts have consistently involved the development of new distributions to be used in simulation contexts. Hence several chapters reflect my naturally biased disposition toward certain distributions (Pearson Types II and VII elliptically contoured distributions, Khintchine distributions, the unifying class for the Burr, Pareto, logistic distributions). As a reasonable attempt for completeness, various multivariate distributions that are potential (but as yet not

sufficiently developed) competitors to the highlighted distributions are mentioned in the research directions chapter or in the supplementary bibliography.

Although not designed as a text, this book can be used as the primary reference in a graduate seminar in simulation. Exercises could consist of adapting for simulation purposes various references in the supplementary list.

The initial draft of this book was written while I was on sabbatical at the University of Arizona and the University of Minnesota during the academic year 1982–1983. For the Tucson connection I am grateful to John Ramberg, Chairman of the Systems and Industrial Engineering Department, and to Chiang Wang, who gave me considerable support. The Minnesota visit was made possible by the efforts of Dennis Cook, Chairman of the Department of Applied Statistics, and financial support was provided by the School of Statistics under the aegis of Seymour Geisser. The hospitality at both departments is gratefully acknowledged. Valuable insights were provided by Dick Beckman (Los Alamos), Christopher Bingham (St. Paul), Adrian Raftery (Seattle), and George Shantikumar (Berkeley). I am particularly indebted to Sandy Weisberg at Minnesota, whose careful reading of the manuscript led to significant improvements. I am grateful to Myrle Johnson and Geralyn Hemphill for computer graphics support. Finally, this book would not have been possible without the continued support of Larry Booth and Harry Martz, Jr. at Los Alamos.

Skilled typing was performed by Kathy Leis and Kay Woefle (Tucson), Carol Leib and Terry Heineman-Baker (St. Paul), and Kay Grady, Hazel Kutac, Sarah Martinez, Corine Ortiz, and Esther Trujillo (Los Alamos).

MARK E. JOHNSON

Los Alamos, New Mexico
October 1986

Contents

CHAPTER 1

Introduction

Monte Carlo methods are becoming widely applied in the course of statistical research. This is particularly true in small-sample studies in which statistical techniques can be scrutinized under diverse settings. Developments in computing have also encouraged the creation of new methods, such as bootstrapping (Efron, 1979), which exploit this capability. In these respects, statistical research and computing have evolved a symbiotic relationship.

Monte Carlo studies as reported in the statistical literature typically result from the progression of tasks outlined in Figure 1.1. A new statistical technique is first conceived and its associated properties are sought. The main goal of the investigation is probably to collect evidence that will persuade others to employ the method. There are bound to be some characteristics of the new method that resist mathematical analysis, in which case Monte Carlo methods may be used to provide additional knowledge. A preliminary or pilot Monte Carlo study might detect any obvious flaws or possible improvements in the new procedure or suggest numerically efficient shortcuts. Next a larger scale Monte Carlo study is designed in order to address the open questions about the method's properties. A key step in this particular task is the selection of cases, which involves the choice of distributions and their parameters, sample sizes, and so forth. The large-scale study is then conducted and the results synthesized. In the happy situation that the new method performs "well," the investigator can proceed to the publication stage. Otherwise, adjustments to the simulation design or the method itself may be pondered and various tasks repeated.

Some possible purposes of the generic study described in Figure 1.1 might include examination of robustness properties, assessment of small-sample versus asymptotic agreement, or comparison of the new method

with its competitors. This flow chart is generally appropriate for studies that involve either univariate or multivariate distributions. However, the problems in the design stages are vastly different with regard to case selection. For a study in which univariate distributions are used, the problem is to select from among the many distributions available. The set of continuous univariate distributions that can fairly easily be used includes the following:

Beta	Kappa
Burr	Lambda
Cauchy	Laplace
Contaminated normal	Logistic
Exponential power	Normal
Extreme value	Pareto
F	Pearson system
Gamma (including χ^2 and exponential)	Slash
Generalized gamma	Stable
Inverse Gaussian	t
Johnson system (including lognormal)	Weibull

With relative ease, an investigator can accumulate vast quantities of numerical results. However, broad coverage of distributions garnered from the extensive use of the above list can make the subsequent assimilation of results difficult. In particular guiding principles can be lacking as to the effect of distribution on the statistical method under study. Some authors such as Pearson, D'Agostino, and Bowman (1977) have resorted to tabulating distributional results according to the population skewness and kurtosis values. This tactic provides at best a crude ordering for diverse univariate distributions.

