Statistical data analysis by using ranks

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy at Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/32010

Publisher: © B.S. Acar

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 2.5 Generic (CC BY-NC-ND 2.5) licence. Full details of this licence are available at: http://creativecommons.org/licenses/by-nc-nd/2.5/

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

![Creative Commons License](https://creativecommons.org/licenses/by-nc-nd/2.5/)

**Attribution-NonCommercial-NoDerivs 2.5**

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

**BY:** Attribution. You must attribute the work in the manner specified by the author or licensor.

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
STATISTICAL DATA ANALYSIS BY USING RANKS

by

BEHICE SERPIL ACAR
BS, MS (METU, Ankara), MSc (Loughborough), FSS

A DOCTORAL THESIS
Submitted in partial fulfilment of the requirements
for the award of
DOCTOR OF PHILOSOPHY
of the
Loughborough University of Technology

June 1985

© by B.S. Acar, 1985
Anneme ve Babama

To my Mother and Father
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>v</td>
</tr>
<tr>
<td><strong>CHAPTER 1</strong></td>
<td>INTRODUCTION</td>
<td>...</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Rank Order Statistics</td>
<td>...</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>A Review of Methods Using Ranks</td>
<td>...</td>
<td>...</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Ad-hoc Techniques</td>
<td>...</td>
<td>...</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Mathematical Approaches</td>
<td>...</td>
<td>...</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Scope of the Thesis</td>
<td>...</td>
<td>...</td>
<td>7</td>
</tr>
<tr>
<td><strong>CHAPTER 2:</strong></td>
<td>PREDICTION OF A FUTURE RESPONSE BY USING RANKS</td>
<td>...</td>
<td>...</td>
<td>11</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>...</td>
<td>...</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>Predictive Probabilities Using Ranks</td>
<td>...</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Marginal Likelihood of Ranks</td>
<td>...</td>
<td>...</td>
<td>12</td>
</tr>
<tr>
<td>2.2.2</td>
<td>An Approximation to Marginal Likelihood of Ranks for the Linear Model</td>
<td>...</td>
<td>...</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Predictive Probability of a Future Response Having a Specified Rank</td>
<td>...</td>
<td>...</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>Modification for Ties</td>
<td>...</td>
<td>...</td>
<td>18</td>
</tr>
<tr>
<td>2.5</td>
<td>Estimates Based on the Predictive Probabilities</td>
<td>...</td>
<td>...</td>
<td>20</td>
</tr>
<tr>
<td><strong>CHAPTER 3:</strong></td>
<td>TRANSFORMATION OF REGRESSOR VARIABLES</td>
<td>...</td>
<td>...</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>...</td>
<td>...</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Mean Isotonic, Median Isotonic and Rank Transform Estimator</td>
<td>...</td>
<td>...</td>
<td>26</td>
</tr>
<tr>
<td>3.3</td>
<td>Invariance Properties of MNIS, MDIS, RTIC and PRPT Under a Monotone Transformation of Regressor Variables</td>
<td>...</td>
<td>...</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulation Study</td>
<td>...</td>
<td>...</td>
<td>31</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Introduction</td>
<td>...</td>
<td>...</td>
<td>31</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Details of Calculation</td>
<td>...</td>
<td>...</td>
<td>32</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Results</td>
<td>...</td>
<td>...</td>
<td>33</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusions</td>
<td>...</td>
<td>...</td>
<td>35</td>
</tr>
<tr>
<td>CHAPTER 4: TRANSFORMATION OF RESPONSE VARIABLES</td>
<td>Page No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2 Bayesian Predictor and the Maximum Likelihood Predictor Based on Box-Cox Transformations</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3 Suggestions to Overcome Impracticalities of Box-Cox Transformations</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4 Simulation Study</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4.1 Introduction</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4.2 Details of Calculation</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.5 Results</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.6 Conclusions</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.7 Illustration</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 5: DISCRIMINANT ANALYSIS BY USING RANKS AND A CASE STUDY</th>
<th>Page No</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>65</td>
</tr>
<tr>
<td>5.2 Approximation to the Predictive Estimate of $P_i$ by Using Ranks</td>
<td>67</td>
</tr>
<tr>
<td>5.3 Approximation to the Predictive Estimate of $P_i$ by Using its Distribution</td>
<td>70</td>
</tr>
<tr>
<td>5.4 Use of Pearson Curves</td>
<td>75</td>
</tr>
<tr>
<td>5.5 Illustrations</td>
<td>77</td>
</tr>
<tr>
<td>5.5.1 Illustration I</td>
<td>78</td>
</tr>
<tr>
<td>5.5.2 Illustration II</td>
<td>79</td>
</tr>
<tr>
<td>5.6 Difference Between $\hat{P}_i$ and mean of $P_i$</td>
<td>80</td>
</tr>
<tr>
<td>5.7 A Similarity of Rank and Parametric Techniques</td>
<td>81</td>
</tr>
<tr>
<td>5.8 Error Rate Estimations</td>
<td>82</td>
</tr>
<tr>
<td>5.9 Application of Predictive Techniques to a Medical Data</td>
<td>85</td>
</tr>
<tr>
<td>5.9.1 Application of Predictive Techniques to the Reference Data</td>
<td>87</td>
</tr>
<tr>
<td>5.9.2 Comparison of Predictive Estimates of $P_i$ for Further Observations</td>
<td>89</td>
</tr>
<tr>
<td>5.10 Conclusions</td>
<td>90</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to Dr A.N. Pettitt for bringing the subject to my attention and for his guidance, many helpful comments and encouraging support during the course of this work. I would also like to express my thanks to Mr B.A. Moore for overseeing my work at its early and very late stages.

I would like to thank the Head of Department of Mathematics, Professor C Storey for providing the departmental and central facilities for this research.

Thanks are also due to the staff of the Computer Centre, in particular, Mr G.P. Gerard and Ms M. Hearnden for their friendly and helpful advice on computational problems.

I am grateful to the Ministry of Education of the Turkish Government and to the Academic Initiatives Development (AID) scheme of Loughborough University of Technology for providing financial support.

My thanks also go to Janet Smith for her skilful and careful typing.

Finally, I am most grateful to Memiş and Barış, not only for their enduring patience and support, but also for being my best friends.
ABSTRACT

The research is concerned with the development and implementation of statistical techniques by using approximations to marginal likelihood of ranks.

A brief review of rank techniques to analyse data is made.

For the problem of predicting a future response, when the data can be transformed to a standard linear model, but the transformation is unknown, a technique is introduced. The predictive probabilities for a future response, in terms of the spacings of the order statistics of the sample of responses are found. Various estimates of the future response based on these probabilities are presented. Simulation studies are undertaken in different conditions for comparing these estimates with existing estimates based on ranks and some parametric estimates. It is shown that the new estimates do not suffer the same difficulties as the earlier estimates.

Some techniques are proposed for the two population discrimination problem. Estimates of the predictive probability of a new case being a member of one of the population given the ranks of the observations are obtained by approaching the problem in two different ways. A similarity of rank and parametric techniques and differences between the two rank methods proposed, are highlighted, error rate estimates are obtained. A medical data set is examined.
The approximation to the rank marginal likelihood is extended up to quartic terms and the scores are found where the variables are assumed to have logistic distribution. The asymptotic likelihood is considered for two sample population and trend models to compare the approximation with earlier ones.

Finally rank based techniques are developed for making inferences for a serial correlation model. A computational model is developed to evaluate the integral for the marginal likelihood of ranks and it is used to investigate the performance of the approximation. Simulation studies are undertaken to compare the proposed technique with the earlier rank based methods and to show the effect of marginalization over the ranks of the serially correlated random variables on the curvature of the resulting likelihood function. A test of location is proposed and assessed where the observations are related by the linear model with serially correlated errors.
CHAPTER 1
INTRODUCTION

Techniques to analyse data without making any assumptions about the distribution of the sample observations, known as non-parametric or distribution free techniques, can be traced back as far as the eighteenth century. However, publication of the article by Wilcoxon (1945), is often considered as the beginning of the full scale development of such techniques. This article proposed sign tests which became the most widely used non-parametric techniques in existence.

The development of distribution free techniques was mostly motivated by problems in the analysis of data in behavioural and medical sciences. It became a complementary branch of statistics. The theory of parametric methods is generally concerned with the development of the techniques where observations are assumed to come from a known population such as normal. The use of parametric techniques may be subject to criticism in practice, when the assumptions regarding the underlying population are not satisfied.

