UNIVERSITY OF LONDON

ACCELERATED RESAMPLING CODES WITH APPLICATION TO LIKELIHOOD

by

SUNDAY MARTINS OGBONMWAN

A thesis submitted for the
DEGREE OF DOCTOR OF PHILOSOPHY
and the
DIPLOMA OF IMPERIAL COLLEGE

Department of Mathematics
Imperial College of Science &
Technology
London SW7 2BZ

December 1985

DEDICATED TO

My Parents

Osawe Ogbonmwan
Ugheghe Ogbonmwan

My Wife

Juliet Imuetinyan Ogbonmwan

and my children

Nosakhare
Osamudiamen
Osarogie

ACKNOWLEDGEMENTS

I should like to thank Professor H.P. Wynn, my supervisor, for his guidance throughout the period for which this work was done. His useful suggestions and comments helped with general improvements of this thesis. Many thanks due to all staff (academic and nonacademic) especially Professor Sir David Cox, Professor A.C. Atkinson, Professor J.T. Stuart and all my fellow research students of the Department of Mathematics, Imperial College, who either made useful comments during research students seminars or gave me some encouragement and useful support.

I am grateful to S.W. Wynn who provided the test data on A1 Duracell batteries for my examples. Special thanks are due to the Federal Government of Nigeria for the postgraduate scholarship award and Mr B.O. Omokaro for his support at the very beginning of my academic career. I am especially grateful for the encouragement received from my relations Avanrenren, Osemwegie & Aiyanyo Ogbonmwan and my colleagues Godwin Obodi, Midhat Abdelmagied and Moses Amadasun. Finally, I should like to thank all staff at the Advisory Desk at the Computer Centre, Imperial College, the Mathematics Department Librarian, Miss J. Pindelska who found many papers (some rare) needed for this work. I should also like to thank Dr J. Beck, Dr W.W.L. Chen and Professor K.F. Roth for their seminar series on "Irregularities of Distribution" which helped our discussion on Discrepancy.

ABSTRACT

A problem of rerandomization/resampling is the enumeration of all the configurations in the reference set. A suggestion to restrict the number of configurations in the reference set S to a much smaller set $S' \subseteq S$ has been proposed by Tukey et al. (1978). A method called Accelerated Resampling Technique (ART), for cutting down on the number of simulations and the number of bootstrap samples is developed in this thesis. The ART translates "randomly" generated coordinates of δ separated points in $\mathbb{R}^n$ to resampled configurations by a coordinate-data matching code to form the restricted reference set. The ART is technically equivalent to the bootstrap resampling procedure. The balanced array δ-method ($\delta \geq 0$) form of the ART performs well as measured by "discrepancy". Some theoretical results justify the use of balancing and computational results from examples point to about 25% improvement.

A new method of generating likelihood is introduced providing a general application for the proposed ART and the bootstrap. The method is a reasonable and cheap way of obtaining inferences in complex situations with only a few distributional assumptions. A realized data set itself is used to generate the likelihoods and we hope the method will provide the necessary framework for "sensitivity analysis".

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

The tedium of computational workload common to all re-randomization/resampling methods is tackled in this work. Most methods based on resampling produce less accurate results if the number of configurations is small, particularly in complex situations. To improve the accuracy of results, the suggestion has been to increase the number of configurations. Another form of solution to the problem is to adopt the variance reduction technique like in the bootstrap application of Therneau (1983). Generally the fundamental issues are:

(1) How many simulations/bootstrap configurations do we need for accurate results?

(2) Is it possible to risk a cutdown on the number of configurations without risking accuracy of the results? Can we find a simulation design that will perform the task of reducing the number of configurations and still achieve accuracy?

(3) If we succeed in (2), can we examine the properties of such a design in relation to existing designs? Has such a design any other beneficial statistical applications?

The main objective of this thesis is to examine the above issues and provide some useful solutions. The motivation for this comes from two main sources. Tukey et al. (1978) suggested a restriction of the reference set of configurations S, to a much smaller set $S' \subseteq S$ with $\#(S') = N' << N = \#(S)$ in such a way that S' mimics S sufficiently well to obtain accurate permutation test confidence levels. The second source is the work of Therneau (1983) on the

"Bootstrap approach to variance reduction techniques" which has the same objective as ours. Our method draws on the literature on Monte Carlo methods and number theory in which low discrepancy sequences (of points - configurations) are sought to improve numerical integration over simple Monte Carlo. Chapter 2 stresses the main problem of computational workload usually encountered in rerandomization/ resampling procedures. A new resampling design called Accelerated Resampling Technique (ART) is formally introduced in Chapter 3. ART requires the translation of coordinates of points in $\mathbb{R}^n$ to resampled configurations by a very natural coordinate - data matching code. The new technique selects "refined" configurations and with these it produces reasonably accurate results. This resampling technique could be thought of as a system that sieves out good "products" to form a restricted set S' from a reference set S with $N' << N$. ART consists of two main designs; viz: (i) the Randomized δ-method and (ii) the Balanced Array δ-method. δ is a spacing parameter adopted by the technique for uniformly spreading out selected points in S. Good estimates are realized generally from only n^2 configurations generated from a data set of size n. The balanced arrays design does better by giving lower discrepancy for the restricted $S' \subseteq S$. The detail properties of discrepancy are examined in Chapter 4. Ours is a discretized version of the works of Warnock (1972) and Zaremba (1968). The purpose is essentially to approximate an integral of a function $f(\cdot)$ over S by the integral over $S' \subseteq S$, the integral being with respect to the uniform measure in each case.

In Chapter 5, the bootstrap/resampling procedure is used repeatedly as a tool to generate likelihoods in a special way by making use of the realized data set itself. The technique requires the use of kernel density estimation without major theoretical development.

The amount of bootstrapping the construction of the likelihoods demands is so large that ART will certainly find ready application in such complex situations.

There are built-in computer Library and NAG routines for generating uniformly distributed random numbers (points). For this work, a NAG routine GØ5CBF, that produces uniformly distributed random numbers (with repeatable initial state) in (0,1) was used. Also Fortran 77 graphical routines PLTON, PLTLIN, PT and CONTR of the Imperial College, Department of Mathematics, graphical packages were used for all the graphs. All computer programs were run on the Imperial College CDC cyber 174 and 855 computer.

CHAPTER TWO

GENERAL NONPARAMETRIC STATISTICAL INFERENCE

2.1 Introduction

This chapter begins with nonparametric aspects of multiple comparisons. In particular, an extension of Sen's (1966) testing procedure for the one-way classification is given. Unlike Sen (1966), our nonparametric statistical test of significance treats all the observations together as a whole (viz: the rank of an observation is determined by its ordering with respect to all observations in all the samples). The test statistic W_N is based on functions of ranks and serves as a generalized nonparametric analogue of the Duncan Multiple Range test. The test statistic has some computational simplicity over most other multiple comparison tests. In section 2.5 we discuss multiple comparisons or indeed statistical inference in general in the context of rerandomization. Rerandomization/Resampling, in our context, compares the evaluation of data from a randomized experiment with an analogous evaluations of the same data calculated as though another experimental design had been used. The problem of the tedium of computational workload common in any rerandomization set-up is stressed. This thesis seeks to find some solutions to this problem and in particular find a simulation design that might lead to accurate results for a reduced number of resampling configurations.

2.2 Nonparametric Multi-Sample Statistical Inference

It is generally accepted that Nonparametric Statistical Techniques are still appropriate when there is doubt about the normal theory procedures leading to some inexplicable results which may be

false. For the p-sample problem, there are $p \geqslant 2$ samples which are assumed to have continuous cumulative distribution functions (cdfs) $F_i(x)$, $i = 1,2,\ldots,p$. Each ith sample has n_i identically independent distributed random variables $X_{i1},X_{i2},\ldots,X_{in_i}$. Interest is usually centred on the testing of the null hypothesis

$$H_0 : F_1 = F_2 = \ldots = F_p = F \qquad (2.2.1)$$

against any other alternative that the F_i's are not all equal. Thus, H_0 might be interpreted as that the sample are all from the same population. Consequently, if we consider the model to be of the form:

$$H_1 : F_i(x) = F(x-\vartheta_i) \qquad i = 1,2,\ldots,p \qquad (2.2.2)$$

and for ϑ_i not all equal and $\vartheta = (\vartheta_1,\vartheta_2,\ldots,\vartheta_p)$ being a real vector. Thus, all the distributions being compared are equal except for the shifts ϑ_i in location. Then, the null hypothesis (2.2.1) to be tested will be equivalent to

$$H_0 : \vartheta_1 = \vartheta_2 = \ldots = \vartheta_p = 0 \qquad (2.2.3)$$

Various solutions have been offered to the above problem. A more exhaustive review of this problem as well as some possible solutions are in Kruskal & Wallis (1952), Scheffé (1959), O'Neil & Wetherill (1971), Hollander (1966), Miller (1981) and So & Sen (1982/1983). For two sample cases, the works of Wilcoxon (1945), Mann & Whitney (1947), Steel (1960), Hodges & Lehmann (1963) and Sen (1966)

among others have been useful and satisfactory within the limits of the derivation of their test statistics. Kruskal & Wallis (1952), gave a nonparametric analogue of the Snedecor's F-test. Their test statistic is

$$H = \left(\frac{12}{N(N+1)} \sum_{j=1}^{p} \frac{R_j^2}{n_j} \right) - 3(N+1) \qquad (2.2.4)$$

for the cases without ties and,

$$H' = \frac{H}{1-\Sigma T/N(N^2-1)} \qquad (2.2.5)$$

for the cases with ties. Here, N is the total number of the combined observations, R_j is the sum of the ranks received by the jth treatment, n_j is the number of observations in the jth treatment and $\Sigma T = \sum_{k=1}^{g} T_k$ for which $T_k = (t-1)(t)(t+1)$ and t is the number of observations in the kth group of tied observations. These have proved quite useful in making a general decision about the null hypothesis (2.2.1) or (2.2.3). For ordered alternatives of the form:

$$H_1 : \vartheta_1 \leqslant \vartheta_2 \leqslant \ldots \leqslant \vartheta_p \qquad (2.2.6)$$

where at least one of the inequalities is strict, Jonckheere (1954) gave a test statistic

$$J = \sum_{u<v}^{p} U_{uv} = \sum_{u=1}^{p-1} \sum_{v=u+1}^{p} U_{uv} \qquad (2.2.7)$$

which is simply the sum of p(p-1)/2 Mann & Whitney counts U_{uv}, $u < v$ where

$$U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \varphi(X_{iu}, X_{jv}) \qquad (2.2.8)$$

and $\varphi(a,b) = 1$ if $a < b$

$= 0$ otherwise (2.2.9)

For the $r \geqslant 2$ - way classification, the Friedman's (1937) type of simultaneous rank test statistic,

$$\chi_r^2 = \frac{12n}{p(p+1)} \sum_{i=1}^{p} \bar{R}_{i\cdot} - 3n(p+1) \qquad (2.2.10)$$

where, $\bar{R}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} R_{ij}$ (2.2.11)

and R_{ij} is the rank of X_{ij}, is applied to any two-way classification with one observation per cell.

Also, Sen (1968) gave the following test statistic for the two-way classification problem:

$$S_N = n \sum_{j=1}^{p} \left\{ T_{N,j} - \bar{E}_N \right\}^2 / \sigma_N^2 \qquad (2.2.12)$$

where,

$$T_{N,j} = \frac{1}{n} \sum_{\alpha=1}^{N} E_{N,\alpha} Z_{N,\alpha}^{(j)} \qquad (2.2.13)$$

in which

$$E_{N,\alpha} = J_N\left(\frac{\alpha}{N}\right) \qquad (1 \leqslant \alpha \leqslant N) \qquad (2.2.14)$$

$Z_{N,\alpha}^{(j)}$ is an indicator, J_N is as defined in Chernoff and Savage (1958) and n is the number of observations in each treatment.

$$\sigma_N^2 = \left\{ \frac{1}{n}(p-1) \right\} \sum_{i=1}^{n} \sum_{j=1}^{p} \left\{ E_{N,R_{ij}} - E_{N,R_{i\cdot}} \right\}^2 \qquad (2.2.15)$$

in which,

$$E_{N,R_{i\cdot}} = \frac{1}{p} \sum_{j=1}^{p} E_{N,R_{ij}} \qquad (2.2.16)$$

and R_{ij} is the rank of X_{ij} observation, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,p$.

However, for all the $p \geqslant 2$ sample cases, a natural question, viz: "Which is the offending sample(s)?" arises whenever the null hypothesis (2.2.1) or (2.2.3) is rejected. Obviously, the need to know the reasons leading to the rejection of the null hypothesis is of prime importance to the research statistician and his clients. We shall consider a solution to the above problem for the one-way layout in the context of multiple comparison. Our solution is motivated by the work of Sen (1966) on Simultaneous Confidence Regions and Test for the One-Criterion Analysis of Variance.

2.3 Assumptions and Notations

Let n_i, $i = 1,2,\ldots,p$ be the number of observations in the ith treatment. Let us consider the case of equal sample sizes

$$n_1 = n_2 = \ldots n_p = n \qquad (2.3.1)$$

Let $X_{i1}, X_{i2},\ldots,X_{in_i}$, $i = 1,2,\ldots,p$ be the ordered observations of a sample from a population with continuous cumulative distribution function (cdf) $F_i(x)$, $i = 1,2,\ldots,p$.

Let $\lambda_{Ni} = n_i/N \quad (= n/N) \qquad (2.3.2)$

where $N = \sum_{i=1}^{p} n_i \quad (= np) \qquad (2.3.3)$

Let,

$$F_{n_i}^{(i)}(x) = \frac{1}{n_i} \quad (\#\text{ of } X_{ij} \leqslant x,\ j = 1,2,\ldots,n_i,\ i = 1,2,\ldots,p) \tag{2.3.4}$$

be the sample cdf of the n_i observations in the ith sample.

Define,

$$H_N(x) = \lambda_{N1}F_n^{(1)}(x) + \lambda_{N2}F_n^{(2)}(x) + \ldots + \lambda_{Np}F_n^{(p)}(x) = \sum_{i=1}^{p} \lambda_{Ni}F_n^{(i)}(x) \tag{2.3.5}$$

Thus, H_N is the combined sample cdf. The combined population cdf is

$$H(x) = \lambda_{N1}F_1(x) + \lambda_{N2}F_2(x) + \ldots + \lambda_{Np}F_p(x) = \sum_{i=1}^{p} \lambda_{Ni}F_i(x) \tag{2.3.6}$$

For the formulation of a class of rank order statistics which we shall use in the work, let

$$X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i}) \qquad i = 1,2,\ldots,p \tag{2.3.7}$$

$$E_N = (E_{N,1}, E_{N,2}, \ldots, E_{N,p}) \tag{2.3.8}$$

where,

$$E_{N,\alpha} = J_N\left(\frac{\alpha}{N+1}\right) \qquad (1 \leqslant \alpha \leqslant N) \tag{2.3.9}$$

and J_N is defined on the same line as in Chernoff & Savage (1958) and satisfies all the regularity conditions of Theorem 3.6.1 of Puri & Sen (1971).

Let us write,

$$\bar{E}_N = \frac{1}{N} \sum_{\alpha=1}^{N} E_{N,\alpha} \tag{2.3.10}$$

and

$$A_N^2 = \frac{1}{N} \sum_{\alpha=1}^{N} E_{N,\alpha}^2 - \bar{E}_N^2 \tag{2.3.11}$$

Let $Z_{N,\alpha}^{(i)} = 1$ if the αth smallest observation in the combined ranking of all the N observations is from the ith sample and let $Z_{N,\alpha}^{(i)} = 0$ otherwise for $\alpha = 1,2,\ldots,N$.

Let us now consider (for the ith sample) the following class of rank order statistic

$$h_N(X_i) = \frac{1}{n_i} \sum_{\alpha=1}^{N} E_{N,\alpha} Z_{N,\alpha}^{(i)} \qquad i = 1,2,\ldots,p \tag{2.3.12}$$

Our $h_N(X_i)$, $i = 1,2,\ldots,p$, given in (2.3.12) are asymptotically distributed as normal. This conclusion follows from the earlier works of Chernoff & Savage (1958), Puri (1964) and Hajek (1968).

For $i = 1,2,\ldots,p$ samples and for the special case of equal sample sizes $n_1 = n_2 = \ldots = n_p = n$, let us consider the following test statistic

$$W_N = \max_{1 \leq i,j \leq p} \left[n^{\frac{1}{2}} A_N^{-1} \left| h_N(X_i) - h_N(X_j) \right| \right] \tag{2.3.13}$$

or

$$W_N = n^{\frac{1}{2}} A_N^{-1} [\max h_N(X_i) - \min h_N(X_i)] \qquad i = 1,2,\ldots,p \tag{2.3.14}$$

For small samples, a permutation method of evaluating the exact null hypothesis of W_N is as follows:

Recall that $X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i})$ $i = 1,2,\ldots,p$.

Let us write,

$$X_N = (X_1, X_2, \ldots, X_p)$$

to represent the actual data set for the one-way layout. Under the null hypothesis (2.2.1) or (2.2.3), X_N is composed of N i.i.d random variables and hence conditioned on any given X_N, all the possible (N!) permutations of the variates among themselves are equally likely. There are altogether $N!/\prod^{p}(n_i)!$ possible configurations of N variables each occurring with probability $(N!/\prod_{i=1}^{p}(n_i)!)^{-1}$. Thus, if we do a complete enumeration of all the configurations and for each compute the value of W_N we will get at the permutation distribution of W_N in the following manner. After computing the values of W_N for all the possible configurations, we then consider the exact probability of the different values of $W_{N,p}(\alpha_0)$ of W_N corresponding to the pre-assigned α_0 level of significance. A point to worry about in the above discussion of the exact distribution of W_N is that the amount of computation needed for the implementation is considerably enormous. For example, a data set consisting of only three treatments with three observations per treatment demands as much as 1680 configurations for a complete enumeration. For large samples the cost of generating the configurations is very much even with improved high speed computers. For such large samples, the limiting (asymptotic) distribution of W_N is discussed fully in Ogbonmwan (1983).

When no ties exist in the distribution, we shall have that

$$E_{N,\alpha} = \frac{\alpha}{N+1} \qquad 1 \leqslant \alpha \leqslant N \qquad (2.3.15)$$

and hence get

$$\bar{E}_N = \frac{1}{N} \sum_{\alpha=1}^{N} \frac{\alpha}{N+1} = \tfrac{1}{2} \tag{2.3.16}$$

Also,

$$\sum_{\alpha=1}^{N} E_{N,\alpha}^2 = \frac{1}{(N+1)^2} \sum_{\alpha=1}^{N} \alpha^2 = \frac{N(2N+1)}{6(N+1)} \tag{2.3.17}$$

Substituting in (2.3.11), we get

$$A_N^2 = \frac{N-1}{12(N+1)} \tag{2.3.18}$$

and W_N becomes,

$$W_N^* = \max_{1 \leqslant i,j \leqslant p} \left[\left\{ \frac{12n(N+1)}{N-1} \right\}^{\frac{1}{2}} \left| h_N(X_i) - h_N(X_j) \right| \right] \tag{2.3.19}$$

When ties do occur in the distribution, the variance component A_N^2 will be reduced by

$$\frac{1}{N} \frac{1}{(N+1)^2} \sum \frac{T}{12} \qquad \left(= \frac{\Sigma T}{12N(N+1)^2} \right) \tag{2.3.20}$$

where,

$$T = (t-1)(t)(t+1) \tag{2.3.21}$$

and t is the number of tied observations. The summation Σ covers all groups of ties.

Subsequently, we have,

$$A_N^2 = \frac{N(N^2-1) - \Sigma T}{12N(N+1)^2} \tag{2.3.22}$$

and the test statistic W_N becomes

$$W_N^{**} = \max_{1\leqslant i,j\leqslant p} \left[\left\{ \frac{12nN(N+1)^2}{N(N^2-1)-\Sigma T} \right\}^{\frac{1}{2}} \left| h_N(X_i) - h_N(X_j) \right| \right] \quad (2.3.23)$$

2.4 Decision Rule

The multiple comparison test for $p \geqslant 2$ treatments may be formulated as follows:

(1) Compute,

$$W_N = n^{\frac{1}{2}} A_N^{-1} [\max h_N(X_i) - \min h_N(X_i)]$$

(2) Compute the value of $W_{N,p}(\alpha_0)$ corresponding to the preassigned α_0-level of significance. More precisely, due to Theorem 1 of Ogbonmwan (1983), $W_{N,p}(\alpha_0) \longrightarrow R_p(\alpha_0)$ in distribution, read the value of $R_p(\alpha_0)$ from the table of upper $100\alpha_0\%$ point of the exact null distribution of $\chi_p(t)$, where $\chi_p(t)$ is the cdf of the sample range in a sample of size p drawn from a standardized normal distribution.

(3) Reject the null hypothesis (2.2.1) or (2.2.3) if

$$W_N \geqslant W_{N,p}(\alpha_0)$$

or if

$$W_N \geqslant R_p(\alpha_0) \quad .$$

Subsequently, we replace W_N with the simplified version W_N^* for the cases without ties or with W_N^{**} for the cases with ties.

REMARK:

In step (3) of the decision rule, we reject any set of $k \leqslant p$ treatments if

$$W_{N,k} \geqslant W_{N,p}(\alpha_0)$$

or if

$$W_{N,k} \geqslant R_p(\alpha_0)$$

If any set of $k \leqslant p$ treatments is found not to be significantly different, then any subsets of the k treatments are also not significantly different and there might be no need to carry out separate tests on the subsets of the k treatments.

Our test statistic W_N is quite close in form to equation 6.8.4 of Puri & Sen (1971) which is designed for nonparametric simultaneous regions and tests for the one-way criterion analysis of variance. Ours is not only an extension of Sen's (1966) but it is the nonparametric analogue of the Duncan (1955) multiple range test.

EXAMPLE:

In a laboratory experiment concerning Life Times, five grades of A_1 Duracell batteries were tested.

The grades of the batteries had been determined by the time in hours to reach a specified End Point Voltage (EPV). The five grades were produced under exactly the same experimental conditions. This means that any observable difference in the life times can only be due to the EPV level if it has any effect on the life times of the A_1 Duracell batteries. So, our null hypothesis is that the EPV has no effect on the life times of the A_1 Duracell batteries. This is in agreement with our null hypothesis (2.2.1) or (2.2.3). The results obtained are given in Table 2.

The decisions made by the use of the Snedecor's F-test (see Scheffé (1959)), the Kruskal & Wallis (1952) H-test, i.e. by making use of equation (2.2.5), and the proposed testing procedures, using equation

(2.3.23), are in agreement. They show that there are significant differences among the p (=5) samples (grades of the A_1 Duracell batteries) - see Table 1 below.

Table 1

Decision Made Using the Proposed Testing Method, the Snedecor's F-test and the Kruskal & Wallis H-test.

Testing Procedure	Calculated Test Statistic	Decision
Proposed Method	$W_N = 6.4886$	S
F-test	$F_{ratio} = 8.9555$	S
Kruskal & Wallis H-test	$H' = 30.1991$	S

Note: (1) $R_5(0.05) = 3.86$, $F^4_{45}(0.025) = 3.10$ and $\chi^2_4(0.05) = 9.488$ are the critical values of the upper $100\alpha_0\%$ point of the exact null distribution of the cdf of the sample range in a sample of size p (=5) drawn from the standardized normal distribution, the F-distribution and the χ^2-distribution respectively.

(2) S ≡ significantly different.

(2.3.23) was used to test for significant differences among subsets $k < p = 5$ samples (grades of A_1 Duracell batteries). The tests show that the five grades of the A_1 Duracell batteries could be put in two main subclasses; viz. B, D, A and C, E. Arranging the grades of the batteries in order of magnitude of the values of $h_N(X_i)$ and underlining the batteries that are not significantly different, we get,

<u>B D A</u> <u>C E</u>

Table 2

Recorded Life Times of Five Grades, A-E, of A_1 Duracell Batteries The Grades are Determined by the EPV levels put in parenthesis

GRADE OF A_1 DURACELL BATTERIES				
A (0.8)	B (0.9)	C (1.0)	D (1.1)	E (1.2)
46.7	42.6	45.4	41.5	48.3
44.1	40.8	45.5	41.5	41.7
46.9	42.7	45.9	42.0	50.6
44.5	41.5	45.6	41.3	49.2
46.5	42.2	46.3	42.2	50.2
35.9	35.0	46.9	42.8	47.7
45.6	41.9	47.0	42.9	50.8
29.0	28.9	46.3	42.1	44.6
45.9	41.9	45.7	41.7	51.5
44.4	41.6	46.2	41.9	50.2

2.5 Rerandomization

"Instead of enlarging the experiment we may attempt to increase its sensitiveness by qualitative improvement; and these are generally speaking, of two kinds: (a) the reorganization of its structure, and (b) refinements of technique. To illustrate a change of structure we might consider that, instead of fixing in advance that 4 cups should be of each kind, determining by a random process how the subdivision should be effected, we might have allowed the treatment of each cup to be determined independently by chance, as by the toss of a coin, so that each treatment has an equal chance of being chosen. The chance of classifying correctly 8 cups randomized in this way, without the aid of

sensory discrimination, is 1 in 2^8, or 1 in 256 chances, and there are only 8 chances of classifying 7 right and 1 wrong; Consequently the sensitiveness of the experiment has been increased, while still using only 8 cups, and it is possible to score significant success, even if one is classified wrongly. In many types of experiment, therefore, the suggested change in structure would be evidently advantageous. For the special requirements of a psycho-physical experiment, however, we should probably prefer to forego this advantage, since it would occasionally occur that all the cups would be treated alike, and this, besides bewildering the subject by an unexpected occurrence, would deny her the real advantage of judging by comparison....." .

Sir Ronald A Fisher (1935, pp.22-23)

A computer aid methodology for doing multiple comparison or statistical inference can be based on Rerandomization. Rerandomization was basically introduced by Fisher (1935) who at that time thought it was not quite practical. Consider an experiment involving N units and p treatments in which each unit u has an equal chance of being assigned a measurable characteristic $X_{k,u}$ of the kth treatment. In this context, we define:

Definition: A random assignment of units to treatments is called an experimental design.

Thus for a given design, each unit u = (1,2,...,N), has an equal chance of being assigned the measurable characteristic $X_{k,u}$. This would mean that a randomized experiment of a particular design is assumed to have been chosen at random from a reference set S of

N_S possible designs. Estimates of parameters are based on the $X_{k,u}$'s. For instance, an estimate of the mean could be expressed as

$$\bar{X} = \sum_{u=1}^{N} X_{k,u}/N \qquad (2.5.1)$$

and for the variance, we could write

$$V = \sum_{u=1}^{N} (X_{k,u}-\bar{X})^2/N \qquad (2.5.2)$$

Other estimates could be defined in the same line. The crucial point of any design is the process of the assignment of units.

Definition: Rerandomization is the process of assigning experimental units to treatments by making use of the elements of the same data set.

Thus an observed data set is simply thought to be a single realization of an appropriate rerandomization process. Various rerandomization/resampling procedures have been discussed in literature. We shall discuss some of them shortly. What is commonly done is to start with a set of configurations and on this set, generate other sets of configurations at random based on a chosen design. For each newly generated configuration, calculate a statistic T. Since, the new configurations are generated at random, the calculated statistics become random variables on which further analysis could be carried out. Miller (1981) gives a good discussion on a permutation method for multiple comparison of p treatments for the one-way layout of which $\{X_{ij}, i = 1,2,...,p, j = 1,2,...,n\}$ is the data set.

The variable

$$M_X = \max_{i,j} \left\{ |\bar{X}_{i\cdot} - \bar{X}_{j\cdot}| \right\} \qquad (2.5.3)$$

based on the treatment means $(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)$ is calculated for each rearrangement (configuration). Treatments i and j are said to be significantly different if

$$|\bar{X}_{i\cdot} - \bar{X}_{j\cdot}| \geq R_c(\alpha_0) \qquad (2.5.4)$$

where, $R_c(\alpha_0)$ is the α_0 quantity of the distribution of range. However, other than for very small values of n and p, Miller (1981) stated "Nevertheless, except in rare circumstances, the amount of labour required to actually carry out the permutation test renders it impractical". A permutation methodology for multiple comparisons for treatment effects in r-way mixed model was constructed by Shuster & Boyett (1979). In their approach, for each M independent simulations chosen by simple random sampling with replacement (constituting a configuration) from all the

$$\prod_{j=1}^{b} \binom{n_{\cdot j}}{n_{1j}, n_{2j}, \ldots, n_{pj}} \quad \text{possible allocations}$$

compatible with their design, the following steps are taken:

(1) Independently, for each block B_j, assign (by simple random sampling without replacement) n_{ij} subjects to treatment T_i.

(2) Compute the desired test statistic, R, for this allocation of subjects. Denote these values by γ_ℓ, $\ell = 1,2,\ldots,M$.