In contrast to the univariate setting in which many distributions are available, the multivariate setting offers relatively few distributions that are suitable in Monte Carlo contexts. Although there are many multivariate distributions—the texts by Johnson and Kotz (1972) and Mardia (1970) attest to this—the key word is suitable. More recent advances, as can be found in the NATO conference volumes following international meetings in Calgary (Patil, Kotz, and Ord, 1975) and Trieste (Taillie, Patil, and Baldessari, 1981), tend to be inappropriate or incomplete for application in Monte Carlo studies. Some current limitations of many of these multivariate distributions with respect to Monte Carlo work include the following:

1. Many distributions are tied directly to sampling distributions of statistics from the usual multivariate normal distribution. Outside this

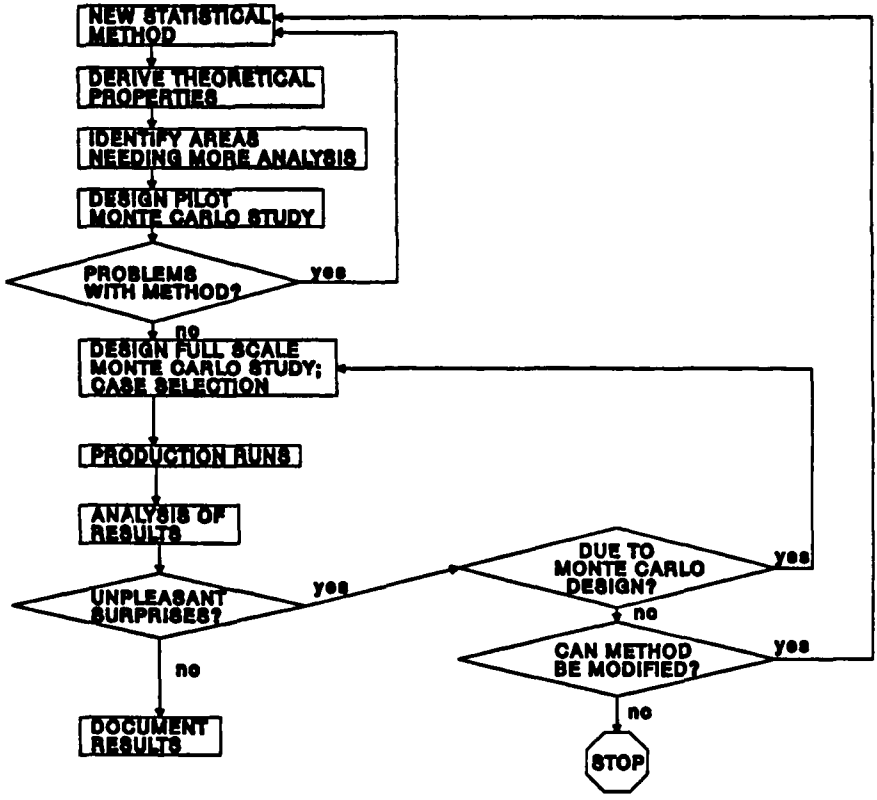

Figure 1.1. Generic Monte Carlo study.

realm of normal theory inference or estimation, the distributions may have little to offer.

2. Other distributions present formidable computational problems (e.g., Bessel function distributions).

3. The support of some of the distributions is too restrictive to be of general interest. Possible examples include the beta-Stacy distribution (Mihram and Hultquist, 1967) and some of the distributions developed by Kimeldorf and Sampson (1978).

4. Some distributions are limited to modeling only weak dependence. Morgenstern's distribution (Section 10.1) is such an example, since by any measure of association, its intrinsic dependence is restrictive. Also, the trivial case of multivariate distributions constructed with independent uni-

variate components has this obvious shortcoming. Multivariate distributions with independent components are, however, important as a baseline for assessing the effects of nontrivial dependence. This issue is explored in detail in some specific contexts later.