1.1 Rank Order Statistics
The main subject of this thesis will be the development and implementation of statistical techniques based on ranks of the observations.

Rank techniques are usually considered as a sub-family of non-parametric methods by statisticians, however we prefer to call these 'semi-
parametric techniques' due to the reasons explained later, in Section 1.3. First it would perhaps be appropriate to explain what is meant by 'rank techniques'.

Rank techniques are statistical procedures which are based only on the relative position of the observations in the sample. Suppose \( Y_1, \ldots, Y_N \) are the random variables and are re-arranged in ascending order of magnitude, \( Y_{a_1} \leq Y_{a_2} \leq \ldots \leq Y_{a_N} \). If \( Y_j \) is the \( i \)th smallest in the sample, i.e. \( j = a_i \), then the rank of \( Y_j \) can be defined by function \( r(Y_j) = i \). Thus, rank order statistics involve the ranks of the ordered observations only, not their actual values. Therefore rank techniques require the actual values of the observations, only to obtain the ranks of the observations.

The main objection to rank order statistics is that these methods only use ordinal structure in the data thus their performance is inferior to the performance of the parametric procedures. This is true when the distributional assumptions of the classical parametric procedures are fully satisfied. The rank techniques are not offered as the best theoretical choice to make statistical inference, but their use in practice is recommended for analysing samples arising from a population where the assumption of normality is dubious. High asymptotic efficiency of some rank tests, for example asymptotic efficiency of Wilcoxon signed rank test when the underlying distribution is double exponential or Cauchy, is a very supportive argument in this context.
Another advantage of rank techniques is their robustness: they are insensitive to outliers. A similar objection arises again, since there are robust techniques, such as Huber type M-estimates, trimmed mean; which ignore the extreme observations but use all the information provided by the rest of the data. Bickel (1965) compared a rank estimate of location with the trimmed mean and showed that the rank technique is better than the trimmed mean with a suitable choice of population distribution.

Most of the rank techniques are quick and simple to apply. Furthermore they can also be applied when the actual values of the observations are difficult, uneconomical and sometimes impossible to obtain, in other words when the data consist of measurements which assess each individual relative to others.

1.2 A Review of Methods Using Ranks

A comprehensive review of early work in rank order statistics can be found in Bradley (1968), Lehman (1975), Gibbons (1971) and the theory of a selection of such tests in Hajek and Sidak (1967). Making comparisons with the other branches of statistics, very little has been published on the analysis of data based on rank order statistics.

A selection of the recent works relevant to the work reported in this thesis are reviewed below briefly, although more detailed descriptions of some of the techniques are given in the relevant Chapters.
1.2.1 Ad-hoc Techniques

One way to develop rank techniques is to replace the data with the ranks of the observations and then to deploy usual parametric methods. Although the mathematical theory has not yet been developed for such an approach, several statisticians used this method, for example Shirley (1980), Iman and Conover (1979), Bartels (1982). In fact in a series of articles Iman and/or Conover applied the techniques of usual least squares regression analysis (1979), discriminant analysis (1981), analysis of experimental design (1974, 1976) to rank transformed data and undertook Monte Carlo studies to show the power of the techniques. They concluded:

"the results of simulation experiments show that this approach is probably a valid procedure".

Bartels followed a similar trend to obtain rank version of von Neuman correlation coefficient.

1.2.2 Mathematical Approaches

Publications by Randles, Broffitt, Ramberg and Hogg (1978a,b) in discriminant analysis, propose use of ranks in a different fashion. They used the ranks of the resulting discriminant functions and the distribution of these ranks to describe a classification rule. Simulation studies were undertaken to compare the techniques with other methods under various circumstances. Ng and Randles (1983) extended the technique to cover settings involving more than two populations.
Following Hajek and Sidak's (1967) theory, Gupta and Govindarajulu (1980) proposed a rank test for randomness against first order auto-regressive normal alternative and undertook Monte Carlo studies to compare the technique with another rank test (Wald and Wolfowitz) and parametric alternative.

Kalbfleisch (1978) suggested the use of marginal likelihood of ranks in developing techniques. Thus, the inference for a parameter \( \theta \) can be made by using likelihood methods where the rank marginal likelihood is defined by the probability of the ranking \( r_1, \ldots, r_N \) a presentation of \( N \) numbers. This probability is given by an \( N \)-fold integral of the joint distribution of the sample where the integral is taken over the ordered observations.

The probability distribution of \( N \) objects over \( N! \) permutations is considered by Plackett (1975) and by Tallis and Dansie (1983). Plackett's model calculates the probability of a ranking of \( N \) objects in terms of the probability that object \( i \) is ranked first for \( i = 1, \ldots, N \), thus it involves \( N-1 \) parameter in the first stage. In second and higher stages, new parameters are added since the probabilities that each pair (triplets) of objects are ranked first and second (and third...) are considered. In their article Tallis and Dansie suggested an alternative, relatively simple method as an 'exploratory technique' which involves up to \( (N!-1) \) parameters.

Numerical solution to the \( N \)-tuple integral to obtain the probability of a ranking was considered and an algorithm was given by Milton (1970).
for independent random variables with continuous density.

Brooks (1978) considered an approximation of the form
\[ \exp(\theta y_1 - \frac{1}{2} \theta^2 y_2) \]
for the two sample problem where \( y_1 \) and \( y_2 \) are rank statistics and the underlying distribution is normal or logistic. Henery (1981) gave approximate probabilities by using Taylor series expansion as far as the linear terms for multisample problems where the distribution of random variables is normal with different location parameters which are in a small neighbourhood of zero and variance one.

Pettitt (1982a) obtained an approximation to the rank marginal likelihood for multisample problems and for various underlying distributions. This approximation is based on the Taylor series expansion of the logarithm of the rank marginal likelihood about zero. In the same article this approximation is used to estimate the location parameter of independently normally distributed random variables which differ only in their location parameters and are related by a linear model. Pettitt (1982b, 1983a,b) can be given as examples of his various studies to make inference based on this approximation. In 1982b article, cubic approximations are considered for normal density to approximate probability of rankings and used in Judges problem. In the 1983a and b articles, scores were modified to analyse incomplete ranked data and various explicit scores were given for right censored data respectively. Real life examples were given to compare the techniques with alternative methods in all the articles mentioned in this paragraph.
1.3 Scope of the Thesis

This thesis is concerned with the development and implementation of rank techniques based on approximations to the marginal likelihood of ranks. As mentioned in Section 1.1, we classify our techniques as 'semi-parametric' since like all the other techniques based on the rank marginal likelihood, the underlying population distributions need to be specified. However, an outstanding property of ranks, being invariant under monotone increasing transformation of the observations, make the assumptions made for the underlying populations rather weak. That is, the techniques developed for the density $f(.)$ can be applied to data when the density of the population of the transformed observations is assumed to be $f(.)$.

One of the problems considered in this thesis is predicting a future response, given a vague or partly unspecified relationship between response variables and explanatory variables. A technique that uses ranks of the observations is introduced, when the data can be transformed to a standard linear model by an unknown monotone transformation. The technique is based on the approximation to the rank marginal likelihood of Pettitt (1982a). The predictive probability distribution of a future response $Y^*$ having a specified rank is obtained and various estimates of the value of $Y^*$ based on these probabilities and order statistics are presented in Chapter 2. Monte Carlo studies are undertaken to compare our technique with some non-parametric and parametric techniques. In Chapter 3, first investigations of the performance of the methods are made when the
required conditions for our technique are not provided, i.e. the error distribution is mis-specified and assumptions about expected values are not satisfied. Secondly, the transformation of regressor variables is considered to provide the required conditions for our technique at least approximately; the effect of this kind of transformations on the other estimates is discussed and the performance of the techniques is investigated. The transformation of response variables is considered in Chapter 4 and techniques are compared when the expected values are as required while the error distribution is either correctly specified or mis-specified (Acar and Pettitt, 1984). Techniques are also applied to a practical example of fitting a monotone relationship for some data from the social sciences.