Let,

$$R(\alpha_0) = \inf\{ x:F_M(x^-) \geq 1-\alpha_0 \} \qquad (2.5.5)$$

where F_M is the emperical cumulative distribution of $\{R,r_1,r_2,\ldots,r_m\}$ and x^- is the limit as y approaches x from below. Then for multiple comparison of all treatments, T_i and T_k are significantly different if and only if

$$|Q_i - Q_k| \geq R(\alpha_0) \qquad (2.5.6)$$

where Q_i is defined as

$$Q_i = \sum_{j=1}^{b} n_{ij}d_{ij} \qquad (2.5.7)$$

$$\left.\begin{aligned} d_{ij} &= \sum y_{ijk}/n_{ij} \quad \text{if} \quad n_{ij} > 0 \\ &= \bar{y}_{\cdot j\cdot} \quad \text{if} \quad n_{ij} = 0 \end{aligned}\right\} \qquad (2.5.8)$$

$n_{\cdot j} = \sum_{i=1}^{b} n_{ij}$.

n_{ij} = # of elements from block B_j assigned to treatment T_i.

y_{ijk} = measurable characteristic associated with the kth subject of block B_j assigned to treatment T_i.

$$R = \max_{1\leq i,j\leq p} |Q_i - Q_j| \qquad (2.5.9)$$

By way of reducing the amount of computational workload Shuster & Boyett (1979) suggested that "simulation, rather than complete enumeration will be required in most practical situations".

Gabriel & Hall (1983) and Gabriel & Hsu (1983) have made use of rerandomization (permutation) in making inference statements about regression and shift effects and in evaluating the power of rerandomization tests. They developed methods that reduce the computational workload. Gabriel & Hall (1983), observed that "the tedium of computation for

randomization test is considerable" for the evaluation of power in the sense that the volume of computation become compounded, since each random plan (design), after being subjected to the effects has to be tested against all rerandomization of the outcomes. Petrondas & Gabriel (1983) do not seem to agree generally with inferences concerning pairs of treatments (and subsets) among p treatments as given by Miller (1981), Gabriel & Hsu (1983), Shuster & Boyett (1979) and some others. Petrondas & Gabriel (1983) warned that methods of simultaneous inference cannot simply be transplanted from distribution bound setups to those of rerandomization without risking excessive type I errors. Petrondas & Gabriel (1983) therefore provided a stepwise rerandomization procedure which remains valid within a rerandomization test. We shall not in this thesis contribute in the debate on how best to do multiple comparison in the context of rerandomization. Our interest is mainly focussed on the amount of computational workload required for a rerandomization exercise and on a simple likelihood based model as an alternative to the inversion of a permutation tests needed for some inference problems. Hartigan (1969) used subsamples as typical values to form a coverage probability for any parameter of interest ϑ. This provides another interesting approach to resampling. For a real valued statistic $T(\underset{\sim}{X})$, values of $T(\underset{\sim}{X})$ are recomputed for the different subsamples of a balanced set to provide approximate confidence interval for the "true value" of $T(\underset{\sim}{X})$.

In this approach, the 2^n possible subsamples of the reference set $X_1, X_2, \ldots, X_n$ are the rerandomized new configurations. But generating the 2^n subsamples for large values of n and picking the suitable ones, demands a lot of computation workload which might not be cheap. Tritchler (1984) makes the important connection with the rerandomization likelihood. Efron (1979a, 1979b, 1982) Bootstrapping

is a Monte Carlo resampling method. Consider the random sample X of size n from an unknown probability distribution F.

i.e. $$X = (X_1, X_2, \ldots, X_n) \sim_{iid} F \qquad (2.5.10)$$

Let $x = (x_1, x_2, \ldots, x_n)$ denote the observed realization of the random sample X. Usually, the bootstrap solves the problem of estimating the sampling distribution of a random variable $T_F(X)$ (a statistic) depending on X and the unknown distribution F. The bootstrap adopts the following algorithm:

(1) Construct an estimate of the unknown probability distribution F by putting mass 1/n at x_i, i = 1,2,...,n,

i.e. $\hat{F}$: mass $\frac{1}{n}$ at x_i i = 1,2,...,n.

(2) With $\hat{F}$ fixed draw a "bootstrap sample" of size n,

$$X^* = (X_1^*, X_2^*, \ldots, X_n^*) \text{ from } \hat{F}$$

i.e. $X_i^* \sim_{iid} \hat{F}$.

Calculate $T_{\hat{F}}(X^*) = T_{\hat{F}}(X_1^*, X_2^*, \ldots, X_n^*)$.

(3) The Monte Carlo step: Repeat Step (2) independently for a large number B of times to obtain, $T_{\hat{F}}^{(1)}(X^*), T_{\hat{F}}^{(2)}(X^*), \ldots, T_{\hat{F}}^{(B)}(X^*)$. Approximate the sampling distribution of $T_F(X)$ by the bootstrap distribution of $T_{\hat{F}}(X^*)$.

In Step (3) of the above algorithm, the histogram of the values $T_{\hat{F}}^{(1)}(X^*), T_{\hat{F}}^{(2)}(X^*), \ldots, T_{\hat{F}}^{(B)}(X^*)$ is taken as the approximation of the bootstrap distribution of $T_{\hat{F}}(X^*)$. If Step (3) is omitted, one can still approximate the distribution of $T_{\hat{F}}(X^*)$ by either a direct

theoretical calculation or by Taylor series expansion (see Efron, 1979a). However, the crucial point in the use of the bootstrap algorithm is that, once the data x is observed the distribution of $T_{\hat{F}}(X^*)$ can be calculated. But how well this distribution of $T_{\hat{F}}(X^*)$ approximates the distribution of $T_F(X)$ depends on the nature of $T_F(X)$ itself and on the size of B. The bootstrap resampling procedure has received various applications in several sections of Statistics. Efron (1979a, 1979b, 1982) contain the fundamental introduction and use of bootstrap for the estimation of the distributions of basic statistics like; viz: the mean, the standard deviation, the median, bias, correlation coefficient, confidence interval, etc. Due to the important connection the bootsrap has with our proposed Accelerated Resampling Technique to be discussed in Chapter 3, we shall discuss the bootstrap confidence interval shortly. Therneau (1983) gives a detailed discussion on variance reduction techniques for the bootstrap. Freedman (1981) makes use of bootstrapping in regression analysis. Also in a recent monograph, Wu (1985) gives a full length treatment of bootstrap regression. These papers cited above contain excellent discussions on the methodology for the use and give good guidelines for the adoption of the bootstrap resampling procedure for the various problems tackled. We shall consider the application of bootstrap procedure for estimating confidence interval in this work.

Suppose a parameter of the underlying distribution function of X is ϑ. The approximate confidence interval (usually based on the Maximum Likelihood Estimation (MLE)) is given (see Efron, 1985) as:

$$\vartheta \in \hat{\vartheta} \pm \hat{\sigma} z^{(\alpha)} \qquad (2.5.11)$$

where $\hat{\vartheta}$ is the MLE of ϑ, $\hat{\sigma}$ is an estimate of the standard deviation of

$\hat{\vartheta}$, and $Z^{(\alpha)}$ is the α point of the standard normal variate.

Efron's (1982) bias-corrected percentile method constructs approximate confidence intervals for ϑ in the following way:

(i) Construct an MLE $\hat{F}$ for the unknown density function F,

(ii) the bootstrap data vectors, $X_1^*, X_2^*, \ldots, X_B^*$ are obtained by i.i.d sampling from $\hat{F}$;

(iii) the corresponding bootstrap MLE estimates $\hat{\vartheta}^{*1}, \hat{\vartheta}^{*2}, \ldots, \hat{\vartheta}^{*B}$ are calculated giving the cdf of

$$\hat{G}(t) \equiv \#\{\hat{\vartheta}^{*(i)} < t\}/B \qquad i = 1,2,\ldots,B \ . \tag{2.5.12}$$

(iv) Calculate,

$$z_0 = \Phi^{-1}\{\hat{G}(\hat{\vartheta})\} \tag{2.5.13}$$

where Φ is the standard normal cdf

(iv) the central $1-2\alpha$ interval for ϑ is then taken to be

$$\vartheta \in [\hat{G}^{-1}\{\Phi(2z_0+z^{(\alpha)})\}, \quad \hat{G}^{-1}\{\Phi(2z_0+z^{(1-\alpha)})\}] \ . \tag{2.5.14}$$

If $\hat{G}(\hat{\vartheta}) = \frac{1}{2}$, then $z_0 = 0$ and the interval in (2.5.14) becomes $[\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)]$ which is the α and $1-\alpha$ percentile of the bootstrap distribution. We have more on the percentile method of bootstrap confidence interval in §5.5 of Chapter 5.

We remark that step (ii) above can be, but does not always have to be by the Monte Carlo approach. However, when the Monte Carlo approach is adopted in this step (ii), Efron (1982) suggests that a large number of bootstrap replications, say, $B \geq 1000$ is necessary in order to get reasonable accuracy in the tails of the distribution.

A study using the percentile methods has recently been carried out by Schenker (1985). Schenker (1985) concludes that "If little is known about a problem, it is very difficult or impossible to check the assumptions underlying the use of bootstrap confidence intervals. Therefore, they should be used with caution in complex problems". We conclude this section by observing that in a general rerandomization procedure, given are a data set $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ and a statistic $T(\underset{\sim}{X})$. Interest is usually centred on finding estimates of the parameter $\vartheta(\underset{\sim}{X})$ of the population from which $\underset{\sim}{X}$ has been taken as well as the general nature of the distribution of $T(\underset{\sim}{X})$. Earlier work in the literature (some of which have been cited above) have touched on the methods or procedures leading to the estimation of $\vartheta(\underset{\sim}{X})$ and $T(\underset{\sim}{X})$. Evidently, in all the cases cited so far, there is the worry over the tedium of computational workload. This worry at times might even result in modifications of the entire design model in some complex situations. At times a preferred permutation method which would give exact results might not be employed owing to computational considerations. The questions arising therefore are: How many simulations do we need or can one cut down on the number of simulations for an accurate estimate of $\vartheta(\underset{\sim}{X})$ and distribution of $T(\underset{\sim}{X})$? In particular, how many bootstrap configurations do we need for accurate results? This thesis seeks solutions to the second question in particular by making use of a new resampling design technique.

CHAPTER THREE

ACCELERATED RESAMPLING TECHNIQUE:
The Randomized and Balanced Array δ-Methods

3.1 Introduction

This chapter tackles the problem of cutting down the number of Bootstrap configurations or the number of simulations in the context of resampling although this became apparent more from a reading of the rerandomization (permutation) literature which was discussed in Chapter 2. The basic framework for many simulation exercises for discrete problems can be reduced to the following:

(1) Select with probability 1/N one of a number of N possible configurations. We shall call this one simulation and call the set of all configurations the reference set.

(2) Perform some computations on the selected configuration.

(3) Repeat (1) and (2) M times independently to obtain M sets of numerical results one for each of the M configurations.

(4) Perform some kind of averaging or accumulation on the results in (3).

For the simple Bootstrap case, given $X_1, X_2, \ldots, X_n \sim F$, we draw dependent samples of size n, B times. By putting probability mass 1/n at each point x_i, i = 1,2,...,n we can construct the sample cdf $\hat{F}$. Furthermore, $\hat{F}$ is the fitted nonparametric MLE of the unknown F. Since each sample has size n, each bootstrap (simulation) can be

considered as drawing one configuration with probability n^{-n} from the set S of all $N = n^n$ possible configurations. In a rerandomization problem, we may consider S to be all permutations of some basic configurations restricted possibly to satisfy some marginalization conditions. Both these ideas of bootstrapping and rerandomization are carefully "knitted" together in a new resampling design in this chapter. The randomized and balanced array forms of the new resampling design are fully discussed in §§3.3 and 3.5 and an example is considered in §3.7.

3.2 Some Background Ideas

We discuss briefly some motivating works. Kakutani (1975) proposed or conjectured that "if the unit interval (0,1) is randomly split into n points, then, the splitting points or the decomposition points are uniformly distributed as n tends to infinity". Van Zwet (1978) gave an expository algorithm and proof of the Kakutani's (1975) conjecture and introduced randomness into the selection procedure.

Points are chosen at random in the largest interval. At the nth step, choose a point at random in the largest of the n subintervals into which the first (n-1) points subdivide (0,1). Let X_1 be uniformly distributed on (0,1). For n = 2,3,4,.... the conditional distribution of X_n given $X_1, X_2, \ldots, X_{n-1}$ is uniform on the largest of the n subintervals into which $X_1, X_2, \ldots, X_{n-1}$ subdivide (0,1). Let F_n denote the emperical distribution function (df) of $X_1, X_2, \ldots, X_n$.

Thus,

$$F_n = n^{-1} \sum_{t=1}^{n} 1_{\{X_i \leq x\}} \qquad (3.2.1)$$

Van Zwet (1978) then proved that:

With probability 1,

$$\lim_{n\to\infty} \sup_{x\in(0,1)} |F_n(x)-x| = 0 \qquad (3.2.2)$$

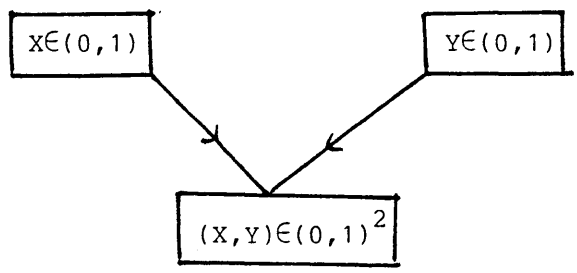
i.e. $F_n(x) \xrightarrow{p} x$ as $n \to \infty$ where $x\in(0,1)$.

Obviously, suppose $\{x_i\}$ $i = 1,2,\ldots,n$ are the first n randomly selected points on the unit interval (0,1) according to the Van Zwet (1978) splitting procedure. Let N be a fixed integer and let $y_i = Nx_i$ such that $\{y_i\}$ $i = 1,2,\ldots,n$ are the projections of the points $\{x_i\}$, $i = 1,2,\ldots,n$ of the unit interval (o,1) onto the interval (o,N). Clearly, for large n, the projected points $\{y_i\}=\{Nx_i\}$, $i = 1,2,\ldots,n$ are also uniformly distributed on the interval (o,N). Let X and Y be two independent random variables (numbers) drawn from the interval (0,1) on the real line; all numbers in the interval being equally likely to be drawn. Let (X,Y) be the two dimensional random variables such that X and Y denote respectively, the first and second numbers drawn from the interval (0,1). We could regard X and Y to be the numbers drawn from (0,1) on the X- and Y-axis respectively. For X and Y independent, each with uniform distribution on (0,1), their joint density f(x,y) given as

$$f(x,y) = \begin{cases} 1 & (x,y)\in(0,1)^2 \\ 0 & \text{otherwise} \end{cases} \qquad (3.2.3)$$

is uniform on $(0,1)^2$. Now suppose values of X and Y are selected independently according to the Van Zwet's (1978) procedure and $F_n(x)$ and $F_n(y)$ are the respective emperical distribution functions of $X_1,X_2,\ldots,X_n$ and $Y_1,Y_2,\ldots,Y_n$. From Van Zwet (1978), $X_1,X_2,\ldots,X_n$ and $Y_1,Y_2,\ldots,Y_n$ are uniformly distributed respectively on (0,1). From the above consideration it is clear that (X_1,Y_1),

$(X_2,Y_2),\ldots,(X_n,Y_n)$ in the unit square $(0,1)^2$ are pairs of independent random variables in (0,1) for the X_i's depend only on the characteristics of X while the Y_i's depend only on the characteristics of Y and there is presumably no reason to assume that the two sources (selections) influence each other's behaviour in any way. Thus, for any two independent ith points on the unit interval (0,1), there is only one corresponding ith point on the two dimensional unit square $(0,1)^2$.



Let us now consider a sequence of bivariate random variables $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ which satisfy our construction. Let their corresponding density functions be $F_1(x,y),F_2(x,y),\ldots,F_n(x,y)$.

$$F_n(x,y) \xrightarrow{a.s} F(x,y) \qquad (3.2.4)$$

(in the limit)

But $$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(t,s)dtds \qquad (3.2.5)$$

From (3.2.3), we have (for X ~ uniform on (0,1), and Y ~ uniform on (0,1)) that

$$F(x,y) = \int_0^x\int_0^y f(t,s)dtds \qquad (3.2.6)$$

which implies that

$$F(x,y) = xy \qquad (3.2.7)$$

Hence

$$F_n(x,y) \xrightarrow{a.s} F(x,y) = xy \qquad (3.2.8)$$

$$\Rightarrow \quad F_n(x,y) \xrightarrow{a.s} xy \qquad (3.2.9)$$

Thus by our construction, (X_1,Y_1), $(X_2,Y_2),\ldots,(X_n,Y_n)$ is "uniformly" distributed in $(0,1)^2$ and we can therefore state that:

With probability 1

$$\lim_{n\to\infty} \sup_{(x,y)\in(0,1)^2} |F_n(x,y)-xy| = 0 \; . \qquad (3.2.10)$$

The above construction based on Van Zwet's (1978) splitting procedure can easily be extended to higher dimensions. We shall make use of the fundamental idea of the splitting process in separating points in $\mathbb{R}^n$ which we need for the development of our new resampling design. The Kakutani (1975) splitting procedure for the unit interval (0,1) serves only as a motivation for the development of our proposed procedure. We shall not split unit intervals or unit squares or unit cubes as done by Kakutani (1975) or Van Zwet (1978), rather we shall separate points within hypercubes by a rejection process.

3.3 The Proposed δ-Method

Our idea is to restrict the reference set S and sample randomly from the restricted reference set. The problem thus becomes one of simulation design. The idea of our design is to separate the members of the restricted set S' by a fixed amount δ, where for points (in $\mathbb{R}^n$), $\omega^{(\ell)}$, $\omega^{(m)} \in S'$, $\delta(\omega^{(\ell)},\omega^{(m)})$ is some simple metric on S. We make use of the Euclidean distance.

Thus,

$$\delta(\omega^{(\ell)},\omega^{(m)})^2 = \sum_{i=1}^{n} (\omega_i^{(\ell)}-\omega_i^{(m)})^2 \qquad (3.3.1)$$

where the suffix i indicates the coordinate. Other metric could be considered as well. For instance, one could consider the r-norm metric:

$$\delta_r(\omega^{(\ell)},\omega^{(m)}) = \left(\sum_{i=1}^{n} |\omega_i^{(\ell)}-\omega_i^{(m)}|^r \right)^{1/r}, \quad r \geqslant 1 \quad . \qquad (3.3.2)$$

For the unbalanced (or randomized) version of the proposed method, which we shall call the randomized δ-method, we make use of the following algorithm in selecting distinct points in $\mathbb{R}^n$.

(1) Choose a fixed interpoint spacing parameter, δ.

(2) Select a first point $\omega^{(1)}$ at random in $\mathbb{R}^n$.

(3) Continue to draw points $\omega^{(t)}$ interdependently until for some t and some $s < t$,

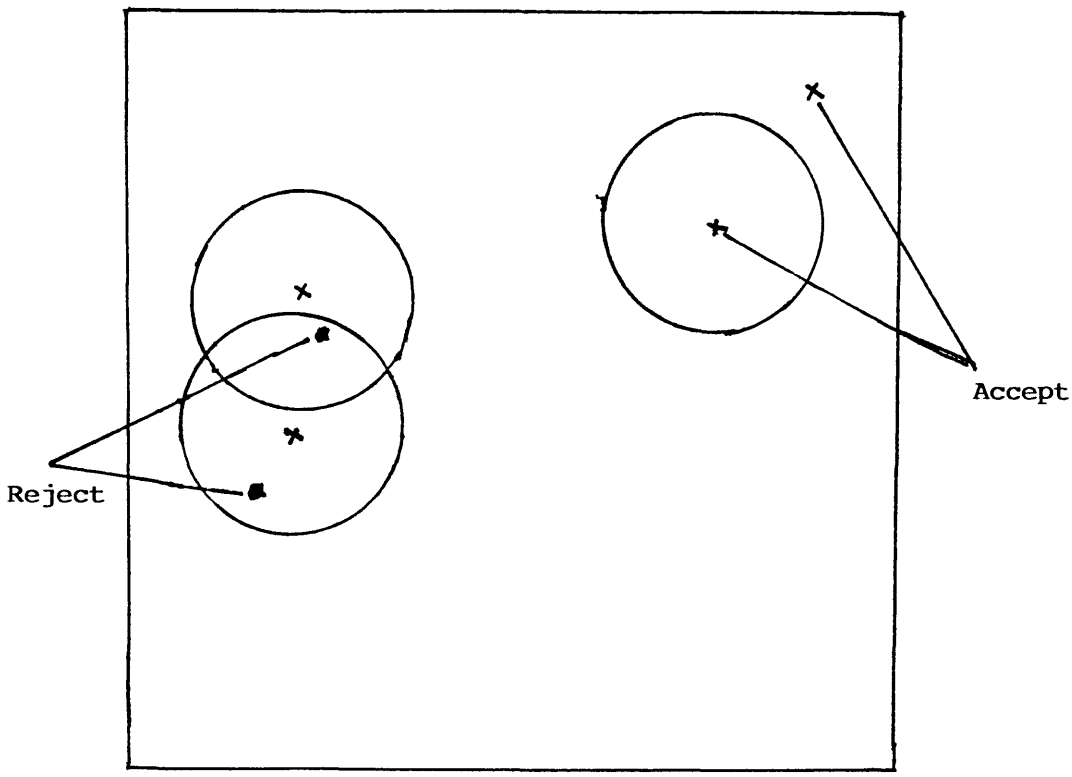
$$\delta(\omega^{(t)},\omega^{(s)})^2 \leqslant \delta \qquad (3.3.3)$$

Reject any such point $\omega^{(t)}$ and continue sampling independently and repeatedly.

Steps (2) and (3) simply mean that we choose the first point at random in the hypercube and create a δ-neighbourhood around this first point. Get a second point at random and check if this second point lies within the δ-neighbourhood of the first point. If so, reject it, otherwise accept it and also create a δ-neighbourhood around it. The process is then repeated for any new point randomly generated.

The spacing parameter δ, could also be changed either during

the procedure or separately. The sketch below clarifies the selection/rejection procedure.



The larger the value of δ the fewer the number of points to be selected. By the use of the spacing parameter δ, the randomized δ-method ensures that the points selected in $\mathbb{R}^n$ do not cluster together. But unfortunately, the selection process itself destroys independence which will make analytical results much more difficult. An upper bound is placed on the total number of simulations (configurations) and the procedure terminates when either this number is reached or a given number of unrejected (selected) points is found. For example, we may take 1000 (= S) simulations to find just 100 (= S′) "good" separated points in $\mathbb{R}^n$. The method is equivalent to a

sequential random packing of spheres and is not unconnected with the idea of separating pieces of information in the theory of error correcting codes. The design of signaling system can be reduced to placing points inside a region of space while constraining them not to be too close together (see Sloane, 1984). On a similar line, we recall that one of the usual problems in sphere packing is to choose the centres so as to maximize the packing density. In our randomized δ-method, we define (n,N,δ,N_δ) packing code as trying to drop N uniform balls (spheres) into an n-dimensional hypercube with a minimum packing (spacing) distance δ, and getting $N_\delta \leq N$ distinct balls actually packed in the hypercube. We repeat: Select N_δ distinct points as possible from a total of N points in an n-dimensional hypercube while maintaining a minimum Euclidean distance of $\sqrt{\delta}$ between any two points. Note that the volume of the hypersphere of radius γ $(= \delta^{\frac{1}{2}})$ is

$$V_{n,\gamma} = \frac{\pi^{n/2}\gamma^n}{\Gamma(n/2+1)} \qquad \left(= \frac{\pi^{n/2}\delta^{n/2}}{\Gamma(n/2+1)} \right)$$

i.e. $$V_{n,\delta} = \frac{\pi^{n/2}\delta^{n/2}}{\Gamma(n/2+1)} \quad .$$

For a 10-dimensional hypercube $[0,1]^{10}$ of unit length 9 the volume of the hypercube is 9^{10}. If we let $\delta = 20$, then the volume of the hypersphere of radius $\sqrt{20}$ will be approximately 1.36×10^6. It is observed that the rejection rate increases as δ takes values greater or equal to 40. In the particular situation of $\delta = 44$, $N' = 204$ distinct points (spheres) are actually selected from $N = 2000$ points. The selected number $N' = 204$ of distinct points is only just about 10% of the total number of points $N = 2000$. The volume of a

10-dimensional hypersphere of radius $\sqrt{44}$ is approximately 7.02×10^7 and hence the total volume of 204 such hyperspheres is 1.43×10^{10} which is surprisingly greater than the volume of the 10-dimensional hypercube of 9^{10}.

The reason for this is that some of the points lie on the edge or on the boundary of the hypercube and therefore have major part of their volumes lying outside the hypercube. The distribution of the coordinates for the $N' = 204$ selected distinct points for $\delta = 44$ as well as the coordinates for the $N = 2000$ points for $\delta = 0$ is given below in Table 3. Clearly, there are more 1's and 10's than any other coordinate. The χ^2-test for the null hypothesis that each coordinate has equal probability of occurring in the distribution reveals that there <u>is</u> significant differences between the coordinates of the $N' = 204$ selected distinct points for $\delta = 44$, while there <u>is no</u> significant differences between the coordinates of the $N = 2000$ points for $\delta = 0$. The distortion is corrected by "balancing" in section 3.5.

3.4 Coordinate-Data Matching Code

We consider points in $\mathbb{R}^n$, $n = 2,3,\ldots$, specified by n coordinates each taking values in $[1,n]$. The proposed δ-method considers only points with integer coordinates. This is however a restriction. Let us define a k-lattice S in n-dimensions. Thus we define

$$S \equiv \{(i_1,i_2,\ldots,i_k) \mid i_j \in (1,2,\ldots,k), \quad i = 1,2,\ldots,n\} \tag{3.4.1}$$

to be our reference set. Given a realized data set $\{x_1,x_2,\ldots,x_n\}$ of size n and a point in S, our code is to match the jth element of

Table 3
Distribution of Coordinates of Selected Points

	FREQUENCY	
Coordinate	N′=204 from N=2000, δ=44	N=2000 for δ = 0
1	271 (+67)	1991 (-9)
2	208 (+4)	2081 (+81)
3	200 (-4)	2085 (+85)
4	177 (-27)	1982 (-18)
5	168 (-36)	1964 (-36)
6	139 (-65)	1934 (-66)
7	190 (-14)	1972 (-72)
8	208 (+4)	2040 (+40)
9	191 (-13)	1975 (-25)
10	288 (+84)	1976 (-24)

Note: The values in parenthesis are the differences between the observed and the expected frequencies.

the data set with the coordinate having the value j. For example, in $\mathbb{R}^5$ the ith point in S could be specified as (i_2,i_2,i_1,i_5,i_4). Using the matching code, we shall have the coordinates of the given point matched as:

$$\begin{array}{ccccc} i_2 & i_2 & i_1 & i_5 & i_4 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ x_2 & x_2 & x_1 & x_5 & x_4 \end{array}$$

The new set $\{x_2,x_2,x_1,x_5,x_4\}$ arising from the coordinate-data matching in our illustration is a resampled set from the original data set $\{x_1,x_2,x_3,x_4,x_5\}$. Consequently, generating points at random and transforming their coordinates leads to a resampling technique which is identical to Efron (1979a, 1979b, 1982) bootstrap resampling method. For instance, if we do a bootstrap of the given data set $\{x_1,x_2,x_3,x_4,x_5\}$ we could have the resultant transformed coordinates (configurations) as follows:-

Sample	**Point**
(x_1,x_1,x_5,x_1,x_1)	(1,1,5,1,1)
(x_3,x_2,x_2,x_2,x_4)	(3,2,2,2,4)
(x_4,x_1,x_2,x_5,x_5)	(4,1,2,5,5)
.	.
.	.
.	.

Applying the δ-method to this case, when $\delta = 0$, we have the standard bootstrap. For $\delta > 0$, we define the method as the Accelerated Resampling Technique (ART). This gives a reduced number of selected distinct points (configurations) in $\mathbb{R}^n$ from the

total number of possible $N = n^n$ points that constitutes the reference set. When $\delta = 1$, we have a sampling without replacement version of the bootstrap.

Therneau (1983) considered a form of transformation linking bootstrap samples (configurations) with sample proportions. Suppose, we have a data set $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ from a distribution function F and we wish to consider the statistic $T\ (=T(\underset{\sim}{x}))$. Usually in the bootstrap practice, F is unknown and it is simply replaced with the non-parametric MLE estimate $\hat{F}$ which is discrete. Since $\hat{F}$ is the cdf of X, then every bootstrap sample $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is treated as the pair $(\underset{\sim}{P}^*, \underset{\sim}{X})$, where $\underset{\sim}{P}$ is the vector of sample proportions. For example, if a realized data set is x = (6.4,9.1,9.9,8.1,5.8) then some possible bootstrap samples from $\hat{F}$ along with their corresponding sample proportions could be:

x^*	P^*
(5.8,5.8,5.8,9.1,9.1)	(0,2/5,0,0,3/5)
(9.1,5.8,9.9,9.9,6.4)	(1/5,1/5,2/5,0,1/5)
(5.8,8.1,9.9,9.1,6.4)	(1/5,1/5,1/5,1/5,1/5)
.	.
.	.
.	.

Thus, sampling x^* from $\hat{F}$ is equivalent to sampling P^* from a multinomial distribution of t choices on n categories, which would imply that

$$\text{Prob}\{\underset{\sim}{P}^* = P\} = \begin{cases} \binom{t}{t\underset{\sim}{p}} \left(\frac{1}{n}\right)^t & \underset{\sim}{tp} \text{ a vector of integers, } \sum P_i = 1 \\ 0 & \text{Otherwise} \end{cases} \qquad (3.4.2)$$

However, there are difficulties in describing the bootstrap samples in terms of distribution because of the unequal weights attached to the different configurations. This difficulty is not present in our proposed δ-method. For instance, let us consider a situation where t = 3 and n = 3. Table 4 below gives the actual bootstrap configurations and the proportions along with the corresponding probabilities.