5. Computational support for some distributions is lacking. For example, no method may have been published for generating variates from the distribution, or if a method is known, the required univariate generation routines are unavailable.

These limitations of some existing distributions should not be viewed as grounds to abandon those distributions entirely. The limitations are cited to explain their rare use, which might be remedied given particular advances in research. Morgenstern's distribution, which has limited dependence structure in its own right, can be incorporated neatly with the Burr, Pareto, and logistic distribution of Section 9.1 to yield a valuable general distribution developed in Section 9.2.

Deficiencies in currently available distributions further point to the general issue concerning the purposes of Monte Carlo studies and the role of multivariate distribution selection to accommodate these purposes. In the absence of a particular investigation, general recommendations for distribution selection are difficult to provide. Most new statistical techniques have some basic characteristics or nuances that can influence case selection and the design of the Monte Carlo study. To illustrate this point, three distinct research topics are outlined in Sections 1.2–1.4. These discussions are intended to illustrate the potential benefits of problem analysis prior to the execution of the study. In addition these sections provide more justification for the developments given in subsequent chapters. The three topics are the robustness of Hotelling's T^2-test (Everitt, 1979), error rates in partial discriminant analysis (Beckman and Johnson, 1981), and a new multivariate goodness-of-fit test (Foutz, 1980).

These topics are described in some detail, as they may be of independent interest and they may provide an appreciation for the problems and challenges awaiting future investigations and developments in Monte Carlo studies. These are certainly the types of studies that have motivated the writing of this text and have influenced the coverage of distributions given in subsequent chapters.

1.1. ROBUSTNESS OF HOTELLING'S T^2 STATISTIC

A standard problem in multivariate analysis is to test the equality of an unknown population mean vector μ and a specified mean vector μ_0. This test can be conducted using a random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ from a

multivariate normal distribution denoted $N_p(\mu, \Sigma)$ where p is the dimension and Σ is a $p \times p$ covariance matrix. The appropriate test statistic is Hotelling's T^2, computed as

$$T^2 = n(\bar{\mathbf{X}}-\mu_0)'S^{-1}(\bar{\mathbf{X}}-\mu_0),$$

where

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \qquad \text{and} \qquad S = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'.$$

Under the null hypothesis that $\mu = \mu_0$, the statistic $(n-p)T^2/p(n-1)$ has an F distribution with p and $n-p$ degrees of freedom. Hotelling's T^2 is the basis of a uniformly most powerful test against the alternative $\mu \neq \mu_0$ and is invariant under nonsingular linear transformation (Muirhead, 1982, pp. 211–215). In simple terms, if the assumptions are valid that the $\mathbf{X}_i$'s are independent and identically distributed $N_p(\mu, \Sigma)$, then T^2 should be used for inference.

An important practical issue concerns the performance of T^2 if the underlying assumptions are incorrect. There are a variety of ways in which the assumptions could go awry, only a few of which have been addressed in the literature (Everitt, 1979 and Nath and Duran, 1983). The typical situation involves $\mathbf{X}_i$'s that have independent and identically distributed components following a simple univariate distribution such as rectangular, exponential, or lognormal. Of course, the normal distribution is usually included in these studies as a check on the computer program. Alternatives with independent components are not so restrictive as might be surmised, since the results for a given set of $\mathbf{X}_i$'s would be identical as those for $A\mathbf{X}_i$, $i = 1, 2, \ldots, n$, for a nonsingular $p \times p$ matrix A. However, it is not sufficient to consider only random vectors with independent components. Some distributions such as the multivariate Pearson II and VII distributions (Section 6.2) cannot be obtained in this manner. The following comments outline possible areas of research on the performance of T^2 when the assumptions are violated. Some of these issues can be handled readily with the distributions described in later chapters of this book.

1. Suppose the assumption of independence in the random sample is invalid? This would mean the $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ could be thought of as one realization from an $n \times p$ dimensional multivariate distribution. How does this affect T^2?

2. Using the independent components model for the $\mathbf{X}_i$'s, what sort of problems arise if the components are not identically distributed?