Chapter 5 is concerned with the application of the approximation to the rank marginal likelihood to a two sample population for independently distributed normal random variables which differ only in their location parameter. The problem of predicting the probability $P_i$ of which of the possible two populations a new case belongs to, given the ranks of the observations is approached either by using the predictive probabilities obtained in Chapter 2 or by treating $P_i$ as a random variable and obtaining an approximate distribution of it; and differences between these two methods are pointed out. In addition error rate estimates are obtained and a similarity of a classical discriminant analysis technique and rank based technique is highlighted over the region of most uncertain cases. A medical data set, consisting of some reference data and further undiagnosed observations, is examined, by obtaining various estimates of $P_i$ based on our techniques and a predictive parameter technique (Acar and Pettitt, 1985a).
The subject of Chapter 6 is the approximation to the rank marginal likelihood where the underlying distribution is logistic. The scores of the approximation which are based on a Taylor series expansion including quartic terms are obtained. The two sample population and a trend model are considered and quartic approximation to the asymptotic likelihood of ranks is compared with cubic, quadratic approximations, a quadratic approximation to the quartic approximation, and the 'exact' values.

Dependent variables where the joint density is multivariate normal with zero mean and covariance matrix $\Sigma_p$, depending on a single parameter are considered and the rank marginal likelihood is approximated in Chapter 7. An estimate of a serial correlation parameter $\rho$ is proposed as well as a statistic to test the hypothesis $H_0: \rho=0$. The approximation is extended for the random variables when the means are related by the linear model and the errors are serially correlated. A test of location is proposed. Chapter 8 is devoted to the investigation of the performance of the approximations obtained in the previous chapter. A computational method is developed to evaluate the complex integral for the marginal likelihood of ranks. Approximated values are compared with various other tests based upon ranks of the observations. Curvature values are computed and compared with suggested values for a good approximation. Various designs for the two sample population are also considered and simulation studies are undertaken to investigate the goodness of the approximation by finding the null distribution of the test of location parameter for simple regression models with serially correlated errors. (Acar and Pettit 1985b).
Finally Chapter 9 summarises the conclusions arrived at in previous chapters.
CHAPTER 2
PREDICTION OF A FUTURE RESPONSE BY USING RANKS

2.1 Introduction

The smoothing of data by the fitting of a regression equation, or some other similar technique, can also be thought of as a means of predicting a future response $Y^*$ at some level of an explanatory variable. Obviously for some historical data this point of view is not possible but, for most experimental data, the object of the experiment is to be able to predict a future response.

The problem of making inferences about the future response $Y^*$ is often approached by first assuming that the response variables $Y_1, \ldots, Y_N$ and regressor variables $x_1, \ldots, x_N$ are related by the model $G(Y_j) = H(x_j) + e_j$, where $H(\cdot)$ is a function of the regressor variables, $G(\cdot)$ is a monotonic function and the distribution function of $e_j$ is $F(\cdot)$. The prediction problem can then be approached by estimating the regression parameters implied by $H(\cdot)$. We are interested in the problem where $G(\cdot)$ and, to a lesser extent, $H(\cdot)$ remain unspecified.

In this Chapter a semiparametric technique based on the ranks of the observations is proposed to infer the rank of a future response $Y^*$, where it is assumed that the observations, after some arbitrary monotonic transformation, are independently normally distributed, with means related by a linear model, that is $G(Y_j) = x_j^T \beta + e_j$, where $\beta$
is the vector of unknown parameters. The value of $Y^*$ can also be predicted by using predictive rank probabilities and proposed 'predictors' of Section 2.5.

2.2 Predictive Probabilities Using Ranks

2.2.1 Marginal Likelihood of Ranks

It is assumed that random variables $Y_1, \ldots, Y_N$, after some arbitrary monotone transformation $G(\cdot)$ are independently distributed with density function $f(y-\theta)$ which satisfies the regularity conditions so that the asymptotic theory for maximum likelihood estimation of $\theta$ holds.

One can make inferences about $\theta$ based on the marginal likelihood of the $Y_j$'s given their rank $r_j$, $j=1, \ldots, N$. Assuming $G(\cdot)$ is unknown, makes the ranks marginally sufficient for $\theta$, Kalbfleish (1978).

The marginal likelihood generated by the probability function of ranks $\mathbf{r} = (r_1, \ldots, r_N)$ is defined as

$$f(\mathbf{r}|\theta) = \Pr(Y_{\alpha_1} < Y_{\alpha_2} < \ldots < Y_{\alpha_N}|\theta)$$  \hspace{1cm} (2.1)

where $\alpha_i$ is the anti-rank of $Y_j$ i.e. $\alpha_i=j$ if and only if $Y_j$ is the $i^{th}$ smallest of $Y_1, \ldots, Y_N$ that is $r_j=i$.

The probability in (2.1) is given by
The integral is not manageable for most choices of $f(\cdot)$, but exact numerical results for (2.2) are available for the two sample problems in the literature (Milton (1970)). Approximations to (2.2) are also considered for some cases, for example Brooks (1978) suggests an approximation for the two sample problems when the distribution function is assumed to be normal and logistic. Henery (1981) also considers Taylor series approximation in $\theta$ as far as the linear terms. Pettitt (1982a) extends Brooks' results for the multi-sample problem for any distribution which is unimodal and satisfies various regularity conditions.

2.2.2 An Approximation to Marginal Likelihood of Ranks for the Linear Model

Assume that $f(\cdot)$ mentioned in Section 2.2.1 is standard normal density function and $\theta = X\beta$, that is, random variables $Y_1, \ldots, Y_N$ after arbitrary monotone $G(\cdot)$ are independently normally distributed with unit variances and are related to the regressor variables $x_1, \ldots, x_N$ by the linear model with means $x_i^T \beta$, $i=1, \ldots, N$ where $\beta$ is the vector of $k$ unknown parameters.

Marginal likelihood for ranks, given by
\[ f(\Omega) = (2\pi)^{-N/2} \exp \left\{ -\frac{1}{2} (y-X\beta)^T (y-X\beta) \right\} dy \]

in this case, where \( \Omega \) is the region \( y_{\alpha_1} < ... < y_{\alpha_N} \), can be approximated by using Taylor expansion of support function for ranks (Pettitt, 1982a) to give

\[ f^*(z|\beta) = \frac{1}{N!} \exp \left\{ \beta^T X z - \frac{1}{2} \beta^T C X \beta \right\} \tag{2.3} \]

where the matrix \( C \) and the vector \( a \) are defined by

\[
\begin{align*}
a &= E(Z) \\
C &= I - A
\end{align*}
\]

where \( Z_{\alpha_1} < ... < Z_{\alpha_N} \) are the order statistics of a sample of size \( N \) from the standard normal distribution \( \zeta^T = (Z_{\alpha_1}, ..., Z_{\alpha_N}) \), and \( A = \text{var}(Z) \). So if \( \xi \) is the vector of expected values and \( \Xi \) is the covariance matrix of normal order statistics in a random sample of size \( N \) then \( a_j = \xi r_j \) and \( (A)_{i,j} = (\Xi)_{r_i,r_j} \), so \( (C)_{i,j} = \delta_{i,j} - (\Xi)_{r_i,r_j} \)

where \( \delta_{i,j} \) is the Kronecker delta function.

Equation (2.3) can be written as a function which is proportional to a multivariate density function

\[ f^*(z|\beta) = \text{const.} \times \exp \left\{ \frac{1}{2} m^T M^{-1} m \right\} \times \exp\left\{ \frac{1}{2} (\beta-m)^T M^{-1} (\beta-m) \right\} \tag{2.4} \]
where \( M^{-1} = X^T C X \)

and \( \mu = M x^T \alpha \)

2.3 Predictive Probability of a Future Response Having a Specified Rank

In this section the problem of making inferences about the rank of a future response \( Y^* \) at the value \( x^* \) of the regressor variable \( x \) is considered. It is assumed that the relationship between response variables and explanatory variables is a vague one or partly unspecified, however after an unknown monotone transformation response variables are independently normally distributed with mean \( \mu_x \) and variance one.

If we were using a standard Bayesian analysis based on observations, rather than their ranks, the predictive density \( f(y^*|y) \) of \( Y^* \) given the data \( y \) is given by

\[
f(y^*|y) = \int f(y^*|\beta) \ p(\beta|y) \ d\beta,
\]

where \( p(\beta|y) \) is the posterior density of \( \beta \) and \( f(y^*|\beta) \) is the conditional density of \( Y^* \) given \( \beta \).