For a given data set $x = (x_1, x_2, \ldots, x_n)$ the function which takes the point ω in S into as explained in (3.4.1) we refer to as $x(\omega) = (x_1(\omega), x_2(\omega), \ldots, x_n(\omega))$. Thus in the usual bootstrap notation we would have $x^* = x(\omega)$.

3.5 The Balance Array Design

Instead of clinging to the idea of complete randomness in the simulation procedure for the generation of the points in $\mathbb{R}^n$ needed in our proposed ART, we shall deviate slightly and consider a Quasi-Monte Carlo method that produces balanced array of coordinates of the selected points. The new design is therefore called the Balanced Array design. It can also accommodate the δ spacing procedure. The motivation for this new Balanced Array δ-method is to get a design that gives low discrepancy (see Chapter 4) for the selected points on which our estimates are to be based. We shall now consider a version of the ART which produces balanced array of coordinates. This is first-order balance (or simply balance). Let $x = (x_1, x_2, \ldots, x_n)$ be the data set. Let $x(\omega) = (x_1(\omega), x_2(\omega), \ldots, x_n(\omega))$ be the vector of values for a given point $\omega \in S$. Define a general class of functions

$$t(\omega) = \sum_{i=1}^{n} t_i(x_i(\omega))$$

Table 4

Configurations, Probabilities and Proportions of Proposed Method Compared with Therneau's (1983) Method, n = 3, t = 3.

Proposed δ-Method		Therneau's (1983) Method	
Configuration	Probability	P(distribution)	Probability
(1,1,1)	1/27	(1,0,0)	1/27
(2,2,2)	1/27	(0,1,0)	1/27
(3,3,3)	"	(0,0,1)	1/27
(1,1,2)	"	(2/3,1/3,0)	3/27
(1,2,1)	"		
(2,1,1)	"		
(1,1,3)	"	(2/3,0,1/3)	3/27
(1,3,1)	"		
(3,1,1)	"		
(1,2,2)	"	(1/3,2/3,0)	3/27
(2,1,2)	"		
(2,2,1)	"		
(1,3,3)	"	(1/3,0,2/3)	3/27
(3,1,3)	"		
(3,3,1)	"		
(2,2,3)	"	(0,2/3,1/3)	3/27
(2,3,2)	"		
(3,2,2)	"		
(2,3,3)	"	(0,1/3),2/3)	3/27
(3,2,3)	"		
(3,3,2)	"		
(1,2,3)	"	(1/3,1/3,1/3)	6/27
(2,1,3)	"		
(2,3,1)	"		
(1,3,2)	"		
(3,2,1)	"		
(3,1,2)	1/27		

More generally, ignoring the actual data set x we can write

$$t(\omega) = \sum_{i=1}^{n} t_i(\omega_i)$$

acting directly on S. The bootstrap expectation of $t(\omega)$ can be written as

$$E_{\hat{F}}(t(\omega)) = \frac{1}{n^n} \sum_{S} t(\omega)$$

where the summation is over all ω in S. We can also define expectation with respect to a restricted set of configurations $S' \subseteq S$:

$$E_{S'}(t(\omega)) = \frac{1}{N'} \sum_{S'} t(\omega)$$

We are interested in conditions under which

$$E_{\hat{F}}(t(\omega)) = E_{S'}(t(\omega)) \quad .$$

Definition: A set of configurations $S' = \{\omega^{(1)},\ldots,\omega^{(N')}\}$ is said to be (first order) balanced if the quantities

$$M_{jk} = \#\{i \mid \omega_k^{(i)} = j,\ i = 1,2,\ldots,N'\}$$

are equal for all $k = 1,2,\ldots,n$, $j = 1,2,\ldots,n$. Here M_{jk} is actually the number of j's in column k.

The common value of the M_{jk} is $M_{jk} = N'/n$.

If we arrange the points of S' as rows in an $N' \times n$ array then the condition merely says that each $j = 1,2,\ldots,n$ appears the same number of times in each column $k = 1,2,\ldots,n$.

THEOREM 1:

The set of configurations S' is first order balanced if and only if

$$E_{\hat{F}}(t(\omega)) = E_{S'}(t(\omega))$$

for every choice of the functions $t_1, t_2, \ldots, t_n$.

Proof:

Assume first order balance. Then,

$$E_{S'}(t(\omega)) = \frac{1}{N'} \sum_{S'} \sum_{i=1}^{n} t_i(\omega_i)$$

$$= \frac{1}{N'} \sum_{i=1}^{n} \sum_{S'} t_i(\omega_i)$$

$$= \frac{1}{N'} \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} t_i(j)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} t_i(j)$$

$$= \frac{1}{n^n} \sum_{i=1}^{n} \sum_{j=1}^{n} n^{n-1} t_i(j)$$

$$= \frac{1}{n^n} \sum_{i=1}^{n} \sum_{S}^{n} t_i(\omega_i)$$

$$\Rightarrow E_{S'}(t(\omega)) = \frac{1}{n^n} \sum_{S} \sum_{i=1}^{n} t_i(\omega_i) = E_{\hat{F}}(t(\omega)) \quad .$$

Now assume that for fixed j

$$t_i(r) = 0 \qquad i \neq k$$

$$= 1 \qquad i = k,\ r = j$$

$$= 0 \qquad \text{otherwise} \quad .$$

Then if $E_{S'}(t(\omega)) = E_{\hat{F}}(t(\omega))$

we have,

$$\frac{1}{N'} \times M_{jk} = \frac{1}{n}$$

repeating this for all j and k we are done.

Notice that we suppress the data in our argument. In actual practical situation in the ART, balanced sets are generated in blocks of n samples so that within each block there is first order balance. Consider Table 5. This consists of 100 samples (configurations) each of size 10 made of 10 blocks. Observe that no digit appears more than once in any column of a block. In fact, the entire (100 × 10) matrix of Table 5 is simply made of 10 (10 × 10) matrices. The δ-spacing is $\delta = 25$, is maintained across all the blocks but sampling is restricted so that balance is achieved by the end of a block. If by any chance, the δ-separation rule is violated at a point within a block, the whole block is rejected so that first order balance is maintained. A practical way of generating the blocks and eventually constructing an $(N' \times n)$ balanced array matrix is the following stages. Here is the complete generation of a block for $n = 4$.

Stage 1:

For the purposes of illustration let us suppose that our first point is (4,1,2,2). We shall retain this point and generate the first coordinate of the second point by choosing an integer number

from {1,2,3,4}. If this first number is 4, reject it and choose another to replace it. If the first number selected is not 4, retain it and go to Stage 2.

Stage 2:

For the second coordinate of the second point, choose a number from {1,2,3,4,}. If this number is 1, reject it and choose another number. But if the number selected is not 1, retain it. So for the second stage we may have

Stage 3:

For the third coordinate of the second point, choose a number from {1,2,3,4} and retain it if it is not 2, otherwise reject it and choose another number to replace it.

Stage 4:

The fourth coordinate is generated in the same manner as described for the other coordinates above.

At the end of Stage 4, we may have our first two points to be

4 1 2 2

1 3 3 3 .

Stage 5:

For the coordinates of the third point, start by selecting

the first coordinate from {1,2,3,4}. If this is 4 or 1 reject it and choose another to replace it. Treat the other coordinates in the same way by generating and checking with the corresponding coordinates of the first two points in the block. Coordinates for the fourth point are generated in the same manner.

Finally we may have our first block to be

4	1	2	2
1	3	3	3
2	2	1	4
3	4	4	1

Other points for the other blocks are then generated in the same way until the required number of points (configurations) is achieved. In a sense, this could be equivalent to first generating the points for each block and then mounting them on top of the other sequentially thereby forming the balanced array matrix. It is easy to see that first order balance is maintained in each block. While generating the points of each block, the spacing parameter δ could be introduced and maintained across all the blocks. The above method of constructing balanced array matrices by a rejection rule on the coordinates of the points is obviously expensive. It is cheaper if selection at each stage is made from only the integers actually involved. For example, in Stage 1, of the first block, we need only for the first coordinate (of the second point) to select a number from {1,2,3} and not from {1,2,3,4} since if 4 is selected, it will eventually be rejected. The same is true of the other coordinates of the second point. For the first coordinate of the third point of the first block, we only need to select a number from {2,3} and not from {1,2,3,4} which might lead to more than one rejection and so on.

Clearly, by the time the last row of a block is reached, it is completely determined. For example if n = 4 and we have generated

4	1	2	2
1	3	3	3
2	2	1	4

then the fourth point must be (3,4,4,1).

3.6 Estimates of Mean, Standard Deviation and MSE

The δ-method makes use of the following algorithm for estimating a parameter ϑ of a distribution.

(1) for a data set $X = (X_1, X_2, \ldots, X_n)$ from a distribution F (unknown), construct a nonparametric MLE $\hat{F}$ for F by putting mass 1/n at X_i, i = 1,2,...,n.

(2) Select a point at random in $\mathbb{R}^n$ and regard the resultant transformed sample as a sample drawn with replacement from $\hat{F}$ and calculate a statistic

$$\hat{\vartheta}^* = \hat{\vartheta}(X_1^*, X_2^*, \ldots, X_n^*) \quad . \tag{3.6.1}$$

(3) All other selected points are independent repetitions of (2), obtaining the replicates (statistics) $\hat{\vartheta}_1^*, \hat{\vartheta}_2^*, \ldots, \hat{\vartheta}_{N'}^*$.

For the estimate of the mean, we calculate

$$\hat{\vartheta}_\cdot^* = \frac{1}{N'} \sum_{i=1}^{N'} \hat{\vartheta}_i^* \tag{3.6.2}$$

For the estimation of the standard deviation, we calculate

$$\hat{\vartheta}^*_{STD} = \left\{ \frac{1}{N'-1} \sum_{i=1}^{N'} (\hat{\vartheta}_i^* - \hat{\vartheta}_\cdot^*)^2 \right\}^{\frac{1}{2}} \qquad (3.6.3)$$

For comparing the boostrap sample results with the results obtained for the proposed ART, a useful criterion is the Mean Squared Error (MSE) of the estimators. Usually the theoretical population MSE is defined as

$E_F[(\hat{\vartheta}-\vartheta)^2]$ where $\hat{\vartheta}$ is the estimate for

the parameter ϑ and E_F is expectation.

Now,

$$E_F[(\hat{\vartheta}-\vartheta)^2] = E_F[\{(\hat{\vartheta}-E_{\hat{F}}(\hat{\vartheta})) + (E_{\hat{F}}(\hat{\vartheta})-\vartheta)\}^2] \qquad (3.6.4)$$

which simplifies to

$$E_F[(\hat{\vartheta}-\vartheta)^2] = var_{\hat{F}}(\hat{\vartheta}^*) + (bias)_F^2 \qquad (3.6.5)$$

where $E_F(\hat{\vartheta}) - \vartheta = bias.$ (3.6.6)

If the true value of the parameter ϑ is known, then the observable MSE for ART (or the bootstrap) could be defined as

$$MSE^* = \frac{1}{N'} \sum_{i=1}^{N'} (\hat{\vartheta}_i^* - \vartheta)^2 \qquad (3.6.7)$$

which simplifies to

$$MSE^* = \frac{1}{N'} \sum_{i=1}^{N'} (\hat{\vartheta}_i^* - \hat{\vartheta}_\cdot^*) + (\hat{\vartheta}_\cdot^* - \vartheta)^2 \qquad (3.6.8)$$

i.e. $MSE^* = var_{\hat{F}}(\hat{\vartheta}_i^*) + (bias^*)^2$ (3.6.9

where $\hat{\vartheta}_{\cdot}^{*} - \vartheta = \text{bias}^{*}$ (3.6.10)

If on the other hand, the true value of the parameter ϑ is not known, we estimate it by $\hat{\vartheta}$ and then the observable MSE will be defined as

$$\hat{\text{MSE}}^{*} = \frac{1}{N'} \sum_{i=1}^{N'} (\hat{\vartheta}_i^{*} - \hat{\vartheta})^2 \tag{3.6.11}$$

which simplifies to

$$\hat{\text{MSE}}^{*} = \text{var}(\hat{\vartheta}_i^{*}) + (\hat{\text{bias}}^{*})^2 \tag{3.6.12}$$

where $\hat{\vartheta}_{\cdot}^{*} - \hat{\vartheta} = \hat{\text{bias}}^{*}$. (3.6.13)

In the balance array design, by Theorem 1, we always have that $\hat{\vartheta}_{\cdot}^{*} = \hat{\vartheta}$ (estimates of the mean) and so we will always get a $\hat{\text{bias}}^{*}$ of zero.

3.7 Example

A data set X = {17.3, 9.6,33.8,10.4,24.0,13.0,15.0,21.8,16.6, 17.2} of size 10 was taken at random from a certain test data set of size 90 (Rainfall in inches in Sacramento, California, 1854-1944, see Table 2.3, Alder & Roessler (1964)). The "true value" of the mean, μ, is 17.9378. For various values of δ and various "length" N′, sequence of points (configurations) were generated. The value of $\delta = 0$ corresponds to the pure Monte Carlo (i.e. the standard bootstrap in this case). Table 6 contains the estimated results for the mean and standard deviation. The quantity $\hat{F}$ of the results in Table 6 is the sample cdf of the mean described above in §3.6. The graphs of some of these cdf's are in Figure 1. We considered both the Balanced

Array and the Randomized Unbalanced δ-methods. The Monte Carlo simulation (the standard bootstrap configurations) size is N = 2000 providing a benchmark as a close approximation to complete numeration. For the cdf's, the results given in Figure 1 are typical of a large number of results obtained for all other examples tried in this work (which are however not reported here for want of space). The best "fit" to the Monte Carlo (standard bootstrap) with N = 2000 is provided by the Balanced Array design with $N' \ll N$ for $\delta \geqslant 0$. They show more regular and more uniform jumps (variations). For the simple bootstrap confidence intervals in which tail areas are important, the Balanced Array with $\delta > 0$ always do very well by guaranteeing some points in the tails (especially the upper tail). A rough recommendation is that for a data set of size n, a Balanced Array method of "length" n^2 gives quite a reasonable fit. First order balance will favour results on the mean since the observable bias will be zero. The poor results from the unbalanced randomized δ-method are presumably due to edge effects of using Euclidean distance. However in all examples tried, results for the unbalanced randomized δ-method for $\delta > 0$ seem quite reasonable as well. A point of interest is that there is a very good gain in time and cost as a result of the cutdown of the number of configurations achieved by using the proposed ART. While, for the Monte-Carlo (the standard bootstrap) it took 24.424 cp seconds to generate points, transform their coordinates and do the analysis for the mean, the standard deviation and the square root of MSE for N = 2000, it took 3.983 cp seconds to do the same by the Randomized δ-method (δ=44) for $N' = 200$ and only 2.260 cp seconds for $N' = 100$ for δ=52 and yet the cdf's in Figure 1 for the 200 and 100 selected configurations do comparatively well as the 2000 Monte Carlo benchmark.

We are aware that many of the computation times could be lowered with more work on the sampling procedures in particular by

by replacing the use of the NAG routine by "internal" random number generation. But we feel that the relative times given above are a good indication of the likely results. Furthermore the approach that maybe adopted is to store short codes for use in special problems - rather as an experimental design.

The choice between storage and generation of simulation codes points towards the need for much further work in the speed and "computability" of our methods.

Table 5

Example of resampling code

8	3	4	3	9	1	2	5	1	6
10	7	7	7	3	9	4	4	2	8
9	1	1	1	1	5	7	9	5	9
3	5	9	2	5	7	5	2	7	1
1	8	10	5	6	4	6	7	10	5
5	4	6	10	7	6	3	6	9	2
7	2	3	9	8	8	1	3	6	10
6	10	5	8	4	3	10	1	8	4
4	9	2	6	10	10	8	10	4	7
2	6	8	4	2	2	9	8	3	3
6	7	7	3	3	1	3	5	5	3
5	6	4	2	8	4	1	9	9	7
7	5	8	6	1	5	7	8	8	2
2	1	9	10	4	3	2	7	3	1
8	9	3	8	2	9	9	1	10	4
4	3	2	7	10	8	10	4	1	10
1	10	1	4	5	6	8	6	2	8
10	4	10	5	9	7	4	2	7	5
9	2	5	1	6	10	6	10	4	6
3	8	6	9	7	2	5	3	6	9
4	8	3	9	2	4	9	3	3	7
3	7	4	7	7	9	8	2	2	4
5	5	1	5	6	7	7	5	1	3
1	10	6	6	1	10	10	9	7	2
8	9	2	8	8	3	5	7	4	1
10	3	5	3	3	1	2	8	9	9
7	4	10	2	10	5	3	6	10	6
9	2	8	1	5	2	4	1	8	5
2	1	9	4	9	8	1	10	5	10
6	6	7	10	4	6	6	4	6	8
8	9	6	4	9	4	5	4	9	5
1	8	5	7	7	7	8	3	5	9
10	4	3	5	6	3	9	8	2	6
7	5	2	2	5	6	10	10	10	3
9	7	4	8	2	1	6	1	8	7
3	10	8	10	4	8	4	7	4	2
2	6	1	9	10	10	2	5	1	10
5	2	7	6	8	2	1	9	7	1
4	3	9	1	1	5	7	6	3	8
6	1	10	3	3	9	3	2	6	4
4	8	7	6	8	2	3	1	5	4
6	3	8	8	9	9	9	3	9	8
5	10	2	10	1	6	5	5	6	2
8	6	5	2	7	1	2	4	8	9
3	9	10	9	4	10	1	8	10	5
1	2	3	3	5	5	4	6	3	10
10	7	1	4	10	8	6	9	7	7
7	5	6	5	3	3	10	2	4	1
9	4	9	7	2	4	7	10	1	6
2	1	4	1	6	7	8	7	2	3

3	1	8	3	1	8	6	6	9	7
7	5	3	5	7	5	8	5	1	8
2	6	2	8	8	6	1	1	2	10
6	7	10	9	3	2	2	7	3	6
5	9	7	1	10	3	3	2	7	4
10	10	4	4	6	1	9	4	6	1
1	4	5	2	9	7	4	10	8	2
4	2	6	7	2	4	7	8	4	5
9	3	9	10	5	10	5	3	10	3
8	8	1	6	4	9	10	9	5	9
6	5	8	2	3	10	1	7	2	3
4	9	4	10	2	7	9	6	4	9
3	6	6	3	4	5	7	3	10	10
1	4	1	4	5	1	3	2	8	8
10	3	7	1	10	8	8	4	3	2
5	10	2	9	8	9	5	10	6	5
2	7	5	7	9	6	10	1	5	1
8	1	10	8	7	2	6	5	7	6
9	2	3	5	6	4	4	9	9	7
7	8	9	6	1	3	2	8	1	4
2	7	9	4	3	7	5	9	3	10
10	1	5	2	1	2	9	6	1	1
6	2	7	5	8	1	8	10	8	5
5	4	1	7	9	9	1	8	9	8
3	9	8	3	2	10	2	4	2	7
7	5	3	8	10	5	10	2	7	9
8	8	6	1	6	6	3	1	4	3
4	3	4	9	5	4	7	5	6	2
1	10	2	6	4	8	4	3	10	6
9	6	10	10	7	3	6	7	5	4
4	2	8	10	8	7	10	5	6	2
2	7	4	1	3	1	3	1	5	8
6	6	5	5	6	2	7	8	4	7
3	10	10	2	10	4	9	10	3	4
10	1	3	9	7	8	8	7	9	10
9	8	6	7	4	9	2	9	1	9
5	3	2	3	5	6	4	2	2	5
1	9	1	8	9	10	6	4	10	3
7	4	7	6	2	3	5	3	7	1
8	5	9	4	1	5	1	6	8	6
7	3	4	4	2	2	3	5	1	2
10	8	3	8	3	10	1	8	4	5
2	1	1	7	7	4	2	3	6	8
3	7	6	5	1	3	8	1	2	10
5	2	8	2	9	6	4	9	7	7
4	6	2	1	8	1	5	2	3	4
1	4	10	10	5	8	10	7	8	1
6	10	5	6	6	7	7	6	5	6
9	5	9	9	10	9	6	4	10	3
8	9	7	3	4	5	9	10	9	9

BALANCED ARRAY METHOD

100 points in 10-dimensional space with δ=25

Table 6

Results of the mean and standard deviation for the Monte Carlo (standard bootstrap, $\delta = 0$), the randomized and the balanced array δ - method. The true mean $\mu_{90} = 17.9378$

Number of Simulations (Configurations) in the Reference Set (N)	δ	Number of Selected Configurations in the Restricted Set (N′)	MEAN		STANDARD DEVIATION	
			Randomized	Balanced	Randomized	Balanced
2000	0	2000	17.9305	17.8700	2.2217	2.1541
1500	0	1500	17.9091	17.8700	2.2366	2.1787
1000	0	1000	17.9058	17.8700	2.2251	2.1769
900	0	900	17.9174	17.8700	2.1977	2.1936
800	0	800	17.9286	17.8700	2.1900	2.2164
700	0	700	17.8313	17.8700	2.1615	2.2106
600	0	600	17.8572	17.8700	2.1696	2.2143
500	0	500	17.9032	17.8700	2.1761	2.1783
400	0	400	17.9446	17.8700	2.1864	2.1902
300	0	300	17.9910	17.8700	2.1360	2.2150
200	0	200	18.0923	17.8700	2.1115	2.2015
2000	44	200 (204)	17.9864		1.9824	
	20	200		17.8700		2.2151
100	0	100	18.2284	17.8700	2.2482	2.2280

Table 6 continued...

Table 6 continued...

Number of Simulations (Configurations) in the Reference Set (N)	δ	Number of Selected Configurations in the Restricted Set (N´)	MEAN		STANDARD DEVIATION	
			Randomized	Balanced	Randomized	Balanced
2000	52	100 (109)	18.0818		1.8801	
	25	100		17.8700		2.1567
90	0		18.3219	17.8700	2.2252	2.2562
2000	53	90 (99)	18.1754		1.9053	
	35	90		17.8700		2.1974
80	0	80	18.3466	17.8700	2.2301	2.2457
2000	55	80 (81)	18.1368		1.9585	
	36	80		17.8700		2.2600
70	0	70	18.4510	17.8700	2.2334	2.2991
2000	57	70	18.1254		1.9520	
	38	70		17.8700		2.3238
60	0	60	18.4458	17.8700	2.2552	2.2670

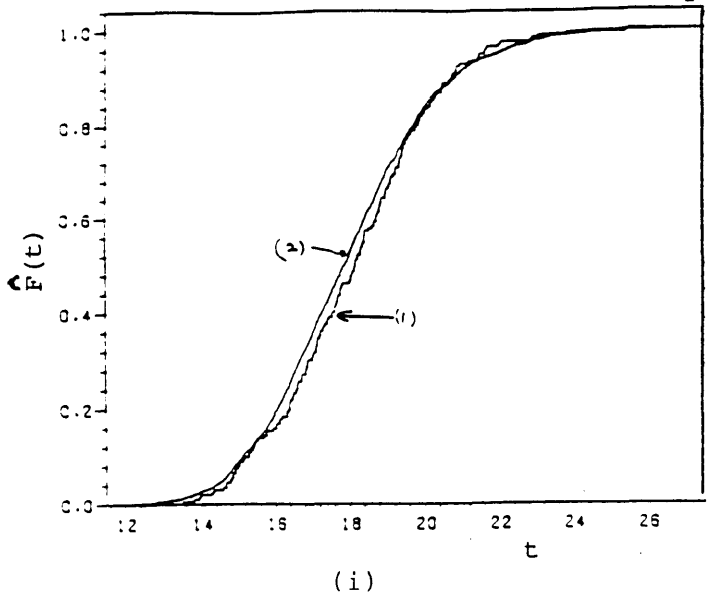
Table 6 continued...

Table 6 continued...

Number of Simulations (Configurations) in the Reference Set (N)	δ	Number of Selected Configurations in the Restricted Set (N´)	MEAN		STANDARD DEVIATION	
			Randomized	Balanced	Randomized	Balanced
2000	58	60 (62)	18.3197		2.0023	
	38	60		17.8700		2.2476
50	0	50	18.2126	17.8700	2.1856	2.2524
2000	59	50 (54)	18.3210		1.9442	
	38	50		17.8700		2.2324
40	0	40	18.3328	17.8700	2.1954	2.3653
2000	63	40	18.1680		1.8992	
	38	40		17.8700		2.1904
30	0	30	18.3073	17.8700	1.8240	2.2593
2000	68	30 (31)	18.0490		1.9930	
	40	30		17.8700		2.2976
20	0	20	17.9545	17.8700	1.6364	2.1565
2000	76	20 (21)	17.9295		2.0373	
	45	20		17.8700		2.1565

NOTE: The numbers in parenthesis are the actual numbers selected.

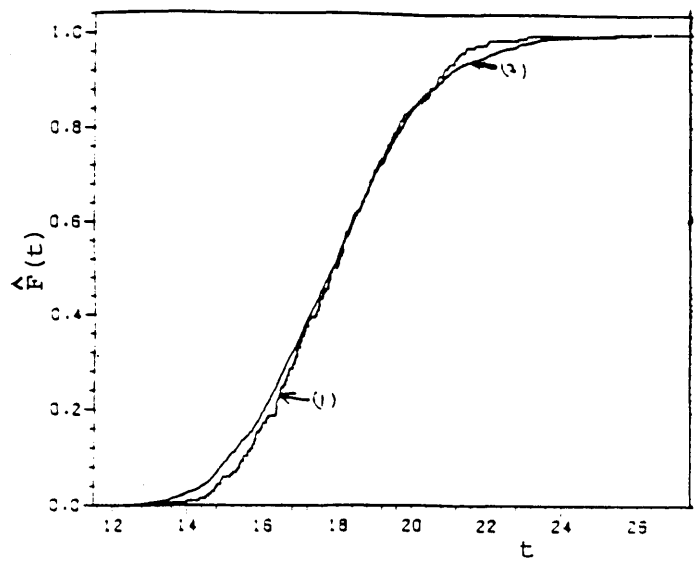
Figure 1: C.D.F. of the mean. $\delta = 0$ is the Monte Carlo.
Randomized means pure Monte Carlo



(i)

(1) Randomized, N´ = 200, $\delta = 0$

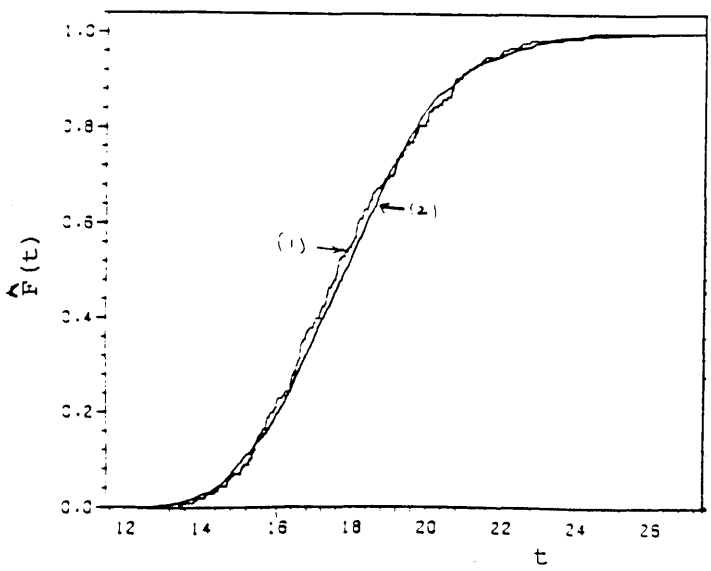
(2) Randomized, N = 2000, $\delta = 0$



(ii)

(1) Randomized, N´ = 200, $\delta = 44$

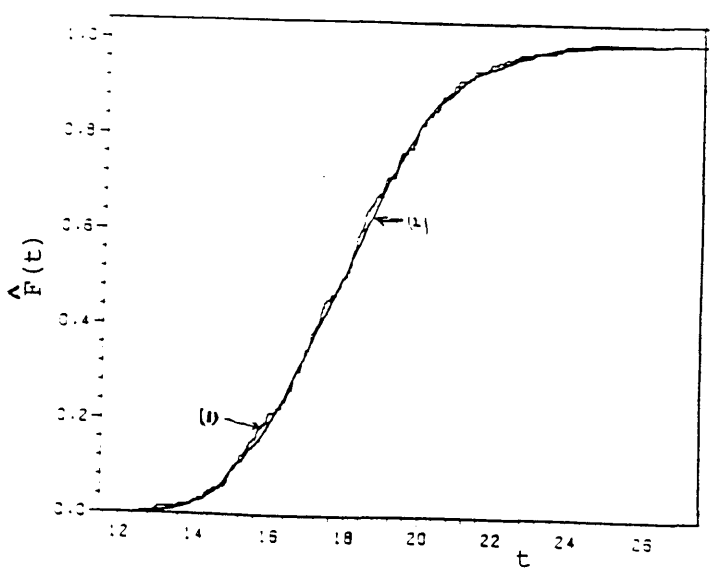
(2) Randomized, N = 2000, $\delta = 0$



(iii)

(1) Balanced, N´ = 200, $\delta = 0$

(2) Randomized, N = 2000, $\delta = 0$

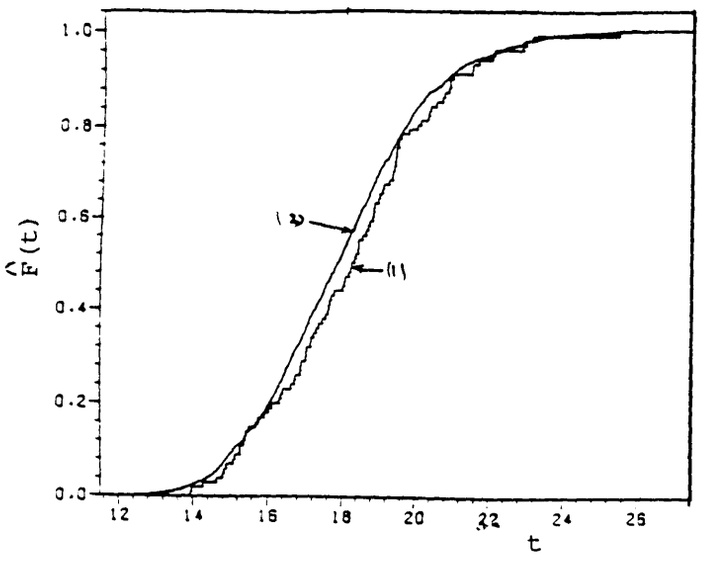


(iv)

(1) Balanced, N´ = 200, $\delta = 20$

(2) Randomized, N = 2000, $\delta = 0$

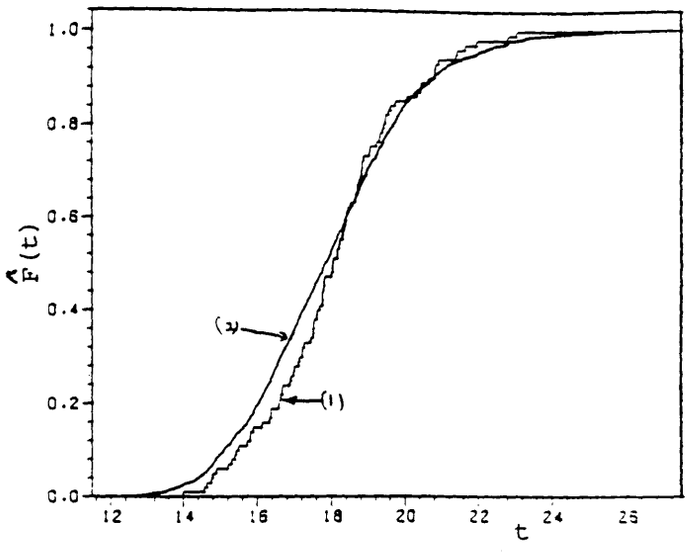
Figure 1 continued...