3. Moreover, is it really reasonable to assess the effects of the component distributions in terms of the univariate population skewness and kurtosis

values, as has been done by Everitt, for example? Is it possible to isolate these effects to avoid the usual confounding? Everitt demonstrated some cases in which lognormal components evidently degraded T^2 performance more than exponential components. Since the skewness of the lognormal is greater than the skewness of the exponential, he argued that skewness was the culprit. However, this argument can as well be applied using kurtosis instead of skewness, so that a more controlled experiment seems warranted.

4. For extremely non-normal cases that give terrible results with T^2, the non-normality is probably apparent, in which case a transformation to normality could be sought. Can this idea be used to ameliorate the performance of T^2?

5. The performance of T^2 has been viewed primarily in terms of holding the α-level or Type I error rate at a nominal prespecified value. A possible extension is that in cases where this robustness holds, what effects, if any, can be observed in terms of power?

6. Consideration of dimension and sample size are critical for any of the above topics. The general question related to each of these two factors is: Do the results improve, degrade, or remain unaffected as these factors vary?

With this vast set of factors of interest, Monte Carlo experiments obviously should not be conducted without considerable planning. Relatively little attention has been given to experimental design principles in the context of Monte Carlo studies, although there are exceptions (Margolin and Shruben, 1978). Additional work in this area would be welcomed.

1.2. ERROR RATES IN PARTIAL DISCRIMINANT ANALYSIS

Discriminant analysis involves techniques for classifying individuals into one of several populations on the basis of vectors of observations taken on the individuals and on the constituents from each population. Many discriminant analysis techniques implicitly assume *forced* discrimination, in which every "new" individual is to be classified. Broffitt, Randles, and Hogg (1976) described a method for partial discrimination in which an additional option—do not classify—is allowed. Subsequently, Beckman and Johnson (1981) advocated a related partial discriminant analysis method appropriate in the two-population case. This method is first described briefly and then its performance from previous Monte Carlo work is surveyed. In keeping with the spirit of the Hotelling's T^2 example above, a number of research questions are then posed. Some of these questions could be addressed through Monte Carlo studies employing the techniques and the multivariate distributions given in this book.

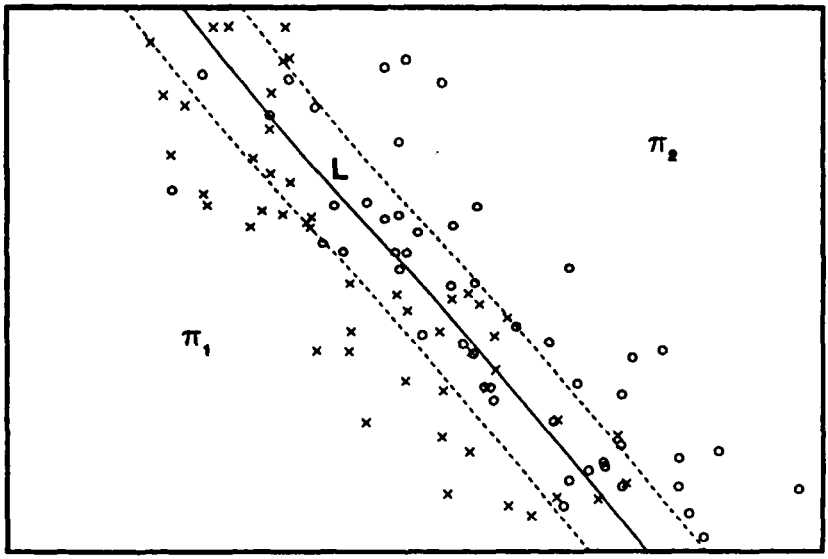

Figure 1.2. Forced and partial discrimination.

Figure 1.2 is useful for describing a simple discriminant rule for partial classification in bivariate populations. The available data on the two populations, denoted π_1 and π_2, is represented by the X's and O's in the figure. In forced discrimination, a new observation Z would be classified in π_1, for example, if it were to the left of the solid line L, and in π_2 otherwise. It should be apparent that many errors in classification will be made because of the considerable overlap in the populations. On the other hand, with partial discriminant analysis, a third region—the do-not-classify area—is included; it is the area enclosed by the dashed lines. Only new observations outside this region will be classified, and then presumably with a high probability of success.