If we base our inferences on ranks of the observation, then we can make predictions about the rank of \( Y^* \) amongst the \( Y_1, \ldots, Y_N \), \( Y^* \) given the ranks of the \( Y_1, \ldots, Y_N \). Let \( B_i \) be the event \( (Y^* \) has rank \( i \) amongst
$Y_1, \ldots, Y_N, Y^*$, and $D_N$ be the event ($Y_j$ has rank $r_j$ amongst $Y_1, \ldots, Y_N^*$, $j = 1, \ldots, N$) where ranks of $Y_1, \ldots, Y_N$ amongst $Y_1, \ldots, Y_N, Y^*$ are given in the following manner

$$r_j(i) = \begin{cases} 
  r_j & \text{if } i > r_j \\
  r_{j+1} & \text{if } i \leq r_j 
\end{cases} \text{ for } j = 1, \ldots, N$$

and $r_{N+1}(i) = i$, the rank of $Y^*$. Let $E_i$ be the event that the $Y_1, \ldots, Y_N$ have ranks specified by the data and $Y^*$ lies between the $(i-1)$th and $i$th smallest of the $Y_1, \ldots, Y_N$, i.e. $E_i = B_i \cap D_N$. The conditional probability $\Pr(B_i \mid D_N)$ corresponds to the predictive probability that $Y^*$ has rank $i$ amongst $Y_1, \ldots, Y_N$, $Y^*$ given the ranks of the $Y_1, \ldots, Y_N$ amongst themselves, that is $\Pr(B_i \mid D_N)$ is the predictive probability distribution of $Y^*$.

Therefore $\Pr(B_i \mid D_N)$ can be obtained by finding unconditional probabilities $\Pr(E_i)$ since

$$\Pr(B_i \mid D_N) = \frac{\Pr(E_i)}{\Pr(D_N)} \propto \Pr(E_i),$$

$$\sum_{i=0}^{N+1} \Pr(B_i \mid D_N) = 1$$

and $D_N$ does not depend on $i$. $\Pr(E_i)$ can be found by

$$\Pr(E_i) = \int \Pr(E_i \mid \beta) \Pr(\beta) \, d\beta$$
where $\beta$ is assumed to have locally uniform prior and $\text{pr}(E_i|\beta)$ is approximated by (2.4) for the rank vector of $N+1$ ranks,

$(r(i) = (r_1(i), \ldots, r_{N+1}(i))^T$ as defined above, that is

$$\text{pr}(E_i|\beta) = \text{const.} \times \exp \left( \frac{1}{2} \text{m}^T(i) M^{-1}(i) \text{m}(i) \right) \times \exp \left( - \frac{1}{2} \left( \beta - \text{m}(i) \right)^T M^{-1}(i) (\beta - \text{m}(i)) \right)$$

(2.6)

where

$$M^{-1}(i) = X^T C(i) X^*$$

$$\text{m}(i) = M(i) X^T a(i),$$

$$X^* = \begin{bmatrix} X & X^T \end{bmatrix}$$

where $X^*$ is the $(N+1) \times k$ matrix of regressor vectors $x_i^T$ having $x_i^* T$ as the $(N+1)$th row. The vector $a(i)$ and matrix $C(i) = I - A(i)$ are similar to the vector $a$ and matrix $C$, equation (2.4), but they are of dimension $(N+1)$ rather than $N$ and are obtained by taking

$r(i) = (r_1(i), \ldots, r_{N+1}(i))$ as the vector of $N+1$ ranks.

Integrating (2.6) gives

$$\text{pr}(E_i) \propto \exp \left\{ \frac{1}{2} \text{m}^T(i) M^{-1}(i) \text{m}(i) \right\} \{\det M(i)\}^{\frac{1}{2}}$$

The proportionality constant can be obtained from

$$\sum_{i=1}^{N+1} \text{pr}(B_i|D_N) = 1.$$
Therefore, the predictive probability of $Y^*$ having rank $i$ amongst $Y_1, \ldots, Y_N, Y^*$ given the ranks of $Y_1, \ldots, Y_N$ amongst themselves is

$$p_i = \frac{\exp \left\{ \frac{1}{2} m^T(i) M^{-1}(i) m(i) \right\} \det M(i)}{\sum_{i=1}^{N+1} \exp \left\{ \frac{1}{2} m^T(i) M^{-1}(i) m(i) \right\} \det M(i)}$$

(2.7)

Note that the probabilities are based only on the ranks of the $Y_1, \ldots, Y_N$. The $p_i$'s are invariant under monotonic (increasing) transformations of the $Y_j$'s since the ranks are. Note also that if $pr(\beta=0)=1$, a priori, then $p_i = (N+1)^{-1}$, that is the distribution of $Y^*$ has a uniform probability distribution associated with the spacings of the sample.

2.4 Modification for Ties

In reality ties may occur in the data. To see how ties are dealt with consider a group of $(k+1)$ tied values $Y_{\alpha_1}, \ldots, Y_{\alpha_{S+k}}$. In this case, the vector $\alpha$ and matrix $A$ have to be modified for order statistics of a sample of size $N$ from standard normal distribution such that $Z_{\alpha_1} < Z_{\alpha_2} < \ldots < Z_{\alpha_S} = \ldots = Z_{\alpha_{S+k}} < Z_{\alpha_{S+k+1}} < \ldots < Z_{\alpha_N}$. In general, the marginal density of any tied statistics and joint density of two tied statistics are given by a mixture of corresponding densities when there is no tie. When the sample is taken from standard normal distribution the mean $\tau \xi$ of any tied statistics, covariance
between two tied and one tied one untied statistics are given as

\[ \xi_i = \frac{1}{k+1} \sum_{i=s}^{s+k} \xi_i \quad \text{for } i = s, \ldots, s+k \]

\[ \Xi_{i,i} = \frac{1}{k+1} \sum_{i=s}^{s+k} (\Xi)_{i,i} + \frac{1}{k+1} \left( \sum_{i=s}^{s+k} \xi_i^2 \right) - \frac{k}{k+1} \sum_{i=s}^{s+k} \xi_i^2 \quad \text{for } i = s, \ldots, s+k \]

\[ \Xi_{i,j} = \frac{1}{k(k+1)} \sum_{i \neq j}^{s+k} (\Xi)_{i,j} + \frac{k}{k+1} \sum_{i=s}^{s+k} \xi_i^2 - \frac{k}{k+1} \sum_{i=s}^{s+k} \xi_i^2 \quad \text{for } i = s, \ldots, s+k \]

\[ \Xi_{i,j} = \frac{1}{k+1} \sum_{i=s}^{s+k} (\Xi)_{i,j} \quad \text{for } i = s, \ldots, s+k \text{ but } j \neq s, \ldots, s+k \]

i.e. \( \alpha_j \) is not a member of tied group.

When there is more than one tied group, the expected value and covariances of tied statistics in the second group are found as above while the summations are taken over appropriate indices. Suppose \( Z_{a_{s'}}, \ldots, Z_{a_{s'+k'}} \) form the second tied group, then the element of vector \( \xi \) and matrix \( \Xi \) should be revised to modify components \( \xi_i, \Xi_{i,i}, \)

\( i = s', \ldots, s'+k' \) etc. In addition covariance between two statistics of two different tied groups is

\[ T^{-1}_{i,j} = \frac{1}{(k+1)(k'+1)} \sum_{i=s}^{s+k} \sum_{j=s'}^{s'+k'} \Xi_{i,j} \quad \text{for } i = s, \ldots, s+k, \quad j = s', \ldots, s'+k' \]

Using this modification procedure, followed by calculation of \( p_i \).
(equation 2.7) as usual, will produce the same value of \( p_i \) for \( i=s+1, \ldots, s+k \); that is

\[
\begin{align*}
\Pr(y^{*}_{s} < Y^{*} < y^{*}_{s+1}) &= \Pr(y^{*}_{s+1} < Y^{*} < y^{*}_{s+2}) = \ldots = \Pr(y^{*}_{s+k-1} < Y^{*} < y^{*}_{s+k}) \\
\end{align*}
\]

and \( \sum_{i=s+1}^{s+k} p_i = kp_{s+1} \) is interpreted as being 'the predictive probability of \( Y^{*} \) taking the value \( y^{*}_{s} \).

To see that this is sensible, we can consider an extreme case. Suppose the sample consists of one tied value and \( S = 0 \) with probability 1, then the probability that \( Y^{*} \) takes the tied value is \((N-1)/(N+1)\), that it takes less than or greater than the tied value is \( \frac{1}{N+1} \).

2.5 Estimates Based on the Predictive Probabilities

Estimates of the future response \( Y^{*} \) at the value \( x^{*} \) can be obtained by using the predictive probabilities, \( p_i \) given by equation (2.7). The performances of some estimates discussed in this section will be investigated under various conditions in Chapters 3 and 4.