(v)

(1) Randomized, N´= 100, δ = 0

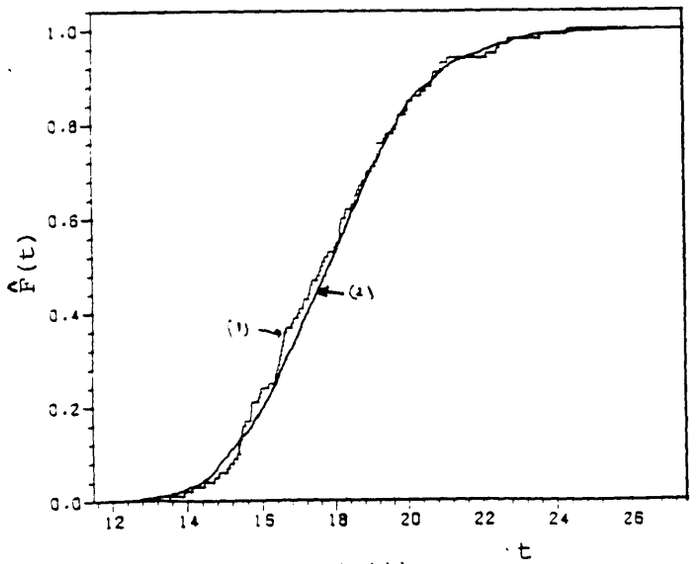
(2) Randomized, N = 2000, δ = 0



(vi)

(1) Randomized, N´ = 100, δ = 52

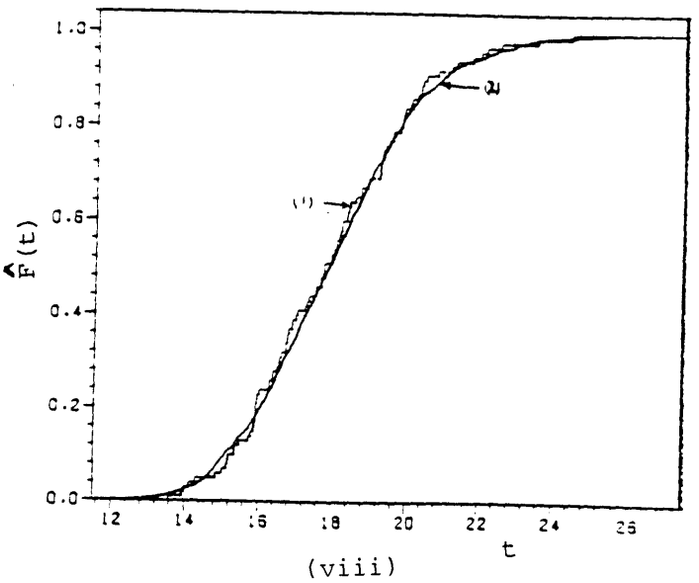
(2) Randomized, N = 2000, δ = 0



(vii)

(1) Balanced, N´ = 100, δ = 0

(2) Randomized, N = 2000, δ = 0



(viii)

(1) Balanced, N´ = 100, δ = 25

(2) Randomized, N = 2000, δ = 0

CHAPTER FOUR

DISCREPANCY

4.1 Introduction

Many studies in the literature reveal that the pseudo-random number generators commonly used in the Monte Carlo method have poor distribution properties in spaces of many dimensions especially in cases where few observations are made in some complex situations. We discuss the distributional properties of the sets (or sequences) of points at which the measurements of the statistical characteristics of interests are computed. There is a need to measure a quantity "discrepancy" which gives a measure of the lack of equidistribution (uniform spread) of the selected (distinct) points. Even spread of selected points gives low measures of discrepancy (see Zaremba, 1968). Of course one of the aims of the proposed ART is to spread out the points (observations) evenly and decrease errors in the quadrature. The success of this gives a guaranteed advantage over the usual Monte Carlo or indeed the standard bootstrap resampling procedures. In this chapter, we follow the approaches of Braaten & Weller (1979), Zaremba (1966, 1968, 1972), Halton (1972) and Warnock (1972) in which low discrepancy point sets are suggested for numerical integration as an alternative to direct application of the Monte Carlo procedure. There are various measures of discrepancy in the literature. We consider below a form of discrepancy which is adaptable to the general bootstrap selection procedure.

4.2 General Methodology and Computation of Discrepancy

Consider a finite sequence $R = \{r_1, r_2, \ldots, r_N\}$ of numbers contained in the interval [0,1]. One measure of discrepancy, denoted by D(R) of R, is the least upper bound for all the intervals I in [0,1],

$$\left| N^{-1} \nu(I,R) - \mu(I) \right|$$

where $\nu(I,R)$ denotes the number of elements of R contained in I and $\mu(I)$ is the length of the interval. Thus D(R) in this case, is the measure of the lack of equidistribution of R over the interval [0,1].
For a more general consideration let us suppose that g is the distribution of N points, not necessarily distinct, in the unit square $0 \leqslant \xi_i \leqslant 1$ (i=1,2). For any point $R = (r_1, r_2)$ satisfying $0 \leqslant r_i \leqslant 1$ (i=1,2), let B(R) denote the "box" consisting of all $\psi = (\psi_1, \psi_2)$ such that $0 \leqslant \xi_i \leqslant 1$ (i=1,2). Let $\eta[g,B(R)]$ be the number of points of the distribution g that fall into the "box" B(R).

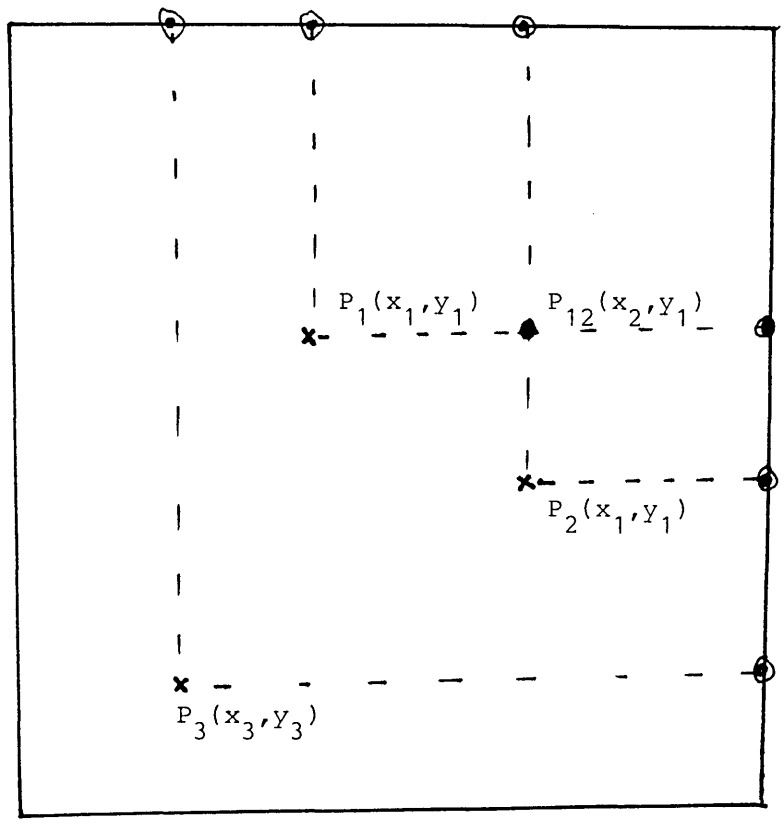
The function,

$$D(R) \leqslant \eta[g,B(R)] - Nr_1.r_2 \qquad (4.2.1)$$

is the measure of the irregularities of the distribution of the sequence of points at R. The irregularities of the distribution g, which is actually the local discrepancy, is measured in a number of ways by considering the behaviour of the function D(R). The above form of local discrepancy, D(R) does not lead to easy computation because the maximum discrepancy may not necessarily occur at any of the selected points of the box B(R).

For example, if we consider the three points P_1, P_2 and P_3

of the 2-dimensional sketch below. The overall maximum discrepancy for the three points may occur at P_1, P_2, P_3 or at the point P_{12} whose coordinates are the maxima of the corresponding coordinates of P_1 and P_2 or it may even occur at the boundary points: $(x_i, 1)$ and $(1, y_i)$. This result can be generalized in an obvious way.



The situation is far worse in higher dimensions. A suitable form of measure of discrepancy considered for the bootstrap and the proposed ART is developed along the lines of Zaremba (1968) but for an easier analytical and computational treatment we adopt Warnock (1972). As in Zaremba (1968), let the domain of integration be a k-dimensional

cube, i.e. $Q^k : 0 \leq \omega^{(i)} < 1 \quad i = 1,2,\ldots,k$ and let

$$S = \{R_0, R_1, \ldots, R_N\} \tag{4.2.2}$$

be any sequence of points of Q^k. Let $\nu(R)$ with

$$R = (r^{(1)}, r^{(2)}, \ldots, r^{(k)}) \tag{4.2.3}$$

be the number of points of S in the k-dimensional interval $0 \leq \omega^{(i)} < 1$, $i = 1,2,\ldots,k$.

The function,

$$g(R) = N^{-1}\nu(R) - r^{(1)}\, r^{(2)} \ldots r^{(k)} \tag{4.2.4}$$

is the local discrepancy of S at R and the norm,

$$D(S) = \sup_{R \in Q^k} |g(R)| \tag{4.2.5}$$

is called the discrepancy or the extreme discrepancy. Furthermore, the norm,

$$T(S) = \left\{ \int_{Q^k} (g(R))^2 dR \right\}^{\frac{1}{2}} \tag{4.2.6}$$

is the mean square (L^2) discrepancy. Upper bounds in terms of either the extreme or mean-square discrepancy (i.e. D(S) or T(S)) are obtained for the error in the appropriate integration of g(R) over Q^k:

$$\text{i.e.} \qquad \text{for} \quad \left| \int_{Q^k} g(R)dR - N^{-1} \sum_{j=0}^{N-1} g(R_j) \right| \quad .$$

In fact, for the one dimensional consideration, Koksma (1960) proved that if the function f(R) has a finite total variation V(f(R)) over [0,1], then for any sequence,

$$R_N = \{r_0, r_1, \ldots, r_{N-1}\} \tag{4.2.7}$$

of points of [0,1], we have,

$$\left| N^{-1}(f(r_0) + \ldots + f(r_{N-1})) - \int_0^1 f(R)dR \right| \leq V(f(R).D(R_N)) \tag{4.2.8}$$

Where V(•) is the total variation.

Hlawka (see Zaremba, 1972) extended Koksma's (1960) idea to an arbitrary number of dimensions. In particular for the two dimensional case, if

$$V = V^{(2)}(f(R)) + V^{(1)}(f(R,1)) + V^{(1)}(f(1,R)) \tag{4.2.9}$$

is finite and where $V^{(1)}$ and $V^{(2)}$ are one and two dimensional variations.

$$R = \{r_0, r_1, \ldots, r_{N-1}\} \tag{4.2.10}$$

is an arbitrary sequence of points of Q^2 subject only to the condition that their coordinates should be smaller than 1, then

$$\left| \int_{Q^k} f(R)dR - N^{-1}(f(r_0) + \ldots + f(r_{N-1}) \right| \leq V.D(R) \tag{4.2.11}$$

The above bound was further extended by Zaremba (1966) to,

$$\left| \int_{Q^k} f(R)dR - N^{-1}(f(r_0) + \ldots + f(r_{N-1}))\right| < V^{(2)}(f(r)).D(R)$$

$$+ V(f(1,R))D(R^{(2)}) + V(f(R,1))D(R^{(1)}) \qquad (4.2.12)$$

where,

$$r_k = (r_k^{(1)}, r_k^{(2)}) \qquad k = 0,1,2,\ldots,N-1 \qquad (4.2.13)$$

$$R^{(i)} = \{r_0^{(i)}, \ldots, r_{N-1}^{(i)}\} \qquad i = 1,2 \qquad (4.2.14)$$

and

$$f(r) = f(r^{(1)}, r^{(2)}) \qquad (4.2.15)$$

There is already a large literature in Number Theory putting upper and lower bounds on discrepancy, D(S) as the number of points $N \to \infty$. For example Roth (1954) in his celebrated work gave a lower bound:

$$D(S) > C_1(k)(\log N)^{\frac{1}{2}(k-1)} \qquad (4.2.16)$$

where k is the number of dimensions. Roth's (1954) bound is an improvement on Van Aardenne-Ehrenfest (1949) result of

$$D(S) > C_2(k) \frac{\log \log N}{\log \log \log N} \qquad (4.2.17)$$

Halton (1960) gave an upper bound of

$$D(S) < C_3(k)(\log N)^{k-1} \qquad (4.2.18)$$

Schmidt (1972) gave a bound of

$$D(S) > C_3(k) \log N \tag{4.2.19}$$

The combination of Halton (1960) and Schmidt (1972) bounds for k = 2 yields

$$\log N\, C_4(k) < D(S) < C_3(k) \log N \tag{4.2.20}$$

which shows that $D(S) \sim O(\log N)$.

Technically, from the basic definition and discussion so far, the local discrepancy (or the extreme discrepancy) is seen to be the difference between the empirical distribution function of the sequence and the theoretical distribution of the uniform distribution.

4.3 Computation of Discrepancy in the Proposed ART

Let our domain of reference be, say, a n-dimensional hypercube $1 \leqslant \xi_i \leqslant k$, $i = 1,2,\ldots,n$.

Let $S' = (\omega^{(1)},\omega^{(2)},\ldots,\omega^{(N')})$. Our purpose is to approximate an integral of a function $f(\cdot)$ over a reference set S by the integral over a restricted set $S' \subseteq S$, the integral being with respect to the uniform measure in each case. In what follows, N or N′ are the number of selected points for the sequence (reference set) S or for the restricted set S′respectively.

Let us consider a k-lattice S in n-dimensions

$$S = \left\{(\omega_1,\omega_2,\ldots,\omega_k) \,|\, \omega_i \in \left\{1,2,\ldots,k\right\} \ , \ i=1,2,\ldots,k\right\} \tag{4.3.2}$$

Let $S' \subseteq S$. We can then define the "empirical distribution function" of S'as

$$V_{S'}(\omega) = \frac{1}{N'} \# \left\{j \,|\, j \in S', \ \omega_i \leqslant j_i \qquad i = 1,2,\ldots,n\right\} \tag{4.3.3}$$

$$V_{S'}(\omega) = \frac{1}{N'} \sum_{j \in S'} I(\omega)_{[j,1]} \tag{4.3.4}$$

where [j,1] is the shorthand notation for the closed rectangle,

$$\{\eta \,|\, j \leqslant \eta_i \leqslant 1, \ i = 1,2,\ldots,n\}$$

For S itself, we may evaluate $V_S(\omega)$ as

$$V_S(\omega) = \frac{1}{k^n} \prod_{i=1}^{n} \omega_i \tag{4.3.5}$$

$V_S(\omega)$ is the c.d.f. for the uniform distribution on S. The local discrepancy of S' at ω is defined to be

$$g(\omega) = \frac{1}{N'} \sum_{j \in S'} I(\omega)_{[j,1]} - \frac{1}{k^n} \prod_{i=1}^{n} \omega_i \tag{4.3.6}$$

i.e. $$g(\omega) = V_{S'}(\omega) - V_S(\omega) \ . \tag{4.3.7}$$

The discrepancy (or the extreme discrepancy) is

$$D(S') = \sup_{\omega \in S'} |g(\omega)| \qquad (4.3.8)$$

and the mean-square discrepancy is

$$T(S') = [\frac{1}{N'} \sum_{\omega \in S'} (g(\omega))^2]^{\frac{1}{2}} \quad , \qquad (4.3.9)$$

(4.3.8) and (4.3.9) are simply the L^∞ and L^2 norms with respect to the uniform measure putting mass $N^{-1} = k^{-n}$ at each point of S. It is easier computationally, to consider $T(S')$ (or $N'^2T^2(S')$) and this is what we shall do shortly.

Let us define,

$$\varepsilon = \frac{1}{N'} \sum_{\omega \in S'} f(\omega) - \frac{1}{N} \sum_{\omega \in S} f(\omega) \quad . \qquad (4.3.10)$$

Following Theorem 1 in §3.5, (4.3.10) in effect means that ε could be written as

$$\varepsilon = E_{S'}(f(\omega)) - E_S(f(\omega)) \quad . \qquad (4.3.11)$$

We shall obtain a bound on $|\varepsilon|$ depending on (i) $T(S')$ and its lower dimension projects and (ii) assessment of the variation in f.

Let us define the backward and forward difference operators on a particular component ω_i for any function $V(\omega)$ on S respectively as

$$\Delta^{1-}_{\omega_i} = V(\omega_1,\omega_2,..,\omega_i,...,\omega_n) - V(\omega_1,...,\omega_i-1,...,\omega_n) \qquad (4.3.12)$$

$$\Delta^{1+}_{\omega_i} = V(\omega_1,\omega_2,...,\omega_i+1,...,\omega_n) - V(\omega_1,\omega_2,...,\omega_i,...,\omega_n) \qquad (4.3.13)$$

More generally, let T(h) be a subset of size n of components of ω and then let us define

$$\Delta^{h\pm}_{T(h)} = \prod_{\omega_i \in T(h)} \Delta^{1\pm}_{\omega_i} \qquad (4.3.14)$$

(4.3.14) is simply the difference analogue of the n-fold partial derivative $\frac{\partial U}{\partial \omega_i \partial \omega_j}$ where $\omega_i, \omega_j, \ldots, \in T(h)$ (see Zaremba, 1972). In general, let us define $\tilde{S}(T(h)) \subseteq S$ as the sublattice with $1 \leqslant \omega_i < k$ for $\omega_i \in T(h)$ and $\omega_i = k$, $\omega_i \notin T(h)$. Let $U(\omega)$ and $V(\omega)$ be functions defined on S (with zero entries adjoined) for which $V(\omega) = 0$ if any $\omega_i = 0$ and $V(k,k,\ldots,k) = 1$. Then Ogbonmwan and Wynn (1985) give the general relation

$$\sum_{\omega \in S} U(\omega)\, \Delta^{n-}_{\omega}\, V(\omega) = \sum_{h=1}^{n} (-1)^h \sum_{T(h)} \sum_{\omega \in \tilde{S}(T(h))} \Delta^{h+} U(\omega) V(\omega) \qquad (4.3.15)$$

We can apply (4.3.15) to our particular problem in which $U = t(x(\omega))$ and $V = g(\omega)$. Define the L^2 (unnormalized) discrepancy of $g(\omega)$ for the sublattice $S(T(h))$ where $n = k$ as

$$\tilde{W}(T(h)) = \left(\sum_{\omega \in \tilde{S}(T(h))} |g(\omega)|^2 \right)^{\frac{1}{2}} \qquad (4.3.16)$$

By taking the modulus of (4.3.16) and using the Cauchy-Schwartz inequality we have the following:

THEOREM 2 :

Let $S' \subseteq S$ and define the local discrepancy as in (4.3.16).

Then for a (bounded) function t on S,

$$|\varepsilon| \leq \frac{1}{n^n} \sum_{k=1}^{n} \sum_{T(h)} \tilde{W}(T(h)) \left[\sum_{\omega \in \tilde{S}(T(h))} \left(\Delta_{T(h)}^{h+} t(x(\omega)) \right)^2 \right]^{\frac{1}{2}} \tag{4.3.17}$$

The full proof of the above theorem is given in the Appendix. (4.3.17) expresses the error in the "bootstrap" estimation of $t(x(\omega))$ over S' rather than S. An obvious consequence of (4.3.17) is that if we know that the variation in $t(x(\omega))$ is small in the sense that $\Delta_{T(h)}^{h+} t(x(\omega))$ is small for larger values of h, we can reduce the error by making the $\tilde{W}(T(h))$ small for smaller values of h. As a consequence we could therefore extend $\tilde{W}(T(h))$ to

$$W(T(h)) = \left(\sum_{\omega \in S(T(h))} (g(\omega))^2 \right)^{\frac{1}{2}} \tag{4.3.18}$$

The only difference being the inclusion of the terms with $\omega_i = k$ rather than $\omega_i < k$ (for $\omega_i \in T(h)$).

Definition: A reduced reference set $S' \subseteq S$ $(n=k)$ is called kth order balance if $W(T(h)) = 0$ for all T(h) with $h = k$

An obvious conclusion from the above definition is that if $W(t(h)) = 0$ for all T(h) with $h = k$, then $W(T(h)) = \tilde{W}(T(h)) = 0$ for all $1 \leq h \leq k$. Thus our definition of kth order balance is enough to make all the kth order and lower order terms zero in (4.3.17). Hence, geometrically kth order balance means merely that every k-dimensional projection of the S' sequence is uniform on the grid $S(T(h))$. Our special interest in this work on discrepancy will again be first order balance which turns out to be a valuable property

(see Chapter 3). In effect, this means that each component ω_i $(1 \leqslant i \leqslant n)$ take the values $1,2,\ldots,n$ the same number of times $\frac{N'}{n}$ in every entry. This is perfectly in line with our earlier discussion on first order balance which is summed up in Theorem 1.

Now, let us redefine the mean square discrepancy of the k-lattice S in n-dimensions as:

$$T(S') = N' \left\{ \frac{1}{k^n} \sum_{i=1}^{n} \cdots \sum_{i=1}^{n} (g(\omega))^2 \right\}^{\frac{1}{2}} \tag{4.3.19}$$

which implies that

$$T^2(S') = N'^2 \frac{1}{k^n} \sum_{i=1}^{n} \cdots \sum_{i=1}^{n} (g(\omega))^2 \tag{4.3.20}$$

In which case, our measure puts mass $\frac{1}{k^n}$ at each of the grid points of S.

Thus,

$$T^2(S') = N'^2 \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} (g(\omega))^2 \quad . \tag{4.3.21}$$

Writing $g(\omega)$ as

$$g(\omega) = N'^{-1} \sum_{m=1}^{N'} \prod_{i=1}^{n} H(\omega_i - \gamma_{mi}) - \prod_{i=1}^{n} \omega_i \tag{4.3.22}$$

where $H(\omega)$ is the Heaviside function, with $H(\omega) = 1$ if $\omega \geqslant 0$ and $H(\omega) = 0$ if $\omega < 0$.

From (4.3.22), we get

$$(g(\omega))^2 = N'^{-2} \sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} H(\omega_i-\gamma_{mi})H(\omega_i-\gamma_{qi})$$

$$- 2N'^{-1} \sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i-\gamma_{mi}) + \left(\prod_{i=1}^{n} \omega_i\right)^2 \qquad (4.3.23)$$

This implies that,

$$N'^2(g(\omega))^2 = \sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} H(\omega_i-\gamma_{mi})H(\omega_i-\gamma_{qi})$$

$$- 2N' \sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i-\gamma_{mi}) + N'^2\left(\prod_{i=1}^{n} \omega_i\right)^2 \qquad (4.3.24)$$

Hence,

$$N'^2T^2(S') = \frac{1}{k^n} \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} H(\omega_i-\gamma_{mi})H(\omega_i-\gamma_{qi}) \right.$$

$$\left. - 2N' \sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i-\gamma_{mi}) + N'^2\left(\prod_{i=1}^{n} \omega_i\right)^2 \right] \qquad (4.3.25)$$

which implies that

$$N'^2T^2(S') = \frac{1}{k^n} \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} H(\omega_i-\gamma_{mi})H(\omega_i-\gamma_{qi}) \right.$$

$$- 2 \cdot \frac{1}{k^n} \cdot N' \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i-\gamma_{mi}) \right]$$

$$\left. + \frac{N'^2}{k^n} \cdot \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\left(\prod_{i=1}^{n} \omega_i \right)^2 \right] \right. \qquad (4.3.26)$$

Simplifying we get

$$\frac{1}{k^n} \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} H(\omega_i-\gamma_{mi})H(\omega_i-\gamma_{qi}) \right]$$

$$= \frac{1}{k^n} \cdot \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} (\varphi)$$

where $\varphi = H(\omega_i-\gamma_{mi})H(\omega_i-\gamma_{qi})$.

Now $\varphi = 1$ if $\omega_i \geq \max(\gamma_{mi},\gamma_{qi})$

$= 0$ otherwise .

$$\sum_{\omega_1=1}^{k} (\varphi) = k-\max(\gamma_{mi},\gamma_{qi}) + 1$$

$$\therefore \quad \frac{1}{k^n} \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} (\varphi) = \frac{1}{k^n} \sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} (k-\max(\gamma_{mi},\gamma_{qi})+1)$$

$$= \sum_{m=1}^{N'} \sum_{q=1}^{N'} (1+ \frac{1}{k} - \frac{\max}{k} (\gamma_{mi},\gamma_{qi})) \quad . \qquad (4.3.27)$$

Simplifying the second term of (4.3.26) we have,

$$\prod_{i=1}^{n} \sum_{\omega_i=1}^{k} \omega_i H(\omega_i-\gamma_{mi}) = \prod_{i=1}^{n} (\gamma_{mi}+(\gamma_{mi}+ \frac{1}{k})+ \ldots+ 1) \quad .$$

Let $\eta = k\gamma_{mi}$.

then $\gamma_{mi} + (\gamma_{mi} + \frac{1}{k}) + \ldots + 1 = \frac{1}{k}\,(\eta+(\eta+1)+\ldots+k)$

$$= \frac{1}{k}\left(\frac{k(k+1)}{2} - \frac{(\eta-1)\eta}{2}\right)$$

$$= \frac{1}{2k}\,(k^2+k-\eta^2+\eta)$$

Hence we have,

$$-2 \cdot \frac{1}{k^n} \cdot N' \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i-\gamma_{mi}) \right]$$

$$= -2^{-n+1} \cdot N' \sum_{m=1}^{N'} \frac{1}{k^2}\,(k^2+k-\eta^2+\eta)$$

$$= -2^{-n+1} \cdot N' \sum_{m=1}^{N'} \left(1 + \frac{1}{k} - \frac{\eta^2}{k^2} + \frac{\eta}{k^2}\right)$$

Now putting back $\eta = k\gamma_{mi}$ we get

$$-2 \cdot \frac{1}{k^n} \cdot N' \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i-\gamma_{mi}) \right.$$

$$= -2^{-n+1} \cdot N' \cdot \sum_{m=1}^{N'} \prod_{i=1}^{n} \left(1 + \frac{1}{k} - \frac{k^2\gamma_{mi}}{k^2} + \frac{\gamma_{mi}}{k}\right)$$

$$= -2^{-n+1} \cdot N' \cdot \sum_{m=1}^{N} \prod_{i=1}^{n} \left(1 + \frac{1}{k} - \gamma_{mi}^{\,2} + \frac{\gamma_{mi}}{k}\right)$$

$$= -2^{-n+1} \cdot N' \cdot \sum_{m=1}^{N} \prod_{i=1}^{n} \left(1-\gamma_{mi}^{\,2} + \frac{\gamma_{mi}+1}{k}\right) \qquad (4.3.28)$$

Simplifying the third term of (4.3.26) we have by using $\sum i^2 = \frac{k(k+1)(2k+1)}{6}$ that

$$\frac{N'^2}{k^n} \cdot \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\left(\prod_{i=1}^{n} \omega_i \right)^2 \right]$$

$$= N'^2 \left(\frac{k(k+1)(2k+1)}{6k^3} \right)^n \qquad (4.3.29)$$

Hence substituting we have, (4.3.26) to be

$$N'^2 T^2(S') = \sum_{m=1}^{N'} \sum_{q=1}^{N'} \prod_{i=1}^{n} \left(1 + \frac{1}{k} - \max(\gamma_{mi}, \gamma_{qi})\right)$$

$$-2^{-n+1} N' \cdot \sum_{m=1}^{N'} \prod_{i=1}^{n} \left(1 - \gamma_{mi}^2 + \frac{\gamma_{mi}+1}{k} \right)$$

$$+ N'^2 \left(\frac{k(k+1)(2k+1)}{6k^3} \right)^n . \qquad (4.3.30)$$

With $k = n$, we obtain the square discrepancy for the n^n grid.