To automate classification, it is convenient to assign a univariate score to each observation. Thus classification decisions can be made by considering certain intervals of the real line, as can be seen from the figure. The scoring function used there is the linear discriminant function given by the projection

$$L(\mathbf{z}) = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})S^{-1}\mathbf{z},$$

where $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ are the sample mean vectors from the training sets of observations and S is an estimated pooled covariance matrix. In particular,

if the training sets are given by $T_1 = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n_1}\}$ for π_1 and $T_2 = \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{n_2}\}$ for π_2, then

$$\bar{\mathbf{X}} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{X}_i$$

$$\bar{\mathbf{Y}} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{Y}_i$$

$$S = \frac{1}{n_1 + n_2 - 2} \left[\sum_{i=1}^{n_1} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' + \sum_{i=1}^{n_2} (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})' \right].$$

Having assigned scores via this or most any other reasonable scheme, the next task is to determine the do-not-classify interval endpoints a and b. If one knew the probability distribution of the scores from each population, represented by the densities f_1 and f_2, then the following optimization problem would need to be solved:

$$\underset{a,b}{\text{maximize}}\ F_1(a) + 1 - F_1(b) + F_2(a) + 1 - F_2(b) \tag{1.1}$$

such that

$$\frac{1 - F_1(b)}{1 - F_1(b) + F_1(a)} \leqslant \alpha_1 \tag{1.2}$$

$$\frac{F_2(a)}{F_2(a) + 1 - F_2(b)} \leqslant \alpha_2, \tag{1.3}$$

where α_i is the specified probability of misclassification of individuals from population i given an attempted classification and F_i is the distribution function corresponding to f_i. The objective function corresponds to the proportion of observations classified. The constraints (1.2) and (1.3) relate to attaining a specified conditional error rate. In any realistic case, the F_i's are unknown but can be estimated by the sample distribution function of scores. The above optimization problem then becomes discrete and can be solved by simple enumeration. An example is provided in Table 1.1. Fifteen observations from two populations were assigned scores, and the nominal error rates are specified as $\alpha_1 = \alpha_2 = 0.10$. The four possible locations for a or b are selected from $\{c_1, c_2, c_3, c_4\}$. Any other choice, such as c_i between the scores of two X's, for example, could be improved. For each of the ten

TABLE 1.1.

Ordering of Scores

XXXXXXXXX O X OOO XXXX OO X OOOOOOOOO

↑ c_1 ↑ c_2 ↑ c_3 ↑ c_4

Enumeration of Possible Solutions

a	b	Number of X's Incorrect / Number of X's Classified	Number of O's Incorrect / Number of O's Classified	Total Number Classified
c_1	c_1	6/15	0/15[a]	30
c_1	c_2	5/14	0/14[a]	28
c_1	c_3	1/10[a]	0/11[a]	21[b]
c_1	c_4	0/9[a]	0/9[a]	18[b]
c_2	c_2	5/15	1/15[a]	30
c_2	c_3	1/11[a]	1/12[a]	23[b]
c_2	c_4	0/10[a]	1/10[a]	20[b]
c_3	c_3	1/15[a]	4/15	30
c_3	c_4	0/14[a]	4/13	27
c_4	c_4	0/15[a]	6/15	30

[a] Observed conditional error rates less than or equal to 0.10.
[b] Both conditional error rates acceptable.

possible pairs (a, b), the error rates on the training sets themselves are computed. Four of the ten cases, namely (c_1, c_3), (c_1, c_4), (c_2, c_3), and (c_2, c_4), attain the specified 10% error rate in each population, at least on the training sets. Of these, the pair (c_2, c_3) classifies the most number of individuals, 23.