The first two estimates are means of \( Y^{*} \):

\[
d_1 = \frac{3y^{(1)} - y^{(2)}}{2} \times p_1 + \sum_{i=2}^{N} \frac{y^{(i-1)} + y^{(i)}}{2} \times p_i + \frac{3y^{(N)} - y^{(N-1)}}{2} \times p_{N+1}
\]

\[
(2.8)
\]
\[
d_2 = \{y(1) - \frac{y(N) - \bar{y}}{2}\} \times p_1 + \sum_{i=2}^{N} \frac{y(i-1) + y(i)}{2} \times p_i + \{y(N) + \frac{\bar{y} - y(1)}{2}\} \times p_{N+1}
\]

These are obtained by assuming that the predictive probabilities \( p_i \) are distributed uniformly over \((y(i-1), y(i))\) for \( i = 1, ..., N+1 \). To deal with the probabilities \( p_1 \) and \( p_{N+1} \), the probability of \( Y^* < y(1) \) and \( y(N) < Y^* \), respectively, given the ranks of the data, 'artificial' end points \( y(0) \) and \( y(N+1) \) are defined as functions of \( y(1), ..., y(N) \), e.g.

\[
y(0) = y(1) - c_1(y(2) - y(1)) \quad , \quad c_1 > 0 \quad (2.10)
\]

\[
y(N+1) = y(N) + c_2(y(N) - y(N-1)) \quad , \quad c_2 > 0.
\]

\( d_1 \) and \( d_2 \) are obtained by putting the probability \( p_i \) at the mass points

\[
y(i-1) + y(i) \quad \text{for} \quad i = 2, ..., N
\]

Two end mass points are chosen as

\[
y(1) - \frac{y(2) - y(1)}{2} \quad \text{for} \quad i = 1
\]

\[
y(N) + \frac{y(N) - y(N-1)}{2} \quad \text{for} \quad i = N + 1
\]
for $d_1$ (i.e. $c = 1$), and

$$
\frac{y(1) + \bar{y} - R}{2} \quad \text{for } i = 1
$$

$$
\frac{y(N) + \bar{y} + R}{2} \quad \text{for } i = N+1
$$

for $d_2$ where $R = y(N) - y(1)$ and $\bar{y}$ is the sample mean.

Finally, the 'median' estimator, $d_3$, is obtained by linear interpolation between two consecutive ordered response values $y(m)$ and $y(m+1)$ for $m=2,...,N$ for $\sum_{i=1}^{m} p_i < 0.5 < \sum_{i=1}^{m+1} p_i$. For the case $p_1 > 0.5$, (similarly for $\sum_{i=1}^{m} p_i < 0.5$) $d_3$ takes the value of the smallest (largest) observation i.e.

$$
d_3 = \begin{cases} 
\begin{array}{l}
 y(1) \\
 y(m) + \frac{y(m+1) - y(m)}{p_{m+1}} \times \frac{0.5 - s_m}{p_m}, \\
 y(N)
\end{array}
\end{cases}
$$

where $y(1), ..., y(N)$ is the ordered sample and $s_m = \sum_{i=1}^{m} p_i$, $m = 1, ..., N$. 


Let us consider the simple case where $\beta = 0$, a priori, with probability one, then

$$\Pr(Y^* < Y(1)) = \Pr(Y(N) < Y^*) = \Pr(Y(i-1) < Y^* < Y(i)) = (N+1)^{-1}$$

for $i = 2, \ldots, N$.

For this particular case $d_1$ would be

$$d_1 = \frac{N}{N+1} \bar{y} - \frac{y(2) + y(N-1) - 2(y(1) + y(N))}{2(N+1)}$$

and choosing end mass points as in (2.11) for $d_2$ is equivalent to choosing $c_1, c_2$ in (2.10) as

$$c_1 = \frac{y(N) - \bar{y}}{y(2) - y(1)}$$

$$c_2 = \frac{\bar{y} - y(1)}{y(N) - y(N-1)}$$
The invariance property of $d_3$ should also be noted, that is if $(Y_{(m-1)}, Y_{(m)})$ is the interval that contains the estimate of $Y^*$, then, after any monotone (increasing) transformation $g(\cdot)$ of responses $(g(Y_{(m-1)}), g(Y_{(m)}))$ is the interval that contains the estimate of $g(Y^*)$.

Computer programs are written to obtain predictive probability of a future response $Y^*$ having rank $i$, amongst $Y_1, \ldots, Y_N$, $Y^*$ given the ranks of $Y_1, \ldots, Y_N$ amongst themselves, and to investigate the performance of $d_1$, $d_2$ and $d_3$. 
3.1 Introduction

The non-parametric techniques to make inference about the future response $Y^*$ often consider the model

$$Y_j = H(x_j) + e_j$$  \hspace{1cm} (3.1)

where $H(\cdot)$ is a monotone function. Brunk (1970) and Cryer et al (1972) proposed some nonparametric estimators of the mean regression function and median regression function where both assumed that (3.1) provides the correct model. One way to make inferences which are insensitive to distributional assumptions is to replace observations by their ranks. Iman and Conover (1979) proposed to perform usual least squares regression analysis on the ranks of both the observations and independent variables. These three nonparametric techniques namely Mean Isotonic (MNIS), Median Isotonic (MDIS), Rank Transform (RTIC) and a parametric technique, Least Squares (LS) are compared with the technique based on predictive rank probabilities (PRPT) proposed in Chapter 2. However, the PRPT has been developed by considering the model

$$G(Y_j) = x_j \beta + e_j$$  \hspace{1cm} (3.2)

where $G(\cdot)$ is an arbitrary monotone transformation and the distribution
of $e_j$ is normal. Therefore the comparison is made by considering the model (3.1) as an approximation to the model $H^{-1}(Y_j) = x_j + \tilde{e}_j$.

In the present investigation of performance of $d_1$, $d_2$, $d_3$ (equations 2.8, 2.9, 2.12 respectively) a simulation study has been undertaken with some monotone transformations of the regressor variables to provide the required conditions for the PRPT model at least approximately. The effect of transformation of regressor variables on MNIS, MDIS, RTIC, PRPT is also discussed briefly.

3.2 Mean Isotonic, Median Isotonic and Rank Transform Estimator

Let $Y_1, \ldots, Y_N$ be independent random variables. Consider the model $Y_i = H(x_i) + e_i$ where $H(\cdot)$ is any monotone non-decreasing function of regressor variables $x_i$. Brunk (1970) and Cryer et al (1972) proposed some non-parametric estimators of the mean and median regression functions.

Let $\{x_i\}_{i=1}^N$ be a sequence of values of $x$ in $[0,1]$, not necessarily distinct, and $\{Y_i\}_{i=1}^N$ be a sequence of independent random variables where the mean and median of $Y_i$ are monotone functions of $x_i$.

Mean isotonic regression function studied by Brunk (1970) uses the estimator

$$\hat{\mu}(x_i) = \max_{r \leq x_i} \min_{s \geq x_i} \left\{ \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{r \leq x_j \leq s} \right\}$$

$$= \min_{s \geq x_i} \max_{r \leq x_i} \left\{ \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{r \leq x_j \leq s} \right\}$$
where \( \text{Av}\{ \} \) is used to denote the arithmetic average of the set of random variables described inside the brackets.

Cryer et al (1972) proposed an estimator of the median isotonic regression function which is given by

\[
\hat{\mu}(x_i) = \max_{r < x_i} \min_{s > x_i} \text{Md} \{ Y_j : r < x_j < s \}
\]

\[
= \min_{s > x_i} \max_{r < x_i} \text{Md} \{ Y_j : r < x_j < s \}
\]

where \( \text{Md}\{\cdot\} \) denotes the median of the set described. Specification of the value of \( \hat{\mu}(x_i) \) or \( \hat{m}(x_i) \) at values of \( x_i \) between observation points can be done in different ways and has been left to the investigator. One possibility is to use linear interpolation between adjacent observation points.

Suppose \( \{x_i\}_{i=1}^N \) is a strictly increasing sequence of values, then the mean and median isotonic estimators simplify to

\[
\hat{\mu}(x_i) = \max_{k < i} \min_{s > i} \text{Av} \{ Y_j : i < N, k < j \leq s \}
\]

\[
= \min_{s > i} \max_{k < i} \text{Av} \{ Y_j : i < N, k < j \leq s \} \quad \text{(3.3)}
\]

\[
\hat{m}(x_i) = \max_{k < i} \min_{s > i} \text{Md} \{ Y_j : i < N, k < j \leq s \}
\]

\[
= \min_{s > i} \max_{k < i} \text{Md} \{ Y_j : i < N, k < j \leq s \} \quad \text{(3.4)}
\]
Cryer et al state that \( \hat{m}(\cdot) \) provides a maximum likelihood estimator when the distribution of \( Y_i \) is double exponential with location parameter \( m(x_i) \). The estimate \( \hat{m}(\cdot) \) provides the nearest non-decreasing function to what we have observed in the \( L_1 \) sense, that is

\[
\sum_{i=1}^{n} |\hat{m}(x_i) - y_i| \leq \sum_{i=1}^{n} |g(x_i) - y_i|
\]

for any non-decreasing function \( g(\cdot) \) where \( y_i, i = 1, \ldots, n \) are the observations.