For pure Monte Carlo, the γ_{mi} are all independent random variables uniformly distributed on $\{1,2,\ldots,k\}$. Explicit computation of the expectation of (4.3.30) yields

$$E(T^2(S')) = \frac{1}{N'} \left[\left(\frac{1}{2} \right)^n \left(1 + \frac{1}{k} \right)^n - \left(\frac{1}{3} \right)^n \left(1 + \frac{3}{2k} + \frac{1}{2k^2} \right)^n \right]$$

(4.3.31)

In the continuous case (as $k \to \infty$), we obtain $E(T^2(S'))$ to be

$$E(T^2(S')) = \frac{1}{N'}\left[\left(\frac{1}{2}\right)^n - \left(\frac{1}{3}\right)^n\right] \qquad (4.3.32)$$

which is formula (21) in Halton (1972). In (4.3.32), the $\frac{1}{k}$ and $\frac{1}{k^2}$ terms arise from the discrete nature of the present problem. We note that when $n = k$ and $n \to \infty$ in (4.3.32),

$$E(T^2(S')) \longrightarrow \frac{1}{N'}\left[\left(\frac{1}{2}\right)^n e - \left(\frac{1}{3}\right)^n e^{3/2}\right] \qquad (4.3.33)$$

which shows that the discretization persists as the same size increases.

When Monte Carlo is carried out which is pure randomization except for the first order balance constraint again $E(T^2(S'))$ can be explicitly evaluated. This follows since $\max(\gamma_{mi}, \gamma_{qi})$ has a distribution which can be evaluated and the different components γ_i, γ_j remain independent. Thus we model this sampling by saying that γ_{mi}, γ_{qi} are sampled without replacement from a list of strata with $\frac{N'}{k}$ in each stratum. For this sampling we obtain

$$E(T^2(S')) = \frac{1}{N'}\left\{\left(\frac{1}{2}\right)^n - \left(1+\frac{1}{k}\right)^n - \left(\frac{1}{3}\right)^n \left(1+\frac{3}{2k} - \frac{1}{2(N'-1)} + \frac{1}{(N'-1)} \cdot \frac{1}{2k^2}\right)^n\right\}$$

for $k > 1$ this is *less than* the pure Monte Carlo value given in (4.3.31). We now summarize this as:

THEOREM 3:

The expected square L^2 discrepancy for first order balanced Monte Carlo is less than for pure Monte Carlo in every dimension greater

than one.

Although the improvement is small for large N' the above Theorem 3 gives a rationale for using balanced sequences rather than unbalanced.

In Table 7, we give the raw discrepancy using equation (4.3.30) for various numbers of selected points (configurations). Surprisingly, the Randomized δ-method for $\delta > 0$, did not do so well. But the Balanced Array δ-method, for $\delta \geq 0$, gave excellent results which are very much better than the results obtained from the straight Monte Carlo or the standard bootstrap procedure. Figure 2 gives the comparative plots of the raw discrepancy for the Balanced Array δ-method versus the Monte Carlo. The $\sqrt{E(T^2(S'))}$ is for the values got using (4.3.28). These values of $\sqrt{E(T^2(S'))}$ seen to be higher are perhaps partly due to the inequality of $(E(T^2(S')))^{\frac{1}{2}} > E(T(S'))$. These results for $\sqrt{E(T^2(S'))}$ had to be used because of the difficult analytical problem of evaluating $E(T(S'))$ directly. This is not an unusual approach. Warnock (1972) adopted the same approach for the comparative studies of the various low-discrepancy sequences. In figure 2(iii) which looks rather crowded, we tried to accommodate the discrepancy for various sizes of "configurations" contained in Table 6. Clearly, we observe that as the number of points (configurations) increases, the discrepancy decreases. For very large number of points, say $N' \geq 1000$, the discrepancy tend to agree for all the methods adopted in this work. This is not true of small number of points as also shown in figure 2(i) and 2(ii). For our range of interest $[50 \leq N' \leq 200]$, we observe that the δ-method, especially the Balanced Array δ-method with $\delta \geq 0$ does very well giving low discrepancy for relatively small numbers of configurations. In comparing the methods adopted, it is not so much the vertical difference in $T(S')$ at a given

N' that is very important, but the number of points (configurations) required to give a given $T(S')$. The plot of figure 3, that is, $N'T^2(S')$, is highly recommended. The relative heights of the curves give a rough estimate of the efficiency of the different methods over the range of interest [50-200] recommended for the proposed ART rather than the usual $\geq$ 1000 common in bootstrap experiments. For the cases considered here, the Balanced Array δ-method, with $\delta \geq 0$ gives a relative improvement of about 25%. In chapter 3, a value of N = 2000 points, was set as our benchmark for the straight Monte Carlo (standard bootstrap, $\delta = 0$). Suppose $\hat{F}_N$ is the empirical cdf for computed N bootstrap means for a reference set S. Let $\hat{F}_{N'}$ be the emperical cdf for N'means for a restricted set S'. We define the "Emperical Bootstrap Bias" of the cdf as

$$EBB_{N'}(\bar{x}^*) = \max \left| \hat{F}_N(\bar{x}^*) - \hat{F}_{N'}(\bar{x}^*) \right| \qquad (4.3.34)$$

$EBB_{N'}(\bar{x}^*)$ gives an indication of the measure of the "error" in using $\hat{F}_{N'}$ as an estimate of $\hat{F}_N$. Figure 4 gives the plot of EBB of the cdf's versus the corresponding discrepancies for various values of N'. The general trend, is that the larger the $EBB_{N'}(\bar{x}^*)$ of the approximating density function, the larger the discrepancy of the approximating restricted set S'. Thus there is a rough linear relationship between $EBB_{N'}(\bar{x}^*)$ and discrepancy. Again the Balanced Array δ-method provides the least values of $EBB_{N'}(\bar{x}^*)$ which gives another clear indication of its good performance as a reasonable resampling design.

Table 7

Computed Discrepancy for Configurations in 10-Dimensional Space

Number of Simulations (Configurations) (N)	δ	Number of Selected Configurations N′	DISCREPANCY T(S′) x 10^2			$N'T^2(S') \times 10^2$			$\text{Max}\lvert\hat{F}_{2000}-\hat{F}_{N'}\rvert$	
			Randomized	Balanced	$\sqrt{E(T^2(S'))}$	Randomized	Balanced	$N'(T^2(S'))$	Randomized	Balanced
2000	0	2000	0.1097	0.1069	0.1109	0.2407	0.2285	0.2461	0.0000	
1500	0	1500	0.1184	0.1182	0.1281	0.2102	0.2095	0.2461	0.0108	
1000	0	1000	0.1392	0.1422	0.1569	0.1938	0.2022	0.2461	0.0155	0.0280
900	0	900	0.1470	0.1498	0.1654	0.1945	0.2020	0.2461	0.0146	0.0301
800	0	800	0.1576	0.1557	0.1754	0.1987	0.1939	0.2461	0.0230	0.0275
700	0	700	0.1728	0.1686	0.1875	0.2090	0.1989	0.2461	0.0229	0.0300
600	0	600	0.1838	0.1876	0.2025	0.2027	0.2112	0.2461	0.0222	0.0323
500	0	500	0.1941	0.2078	0.2219	0.1883	0.2159	0.2461	0.0200	0.0370
400	0	400	0.2253	0.2359	0.2481	0.2030	0.2226	0.2461	0.0250	0.0390
300	0	300	0.2637	0.2565	0.2864	0.2086	0.1974	0.2461	0.0382	0.0565
200	0	200	0.3185	0.3171	0.3508	0.2029	0.2011	0.2461	0.0580	0.0505
2000	44	200 (204)	0.3568		"	0.2546		"	0.052	
	20	200		0.3053	"		0.1864	"		0.0355
100	0	100	0.4399	0.4482	0.4961	0.1935	0.2009	0.2461	0.0770	0.0630
2000	52	100 (109)	0.4725		"	0.2232		"	0.0940	

Table 7 (cont'd..)

Computed Discrepancy for Configurations in 10-Dimensional Space

Number of Simulations (Configurations) (N)	δ	Number of Selected Configurations N′	DISCREPANCY $T(S')\times 10^2$			$N'T^2(S')\times 10^2$			$\text{Max}\lvert\hat{F}_{2000}-\hat{F}_{N'}\rvert$	
			Randomized	Balanced	$\sqrt{E(T^2(S'))}$	Randomized	Balanced	$N'(T^2(S'))$	Randomized	Balanced
	25	100		0.3963			0.1570	"		0.0510
90	0	90	0.4451	0.4077	0.5230	0.1783	0.1496	0.2461	0.0903	0.0708
2000	53	90 (99)	0.4727		"	0.2011		"	0.1047	
	35	90		0.4027	"		0.1459	"		0.0614
80	0	80	0.4658	0.4420	0.5547	0.1736	0.1563	0.2461	0.1025	0.0695
2000	55	80 (81)	0.5220		"	0.2180		"	0.0870	
	36	80		0.4194	"		0.1407	"		0.0670
70	0	70	0.4926	0.4615	0.5930	0.1699	0.1491	0.2461	0.1364	0.0741
2000	57	70	0.5602		"	0.2197		"	0.0799	
	38	70		0.4644	"		0.1510	"		0.0741
60	0	60	0.5396	0.5109	0.6405	0.1747	0.1566	0.2461	0.1317	0.0787
2000	58	60 (62)	0.5828		"	0.2038		"	0.1338	
	38	60		0.4848	"		0.1410	"		0.0845

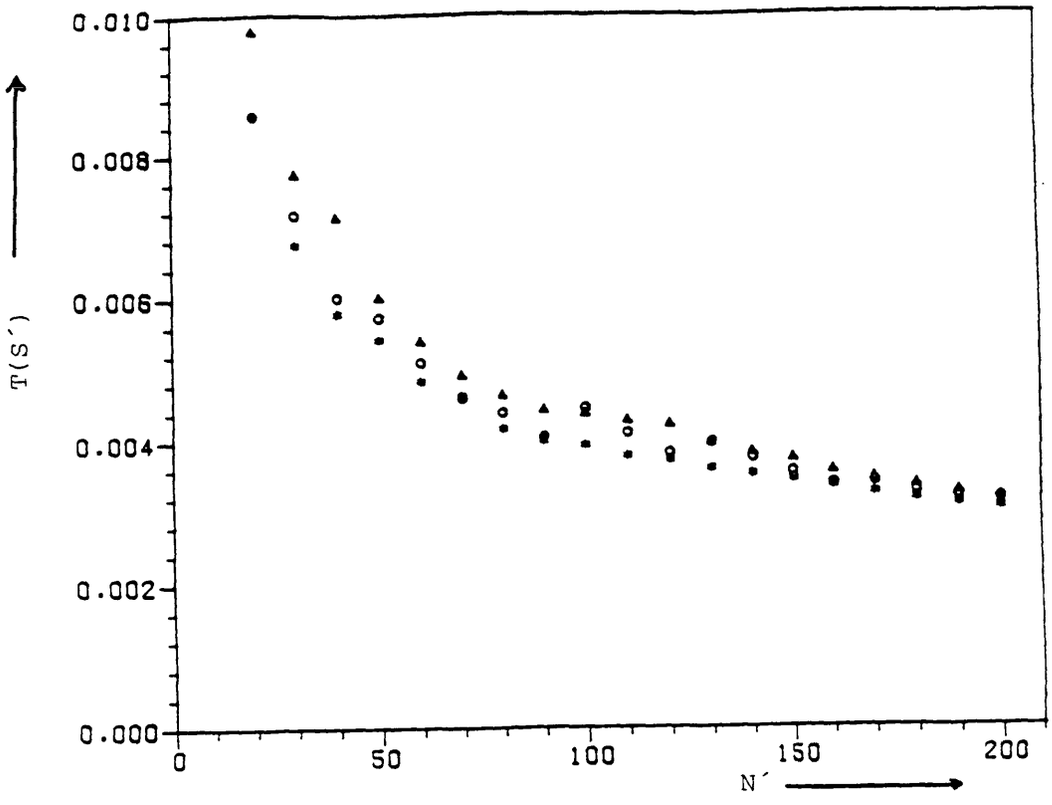
Table 7 (cont'd...)

Computed Discrepancy for Configurations in 10-Dimensional Space

Number of Simulations (Configurations) (N)	δ	Number of Selected Configurations N′	DISCREPANCY $T(S') \times 10^2$			$N'T^2(S') \times 10^2$			$\text{Max}\lvert\hat{F}_{2000}-\hat{F}_{N'}\rvert$	
			Randomized	Balanced	$\sqrt{E(T^2(S'))}$	Randomized	Balanced	$N'(T^2(S')$	Randomized	Balanced
50	0	50	0.6005	0.5732	0.7016	0.1803	0.1643	0.2461	0.0850	0.0920
2000	59	50 (54)	0.6309			0.1990		"	0.1705	
	38	50		0.5432			0.1475	"		0.0945
40	0	40	0.7134	0.6014	0.7844	0.2036	0.1446	0.2461	0.1150	0.1120
2000	63	40	0.6608		"	0.1747		"	0.1325	
	38	40		0.5789	"		0.1341	"		0.0910
30	0	30	0.7754	0.7177	0.9058	0.1804	0.1545	0.2461	0.1650	0.1012
2000	68	30 (31)	0.7865			0.1903		"	0.0943	
	38	30		0.6753			0.1368	"		0.0910
20	0	20	0.9773	0.8588	1.1094	0.1910	0.1475	0.2461	0.1505	0.1170
2000	76	20 (21)	0.9636		"	0.1857		"	0.0760	
	45	20		0.8588	"		0.1475	"		0.1170

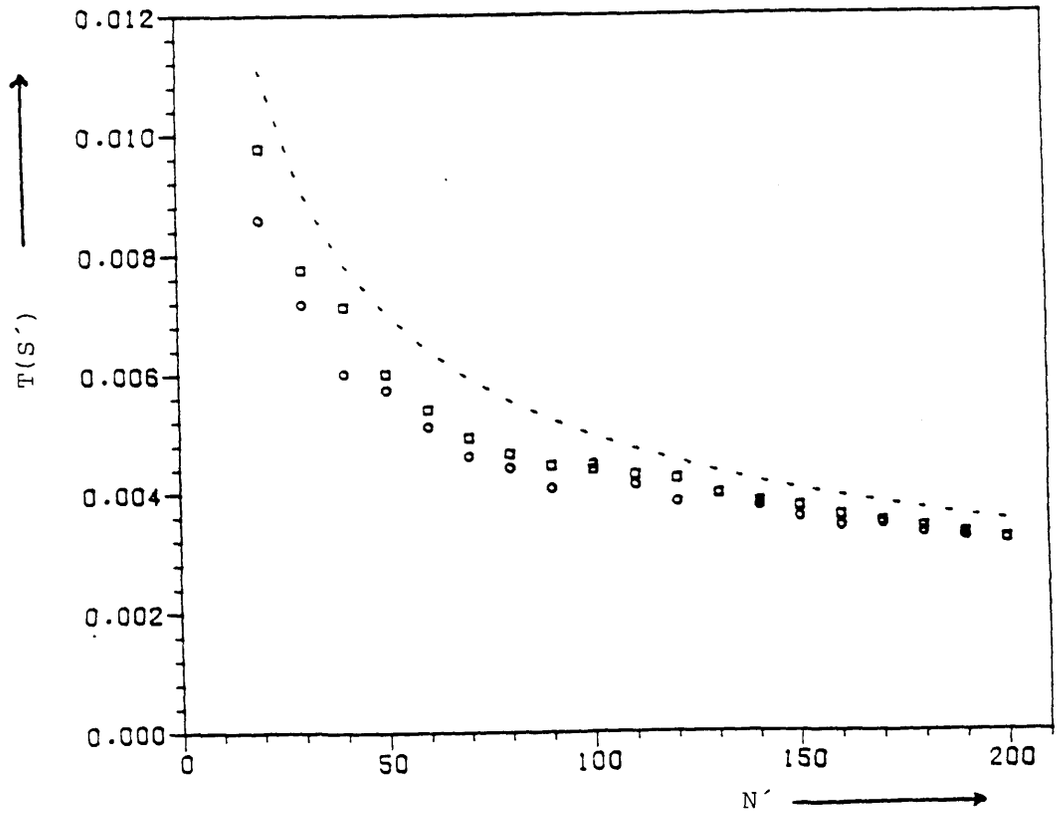
NOTE: The numbers in parenthesis are the actual numbers selected.

Figure 2: Discrepancy, T(S´) versus number of configurations, N´ in 10-dimensions



(i)

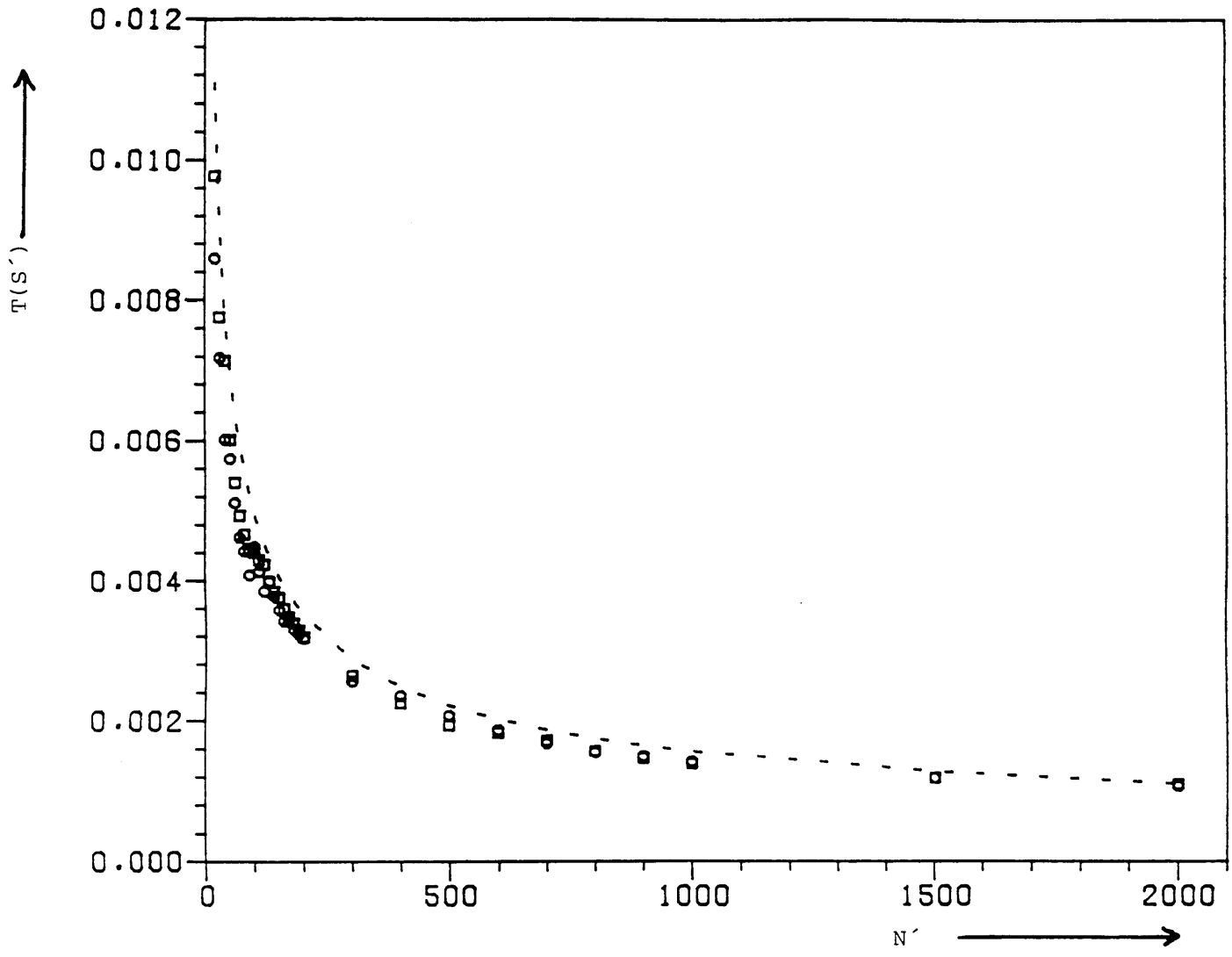
ΔΔΔΔΔΔ Randomized, δ=0; ○○○○○ Balanced Array, δ=0; **** Balanced Array, δ ⩾ 20



(ii)

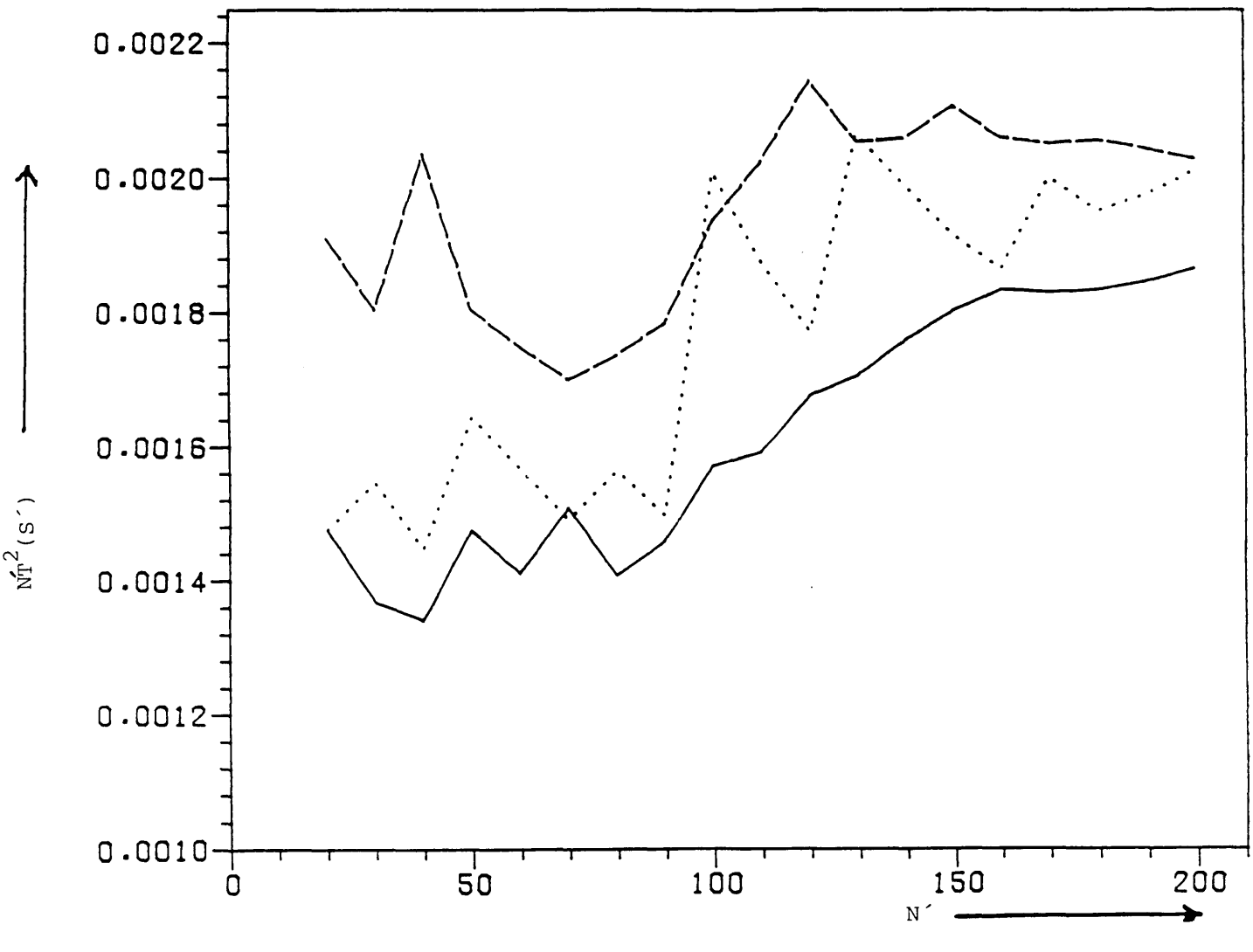
□□□□□ Randomized, δ=0; ○○○○○○ Balanced Array, δ=0, ----- SQRT of $E(T^2(S'))$

(iii)



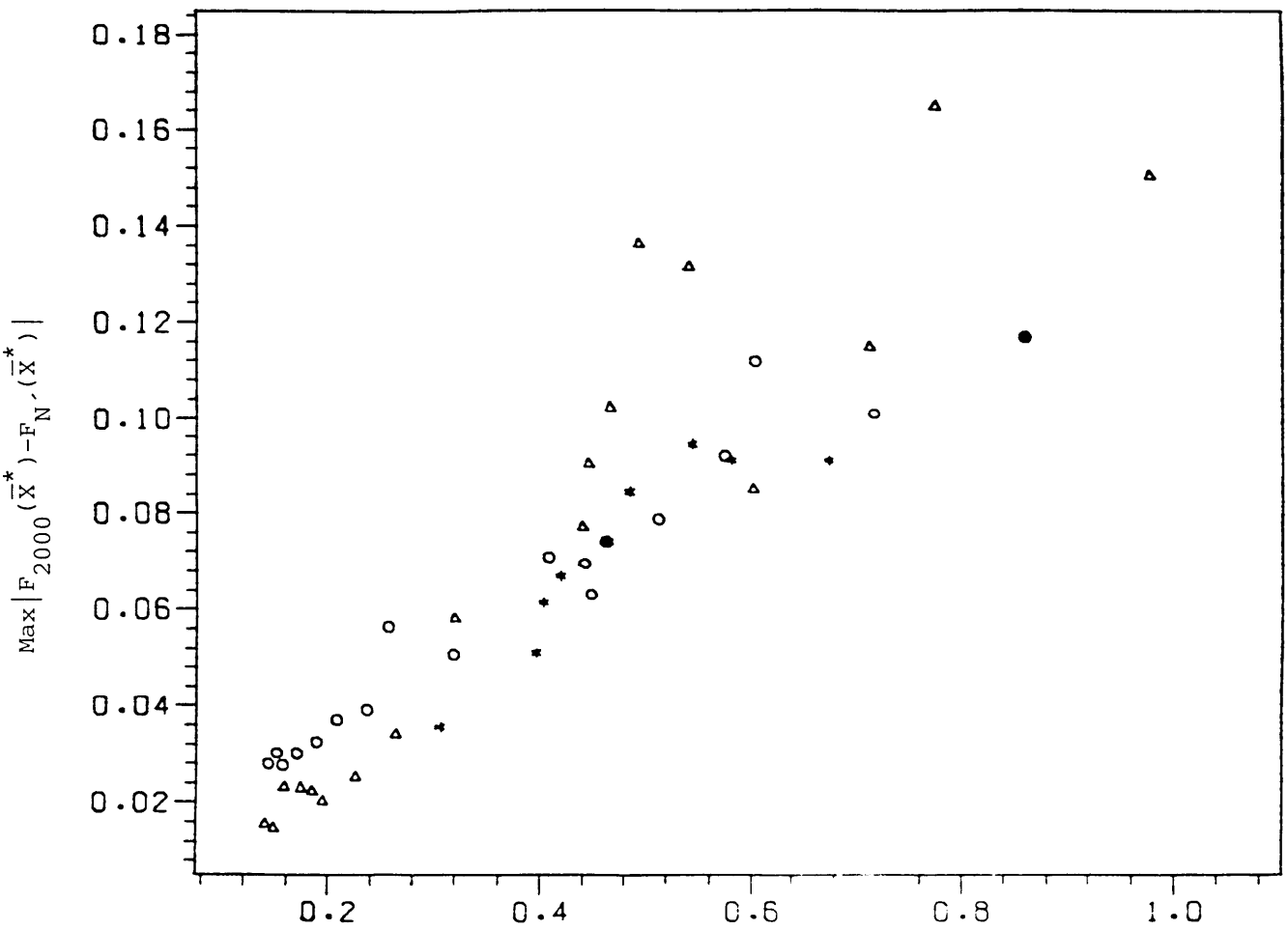
□□□□□□ Randomized, $\delta=0$; OO OOOO Balanced Array, $\delta=0$; ---- SQRT of $E(T^2(S'))$

Figure 3: $N'T^2(S')$ versus number of configurations in 10-dimensions. ($T(S')$) ≡ Discrepancy)



------ Randomized, δ=0; ..... Balanced Array, δ=0, ——— Balanced Array,δ≥20

Figure 4: Graph of the emperical bootstrap bias of cdf versus the corresponding discrepancy for the same number of configurations



Note: ΔΔΔΔΔΔ Randomized, δ=0; OOOOOO Balanced Array, δ=0;

****** Balanced Array with δ≥20

CHAPTER FIVE

GENERATING LIKELIHOOD

5.1 Introduction

In this chapter, we discuss a new application of the bootstrap procedure for the construction of a likelihood, $L(\vartheta)$, of a parameter ϑ. Essentially, we try to combine the resampling ideas of Chapter 2, especially the bootstrapping ideas to create likelihoods with which to do statistical inference. The likelihoods can be used to construct confidence intervals in situations where the underlying density function of the population from which the observed data set is taken is unknown. The method provides an alternative to the exact methods of inverting permutation tests such as the methods of Hartigan (1969), Tritchler (1984) and others. The ideas of our technique has connection with the work of Diggle & Gratton (1984), who used Monte Carlo methods to generate likelihood where the underlying probability density function $f(x,\vartheta)$ is unknown but where the random variable X can be generated using a well specified random mechanism. We shall use the observed data itself to generate the likelihood. The detailed descriptions of the steps of our method shall be given shortly and the theoretical justifications are given in §5.2. The smoothing step is discussed in §5.4. Our smoothing step is a mixture of parametric and nonparametric statistical procedures. In this step, a knowledge of the underlying distribution of the statistic T, is assumed. Examples are considered in §5.7. The example on regression demands much repeated bootstrapping for various values of "ϑ" to generate the normalized likelihood, $L(\vartheta)$. For such complex situations, the proposed ART should serve as a ready

tool. The work of Diggle & Gratton (1984) on curve smoothing in density estimation combined with the work of Tritchler (1984) on inverting permutation tests provides some theoretical justification for this section of the work. Diggle & Gratton's (1984) approach is to estimate the log-likelihood function from simulated realization for an implicit model. It is worth recalling the basic steps taken in their approach:-

(i) Generate data y_i, $i = 1,2,\ldots,n$ as an independent random sample from a distribution with probability density function, $f(y)$ - exponential in their case.