One additional adjustment to this method is necessary to ensure its decent performance in small samples. The reason for an adjustment is that discrimination performance is overly optimistic on the training sets in comparison to new observations. A computing-intensive scheme for reducing this bias was developed by Beckman and Johnson (1981), and a brief description is included here for completeness. Assume a scoring function L whose parameters can be estimated from the training sets T_1 and T_2, as given above. A new observation $\mathbf{Z}$ is to be classified. We first pretend that $\mathbf{Z}$ belongs to population π_1 so that our training sets are $\{\mathbf{Z}\} \cup T_1$ and T_2 of size $n_1 + 1$ and n_2, respectively. A scoring function L_1 is then determined based on these training sets. The discrete optimization problem represented by (1.1)–(1.3) is solved to obtain cutoff points a_1 and b_1 and pertinent intervals $A_1 = (-\infty, a_1]$ and $B_1 = [b_1, \infty)$. Clearly, $L_1(\mathbf{Z}) \in A_2$ is evi-

dence to suspect that $\mathbf{Z}$ belongs to π_1 and $L_1(\mathbf{Z}) \in B_1$ suggests that $\mathbf{Z}$ belongs to π_2. The previous steps are then repeated pretending that $\mathbf{Z}$ belongs to π_2. A revised scoring function L_2 based on training sets T_1 and $T_2 \cup \{\mathbf{Z}\}$ is determined and new cutoff values a_2 and b_2 are calculated leading to intervals $A_2 = (-\infty, a_2]$ and $B_2 = [b_2, \infty)$. If $L_2(\mathbf{Z}) \in A_2$, then perhaps $\mathbf{Z}$ belongs to π_1; if $L_2(\mathbf{Z}) \in B_2$, then possibly $\mathbf{Z} \in \pi_2$. The results of these two exercises lead to the following classification rule:

If $L_1(\mathbf{Z}) \in A_1$ and $L_2(\mathbf{Z}) \in A_2$, then classify $\mathbf{Z}$ in π_1.
If $L_2(\mathbf{Z}) \in B_1$ and $L_2(\mathbf{Z}) \in B_2$, then classify $\mathbf{Z}$ in π_2.
Otherwise, do not classify $\mathbf{Z}$.

This classification rule seems to be asymptotically distribution-free although explicit restrictions on the distributions that underly populations π_1 and π_2 and the scoring functions have not been derived. However, some of the small-sample properties have been considered and no evidence to refute this point has emerged. A brief review of the cases examined by Beckman and Johnson is now given. Of particular interest was the performance of this method for a variety of distributions. Three bivariate distributions for the populations were considered: normal (Section 4.1), t (Section 6.2), and lognormal (Chapter 5). For each distributional case, four subcases set the population covariance matrices equal, $\Sigma_1 = \Sigma_2$ and four subcases have $\Sigma_1 \neq \Sigma_2$. The $\Sigma_1 = \Sigma_2$ cases considered correlations between the two components as 0, $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$. In the $\Sigma_1 \neq \Sigma_2$ case, population π_1 was governed by independence ($\Sigma_1 = I$) and in π_2, Σ_2 had four possibilities obtained from two choices of $\rho(\frac{1}{4}$ or $\frac{3}{4})$ crossed with two choices of (σ_1, σ_2) —either (1, 1) or (1, 2). Finally, sample sizes were varied as 21, 35, 51, 101, 201. Thus there were 120 individual cases considered although only three specific distributions were used. This may be a typical phenomenon—rarely can "distribution" be treated as an isolated treatment in the design sense. Sample size is important in general and covariance structure is particularly important in discriminant analysis studies. The results of the study basically indicated that for sample sizes 51 and larger, the proposed partial discriminant analysis method worked correctly—the nominal conditional error rates were achieved.

The results of this Monte Carlo study were not intended to be extrapolated to cover all multivariate distributions, covariance structures, and their combinations with two populations. The results did, however, appear sufficiently promising to apply the technique in a geological investigation that had previously provided tacit inspiration for the cases in the Monte

Carlo study (Patterson et al., 1981). In fact, there are many additional avenues of investigation that could be pursued, some of which are given below:

1. Does the procedure continue to perform correctly in higher dimensions? It might be surmised that larger sample sizes would be required to attain comparable results.

2. In the previous studies, the distributional forms underlying π_1 and π_2 were the same—only the parameters varied. Are the results different if π_1 and π_2 originate from distinct distributions? For example, suppose π_1 is governed by a normal and π_2 by a lognormal?