Iman and Conover (1979) proposed to use ranks of both the regressors and responses instead of their actual values in the usual least squares regression analysis. A regression equation that expresses ranks of responses in terms of the ranks of regressor variables is obtained. For simple regression it is

\[
R(\hat{Y}) = \frac{n+1}{2} + \hat{\beta} (R(x) - \frac{n+1}{2})
\]

where \( \hat{\beta} \) is Spearman's rho if there are no ties in the assignment of ranks. Otherwise it is

\[
\hat{\beta} = \frac{(4 \sum_{i=1}^{n} R(x_i)R(Y_i) - n(n+1)^2)/(4 \sum_{i=1}^{n} R^2(x_i) - n(n+1)^2)}{(4 \sum_{i=1}^{n} R(x_i) - n(n+1)^2)}
\]

For any future response, \( Y^* \), it assigns a rank \( R(x^*) \) to the regressor variable \( x^* \) (not necessarily an integer) as follows:
Let \( x_1, x_2, \ldots, x_N \) and \( y_1, y_2, \ldots, y_N \) be the ordered values of the original \( x \)'s and \( y \)'s respectively:

1. If \( x^* = x_i \) for some \( i \), then \( R(x^*) = i \).
2. If \( x^* > x_i \) or \( x^* < x_i \), then \( R(x^*) = N \) or \( R(x^*) = 1 \) respectively.
3. If \( x_i < x^* < x_{i+1} \), \( i = 1, 2, \ldots, N-1 \); then \( R(x^*) = i + \frac{x^* - x_i}{x_{i+1} - x_i} \).

Then it finds the rank of response \( R(y^*) \) by equation (3.5) and assigns a value to the observation which has such a rank as follows:

1. If \( R(y^*) = i \) then \( y^* = y_i \).
2. If \( R(y^*) > N \) or \( R(y^*) < 1 \), then \( y^* = y_N \) or \( y^* = y_1 \) respectively.
3. If \( i < R(y^*) < i+1 \), \( i = 2, \ldots, N-1 \); then \( y^* = y_i + (y_{i+1} - y_i) \frac{R(y^*) - i}{x_{i+1} - x_i} \).

3.3 Invariance Properties of MNIS, MDIS, RTIC and PRPT Under a Monotone Transformation of Regressor Variables

The non-parametric isotonic and rank transform techniques are all based on the same model (3.1). However, they do not have the same invariance properties when a monotone transformation of regressor (or response) variables is considered.
Let an arbitrary monotone transformation \( h \) be applied to the regressor variables. If \( x^* \) is taken as one of the observation points, the rank transform estimate of \( Y^* \) obtained by using \( h(x) \)'s would be the same as the one obtained by using untransformed \( x \)'s since: \( R(h(x^*)) \) amongst \( h(x)'s \) is the same as \( R(x^*) \) amongst \( x \)'s (Case 1 of definition of RTIC), and \( R(h(x^*)) \) is used to obtain \( R(Y^*) \) and finally \( Y^* \). Isotonic estimates are obtained by considering the order of \( x \)'s and taking mean or median of the corresponding \( Y \) values. Hence an isotonic estimate of a future response would be the same since the same \( Y \)'s would be corresponding to \( h(x)'s \).

When \( x^* \) is chosen as a point between observation points, to obtain the rank transform estimate of \( Y^* \), \( R(h(x^*)) \) is found by linear interpolation (Case 3 of definition of RTIC). Consequently \( R(h(x^*)) \neq R(x^*) \) unless \( h(\cdot) \) is linear. Hence RTIC is not invariant under a monotone transformation of regressor variables. Isotonic estimates are not defined uniquely for \( x^* \) between observation points. Using linear interpolation between adjacent points is a suggestion given by Cryer et al (1972). We suggest the use of the usual definition of MNIS and MDIS, if \( \{x_i\}_{i=1}^N \) is a strictly increasing sequence then for \( x_i < x^* < x_{i+1} \), \( i=1, \ldots, N-1 \) take \( R(x^*) = i+0.5 \) and apply (3.3) or (3.4). Monotone transformation of \( x \)'s would not make any difference to isotonic estimates if one of the two definitions mentioned above is chosen when \( x^* \) is in between observation points.

On the other hand, 'values of regressor variables' - not the ranks of them - are used to obtain the marginal likelihood of ranks which leads
to calculation of predictive probabilities. Specifying the design vector correctly is, satisfying the assumptions of PRPT, so that better estimates of future responses can be obtained by \( d_1, d_2, d_3 \). Therefore, transformation of regressor variables plays a crucially important role to reduce the function in these variables, \( H(\cdot) \) in (3.1), to as simple a form as possible.

### 3.4 Simulation Study

#### 3.4.1 Introduction

A simulation study is undertaken to investigate the performance of \( d_1, d_2, d_3 \) (Section 2.5). The model \( Y_j = 4\alpha_j + e_j \) is chosen to generate observations, where the distribution of \( e_j \) is double exponential.

Cryer et al. (1972) chose this model in their simulation study to compare MNIS and MDIS with LS where \( x_j = \alpha_j \). This model has also been chosen by Iman and Conover (1979) to compare RTIC with MNIS, MDIS and LS. Therefore, in order to extend these studies, the same model is chosen in this work, to see how well the estimates of \( d_1, d_2, d_3 \) perform, compared with MNIS, MDIS, RTIC and LS. The second reason for choosing this model is:

a) to test the performance of \( d_1, d_2, d_3 \) when the assumptions of PRPT are not satisfied where \( x_j = \alpha_j \); and

b) to see the improvement when some of the assumptions of PRPT are approximately satisfied where a monotone transformation is applied
to the regressor variables $x_j = h(\alpha_j)$, which does not affect the other non-parametric techniques (Section 3.3).

3.4.2 Details of Calculation

The simulation study considers the observations $Y_j$, $j=1, \ldots, 9$ at nine equally spaced points $\alpha_j = 0.1, \ldots, 0.9$ where $Y_j = 4\alpha_j^4 + e_j$. The error distribution is taken as double exponential with mean zero and variance $\sigma^2 = 1$. The errors are generated by using the following routine:

1. Generate two sets $U_1, U_2$ of real random numbers $u_j$, taken from a uniform distribution between 0 and 1, by using NAG(1977) subroutine G05CAF.

2. Take $v_j = \log(u_j)$ if $u_k > 0.5$ and $v_j = -\log(u_j)$ if $u_k < 0.5$, where $u_j$ belongs to $U_1$ and $u_k$ belongs to $U_2$.

3. Take $e_j = 0.707107xv_j$.

Firstly the estimates are observed for $x_j = \alpha_j$ where conditions for the mean of $Y_j$'s are not satisfied and error distribution is miss-specified for PRPT and LS, but the variance of the errors is taken as $\sigma^2 = 1$ as required. Secondly the transformations of the regressor values are considered, to satisfy the expected value requirements partly ($x_j = \alpha_j^2$) and fully ($x_j = \alpha_j^4$) while the error distribution is still miss-specified with variance $\sigma^2 = 1$. 
In the calculation of predictive rank probabilities, the exact values of means and covariances of order statistics, Pearson and Hartley (1972) have been used in (2.7), however good approximations exist, e.g. Davis and Stephens (1978) and Royston (1982). Simulation results are obtained from 10 000 repetitions for case $x_j = \alpha_j$ to make a precise comparison with Cryer et al (1972) and Iman and Conover (1979) since it is the number of repetitions they used. For $x_j = \alpha_j^2$ and $x_j = \alpha_j^4$ 600 repetitions have been made.