(ii) For each chosen ϑ, simulate an independent random sample x_k, $k = 1,2,\ldots,n$ from an appropriate distribution.

(iii) Use a kernel method to estimate each $f(y,\vartheta)$ from the simulated x_k and construct an estimated log-likelihood:

$$L^*(\vartheta) = \sum_{i=1}^{n} \ell n \hat{f}(y_i,\vartheta) \qquad (5.1.1)$$

We list below the main steps of **our** proposed data based methodology for the generation of likelihood.

(1) **The Data:** The data will usually consist of a realized sample $X_1,X_2,\ldots,X_n$. The X_i's , $i = 1,2,\ldots,n$ may be vectors but for the moment let us consider them to be real numbers. The underlying distribution, F, of the random variable X from which the random sample X_i, $i = 1,2,\ldots,n$ is taken is unknown.

(2) **The Statistic:** For any set of values $x_1,x_2,\ldots,x_n$, we consider a function (statistic) possibly vector-valued, $T(x_1,x_2,\ldots,x_n)$ $(= T(\underset{\sim}{x}))$. The choice of the statistic T will be problematical but crucial to the development of the entire technique.

(3) **The Transformation:** We shall assume that we have a family of transformations of the realized data set indexed by a parameter. Thus if,

$$\underset{\sim}{x} = (x_1,x_2,\ldots,x_n)^T \tag{5.1.2}$$

then we can write

$$\underset{\sim}{y} = (y_1,y_2,\ldots,y_n) = g_\vartheta(\underset{\sim}{x}) \tag{5.1.3}$$

In such a situation, $\underset{\sim}{y}$ would usually be an $n \times 1$ vector. An example could be

$$y_i = x_i - \vartheta \tag{5.1.4}$$

where ϑ is a real valued number. This could be the special case of shift parameter problem and it will usually be convenient if $\vartheta = 0$ that

$$g_\vartheta(\underset{\sim}{x}) = \underset{\sim}{x} \tag{5.1.5}$$

namely the identity transformation.

(4) **The Resampling Step:** This will be the point where a random component is introduced. The resampling procedure we adopt is the bootstrap. The ART, could also be adopted. By putting mass 1/n at each point y_i , i = 1,2,...,n do a bootstrap of the y_i , i = 1,2,...,n values

and for each bootstrap sample compute $T(y_1^*, y_2^*, \ldots, y_n^*)$ where * signifies bootstrap values.

(5) **The Smoothing Step:** This step involves the smoothing of a discrete set of values to produce a histogram or a smoothed histogram. By the kernel density estimation procedure, calculate an estimator $\hat{f}(t|\vartheta)$ from the T values in step (4).

(6) Evaluate $\hat{f}(t|\vartheta)$ at $t_o = T(y)$ to obtain $L(\vartheta) = \hat{f}(t_0|\vartheta)$. Thus, $\hat{f}(t_0|\vartheta)$ is our likelihood. From it we estimate the normalized likelihood version (by a numerical integration approach) for each value of ϑ and finally get the distribution of the normalized likelihood. More details showing the connection of Steps 5 and 6 in our proposed technique with the general kernel density estimation are given in §§ 5.4 and 5.6.

We explain the method in more detail.

Suppose that the data sample consists of the values $X_1, X_2, \ldots, X_n$. Let $g_\vartheta(\underset{\sim}{x})$ be a family of transformations where we suppose for simplicity that ϑ is a real number. We shall choose ϑ and generate

$$g_\vartheta(\underset{\sim}{X}) \equiv Y \equiv (Y_1, Y_2, \ldots, Y_n) \tag{5.1.6}$$

From the transformed random sample, $Y_1, Y_2, \ldots, Y_n$, we can construct the emperical cumulative distribution function, for the discrete distribution which attaches probability 1/n to each of $Y_1, Y_2, \ldots, Y_n$. Call this

$$\hat{F}_{\vartheta}(y) = \frac{1}{n} \quad \# \quad \left\{ Y_i \leqslant y \right\} \tag{5.1.7}$$

Since $x_1, x_2, \ldots, x_n$ are the observed (realized) data set and ϑ is chosen deliberately, $\hat{F}_{\vartheta}(y)$ is therefore known, that is $\hat{F}_{\vartheta}(y)$ can be computed.

Let $T(x_1, x_2, \ldots, x_n)$ be a univariate statistic. Let Y^* be a random variable with distribution $\hat{F}_{\vartheta}(y)$ and consider $T(Y^*)$.

Since $\hat{F}_{\vartheta}(y)$ depends on ϑ, so does the distribution of $T(Y^*)$ and we can then write

$$T(Y^*) \sim f_T(t, \vartheta) \tag{5.1.8}$$

Now f_T will invariably be discrete and we shall simply replace it by the smoothed version, namely $\hat{f}_T(t, \vartheta)$. At this stage any of a variety of smoothing methods might be used. The kernel density estimator is our choice of smoothing procedure and we use it throughout in this work. Notice that although $f_T(t, \vartheta)$ is well defined, it may be difficult to compute if n is large. So for such large n, we shall resort to the use of the bootstrap approximation obtained by repeated Monte Carlo sampling from $\hat{F}_{\vartheta}(y)$.

Now $\hat{f}_T(t, \vartheta)$ is our density from which we may obtain a likelihood in the following way:

Let $t_0 = T(y)$ be the value of t for the transformed observed data set $y_1, \ldots, y_n$.

The likelihood for ϑ is

$$L(\vartheta) = \hat{f}_T(t_0, \vartheta) \tag{5.1.9}$$

A simplified method of constructing a likelihood, $L(\vartheta)$ without smoothing

all of $f_T(t,\vartheta)$ is to merely do a count of the number N_ϵ of $T(Y^*)$ from the bootstrap (Monte Carlo) or complete enumeration lying in an interval $[t_0-\epsilon,t_0+\epsilon]$ around t_0 and put

$$L(\vartheta) = \frac{1}{B} \cdot N_\epsilon \tag{5.1.10}$$

where B is the number of Monte Carlo (bootstrap) samples. $L(\vartheta)$ is thus the proportion of $T(Y^*)$ that lie in $[t_0 \pm \epsilon]$. Note that

$$B = n^n \tag{5.1.11}$$

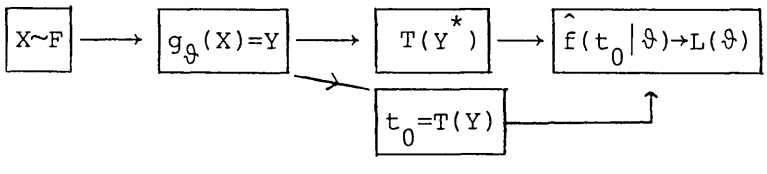
for a complete enumeration.

For this work, we shall simply call the above method, the bootstrap version. If the resampling from $\hat{F}_\vartheta(y)$ is replaced by other distribution for Y^* based on Y, we shall get other versions. For example, if y runs over all 2^n values $(\pm y_1, \pm y_2, \pm \ldots, \pm y_n)$ we can take the distribution for Y^* to be uniform in y_i with

$$\text{Prob}(Y^* = y_i) = \frac{1}{2^n} \tag{5.1.12}$$

This will be the case when the distribution of Y is assumed to be symmetric about the origin.

In general, the sketch diagram of our proposed method is as follows



5.2 The Two Sample Problem

We shall now consider the two sample problem as a test case for our discussion. Suppose we have two data sets (original data sets) of

$$X_1, X_2, \ldots, X_m \qquad \text{and} \qquad Y_1, Y_2, \ldots, Y_n$$

which when combined yields

$$Z = (X_1, X_2, \ldots, X_m; Y_1, Y_2, \ldots, Y_n) \quad .$$

Calculate the statistic of interest. For example, let us consider the statistic

$$t = \sum_{i=1}^{m} x_i/m - \sum_{j=1}^{n} y_j/n \tag{5.2.1}$$

t is simply the difference of means of the two samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$.

Choose a real-valued quantity ϑ which is assumed to be the true difference of means t. Generate the transformed data set

$$Z(\vartheta) = (X_1, X_2, \ldots, X_m, Y_1 - \vartheta, Y_2 - \vartheta, \ldots, Y_n - \vartheta)$$

which we can simply write as

$$Z = (Z_1, Z_2, Z_3, \ldots, Z_N) \qquad N = m + n \ .$$

Bootstrap the transformed data set $Z(\vartheta)$ and for each bootstrap Z_i^*, $i = 1,2,\ldots,B$ calculate

$$t^* = T(Z^*) = \sum_{i=1}^{m} Z_i^*/m - \sum_{j=m+1}^{N} Z_j^*/n \qquad (5.2.2)$$

The basic assumption here is that the two samples $X_1, X_2, \ldots, X_m$ and $Y_1-\vartheta$, $Y_2-\vartheta, \ldots, Y_n-\vartheta$ are from the same population or that infact they have the same distribution. This would imply that the distribution of the differences of the bootstrap values will be centred around zero if ϑ is the actual true difference. By a kernel density estimation method we get the estimator $\hat{f}(t|\vartheta)$ for the probability density function f (for the difference of mean of the two samples considered) for the chosen value of ϑ.

Evaluate $\hat{f}(t|\vartheta)$ at

$$t = t_0 = \sum \frac{x_i}{n} - \frac{\Sigma(y_j-\vartheta)}{n} \quad ,$$

namely the "true" transformed "unbootstrapped" value of t. Subsequently, the normalized likelihood, $L(\vartheta)$ is estimated by a numerical integration approach. The simplified method of getting the likelihood given in (5.1.10) could also be adopted to simply get the proportion that lies in the interval $(t_0 \pm \varepsilon)$ where t_0 is the value of t at $\vartheta = 0$. The above procedure is repeated for various values of ϑ. Finally, we do a plot of the distribution of the normalized likelihood.

5.3 Theoretical Background: The Likelihood

Let X be a random variable with distribution $P_X(x|\vartheta)$, where ϑ is any parameter. Then we can define the likelihood as,

$$\ell_X(\vartheta) = P_X(x_0|\vartheta) \qquad (5.3.1)$$

where x_0 is the observed value of the data.

Suppose we consider a ϑ dependent 1-1 transformation of x say,

$$t = g(\vartheta, x) \tag{5.3.2}$$

then, assuming the usual conditions hold,

$$P_X(x|\vartheta) = P_T(t|\vartheta) \cdot J(\vartheta, x) \tag{5.3.3}$$

where $J(\cdot)$ is the Jacobian of the transformation from $X \rightarrow T$

Then the "true" likelihood is the quantity $P_T(t_0|\vartheta) \cdot J(\vartheta, x_0)$, where $t_0 = g(\vartheta, x_0)$ and x_0 is the observed value of x. If $g_\vartheta(\underset{\sim}{X})$ is linear in ϑ (as for example $\underset{\sim}{Y} = \underset{\sim}{X} - \vartheta_j$) then the Jacobian, $J(\vartheta, x)$ becomes functionally independent of ϑ and therefore

$$L(\vartheta) \propto P_T(t_0|\vartheta) \tag{5.3.4}$$

When $g_\vartheta(\cdot)$ is not 1-1, then $P_T(t_0|\vartheta)$ has to be reinterpreted. The most typical situation is when $g_\vartheta(\underset{\sim}{X})$ is a many-one transformation.

Suppose for example that

$$\underset{\sim}{T} = g_\vartheta(\underset{\sim}{X}) \tag{5.3.5}$$

$$\underset{\sim}{U} = h_\vartheta(\underset{\sim}{X}) \tag{5.3.6}$$

are such that $\underset{\sim}{T}$ and $\underset{\sim}{U}$ are independent (for all values of ϑ), then,

$$P_X(x|\vartheta) = P_T(t|\vartheta) \cdot P_U(u|\vartheta) \cdot J(x, \vartheta) \tag{5.3.7}$$

where J is the Jacobian of the *full transformation*. Then whether or not $P_T(t|\vartheta)$ merely reflects the shape of the likelihood depends on the functional dependence of $P_U(u|\vartheta)$ and J on ϑ. It may be (for example) that $\underset{\sim}{U}$ does not even depend on ϑ at all. That is, $\underset{\sim}{U}$ is a pivotal quantity.

In situations where $\underset{\sim}{X}$ has an unknown distribution it may still be natural to assume that given ϑ, $T = g_\vartheta(\underset{\sim}{X})$ has a known distribution. Our procedure then is to take $\ell_T(\vartheta) = P_T(t_0|\vartheta)$ as the likelihood for inference purposes. For the resampling situation, we generate $P_T(t_0|\vartheta)$ by a resampling procedure. Thus in effect (together with smoothing) we use $\hat{P}_T(t_0|\vartheta)$. We simply perturb a complex system over some selected reference set and look at the behaviour of a selected numerical quantity. We need to investigate the system under different parameter configurations. Thus, the reference set is essentially parameter dependent. All our interest in the outcomes are summarised by $T = g_\vartheta(\underset{\sim}{X})$. The method can be considered as a sensitivity or perturbation method which needs no distributional assumptions but only;

(i) a parameter structure

(ii) a (parameter dependent) reference set.

Item (ii) is taken from the observed data itself. The data eventually serves two purposes:

(a) Construction of the reference sets as well as the evaluation of $\hat{P}_T(t_0|\vartheta)$.

(b) Production of an observed value t_0.

5.4 Connection with Inversion of Permutation Test

"The procedure of significance testing reaches its full utility only when the significance test can be inverted to obtain sets of values of the parameter under test which give the possible significance levels".

Kempthorne and Doerfler (1969,p.246)

Let $X_1,X_2,\ldots,X_m$; $Y_1,Y_2,\ldots,Y_n$ be independently distributed according to continuous distribution $F(x)$ and $G(y) = F(y-\vartheta)$ respectively.

Consider the combined set,

$$(Z_1,Z_2,\ldots,Z_{m+n}) = (X_1,X_2,\ldots,X_m;Y_1-\vartheta,Y_2-\vartheta,\ldots,Y_n-\vartheta)$$

and the $\binom{m+n}{m}$ permutations $i_1 < i_2 < \ldots < i_m$; $i_{m+1} < \ldots < i_{m+n}$ of the integers $1,2,3,\ldots,m+n$.

Without further assumptions concerning F, Lehmann (1959) obtained confidence intervals for ϑ from a permutation test of the hypothesis:

$$H(\vartheta_0):\vartheta = \vartheta_0 \qquad (5.4.1)$$

where the null hypothesis $H(\vartheta)$ is accepted for k of the permutations which lead to the smallest values,

$$\left| \sum_{j=m+1}^{m+n} z_{i_j}/n - \sum_{j=1}^{m} z_{i_j}/m \right| \qquad \text{and}$$

$$k = (1-\alpha)\binom{m+n}{m} \quad . \tag{5.4.2}$$

Hartigan (1969) used balanced sets of subsamples as typical values to form coverage probabilities for a constant, ϑ (a parameter). Hartigan (1969) defines a set of random variables $X_1, X_2, \ldots, X_N$ to be a set of values for ϑ, if each of the intervals between the ordered random variables $(-\infty, X_{(1)})$, $(X_{(1)}, X_{(2)}), \ldots, (X_{(N)}, \infty)$ includes ϑ with equal probability of 1/N+1. Essentially, Hartigan (1969) showed that for a real valued statistic $t(X_1, X_2, \ldots, X_N)$, the recomputed values of $t_n(\cdot)$ for the subsamples of a balanced set provide appropriate confidence intervals for the true value of t where the true value of t is the value of t given as

$$t \equiv \lim_{n\to\infty} t(Y_1, Y_2, \ldots, Y_n) \quad .$$

Tritchler (1984) presents polynomial time algorithms for the inversion of permutation test for the one- and two-sample problems. In line with Tritchler's (1984) approach for the permutation confidence interval for shift in the two sample problem, let us consider two realized samples $x_1, x_2, \ldots x_m$ and $y_1, y_2, \ldots, y_n$ which when combined may be written as $Z_1, Z_2, \ldots, Z_N$ (N = m+n) where,

$$\begin{aligned} Z_i &= x_i \qquad (i = 1,2,\ldots,m) \\ &= y_j \qquad (j = m+1, m+2, \ldots, N \end{aligned} \tag{5.4.3}$$

Define the statistic (for the bootstrap case)

$$S = \sum_{j=1}^{N} Z_j I_j \tag{5.4.4}$$

where I_j is a function representing the number of times Z_j appears in the bootstrap.

In effect,

$$S \equiv \sum_{k=1}^{m}{}^{*} x_k I_k + \sum_{t=m+1}^{N}{}^{*} y_t I_t \tag{5.4.5}$$

such that

$$\sum{}^{*} I_k + \sum{}^{*} I_t = N \quad .$$

By keeping the Z_j's fixed we let $I = (I_1, I_2, \ldots, I_N)$ vary over each of the B equally likely bootstrap samples of fixed size N.

Consider the probability distributions

$$P(X_i \leq x) = F(x) \qquad i = 1,2,\ldots,m \tag{5.4.6}$$

and

$$P(Y_j \leq y) = F(y-\vartheta) \qquad j = 1,2,\ldots,n \tag{5.4.7}$$

where ϑ is the shift parameter.

We shall be interested in testing the null hypothesis

$$H_0: \vartheta = 0 \tag{5.4.8}$$

that is, that X is distributed as Y.

This could be generalized by considering the hypothesis,

$$H_0: \vartheta = \vartheta_0 \tag{5.4.9}$$

and using the significance probability of the observed values,

$$s(\vartheta_0) = \sum_{i=1}^{N} Z_i \tag{5.4.10}$$

of the random variable

$$S(\vartheta_0) = \sum_{j=1}^{N} Z_j(\vartheta) I_j \tag{5.4.11}$$

where,

$$\begin{aligned} Z_j(\vartheta) &= Z_j = x_j , && j = 1,2,\ldots,m \\ &= Z_j - \vartheta = y_j - \vartheta && j = m+1,m+2,\ldots,N \quad . \end{aligned} \tag{5.4.12}$$

We shall now derive an expression for the resampling likelihood.

For a given bootstrap sample, $Z_1^*, Z_2^*, \ldots, Z_N^*$, let us consider,

$$B \cdot \ell_*(\vartheta) = \sum \delta(S^*(\vartheta) - s(\vartheta)) \tag{5.4.13}$$

Now, let

$$t^* = S^*(\vartheta) = \frac{\Sigma_1 Z_i^* I_i}{m} - \frac{\Sigma_2 Z_i^* I_i}{n} \tag{5.4.14}$$

and let,

$$s(\vartheta) = t_0 = \text{"true value"} = \frac{\sum_{i=1}^{m} x_i}{m} - \frac{\sum_{j=m+1}^{N} (y_j - \vartheta)}{n} \tag{5.4.15}$$

We shall examine what happens when $t^* = t_0$, where,

$$\Sigma_1 Z_i^* I_i = \Sigma a_i x_i + \Sigma b_i (y_i - \vartheta) \tag{5.4.16}$$

$$\Sigma_2 Z_i^* I_i = \Sigma c_i x_i + \Sigma d_i (y_i - \vartheta) \tag{5.4.17}$$

Note that,

$$\Sigma a_i + \Sigma b_i = m \tag{5.4.18}$$

$$\Sigma c_i + \Sigma d_i = n \tag{5.4.19}$$

Let us write,

$$\Sigma b_i = m^* \tag{5.4.20}$$

and $$\Sigma d_i = n^* \tag{5.4.21}$$

Then, $t^* = t_0$

$$\Leftrightarrow \frac{\Sigma a_i x_i + \Sigma b_i y_i}{m} - \frac{\Sigma c_i x_i + \Sigma d_i y_i}{n}$$

$$- \left(\frac{m^*}{m} - \frac{n^*}{n} \right) \vartheta = \frac{\Sigma x_i}{m} - \frac{\Sigma y_i}{n} + \vartheta \tag{5.4.22}$$

$$\Leftrightarrow \vartheta = \left(1 + \frac{m^*}{m} - \frac{n^*}{n}\right)^{-1} (S^* - s_0) \stackrel{\text{def}}{=\!=} \tilde{\vartheta} \tag{5.4.23}$$

where S^* and s_0 are the values of $S^*(\vartheta)$ and $s(\vartheta)$ when $\vartheta = 0$.

$$\therefore \quad \vartheta = q^*(S^* - s_0) \tag{5.4.24}$$

where $$q^* = \left(1 + \frac{m^*}{m} - \frac{n^*}{n}\right)^{-1} \tag{5.4.25}$$

Hence the rerandomization likelihood is proportional to

$$\sum \delta(\vartheta - q^*(S_0^* - s_0))$$

and we can therefore state the following:

THEOREM 4:

Let the rerandomization likelihood of S be

$$\ell_*(\vartheta) = \text{Prob}_*(S^*(\vartheta)=s(\vartheta)\,|\,\vartheta) \quad . \tag{5.4.26}$$

Then

$$B.\ell_*(\vartheta) = \sum \delta(\tilde{\vartheta}-\vartheta) \tag{5.4.27}$$

where the summation is over all the B bootstrap configurations, $\delta(x) = 1$ for $x = 0$ and $\delta(x) = 0$ otherwise.

This has an important connection with confidence levels for inversion of permutation test. Suppose the hypothesis is about ϑ. The α-level test using $S^*(\vartheta) > s(\vartheta)$ is

$$\text{Prob}(S^*(\vartheta) > s(\vartheta)\,|\,\vartheta) \leqslant \alpha \quad . \tag{5.4.28}$$

Now, the set of "ϑ"-values for which we will reject the data set at level α is

$$U(\vartheta\,|\,\text{Prob}(S(\vartheta) > s(\vartheta)\,|\,\vartheta) \leqslant \alpha$$

i.e. $$U(\vartheta\,|\,\text{Prob}(q^*(S_0^*-s_0) \geqslant \vartheta|\vartheta) \leqslant \alpha \ .$$

Thus we take the tail of all the points

$$\geqslant q^*(S_0^*-s_0) \qquad \text{(which themselves do not depend on } \vartheta).$$

We note the subtle difference from our simplified (likelihood) method in (5.1.10). We simply consider the proportion of times that

$$(S^*(\vartheta) \in [S(\vartheta)-\varepsilon, S(\vartheta)+\varepsilon]\,|\,\vartheta) \quad . \tag{5.4.29}$$

We can compare this rerandomization likelihood with the application of the simplified method. (5.4.29) is just an application of the method for the two sample problem in which case

$t^* \equiv S^*(\vartheta)$ and $t_0 = S(\vartheta)$ giving the proportion of times

$$(t^* \in [t_0-\varepsilon, t_0+\varepsilon] \mid \vartheta).$$

But the statement

$$S^*(\vartheta) \in [S(\vartheta)-\varepsilon, S(\vartheta)+\varepsilon]$$

$$\Leftrightarrow q^*\vartheta - \varepsilon \leq S_0 - s_0 \leq q^*\vartheta + \varepsilon$$

$$\Leftrightarrow \vartheta - \frac{\varepsilon}{q^*} \leq \frac{S_0-s_0}{q^*} \leq \vartheta + \frac{\varepsilon}{q^*} \quad .$$

Thus we need the proportion of times

$$\frac{S_0-s_0}{q^*} \in \left[\vartheta - \frac{\varepsilon}{q^*} \quad \vartheta + \frac{\varepsilon}{q^*} \right] \quad .$$

The interval which occurs here which is simply a random interval due to the values of q^*, whereas the permutation case would count the number of $\frac{S_0-s_0}{q^*}$ values in a "fixed" interval.

We believe that this subtle difference between the rerandomization likelihood approach and our semi-parametric likelihood approach is critical. Inversion of permutation tests is essentially a hard problem and does not lead to sensible intervals in many cases. We believe that our approach allows an understanding of how to simulate complex parametric problem where there are few distributional assumptions.

5.5 The Smoothing Step: Kernel Density Estimation

The kernel density estimator introduced by Rosenblatt (1956) is one of the nonparametric density estimators of a probability density function $f(\cdot)$. Others are (i) the orthogonal series estimators (Kronmal and Tarter, 1968), (ii) the penalized maximum likelihood estimators (Good & Gaskins, 1971), and (iii) the k-nearest neighbour estimators (Loftsgaarden & Queensberry, 1965). The monograph of Tapia & Thompson (1978) and a book by Wertz (1978) give excellent discussions and reviews on the above topics. Quite recently, Titterington (1985) draws together a variety of smoothing techniques from several areas of statistics and gives a listing of references covering most of the statistical smoothing techniques. The kernel density estimator has in particular attracted a considerable attention to date. For this work, we require it in the context of smoothing. Our method of approach for the univariate case is close in form to the works of Rudemo (1982) and Sheather (1983) and for the multivariate case (bivariate) the works of Silverman (1978b) and Wertz (1978) form a close guide as regards the choice of the window width.

Suppose that the observed data set $X_1, X_2, \ldots, X_n$ are i.i.d. real-valued random variables with probability density function f (unknown). Our interest is to find estimators $\hat{f}$ of f.

Let $I = (I_k)$ be the kth partition of the real line subdivided into disjoint intervals.

Let h_k denote the length of I_k and let

$$N_k = \#\{i : X_i \in I_k,\ 1 \leq i \leq n\} \tag{5.5.1}$$

be the number of observations in I_k.

Let us put $X = (X_1, X_2, \ldots, X_n)$ such that,

$$\hat{f}(x) = \hat{f}_I(x,X) = N_k/nh_k \ , \quad x \in I_k \ . \tag{5.5.2}$$

Then $\hat{f} = \hat{f}_I$ is the histogram corresponding to I and X.

Suppose all the intervals I_k, k = 1,2,..., have equal length h.

Definition:

The non-negative real valued function K is a kernel if

$$\int K(x)dx = 1 \tag{5.5.3}$$

Suppose h > 0, then the kernel density estimator (Rosenblatt, 1956) for the unknown probability density function, f, is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{5.5.4}$$

Relating the above definition to the problem of likelihood, we have that the kernel density estimator of $\tilde{f}(\tilde{x})$ is

$$\hat{f}(\tilde{x}|\vartheta) = \frac{1}{Bh} \sum_{i=1}^{B} K\left\{\left(\frac{\tilde{x}-Y_i^*}{h}\right)\right\} \tag{5.5.5}$$

where, $K\left(\frac{\tilde{x}-Y_i^*}{h}\right)$ is the kernel function, Y_i^* are the bootstrap statistics obtained using the transformed data set $(Y_1, Y_2, \ldots, Y_n)$, $\tilde{x}$ is the computed statistic based on the original data set $(X_1, X_2, \ldots, X_n)$, B is the number of bootstrap configurations, h is the window width which determines the degree of smoothness and ϑ is the quantity chosen for the transformation of the original data set.