3. For cases in which π_1 and π_2 have the *same* mean vector, discrimination could be made on the basis of dispersion about the mean. Assuming adjustments to the basic partial discriminant analysis method can be discerned, what sort of performance is achievable?

4. Certainly other scoring functions could be considered. Also, different estimators of the parameters of a particular scoring function could be tried. For example, robust estimates of the μ's and Σ could be calculated. Is there an increase in the proportion of observations classified and if so, is it sufficient to justify the additional effort in calculating these estimates? Is it possible to recommend particular scoring functions for various classes of distributions?

5. Extensive Monte Carlo studies provide some assurance that the method applied to real data will be satisfactory. A practitioner may want additional guarantees that on the specific data being considered the specified conditional error rates will really be achieved and the maximum number of individuals will be classified. One possible approach to collecting this evidence is to apply Efron's (1979) bootstrapping technique to estimate the standard errors of the estimated conditional error rates and proportion of classifications Efron gives an example of the bootstrap's use in forced discrimination so that only a slight adaptation is necessary for the partial discrimination problem.

1.3. FOUTZ' TEST

Foutz (1980) developed a new general-purpose goodness-of-fit test that can be applied in multivariate situations. This test has a number of intuitively appealing features that encourage a thorough empirical examination. First Foutz' test is described and then, in keeping with the previous two exam-

ples, a set of research questions is posed. The primary intent is to provide additional motivation for having included the material in subsequent chapters.

Suppose $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}$ constitute a random sample distributed according to a probability measure P that is assumed to be absolutely continuous with respect to Lebesque measure. The problem is to test the hypothesis that $P = P_0$ where P_0 is a specified probability measure. In comparing these two measures, it is natural to find the Borel set in R^P, say B^*, for which $|P(B^*) - P_0(B^*)|$ is a maximum. This is a tricky problem to say the least—searching through the Borel sets. Foutz devises an ingenious scheme for conducting a test of $P = P_0$, a scheme that requires the computation of probabilities of only n Borel sets. These sets are generated through the construction of statistically equivalent blocks $B_1, B_2, \ldots, B_n$, which depend on the data in a manner to be described shortly. Given these blocks and the hypothesized measure P_0, Foutz' test statistic F_n is computed as follows:

1. Compute $D_i = P_0(B_i)$, $i = 1, 2, \ldots, n$.
2. Order these probabilities as $D_{(1)} < D_{(2)} < \cdots < D_{(n)}$.
3. Evaluate $F_n = \max_{j=1,2,\ldots,n-1}(j/n - D_{(1)} - D_{(2)} - \cdots - D_{(j)})$.

For very small values of n (< 5) the null distribution of F_n can be worked out exactly. For intermediate values up to possibly 50, a Monte Carlo approach can be used. Franke and Jayachandran (1983) report critical values for $n = 20, 30, 50$ at the significance levels 0.10, 0.05, and 0.01. Asymptotically, F_n is normally distributed with mean e^{-1} and variance $(2e^{-1} - 5e^{-2})/n$. Of particular importance is that the distribution of F_n is the same for any dimension p.

It remains to describe how the statistically equivalent blocks B_1, $B_2, \ldots, B_n$ are constructed. The method is due to Anderson (1966) and described by Foutz. Here the basic setup using Foutz' notation is described and then the mechanics are carried out on a simple example. To start, the $n - 1$ data points $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}$ are in R^P, which in the block notation is $B_{12\ldots n}$. Suppose there are $n - 1$ real-valued functions $h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_{n-1}(\mathbf{x})$ such that $h_i(\mathbf{X})$ is a continuous random variable $i = 1, 2, \ldots, n - 1$, if $\mathbf{X}$ is distributed according to P.

Each function is used to identify one data point for purposes of partitioning R^P or a subset of R^P. These functions are subsequently considered in the order given by a permutation $K_1, K_2, \ldots, K_{n-1}$ of the integers $1, 2, \ldots, n - 1$. At the first stage, the function h_{K_1} is used and the values $h_{K_1}(\mathbf{X}_1), \ldots, h_{K_1}(\mathbf{X}_{n-1})$ are computed. The K_1th smallest such value