Assessment of the techniques is based on Mean Squared Error (MSE), Mean Absolute Error (MAE) and Bias of the estimates which have been calculated at every observation point for each case. Although, the results based on MAE are neither used nor given here, they are almost the same as the results based on MSE. For the LS estimates, the exact values of MSE are computed from

$$E(\hat{x}(x)) = (x-\bar{x}) \left[ \sum_{j=1}^{9} (x_j - \bar{x})^2 \right]^{-1} \sum_{j=1}^{9} (x_j - \bar{x}) \mu(x_j) + \frac{1}{9} \sum_{j=1}^{9} \mu(x_j)$$

$$\text{Var}(\hat{x}(x)) = \left\{ \frac{(x-\bar{x})^2}{\sum_{j=1}^{9} (x_j - \bar{x})^2} + \frac{1}{9} \right\} \text{Var} (Y_j).$$

3.4.3 Results

The results of the simulation study where no transformation is applied to the regressor variables are given in Tables 3.1a and 3.1b and Diagram 3.1. Ratio of MSE of LS to MSE of the other estimates are
tabulated in Table 3.1a and MSE's are plotted in Diagram 3.1. Rankings of estimates are considered in Table 3.1b.

Diagram 3.1 reveals that all of the estimates do much better in the middle than in the tails of the interval $(0,1)$ in the same way that standard least squares estimates perform. RTIC and median estimate $d_3$ appear to have smaller MSE than LS at most of the observation points (Table 3.1a) and $\text{MSE(}LS\text{/MSE(RTIC)} < \text{MSE(}LS\text{)/MSE(}d_3\text{)}$ except at $\alpha = 0.5, 0.6$. The total of the ratios of MSE's are smaller than 9 for MNIS and MDIS whereas it is greater than 9 for RTIC and for the estimates of PRPT. A similar picture emerges if the rankings of the estimates are considered. Tables 3.1a and 3.1b reveal that the estimates RTIC and $d_3$ are much better than LS for $\alpha < 0.7$, but they appear not to be able to predict the rapid increase in the value of $y$ for $\alpha \geq 0.8$ as well as LS. Overall, the estimates $d_3$ and RTIC came out best.

Secondly, simulation is repeated for the same example with transformed regressor variables and ratio of $\text{MSE(}LS\text{)}$ to $\text{MSE}$ of other estimates are tabulated in Tables 3.2a and 3.2b. Taking $x_j = \alpha_j^2$ has improved the performance of LS and PRPT to give smaller MSE than plotted in Diagram 3.1, whereas it has had no effect on MNIS, MDIS and RTIC as expected. In this case, $d_1$, $d_2$ and $d_3$ perform better than LS, MNIS, MDIS and RTIC at most of the observation points (see Total, Table 3.2a): however they do not perform as well as LS, MNIS or MDIS at the end point $\alpha = 0.9$. 
Diagram 3.2 shows MSE of the competing techniques where the required transformation $x_j = \alpha_j^4$ is applied. It reveals that the MSE of $d_1$, $d_2$ and $d_3$ are much smaller than corresponding values of competitors (MNIS, MDIS, RTIC) for all values of $\alpha$ except $\alpha=0.9$ for $d_1$. LS is improved as expected but the ratio of MSE(LS) to MSE(PRPT) shows that $d_2$ is a better estimator than LS and, $d_1$ and $d_3$ are almost as good as LS (Table 3.2b). RTIC is generally more biased than $d_1$, $d_2$, $d_3$.

Diagram 3.3 points out the reduction in MSE, and Diagram 3.4 considers bias, clearly show the improvement of $d_2$ after transformation of regressor variables and confirms the superiority of one of the PRPT to RTIC.

3.5 Conclusion

One of the various conclusions to be drawn from this study is that none of the rank estimates, RTIC, PRPT can follow the rapid increase in values of $y$ for $x=\alpha$ and $\alpha=0.8$, 0.9, so that even LS based on $x=\alpha$ is better than these estimates at $\alpha=0.9$ (Table 3.1a,b). However, when the regressor variable is correctly specified, that is $x=\alpha^4$, or mis-specified, but not as badly as $x=\alpha$, that is $x=\alpha^2$, then the PRPT provide very good estimates even at $\alpha=0.9$, but the estimator RTIC cannot show any improvement over the $x=\alpha$ case. The isotonic estimators (MNIS, MDIS) do not perform well for slowly increasing $y$, that is $\alpha<0.5$, say. However, they perform very well for $\alpha=0.9$ in comparison to the other estimates (Tables 3.1a,b). There is no improvement in their performance if $x$ is correctly specified.
Thus if \( x \) is incorrectly specified, or alternatively the model 
\[ Y_j = H(x_j) + e_j \]
is assumed with \( H(.) \) non-linear, then \( d_3 \) is the 
best estimate overall (totals Tables 3.1a,b) with the provision 
that it does not perform well where \( y \) is rapidly increasing as a 
function of \( x \). A great improvement in the predictive rank tech­
iques can be brought about if \( x \) is less badly mis-specified, see 
Tables 3.2a,b, \( x = \alpha^2, \alpha^4 \). In practical cases it should be possible 
to specify the regressor variable so that the mis-specification is 
not as bad as \( x=\alpha \) compared to the correct \( x=\alpha^4 \) so that the performance 
of \( d_3 \) in practice should be better than that indicated by Table 3.1b 
where \( x=\alpha \), giving a substantial improvement at \( \alpha=0.9 \).

When the regressor variable is correctly specified (\( x=\alpha^4 \)) \( d_2 \) tends 
to be better than \( d_3 \), in fact \( d_2 \) performs better than LS (\( x=\alpha^4 \)) by a 
substantial amount (Table 3.2a). Now the LS (\( x=\alpha^4 \)) estimate is the 
maximum likelihood estimate based on normal errors with zero mean 
and variance \( \sigma_e^2 \) in the model \( y = \beta x + e \), whereas \( d_2 \) is based on the 
marginal likelihood of the ranks of the \( y \)'s assuming the same paramet­
ric model. The data are generated with double exponential errors, 
that is the error distribution in both models is mis-specified. 
However the estimate \( d_2 \) is superior to the LS(\( x=\alpha^4 \)) estimate Table 3.2b 
is showing the robustness of the rank estimate to mis-specification of 
the error distribution. Asymptotic theory, see for example, Cox and 
Hinkley (1974) suggests that a mean absolute deviations estima­
tor with \( x=\alpha^4 \) would have MSE about 0.5 of the LS(\( x=\alpha^4 \)) estimate but 
small samples results indicate this might be more like 0.7 or larger.
The overall conclusion is that the estimates using the rank predictive probabilities provide very good estimates when the relationship between the dependent variable and the regressor variable is not fully satisfied. In addition, of course, a predictive distribution, not just an estimate, is also given which the isotonic and Iman and Conover's technique cannot give. The price that is paid for these benefits is, more computation.
TABLE 3.1a:
A comparison of LS, MNIS, MDIS, RTIC and PRPT for \( y_j = 4a_j \) i.e. for observations \( y_j = 4a_j + e \), where the distribution of errors is double exponential with zero mean and variance \( \sigma^2 = 1 \). Simulation study is undertaken without applying any transformation to regressor variables i.e. \( x_j = a_j \) and based on 10,000 samples.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>MSE(LS)</th>
<th>MSE(MNIS)</th>
<th>MSE(MDIS)</th>
<th>MSE(RTIC)</th>
<th>MSE(D1)</th>
<th>MSE(D2)</th>
<th>MSE(D3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.929</td>
<td>0.948</td>
<td>1.430</td>
<td>1.604</td>
<td>1.243</td>
<td>1.749</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.997</td>
<td>1.075</td>
<td>1.433</td>
<td>1.185</td>
<td>1.119</td>
<td>1.552</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.745</td>
<td>0.844</td>
<td>1.035</td>
<td>0.836</td>
<td>0.994</td>
<td>1.112</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.858</td>
<td>0.995</td>
<td>1.051</td>
<td>0.813</td>
<td>0.981</td>
<td>1.068</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.077</td>
<td>1.236</td>
<td>1.279</td>
<td>0.951</td>
<td>1.054</td>
<td>1.236</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>1.047</td>
<td>1.152</td>
<td>1.563</td>
<td>1.110</td>
<td>1.148</td>
<td>1.359</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.684</td>
<td>0.723</td>
<td>1.215</td>
<td>1.278</td>
<td>1.255</td>
<td>1.260</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.510</td>
<td>0.520</td>
<td>0.570</td>
<td>1.011</td>
<td>1.089</td>
<td>0.719</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.269</td>
<td>1.272</td>
<td>0.634</td>
<td>0.788</td>
<td>0.885</td>
<td>0.693</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3.1b:
Rankings of MSE's of estimates in Table 3.1a.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>LS</th>
<th>MNIS</th>
<th>MDIS</th>
<th>RTIC</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.2</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.3</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.4</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>6</td>
<td>4</td>
<td>2.5</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>0.6</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>0.7</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0.9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL:</td>
<td>41</td>
<td>53</td>
<td>38.5</td>
<td>27</td>
<td>38</td>
<td>34</td>
<td>20.5</td>
</tr>
</tbody>
</table>
A comparison of LS, MNIS, MDIS, RTIC and PRPT for observations

\[ y_j = 4a_j^4 + e_j \] where the distribution of errors is double exponential with zero mean and variance \( \sigma^2 = 1 \). The following monotone transformations are applied to the regressor variables.