The implementation of (5.4.4) and (5.4.5) depends on the choice of the kernel function $K(\cdot)$ and the smoothing parameter h. The choice of the kernel function $K(\cdot)$ does not pose a problem. Sacks & Ylvisaker (1981) have shown that when x_0 is an interior point of the support of f, then the asymptotically optimal kernel does not usually give substantial improvement over many standard kernels. This statement is supported by many authors as well; viz.: Silverman (1978a), Scott & Factor (1981), Diggle & Gratton (1984) and Epanechnikov (1969) who showed that any kernel gives optimal results. The behaviour therefore, of the density estimators in (5.5.4) and (5.5.5) rests on the choice of optimal smoothing parameter h. Several data based algorithms have been offered for the choice of h. Tapia and Thompson (1978) gave an interative procedure for estimating an asymptotically optimal h. The h value h_i after i iterations is used to estimate the density f(x) and $\int(f''(x))^2dx$ from which h_{i+1} is obtained by

$$h_{i+1} = A(K)\cdot B(f_i)\cdot n^{-1/5} \tag{5.5.6}$$

where,

$$A(K) = \left[\frac{\int K^2(y)dy}{2(\int y^2K(y)dy)^2} \right]^{1/5} \tag{5.5.7}$$

and

$$B(\hat{f}_i) = \left[\int |f''(y)|^2 dy \right]^{-1/5} \tag{5.5.8}$$

A data based algorithm which chooses the window width h for a

kernel estimator of the density at a point is described in Sheather (1983). Usually, the general procedure is to obtain an optimal window width h by minimizing the Mean Integrated Square Error (MISE).

For this case, let

$$a = \int (K(x))^2 dx \tag{5.5.9}$$

and

$$b = \int x^2 K(x) dx \tag{5.5.10}$$

For the MISE risk function, an explicit formule for the asymptotically optimal h is given (Rudemo, 1982) as

$$h_{opt} = (a/b^2)^{1/5} [\int (f''(x))^2 dx]^{-1/5} \cdot n^{-1/5} \tag{5.5.11}$$

provided $f''(\cdot)$ is bounded, continuous and square integrable and that $K(\cdot)$ is symmetric and bounded with a finite second moment.

A similar result of

$$h_{opt}(x) = \alpha(K) \cdot \beta(f(x), f''(x)) \cdot n^{-1/5} \tag{5.5.12}$$

was obtained by Sheather (1983) where

$$\alpha(K) = \left[\int_{-\infty}^{\infty} K^2(y) dy \right]^{1/5} \cdot \left[\int_{-\infty}^{\infty} y^2 K(y) dy \right] \tag{5.5.13}$$

and

$$\beta(f(y), f''(y)) = (f(y))^{1/5} \cdot (f''(y))^{-2/5} \tag{5.5.14}$$

The determination of $\beta(f(x),f''(x))$ in (5.5.13) depends on the knowledge of $f(x)$ and $f''(x)$.

In practice, $f(x)$ and $f''(x)$ would be replaced by their respective estimators $f(x)$ and $\hat{f}''(x)$ and so (5.5.12) would become

$$\hat{h}_{opt}(x) = \alpha(K)\cdot\beta(\hat{f}(x),\hat{f}''(x))\cdot n^{-1/5} \qquad (5.5.15)$$

where $\alpha(\cdot)$ and $\beta(\cdot,\cdot)$ retain their usual definitions in (5.5.13) and (5.5.14) respectively.

For the histogram estimator, Scott (1979) obtained an explicit formula

$$h_s = \left[6/ \int (f'(x))^2 dx \right]^{1/3} \cdot n^{-1/3} \qquad (5.5.16)$$

for the MISE risk function. Again to realize the h_s value given in (5.5.16) one has to assume the knowledge of $f(\cdot)$.

For h known, the estimation of our $\hat{f}(x)$ (or $\hat{f}(x|\tilde{\vartheta})$) given in (5.5.4) (or (5.5.5)) becomes feasible. The problem therefore is finding a suitable choice of values for h to be used in constructing estimates for the unknown probability density f. A general result for kernel density estimators of probability density functions is that as n (or B) increases h should decrease essentially as $n^{-1/5}$ (or $B^{-1/5}$) - see Rudemo (1982), Rosenblatt (1971) and Silverman (1978b). For the histogram estimator h decreases as $n^{-1/3}$(or $B^{-1/3}$) - see Scott (1979).

For our work here, which is bootstrap procedure based, we require a smoothing parameter h that could be self up-dating whenever the size of the bootstrap, B, changes. Suppose, we choose the kernel fucntion K(x) in (5.5.4) to be the Gaussian distribution.

That is, we take,

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{5.5.17}$$

From (5.5.17), we get

$$\int (K(x))^2 dx = \int \frac{1}{2\pi} e^{-x^2} dx \tag{5.5.18}$$

which simplifies (5.5.9) to

$$a = \int (K(x))^2 dx = \frac{1}{2\sqrt{\pi}} \tag{5.5.19}$$

and (5.5.10) simplifies to

$$b = \int x^2 K(x) dx = 1 \tag{5.5.20}$$

Let us now consider the normal probability density function, f(x) given as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{5.5.21}$$

That is, we assume that the unknown probability density function, f(x) is normal with mean, μ and variance σ^2.

From (5.5.21), we get

$$f'(x) = -\frac{(x-\mu)}{\sigma^3\sqrt{\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{5.5.22}$$

and

$$f''(x) = \frac{(x-\mu)^2}{\sigma^5\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} - \frac{1}{\sigma^3\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad (5.5.23)$$

from which we get,

$$\int (f''(x))^2 dx = \frac{3}{8\sqrt{\pi}\ \sigma^5} \qquad (5.5.24)$$

Similarly, by making use of (5.5.22), we get

$$\int (f'(x))^2 dx = \frac{1}{4\sqrt{\pi}\ \sigma^3} \qquad (5.5.25)$$

Substituting (5.5.19), (5.5.20) and (5.5.24) in (5.5.11), we get

$$h_{opt} = (a/b^2)^{1/5}[\int (f''(x))^2]^{-1/5} \cdot n^{-1/5}\ (4/3)^{1/5} n^{-1/5}\sigma \qquad (5.5.26)$$

which is formula (5.10) of Rudemo (1982). Simplifications of formula (4.2) of Sheather & Maritz (1983), formula (2.10) of Sheather (1983) and formula (2.4) of Scott & Factor (1981) yield the same result as our (5.5.26).

Substituting (5.5.25) in (5.5.16), we get

$$h_s = [6/\int (f'(x))^2 dx]^{1/3}\ n^{-1/3} = (24\sqrt{\pi})^{1/3}\ n^{-1/3}\sigma \qquad (5.5.27)$$

(5.5.27) is simply formula (5.9) of Rudemo (1982) which is also given in another form as

$$h_s^* = 2\times 3^{1/3}\pi^{1/6} \cdot \sigma \cdot n^{1/3} \qquad (5.5.28)$$

by Scott (1979).

For the bootstrap case, n in (5.5.26) and (5.5.27) is replaced by B.

A slight draw-back of the above objective theoretical approach for estimating the smoothing parameter h is that the estimates derived depend on the knowledge of the unknown probability density fucntion f(x). A suggested approach is to replace f(x) by its nonparametric iterative estimate $\hat{f}(\cdot)$.

The multivariate form of density estimator is

$$f_n(X_1,X_2,\ldots,X_p) = \frac{1}{nh_1,h_2,\ldots h_p} \sum_{i=1}^{n} K_n\left[\left(\frac{x_1-X_{1i}}{h_1} \right), \left(\frac{x_2-X_{2i}}{h_2} \right), \right.$$

$$\left. \ldots, \left(\frac{x_p-X_{pi}}{h_p} \right) \right] \qquad (5.5.29)$$

where $K(.,.,\ldots,.)$ is the multivariate kernel function and h_i, $i = 1,2,\ldots,p$ are the respective window widths.

Hence, the bivariate kernel density estimator for a bivariate probability density function f(x,y) is

$$\hat{f}(x,y) = [nh_x h_y]^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_x}, \frac{x-X_i}{h_y} \right) \qquad (5.5.30)$$

and in particular, for the bootstrap case, we have the bivariate kernel density estimator to be

$$\hat{f}(\tilde{x},\tilde{y}) = \frac{1}{Bh_x h_y} \sum_{i=1}^{B} K\left(\frac{\tilde{x}-X_i^*}{h_x}, \frac{\tilde{y}-Y_i^*}{h_y} \right) \qquad (5.5.31)$$

where X_i^*, Y_i^* and B are bootstrap characteristics.

Suppose we take the bivariate kernel function $K(\cdot,\cdot)$ in (5.5.30) to be the bivariate Gaussian distribution, then we shall have for X and Y independent that

$$K(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \tag{5.5.32}$$

Hence in our bootstrap case we shall have (5.5.31) to be

$$\hat{\tilde{f}}(x,y) = \frac{1}{Bh_x h_y} \sum_{i=1}^{B} \frac{1}{2\pi} e^{-\frac{1}{2}[(\frac{x-X_i^*}{h_x})^2 + (\frac{y-Y_i^*}{h_y})]} \tag{5.5.33}$$

where we take h_x and h_y to be the respective window widths for the marginal kernel density estimators $\hat{\tilde{f}}(x)$ and $\hat{\tilde{f}}(y)$.

Thus, for $X \sim N(\mu_x,\sigma_x^2)$ and $Y \sim N(\mu_y,\sigma_y^2)$ we have

$$h_x \equiv (4/3)^{1/5} B^{-1/5} \sigma_x \tag{5.5.34}$$

$$h_y \equiv (4/3)^{1/5} B^{-1/5} \sigma_y^2 \tag{5.5.35}$$

In certain cases, the kernel density estimators for both the univariate and multivariate cases discussed above can change quickly under very small variations in the smoothing parameters h or h_i, $i = 1,2,\ldots,p$. A subjective method of using test graphs has been developed by Silverman (1978b). Silverman (1978b) considers the problem of choosing the smoothing parameter that is appropriate for any given sample. The approach involves the consideration of test graphs of the second derivative f" of the density estimate for various values of h_n. Then choose the window width which gives fluctuations of the right size in the test graph and use this chosen window width to construct the estimate of the original density. The subjectivity of the method is in the sense of judgement of the right size of fluctuations observed in the test graphs. Silverman (1978b)

obtained the theoretical justification of the above subjective method for both the univariate and multivariate cases. Gratton (1979) generalized Silverman's (1978b) subjective test graph method for choosing the optimal window widths when estimating an unknown probability density function by a kernel method.

5.6 Confidence Intervals

Now that we are in possession of the likelihood $L(\vartheta)$, all the usual statistical techniques may be applied. For example, if $\pi(\vartheta)$ is a prior distribution for ϑ, then we can obtain a posterior in the usual Bayesian approach of

$$\pi(\vartheta|t) \propto L(\vartheta)\ \pi(\vartheta) \qquad (5.6.1)$$

If it is thought natural to take $\pi(\vartheta)$ locally uniform, then the normalized likelihood itself maybe used as a posterior distribution and appropriate Bayes confidence intervals could be constructed.

Following our discussion of the two sample problem in §5.2 above, we could set an appropriate level α confidence interval for the two sample problem by simply calculating the $100(1-\alpha)\%$ confidence interval for the "true" difference of mean t. However for the purposes of comparison, we recall that the $100(1-\alpha)\%$ confidence interval for the difference of two means of samples taken from the normal distribution is given as

$$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2}(\alpha/2) \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (5.6.2)$$

where $$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \qquad (5.6.3)$$

s_1^2 and s_2^2 are the respective sample variances and n_1 and n_2 are the respective sample sizes of the samples taken from the random variables X_1 and X_2. $t_{n_1+n_2-2}(\alpha/2)$ is the $\alpha/2$-level variate of the t- distribution with n_1+n_2-2 degrees of freedom.

Let us also consider the t-distribution confidence curve. We recall that the t-distribution probability density function is given as

$$f_T(t) = \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma\frac{\nu+1}{2}}{\Gamma\,\nu/2} \left(1+\frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})} \qquad -\infty < t < \infty \qquad (5.6.4)$$

where ν is the number of degrees of freedom.

Let

$$Y = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{n_1+n_2-2}(\alpha/2) \qquad (5.6.5)$$

and put $$c = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (5.6.6)$$

and $$t = t_{n_1+n_2-2}(\alpha/2) \qquad (5.6.7)$$

such that

$$Y = ct \qquad (5.6.8)$$

After the usual transformation, we get

$$f_Y(y) = \frac{1}{c} \cdot \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma\ \frac{\nu+1}{2}}{\Gamma\ \nu/2} \left(1+ \frac{y^2}{\nu c^2}\right)^{-\left(\frac{\nu+1}{2}\right)} \tag{5.6.9}$$

Furthermore, let

$$K = \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma\ \frac{\nu+1}{2}}{\Gamma\ \nu/2} \tag{5.6.10}$$

substituting in (5.5.14) we get

$$f_Y(y) = \frac{1}{c} \cdot K \cdot \left(1+ \frac{y^2}{\nu c^2}\right)^{-\left(\frac{\nu+1}{2}\right)} \tag{5.6.11}$$

Hence for a practical problem, we do the plot of $\hat{f}_Y(y)$ versus ϑ and calculate the 100(1-α)% confidence interval for $y(=\tilde{x})$.

Efron (1982) gives a percentile method for finding confidence intervals.

Suppose,

$$\hat{CDF}(t) = \text{Prob}_*\{\hat{\vartheta}^* \leqslant t\} \tag{5.6.12}$$

is the cdf of the bootstrap distribution of $\hat{\vartheta}^*$.

$\hat{CDF}(t)$ as given above could be approximated by

$$\hat{CDF}(t) = \#\ \{\hat{\vartheta}^{*i} \leqslant t\}/B \tag{5.6.13}$$

For a level of significance α, let us define

$$\hat{\vartheta}_L(\alpha) = \hat{CDF}^{-1}(\alpha/2) \tag{5.6.14}$$

and

$$\hat{\vartheta}_U(\alpha) = \hat{CDF}^{-1}(1-\alpha/2) \qquad (5.6.15)$$

Then $[\hat{\vartheta}_L(\alpha), \hat{\vartheta}_U(\alpha)]$ is the $1-\alpha$ central confidence interval for the parameter ϑ.

Examples of these various forms of confidence intervals discussed above are given in the next section.

5.7 Examples

E1 An Example of the Two Sample Problem

We shall now consider a practical problem for our discussion on the two sample problem given previously in §5.2. For this work, we consider X and Y given as

X ≡ 37.50, 34.80, 38.90, 38.60, 37.00, 37.40, 36.50, 38.40, 38.00, 30.70

Y ≡ 37.70, 36.30, 38.00, 37.00, 37.60, 33.20, 36.70, 27.40, 37.10, 37.40

which when combined, could be written as

Z ≡ 37.50, 34.80, 38.90, 38.60, 37.00, 37.40, 36.50, 38.40, 38.00, 30.70, 37.70, 36.30, 38.00, 37.00, 37.60, 33.20, 36.70, 27.40, 37.10, 37.40.

X and Y are samples from a laboratory experiment concerning the life times in hours of certain batteries at EPV levels 1.1 and 1.0 respectively. Our interest is to set a $100(1-\alpha)\%$ confidence interval for the true difference of mean life times of the two grades of Duracell

batteries.

The difference of means t_0 of the original data set is calculated to be -0.94. Following the procedure described in §§5.1 and 5.2, various assumed ϑ were chosen for the "true" difference of means. Subsequently, with each chosen ϑ, the original data set Z was transformed and the transformed data set $Z(\vartheta)$ was bootstrapped and eventually the normalized likelihood $L(\vartheta)$ was calculated using the kernel density estimation. Both the pure Monte Carlo (standard bootstrap, $\delta = 0$) and the balanced array δ-method as well as the t-distribution confidence fit were adopted. In the pure Monte Carlo and the balanced array δ-method adopted, the standard normal kernel fucntion was used and the unknown distribution of the difference of mean was also assumed to be normal. Also Efron's (1982) percentile method for confidence interval described in §5.6 was adopted. In all the procedures adopted, the 95% confidence intervals for the "true" difference of mean. The results got are presented in Table 8 below.

Table 8

Summary of the results for the 95% confidence interval for the "true" difference of mean

Procedure	Interval
The standard t-test	[-3.6486, 1.7686]
Efron (1982) percentile method	[-3.2908, 1.4675]
Proposed method (i) Pure Monte Carlo, N = 1000	[-3.9437, 2.1918]
(ii) Balanced Array δ = 600, N = 200	[-3.7358, 1.8221]

Figure 5 gives the plot of the distribution of the normalized likelihood. In the balanced array approach, we used only $N' = 200$ configurations obtained for $\delta = 600$, while for the pure Monte Carlo (standard bootstrap) and for Efron's method $N = 1000$ configurations were used.

The balanced array method seems to give excellent results especially at the tails where the values are close to the t-distribution confidence curve. It seems quite satisfactory that after two major steps bootstrapping and smoothing this should be the case. The pure (unbalanced) Monte Carlo method gives slightly longer intervals. Using this as a benchmark there is not much difference with the balanced array. Efron's method gives a considerably shorter interval a result which seems to be typical. The Efron method, of course, is based on a *single* bootstrap. The difference can be explained by the observation that our method carries the information that if ϑ is *indeed* the true value of the parameter the *spread* of the transformed data set may depend on ϑ.

E2 Simple Linear Regression

We illustrate the approach adopted in more general regression problems through simple linear regression.

Let the observed data set be

$$x = (x_1, \ldots, x_n)$$
$$y = (y_1, \ldots, y_n) \quad .$$

We consider an underlying regression model of the form

$$y_i = \alpha + \beta x_i \quad .$$

Proceeding as for other examples choose quantities α and β and generate the transformed data set consisting of

$$x \text{ and } y(\alpha,\beta) = (y_1-(\alpha+\beta x_i),\ldots,y_n-(\alpha+\beta x_n))$$

write $y_i(\alpha,\beta) = y_i - (\alpha+\beta x_i) \qquad (i = 1,\ldots,n)$

We consider the extended reference consisting of all vectors

$$(\pm y_1(\alpha,\beta),\ldots,\pm y_n(\alpha,\beta)) \quad .$$

If we sample from this set according to the simple bootstrap we will generate B vectors Y^* for each such vector we perform a simple least squares regression on the observations X. This leads to a pair (one for each bootstrap) $(\hat{\alpha}^*,\hat{\beta}^*)$. Writing

$$t^* = (\hat{\alpha}^*,\hat{\beta}^*)$$

we have B values of t^*. From these values form a bivariate kernel density estimator $\hat{f}(t|\alpha,\beta)$. The likelihood is then

$$L(\alpha,\beta) = \hat{f}(t_0|\alpha,\beta)$$

where $t_0 = (\hat{\alpha}_0,\hat{\beta}_0)$ the values obtained by regressing the $y_i(\alpha,\beta)$ against the x_i values.

We considered as an example

$$X = (0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0,\ 1.2,\ 1.4,\ 1.6,\ 1.8,\ 2.0,$$
$$2.2,\ 2.4,\ 2.6,\ 2.8,\ 3.0,\ 3.2,\ 3.4,\ 3.6,\ 3.8,\ 4.0).$$

Y = (1.8045, 2.9393, 3.3701, 0.9602, 3.2751, 0.3534,
-0.1554, 1.3517, 1.3065, -1.1758, 1.0900, 3.3202, 5.1261, 1.9094,
0.1159, 0.7619, 3.0938, -0.2918, 1.6116, 1.4945).

Table 9 contains the estimated *normalised* likelihood $L(\alpha,\beta)$. The contours and relief plots are given in Figure 6. The usual standardized versions of $\hat{\alpha}$ and $\hat{\beta}$ were used which is why the plots are roughly circular. Note that $(\hat{\alpha}_0,\hat{\beta}_0) = (1.9094, -0.1647)$.

The pure Monte Carlo (standardized bootstrap, $\delta = 0$ on the extended configurations) was used and the randomized δ-method. In both cases the unbalanced cases are presented.

The results in Figure 6 which here are close to the standard normal theory confidence ellipse were obtained using the present technique, with *no distributional assumptions*, but very many simulations. We hope that the methods point the way to more widespread use of simulation techniques with parameter dependent reference sets.

E3 Other Problems

E3.1 The Sign Test with Bootstrap

Let $\underset{\sim}{x} = (x_1,x_2,\ldots,x_n)$ be a realized data set, and let the statistics of interest be

$$t = \sum_{i=1}^{n} x_i \quad .$$

Generate the vector

$$Y = g_{\vartheta}(x) = (x_1-\vartheta,x_2-\vartheta,\ldots,x_n-\vartheta)$$

Let Y^* have the following distribution namely,

$$y = (\pm y_1, \ldots, \pm y_n)$$

and

$$\text{Prob}(Y^*=y) = \frac{1}{2^n} \quad ,$$

then

$$T(Y^*) = \sum_{j=1}^{n} Y_j \quad .$$

Hence by analogy, we define the likelihood as

$$L(\vartheta) = \frac{1}{2^n} \quad \#\{T(y_i) \in [t_0-\varepsilon, t_0+\varepsilon]\}$$

where $t_0 = \sum_{i=1}^{n} (x_i-\vartheta) = \sum_{i=1}^{n} x_i - n\vartheta.$

We can repeat the analysis described above by sampling Y^* independently from all the y-values.

E3.2 The k-Sample Problem

We shall list below the procedures to be followed for the k-sample problem.

(1) Consider the data set

$$X \equiv \begin{array}{cccccc} X_1 & X_2 & X_3 & \cdots & X_k \\ x_{11} & x_{21} & x_{31} & & x_{k1} \\ x_{12} & x_{22} & x_{32} & & x_{k2} \\ x_{13} & x_{23} & x_{33} & & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{1n_1} & x_{2n_2} & x_{3n_3} & & x_{kn_k} \end{array}$$

(2) Calculate the mean of each sample and subsequently calculate the differences of means for the $\binom{k}{2}$ pairs of samples. We shall denote these as t_{ij}, $i \neq j = 1,2,\ldots,k$. That is t_{ij} is the difference of the means of i and j samples. One could also consider the absolute differences of the means.

(3) Consider a set of transformation parameters ϑ_t, $t = 2,3,\ldots,k$ and transform the original data set to get

$$
Y \equiv \begin{array}{cccc}
X_1 & X_2-\vartheta_2 & X_3-\vartheta_3,\ldots, & X_k-\vartheta_k \\
x_{11} & x_{21}-\vartheta_2 & x_{31}-\vartheta_3 & x_{k1}-\vartheta_k \\
x_{12} & x_{22}-\vartheta_2 & x_{33}-\vartheta_3 & x_{k2}-\vartheta_k \\
x_{13} & x_{23}-\vartheta_2 & x_{33}-\vartheta_3 & x_{k3}-\vartheta_k \\
\vdots & \vdots & \vdots & \vdots \\
x_{1n_1} & x_{2n_2}-\vartheta_2 & x_{3n_3}-\vartheta_3 & x_{kn_k}-\vartheta_k
\end{array}
$$

(4) Bootstrap the transformed data set Y. For each bootstrap configuration, calculate the (absolute) differences of the means for all pairs $\binom{k}{2}$ of samples. That is calculate

$$t^*_{ij} = |t^*_{i0}-t^*_{j0}-\vartheta_j| \qquad i \neq j = 1,2,\ldots,k \quad .$$

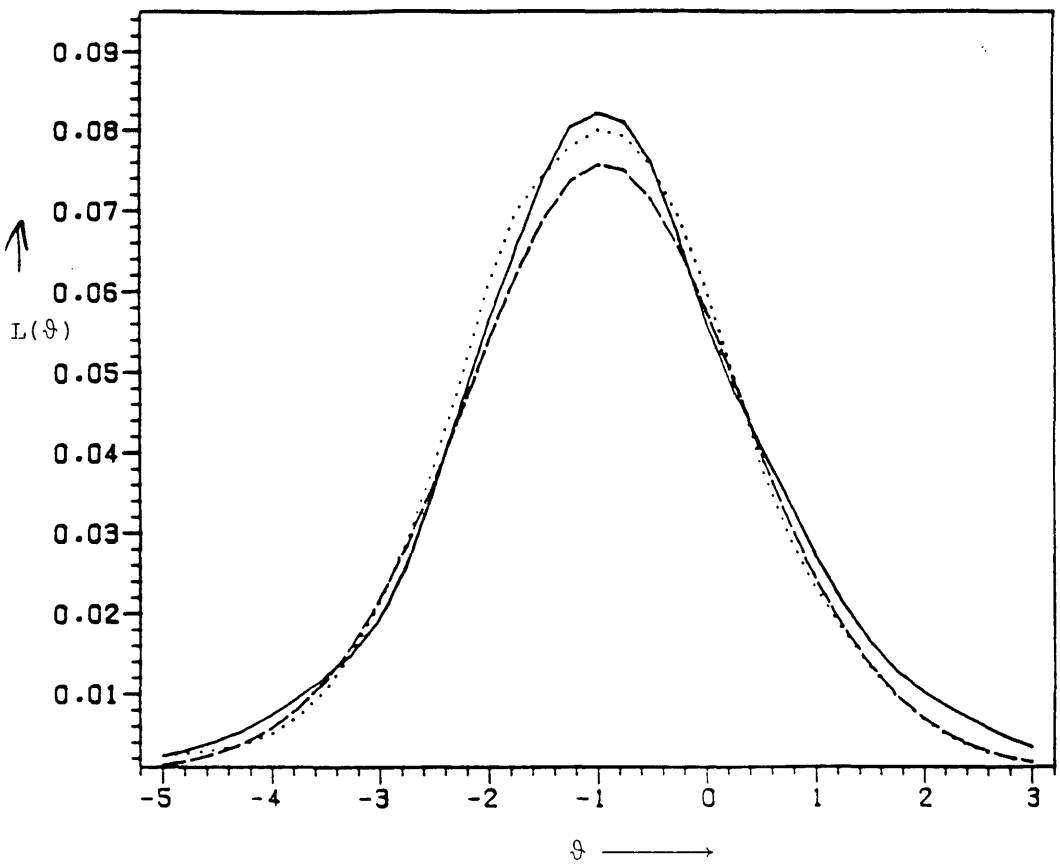
Then treat $\{t_{ij}\}$ and $\{t^*_{ij}\}$ as t and t^* in the two sample problem discussed in §5.2.

A method of considering all the k-samples together once could be to calculate the maximum of the (absolute) differences of means for each of the bootstrap samples. That is, calculate

$$T^{*}(y_{ij}) = \max_{\substack{1 \leq i,j \leq k \\ i \neq j}} (t^{*}_{ij}) \quad .$$

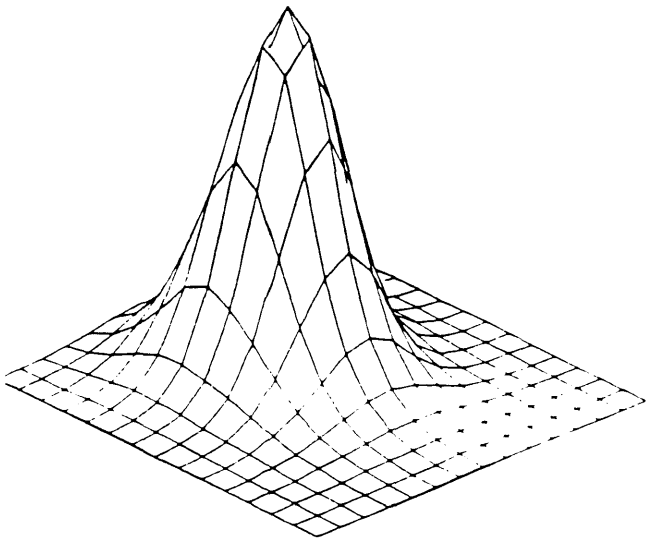
The basic assumption here is that the pairs of samples in the transformed data set are from the same population or indeed that the samples of the transformed data set have the same distribution.

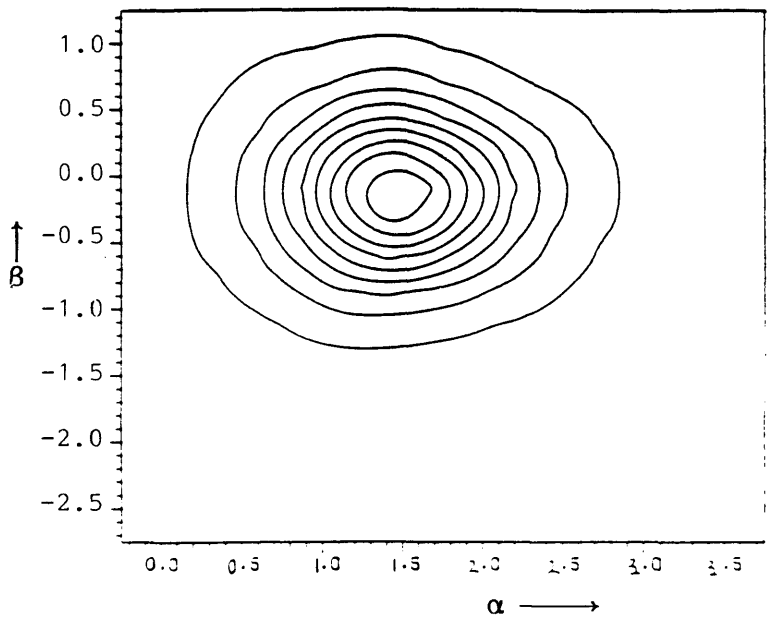
Figure 5: The Distribution of Normalized Likelihood, $L(\vartheta)$, for a Two-Sample Problem

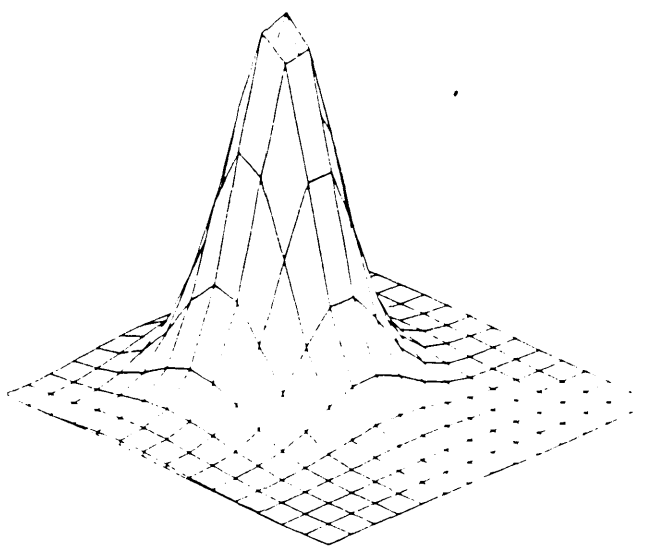


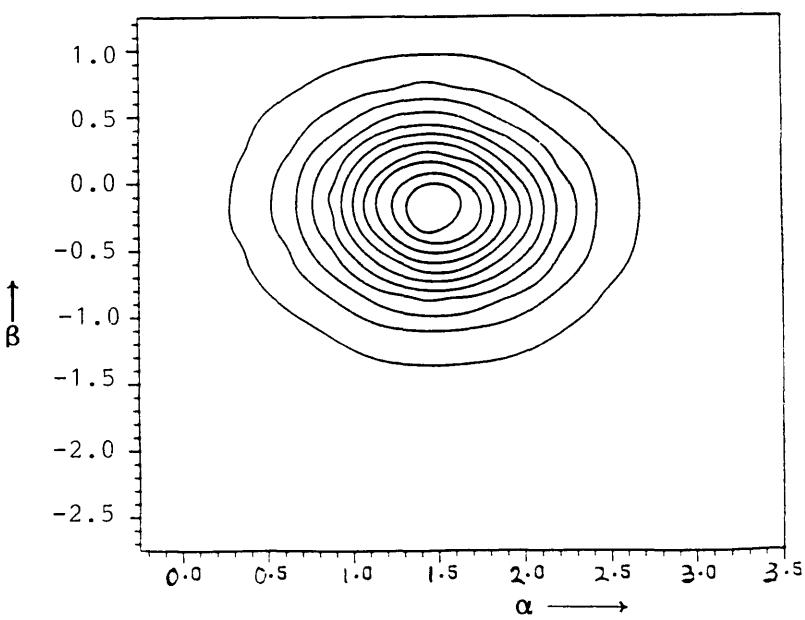
———— Randomized, δ=0,..... Balanced Array, δ = 200,...t-Dist. fit.

Figure 6: Contour Maps for the Regression the Normalized Likelihood, $L(\vartheta_1, \vartheta_2)$ for Problem









β \ α	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50
-1.50	.0000	.0001	.0002	.0005	.0009	.0015	.0018	.0019	.0017	.0014	.0010	.0006	.0004	.0002	.0001
-1.25	.0001	.0002	.0005	.0010	.0020	.0032	.0041	.0043	.0039	.0030	.0020	.0012	.0006	.0003	.0001
-1.00	.0001	.0003	.0008	.0019	.0040	.0065	.0088	.0096	.0082	.0057	.0035	.0018	.0008	.0004	.0001
-0.75	.0001	.0004	.0012	.0032	.0069	.0122	.0169	.0187	.0160	.0099	.0052	.0024	.0010	.0004	.0002
-0.50	.0002	.0005	.0015	.0043	.0103	.0190	.0272	.0301	.0251	.0157	.0072	.0029	.0011	.0004	.0002
-0.25	.0002	.0005	.0015	.0044	.0114	.0233	.0343	.0375	.0304	.0184	.0086	.0032	.0011	.0004	.0001
0.00	.0002	.0005	.0013	.0038	.0101	.0212	.0323	.0354	.0278	.0161	.0074	.0028	.0010	.0003	.0001
0.25	.0001	.0004	.0010	.0031	.0078	.0155	.0233	.0259	.0203	.0113	.0049	.0019	.0007	.0002	.0001
0.50	.0001	.0003	.0008	.0022	.0052	.0095	.0136	.0153	.0124	.0070	.0030	.0011	.0004	.0001	.0000
0.75	.0001	.0002	.0006	.0015	.0029	.0049	.0070	.0077	.0065	.0039	.0018	.0007	.0002	.0001	.0000
1.00	.0000	.0001	.0004	.0008	.0015	.0024	.0033	.0036	.0031	.0021	.0010	.0004	.0001	.0000	.0000
1.25	.0000	.0001	.0002	.0005	.0008	.0011	.0015	.0016	.0015	.0010	.0005	.0002	.0001	.0000	.0000
1.50	.0000	.0001	.0001	.0002	.0004	.0006	.0007	.0008	.0007	.0005	.0003	.0001	.0000	.0000	.0000
1.75	.0000	.0000	.0001	.0001	.0002	.0003	.0003	.0004	.0003	.0002	.0001	.0001	.0000	.0000	.0000
2.00	.0000	.0000	.0000	.0001	.0001	.0002	.0002	.0002	.0001	.0001	.0001	.0000	.0000	.0000	.0000

Table 9

Computed Normalized Likelihood for the Regression Problem : BOOTSTRAP, B = 1000

β \ α	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50
-1.50	.0000	.0000	.0001	.0003	.0006	.0009	.0011	.0011	.0009	.0007	.0005	.0003	.0001	.0000	.0000
-1.25	.0000	.0001	.0002	.0006	.0013	.0022	.0030	.0031	.0026	.0018	.0011	.0005	.0002	.0001	.0000
-1.00	.0001	.0002	.0005	.0014	.0031	.0056	.0079	.0084	.0068	.0042	.0021	.0009	.0003	.0001	.0000
-0.75	.0002	.0005	.0011	.0028	.0067	.0127	.0181	.0195	.0153	.0087	.0038	.0014	.0005	.0002	.0001
-0.50	.0004	.0008	.0018	.0048	.0119	.0227	.0327	.0345	.0267	.0150	.0060	.0020	.0007	.0003	.0002
-0.25	.0006	.0010	.0022	.0060	.0151	.0301	.0434	.0448	.0335	.0183	.0076	.0025	.0008	.0004	.0002
0.00	.0006	.0010	.0021	.0056	.0142	.0282	.0409	.0416	.0299	.0158	.0065	.0023	.0008	.0003	.0002
0.25	.0005	.0008	.0017	.0042	.0102	.0195	.0277	.0278	.0196	.0101	.0041	.0014	.0006	.0003	.0001
0.50	.0003	.0006	.0011	.0026	.0057	.0101	.0138	.0138	.0099	.0052	.0021	.0007	.0003	.0002	.0001
0.75	.0002	.0004	.0007	.0013	.0024	.0041	.0054	.0055	.0042	.0023	.0010	.0003	.0001	.0001	.0000
1.00	.0001	.0002	.0003	.0005	.0009	.0014	.0019	.0020	.0017	.0010	.0004	.0002	.0001	.0000	.0000
1.25	.0001	.0001	.0001	.0002	.0003	.0005	.0007	.0008	.0007	.0004	.0002	.0001	.0000	.0000	.0000
1.50	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0003	.0003	.0002	.0001	.0000	.0000	.0000	.0000
1.75	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000
2.00	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000

Table 9 continued.... ART (200), δ = 600

CHAPTER SIX

GENERAL DISCUSSION

The main purpose of this thesis was to introduce methods to cut down on computational workload in resampling and rerandomization problem. The non-parametric multi-sample test statistic W_N was introduced as an analytic method of cutting down computation (over Sen's 1966 test). Exact distributions are obtained for such problems by considering the correct reference set of permutation configurations.

It is difficult to extend any given non-parametric testing procedure to a confidence interval procedure, that is to give a semi-parametric development. It may be necessary to restructure the simulation design if the number of configurations is large.

The ART method proposed works well in the balanced case for Bootstrapping as measured by "discrepancy". Some theoretical results justify the use of balancing. Direct computational methods leading to balanced low discrepancy codes seems a natural and exciting research direction and would be a pleasing development from this work.

Resampling codes are seen to go hand-in-hand with new methods of semi-parametric inference and one such method is proposed here.

In situations where the corresponding permutation test can easily be inverted the results are similar though subtly different. The "likelihood" method proposed seems to be a reasonable procedure at least as a cheap method of obtaining inferences in situations with few distributional assumptions. It is hoped that the method will provide a framework for "sensitivity analysis" in complex situations.

CHAPTER SEVEN

APPENDIX

We first give a proof of formula (4.3.15). This proceeds by induction on the dimension n. The case n = 1 is elementary. Select one coordinate say ω_n, without loss of generality. Write ω' for ω with ω_n removed. Then

$$\sum_{\omega \in S} U(\omega)\Delta_{\omega} V(\omega) = \sum_{\omega \in S} U(\omega)\Delta_{\omega_n}{}^{1-}\Delta_{\omega'}{}^{(n-1)-} V(\omega) \quad .$$

We can expand the right hand side as

$$\sum_{\omega \in S}{}^{1} U(\omega)\Delta_{\omega'}{}^{(n-1)-} V(\omega) - \sum_{\omega \in S}{}^{2} U(\omega)\Delta_{\omega'}{}^{(n-1)-} V(\omega)$$

where the first summation is with the lost coordinate as ω_n the second with $\omega_n - 1$. Now apply the induction hypothesis to each term to obtain

$$\sum_{\omega \in S}{}^{1} (-1)^h \sum_{T(h)} \sum_{\omega \in \tilde{S}(T(h))} \Delta^{h+} U(\omega)V(\omega) - \sum_{\omega \in S}{}^{2} \sum_{T(h)} \sum_{\omega \in \tilde{S}(T(h))} \Delta^{h+} U(\omega)V(\omega) \quad ,$$

where in both cases the T(h) refer to ω' not ω. We may now collect terms in $V(\omega)$ and extend the T(h) to ω. In doing so we capture the right hand side of (4.3.15).

The proof of Theorem 2 follows a standard route in the discrepancy literature. In (4.3.15) put $U = t(x(\omega))$ and $V = g(\omega)$. The left hand side of (4.3.15) becomes ε. Thus taking modulii and

using (4.3.15) we obtain

$$
\begin{aligned}
|\varepsilon| &= \Big| \sum_{\omega \in S} t(x(\omega)) \Delta^{n-} g(\omega) \Big| \\
&= \Big| \sum_{h=1}^{n} (-1)^h \sum_{t(h)} \sum_{\omega \in \tilde{S}(T(h))} \Delta^{h+} t(x(\omega)) g(\omega) \Big| \\
&\leqslant \sum_{h=1}^{n} \sum_{T(h)} \sum_{\omega \in \tilde{S}(T(h))} \left| \Delta^{h+} t(x(\omega)) g(\omega) \right| \quad .
\end{aligned}
$$

Apply the Cauchy-Schwartz inequality to each interior summation to continue

$$
\leqslant \sum_{n=1}^{n} \sum_{T(h)} \left(\sum_{\omega \in \tilde{S}(T(h))} (g(\omega))^2 \right)^{\frac{1}{2}} \times
$$

$$
\left[\sum_{\omega \in \tilde{S}(T(h))} (\Delta_{T(h)}^{h+1} t(x(\omega)))^2 \right]^{\frac{1}{2}}
$$

which from the definition of $\tilde{W}(T(h))$ gives the result.

REFERENCES

ADLER, H.L. and ROESSLER, B. (1964), *Introduction to Probability & Statistics*, Freeman, San Francisco.

BERGER,J.O. and WOLPERT, R. (1984), *The Likelihood Principle*, IMS Lecture Notes **6**, Hayward, California.

BRAATEN, E. and WELLER, G. (1979), An Improved Low-Discrepancy Sequence for Multidimensional Quasi-Monte Carlo Integration, *Journ. Comp. Physics*, **33**, 249-258.

CHERNOFF, H. and SAVAGE, R.I. (1958), Asymptotic Normality and Efficiency of Nonparametric Tests, *Ann. Math. Statist.*, **29**, 972-994.

DIGGLE, P.J. and GRATTON, R.J. (1984), Monte Carlo Methods of Inference for Implicit Statistical Models, *Journ. Roy. Statist. Soc., (B)*, **46**, 193-227.

DUNCAN, D.B. (1955), Multiple Range and Multiple F-Tests, *Biometrics*, **11**, 1-42.

EFRON, B. (1979a), Bootstrap Methods: Another Look at the Jackknife, *Ann. Statistics*, **7**, 1-26.

(1979b), Computers and the Theory of Statistics: Thinking the Unthinkable, *SIAM Rev.*, **21**, 460-480.

(1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM Regional Conference Series in Applied Mathematics.

(1983), Computer Intensive Methods in Statistics, *Scientific American*, **248(1)**, 116-130.

(1985), Bootstrap Confidence Intervals for a Class of Parametric Problems, *Biometrika*, **72**, 45-58.

EPANECHNIKOV, V.A. (1969), Nonparametric Estimation of Multivariate Probability Density, *Theory of Prob. Appl.*, **14**, 153-158.

FISHER, R.A. (1935), *The Design of Experiments (8th Edition)*, Oliver & Boyd, Ltd, Edinburgh and London.

FRIEDMAN, M. (1937), The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, *Journ. Amer. Statist. Assoc.*, **32**, 675-701.

GABRIEL, K.R. and HALL, W.J. (1983), Rerandomization Inference on Regression and Shift Effects: Computational Feasible Methods, *Journ. Amer. Statist. Assoc.*, **78**, 827-836.

GABRIEL, K.R. and HSU, C.F. (1983), Evaluation of the Power of Rerandomization Tests with Application to Weather Modification Experiments, *Journ. Amer. Statist. Assoc.*, **78**, 766-775.

GOOD, I.J. and GASKIN, R.A. (1971), Nonparametric Roughness Penalties for Probability Densities, *Biometrika*, **58**, 255-277.

GRATTON, R.J. (1979), Generalizing the Test Graph Method for Estimating Functions of an Unknown Density, *Biometrika*, **66**, 663-667.

HAJEK, J. (1968), Asymptotic Normality of Simple Linear Rank Statistic Under Alternatives, *Ann. Math. Statist.*, **39**, 325-346.

HALTON, J. (1960), Note Erratum in Numer. Math., 2, *Numer. Math.*, **2**, 84.

(1972), Estimating the Accuracy of Quasi- Monte Carlo Integration in "*App. of Number Theory to Numerical Analysis*", *(S.K. Zaremba, Ed.)*, Academic Press, New York.

HARTIGAN, J.A. (1969), Using Subsample Values as Typical Values, *Journ. Amer. Statist. Assoc.*, **64**, 1303-1317.

HODGES, J.L. Jr. and LEHMANN, E.L. (1963), Estimates of Location Based on Rank Tests, *Ann. Math. Statist.*, **34**, 957-966.

HOLLANDER, M. (1966), An Asymptotically Distribution-Free Multiple Comparison Procedure - Treatment Versus Control, *Ann. Math. Statist.*, **37**, 735-738.

JONCKHEERE, A.R. (1954), A Distribution-Free k-Sample Test Against Ordered Alternatives, *Biometrika*, **41**, 133-145.

KAKUTANI, S. (1975), A Problem of Equidistribution on the Unit Interval [0,1], *Lecture Notes in Mathematics*, **541**, 369-375, Springer-Verlag, Berlin.

KEMPTHORNE, O. and DOERFLER, T.E. (1969), The Behaviour of Some Significance Tests Under Experimental Randomization, *Biometrika*, **56**, 231-248.

KOKSMA, J.F. (1960), A General Theorem for the Theory of Uniform Distribution Modulo 1; *Mathematica Zutphen BPP11*, 7-11 (Dutch quoted after Math. Rev.).

KRONMAL, R.A. and TARTER, M.E. (1968), The Estimation of Probability Densities and Cumulatives by Fourier Series Methods, *Journ. Amer. Statist. Assoc.*, **63**, 925-952.

KRUSKAL, W.H. and WALLIS, W.A. (1952), Use of Ranks in One-Criterion Variance Analysis, *Journ. Amer. Statist. Assoc.*, **47**, 583-621.

LEHMANN, E.L. (1959), *Testing Statistical Hypothesis*, John Wiley & Sons Inc., New York.

LOFTSGAARDEN, D.O. and QUEENSBERRY, C.P. (1965), A Nonparametric Estimate of a Multivariate Density Function, *Ann. Math. Statist.*, **36**, 1049-1051.

MANN, H.B. and WHITNEY, D.R. (1947), On a Test of Whether One or Two Random Variables is Stochastically Larger than the Other, *Ann. Math. Statist.*, **18**, 50-60.

MILLER, R.G. Jr., (1981), *Simultaneous Statistical Inference (2nd Edition)*, Springer-Verlag, New York.

OGBONMWAN, S.M. (1983), A New Nonparametric Multiple Comparison Test Based on Sen's Nonparametric Generalization of the T-Method of Multiple Comparison for the One-Way Criterion Analysis of Variance, *Bull. of I.S.I.: Proceedings of the 44th Session Vol. L(3)*, 157-168.

OGBONMWAN, S.M. and WYNN, H.P. (1985), Accelerated Resampling Codes with Low Discrepancy: Submitted to *J. Roy. Statist. Soc.* **c**.

O'NEIL, R. and WETHERILL, G.B. (1971), The Present State of Multiple Comparison Methods, *Journ. Roy. Statist. Soc., (B)* **133**, 218-250.

PETRONDAS, D.A. and GABRIEL, K.R. (1983), Multiple Comparisons by Rerandomization Tests, *Journ. Amer. Statist. Assoc.*, **78**, 949-957.

PURI, M.L. (1964), Asymptotic Efficiency of a Class of c-Sample Test, *Ann. Math. Statist.*, **35**, 105-121.

PURI, M.L. and SEN, P.K. (1971), *Nonparametric Methods in Multivariate Analysis*, John Wiley & Sons Inc., New York.

ROSENBLATT, M. (1956), Remarks on Some Nonparametric Estimators of a Density Function, *Ann. Math. Statist.*, **27**, 832-837.

(1971), Curve Estimators, *Ann. Math. Statist.*, **42**, 1815-1842.

ROTH, K.F. (1954), On Irregularities of Distribution, *Mathematika*, **1**, 73-79.

RUDEMO, M. (1982), Emperical Choice of Density Estimator, *Scand. Journ. Statistics*, **9**, 65-78.

SACKS, J. and YLVISAKER, D. (1981), Asymptotically Optimum Kernel for Density Estimation at a Point, *Ann. Statist.*, **9**, 334-346.

SCHEFFÉ, H. (1959), *Analysis of Variance*, John Wiley & Sons Inc., New York.

SCHENKER, N. (1985), Qualms About Bootstrap Confidence Intervals, *Journ. Amer. Statist. Assoc.*, **80**, 360-361.

SCHMIDT, W.M. (1972), Irregularities of Distribution VII *ACTA Arith.*, **21**, 45-50.

SCOTT, D.W. (1979), On Optimal and Data-Based Histograms, *Biometrika*, **66**, 605-610.

SCOTT, D.W. and FACTOR, L.E. (1981), Monte Carlo Study of Three Data-Based Nonparametric Probability Density Estimators, *Journ. Amer. Statist. Assoc.*, **79**, 9-15.

SEN, P.K. (1966), On Nonparametric Simultaneous Confidence Regions and Tests for the One-Criterion Analysis of Variance Problem, *Ann. Inst. Statist.*, **18(3)**, 319-336.

(1968), On a Class of Aligned Rank Order Tests in Two-Way Layout, *Ann. Math. Statist.*, **39**, 1115-1124.

SHEATHER, S. (1983), A Data-Based Algorithm for Choosing the Window Width when Estimating the Density at a Point, *Comp. Statist. & Data Analysis*, **1**, 229-238.

SHEATHER, S.J. and MARITZ, J.S. (1983), An Estimation of the Asymptotic Standard Error of the Sample Median,

SHUSTER, J.J. and BOYETT, J.M. (1979), Nonparametric Multiple Comparison Procedures, *Journ. Amer. Statist. Assoc.*, **74**, 379-382.

SILVERMAN, B.N. (1978a), Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives, *Ann. Statist.*, **6**, 177-184.

(1978b), Choosing the Window Width when Estimating a Density, *Biometrika*, **65**, 1-11.

SLOANE, N.J.A. (1984), The Packing of Spheres, *Scientific American*, **250(1)**, 92-101.

SO, Y.C. and SEN, P.K. (1982/83), Nonparametric Repeated Significance Tests for One-Way ANOVA WITH Adaptation to Multiple Comparison, *Journ. Statist. Planning & Inference*, **7**, 83-96.

STEEL, R.G.D. (1960), A Rank Sum Test for Comparing all Pairs of Treatments. *Technometrics*, **2**, 197-207.

TAPIA, R.A. and THOMPSON, J.R. (1978), *Nonparametric Probability Density Estimation*, John Hopkins Univ. Press, Baltimore.

THERNEAU, T.M. (1983), Variance Reduction Techniques for the Bootstrap. *Technical Report No. 200, Department of Statistics, Stanford University*, Stanford, California.

TITTERINGTON, D.M. (1985). Common Structure of Smoothing Techniques in Statistics, *Int. Statist. Rev.*, **53(2)**, 141-170.

TRITCHLER, D. (1984), On Inverting Permutation Tests. *Journ. Amer. Statist. Assoc.*, **79**, 200-207.

TUKEY, J.W., BRILLINGER, D.R. and JONES, L.V. (1978), *The Management of Weather Resources, Vol. II,* Weather Modification Advisory Board, Statistical Task Force, Washington, U.S. Government Printing Office.

VAN AARDENNE-EHRENFEST (1949), On the Impossibility of a Just Distribution, *Proc. Kon. Ned. Akad. u. Wetensch.*, **52,** 734-739.

VAN ZWET, W.R. (1978), A Proof of Kakutani's Conjecture on Random Subdivisions of Longest Intervals, *Ann. Prob.*, **6,** 133-137.

WARNOCK, K.T. (1972), Computational Investigations of Low-Discrepancy Point Sets, in *Application of Number Theory to Numerical Analysis (S.K. Zaremba, Ed)*, Academic Press, New York.

WERTZ, W. (1978), *Statistical Density Estimation: A SURVEY*, Vanden Loeck and Ruprecht in Göttingen.

WILCOXON, F. (1945), Individual Comparisons by Ranking Methods, *Biometrics*, **1**, 80-83.

WU, C.F.J. (1984), Jackknife and Bootstrap Inference in Regression and a Class of Representatives for the LSE, *MRC Technical Summary Report*, #**2675.**

ZAREMBA, S.K. (1966), Good Lattice Points, Discrepancy and Numerical Integration, *Ann. Math. Pura. Appl.*, **73(4)**, 293-318.

(1968), The Mathematical Basis of Monte Carlo and Quasi-Monte Carlo Methods, *SIAM Rev.*, **10(3)**, 303-314.

(1972), *Application of Number Theory to Numerical Analysis (Ed.)*, Academic Press, New York.

CORRIGENDA

Page	Line/ Equation	Incorrect	Correct
10	20	detail	detailed
12	14	evaluations	evaluation
13	8	sample	samples
17	(2.3.4)	$F_{ni}^{(i)}$	$F_{n}^{(i)}$
17	12	E_{N1P}	E_{N1N}
21	6	max $h_N(Xi)$	max $h_N(Xi)$
21	6	min $h_N(Xi)$	min $h_N(Xi)$
25	14	aid	aided
33	11	have	has (both times)
35	23	imperical	empirical
37	10	density	distribution
39	2	metric	metrics
45	(3.4.2)	P or P_i	p or p_i
46	12	balance	balanced
46	16	balanced	a balanced
47	-2 on right	$_{n}(0, 1/3),2/3)$	(0, 1/3, 2/3)
49	13 & 14	Σ_S	Σ_S
50	8	In actual	The actual
51	6	{1,2,3,4,}	{1,2,3,4}
65	19	at R	R
65	20	is measured	are measured
71	(4.3.2)	ω_k	ω_n
71	(4.3.3)	$\underline{<}$	$\underline{>}$
73	5	$\partial\omega_i \partial\omega_j$	$\partial\omega_i \partial\omega_j \ldots$
74	14	balance	balanced
74	16	W(t(h))	W(T(h))
75	(4.3.21)	ω_k	ω_n
	(4.3.19)	n	k
	(4.3.20)	$\sum_{i=1}$	$\sum_{i=1}$ (in both cases
75	(4.3.22)	$\prod_{i=1}^{n} \omega_i$	$\frac{1}{k^n} \prod_{i=1}^{n} \omega_i$

75	17	if $\omega < 0$.	if $\omega < 0$ and γ_{mi} in the ith coordinate of $\omega^{(m)} \in S$.
76	(4.3.23),(4.3.24) & (4.3.25)	$\sum_{m=1}^{N^1}$	$\frac{1}{k^n} \sum_{m=1}^{N^1}$
76	(4.3.23),(4.3.24) & (4.3.25)	$(\prod_{i=1}^{n} \omega_i)^2$	$\frac{1}{k^{2n}} (\omega_{i=1}^{n} i$
76	(4.3.26)	$- 2. \frac{1}{k^n}$	$- 2. \frac{1}{k^{2n}}$
76	(4.3.26)	$+ \frac{N^{12}}{k^n}$	$+ \frac{N^{12}}{k^{3n}}$
77	11	$\prod_{i=1}^{n} (\gamma_{mi} + (\gamma_{mi} + \frac{1}{k})+\ldots+1)$	$\prod_{i=1}^{n} (\gamma_{mi}+(\gamma_{mi}+1)+\ldots+k)$
77	12	$k\gamma_{mi}$	γ_{mi}
78	Minor corrections in several lines - new page attached.		
79	(4.3.29)	$\frac{N^{12}}{k^n}$	$\frac{N^{12}}{k^{3n}}$
79	(4.3.30)	$\max(\gamma_{mi}, \gamma_{qi})$	$\max(\tilde{\gamma}_{mi}, \tilde{\gamma}_{qi})$
79	(4.3.30)	γ_{mi}	$\tilde{\gamma}_{mi}$ (elsewhere).

80	7	same	sample
82	13	emperical	empirical
91	5	realization	realizations
92	15	shift	the shift
93	21	emperical	empirical
97	11	$\Sigma x_i/n - \frac{\Sigma(y_j-\vartheta)}{n}$	$\sum_{i=1}^{m} x_i/m - \sum_{j=m+1}^{N} (y_j-\vartheta)/n$
99	8	$\ell_T(\vartheta)$	$L_T(\vartheta)$
100	9	distribution	distributions
101	15	permutation	the permutation
109	1	(5.4.4)and(5.4.5)	(5.5.4)and(5.5.5)
110	8	formule	formula
111	4	$f(x)$	$\hat{f}(x)$
117	(5.6.7)	$\frac{\Gamma\frac{\nu+1}{2}}{\Gamma\frac{\nu}{2}}$	$\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)}$
131	3	Problem	Problems
132	5	Δ_ω	Δ_ω^{n-}

132	8	Lost	Last
132	17	modulii	moduli
135	4	GASKIN	GASKINS
136	1	Hypothesis	Hypotheses
136	3	QUEENSBERRY	QUESENBERRY
136	6	or	of
137	17	Median,	Median, Austral.J.Statist 25, 109-122.
137	21	B.N.	B.W.
138	16	(1984)	(1985)

then $\gamma_{mi} + (\gamma_{mi} + 1) + \ldots + k$

$$= \frac{1}{k}\left(\frac{k(k+1)}{2} - \frac{(n-1)n}{2}\right)$$

$$= \frac{1}{2k}(k^2+k-n^2+n)$$

Hence we have,

$$-2 \cdot \frac{1}{k^n} \cdot N' \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_n=1}^{k} \left[\sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i - \gamma_{mi}) \right]$$

$$= -2^{-n+1} \cdot N' \sum_{m=1}^{N'} \prod_{i=1}^{n} \left(\frac{1}{k^2}(k^2+k-n^2+n)\right)$$

$$= -2^{-n+1} \cdot N' \sum_{m=1}^{N'} \prod_{i=1}^{n} \left(1 + \frac{1}{k} - \frac{n^2}{k^2} + \frac{n}{k^2}\right)$$

Now putting back n as $n = k\gamma_{mi}$, we get

$$-2 \cdot \frac{1}{k^n} \cdot N' \sum_{\omega_1=1}^{k} \cdots \sum_{\omega_k=1}^{k} \left[\sum_{m=1}^{N'} \prod_{i=1}^{n} \omega_i H(\omega_i - \gamma_{mi}) \right]$$

$$= -2^{-n+1} \cdot N' \cdot \sum_{m=1}^{N'} \prod_{i=1}^{n} \left(1 + \frac{1}{k} - \frac{k^2\gamma_{mi}}{k^2} + \frac{\gamma_{mi}}{k}\right)$$

$$= -2^{-n+1} \cdot N' \cdot \sum_{m=1}^{N} \prod_{i=1}^{n} \left(1 + \frac{1}{k} - \gamma_{mi}^2 + \frac{\gamma_{mi}}{k}\right)$$

$$= -2^{-n+1} \cdot N' \cdot \sum_{m=1}^{N} \prod_{i=1}^{n} \left(1 - \gamma_{mi}^2 + \frac{\gamma_{mi}+1}{k}\right) \qquad (4.3.28)$$