For Table 2a: \( x = a^2 \)

For Table 2b: \( x = a^4 \)

Results are based on 600 repetitions.

**TABLE 3.2a:**

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>MSE(LS)</th>
<th>MSE(MNIS)</th>
<th>MSE(MDIS)</th>
<th>MSE(RTIC)</th>
<th>MSE(D1)</th>
<th>MSE(D2)</th>
<th>MSE(D3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.479</td>
<td>0.483</td>
<td>0.822</td>
<td>1.348</td>
<td>1.150</td>
<td>1.628</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.896</td>
<td>0.944</td>
<td>1.307</td>
<td>1.213</td>
<td>1.126</td>
<td>1.506</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.883</td>
<td>0.985</td>
<td>1.114</td>
<td>1.016</td>
<td>1.077</td>
<td>1.273</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.681</td>
<td>0.749</td>
<td>0.790</td>
<td>0.859</td>
<td>1.006</td>
<td>0.987</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.644</td>
<td>0.700</td>
<td>0.727</td>
<td>0.866</td>
<td>1.000</td>
<td>0.982</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.577</td>
<td>0.638</td>
<td>0.923</td>
<td>1.046</td>
<td>1.103</td>
<td>1.134</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.533</td>
<td>0.577</td>
<td>0.991</td>
<td>1.365</td>
<td>1.318</td>
<td>1.129</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.562</td>
<td>0.573</td>
<td>0.600</td>
<td>1.233</td>
<td>1.256</td>
<td>0.881</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.943</td>
<td>0.937</td>
<td>0.428</td>
<td>0.715</td>
<td>0.840</td>
<td>0.698</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL:</strong></td>
<td><strong>6.198</strong></td>
<td><strong>6.586</strong></td>
<td><strong>7.702</strong></td>
<td><strong>9.661</strong></td>
<td><strong>9.876</strong></td>
<td><strong>10.218</strong></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>MSE(\text{LS})</td>
<td>MSE(\text{MNIS})</td>
<td>MSE(\text{MDIS})</td>
<td>MSE(\text{RTIC})</td>
<td>MSE(\text{D1})</td>
<td>MSE(\text{D2})</td>
<td>MSE(\text{D3})</td>
</tr>
<tr>
<td>----</td>
<td>----------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>0.1</td>
<td>0.257</td>
<td>0.257</td>
<td>0.441</td>
<td>0.905</td>
<td>1.006</td>
<td>1.117</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.602</td>
<td>0.634</td>
<td>0.878</td>
<td>0.909</td>
<td>1.011</td>
<td>1.118</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.784</td>
<td>0.874</td>
<td>0.989</td>
<td>0.911</td>
<td>1.012</td>
<td>1.115</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.698</td>
<td>0.768</td>
<td>0.810</td>
<td>0.931</td>
<td>1.032</td>
<td>1.095</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.533</td>
<td>0.579</td>
<td>0.602</td>
<td>0.965</td>
<td>1.061</td>
<td>1.037</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.327</td>
<td>0.362</td>
<td>0.523</td>
<td>1.036</td>
<td>1.106</td>
<td>0.906</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.287</td>
<td>0.311</td>
<td>0.535</td>
<td>1.150</td>
<td>1.171</td>
<td>0.764</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.500</td>
<td>0.509</td>
<td>0.533</td>
<td>1.186</td>
<td>1.251</td>
<td>0.830</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.971</td>
<td>0.966</td>
<td>0.441</td>
<td>0.930</td>
<td>1.149</td>
<td>1.038</td>
<td></td>
</tr>
</tbody>
</table>
Diagram 3.1: $\alpha$ vs MSE of some nonparametric and parametric estimates when $X_j = \alpha_j$
Diagram 3.2: $\alpha$ vs MSE of some nonparametric and parametric estimates when $X_j = \alpha_j^*$
DIAGRAM 3.3: $\alpha$ vs MSE of RTIC and $d_2$ (improvement of performance of $d_2$)
Diagram 3.4: $\alpha$ vs bias of RTIC and $d_2$ (improvement of performance of $d_2$)
CHAPTER 4
TRANSFORMATION OF RESPONSE VARIABLES

4.1 Introduction

The technique introduced in Chapter 2 which provides probabilities for the intervals \((y_{i-1}, y_i)\), \(i=1, \ldots, N+1\), \(y(0) = -\infty\), \(y(N+1) = \infty\); to introduce a predicted future response \(Y^*\) at \(x=x^*\), considers the model

\[
G(Y_j) = \beta x + e_j
\]

(4.1)

for an arbitrary monotone \(G(.)\) where \(x\) is the regressor variable and the distribution of \(e_j\) is standard normal. In Chapter 3, comparison of the Predictive Rank Probabilities Technique (PRPT) with Mean Isotonic (MNIS), Median Isotonic (MDIS), Rank Transform (RTIC) and Least Squares (LS) is made where the model that PRPT based on, either mis-specified or, by applying monotone transformation to regressor variables, partly specified. In this Chapter, monotone transformation of response variables is considered.

When the model (4.1) is true, a parametric approach to obtain a future response might begin with specifying the monotone transformation \(G(.)\), by assuming that it belongs to some parametric family of transformations such as that considered by Box and Cox (1964). In many practical situations it makes sense only to make predictions in the original metric rather than in the transformation metric. Therefore,
predictions about the future response $Y^*$ can be made by estimating the transformation, fitting the linear model to the transformed $Y$'s and transforming the predicted response back to original metric.

In this Chapter, the aim is to compare the performance of the 'most non-parametric' PRPT, namely $d_3$, with the performance of some non-parametric and parametric predictors. A Monte Carlo experiment which generates data according to inverse Box-Cox transformations applied to a linear model with normal homoscedastic errors (so that the model (4.1) is correctly specified), and with double exponential heteroscedastic errors.

### 4.2 Bayesian Predictor and the Maximum Likelihood Predictor Based on Box-Cox Transformations

Parametric predictors can be obtained if $G(.)$ of (4.1) is specified. Consider the family of transformations, named Box-Cox transformations introduced in their 1964 paper:

$$y(\lambda) = \begin{cases} 
\frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\
\log y, & \lambda = 0,
\end{cases}$$

for $y > 0$. The transformed observations are assumed to be independently normally distributed with constant variance $\sigma^2$ and $E(y(\lambda)) = x^T\beta$. 
Box and Cox (1964) considered the Maximum Likelihood Estimate (MLE) of \( \lambda \) and found it in two steps. First, for given \( \lambda \), the MLE of the vector of parameters \( \hat{\varphi} \) and of the assumed common variance of \( y(\lambda) \) are determined, i.e.

\[
\hat{\sigma}^2(\lambda) = (\hat{y}(\lambda) - \hat{\varphi}(\lambda))T(\hat{y}(\lambda) - \hat{\varphi}(\lambda))/N = S(\lambda)/N
\]

where \( \hat{y}(\lambda) \) is the least squares estimate of \( y(\lambda) \). Then \( \lambda_0 \) is chosen to maximize \( L_{\text{max}}(\lambda) = -\frac{1}{2} N \log \hat{\sigma}^2(\lambda) + \log J(\lambda; y) \) with respect to \( \lambda \), where \( J(\lambda; y) \) is the Jacobian of the transformation, that is

\[
J(\lambda; y) = \prod_{i=1}^{N} \left| \frac{dy_i(\lambda)}{dy_i} \right| . \text{ Working with the normalized transformation}
\]

\[
z(\lambda) = \frac{y(\lambda)}{\hat{y}^{1/N}} \quad \text{i.e.}
\]

\[
z(\lambda) = \begin{cases} 
\frac{y^{\lambda-1}}{\lambda} & \lambda \neq 0, \\
\hat{y} \log y, & \lambda = 0,
\end{cases}
\]

where \( \hat{y} = \left( \prod_{i=1}^{N} y_i \right)^{1/N} \), simplifies \( L_{\text{max}} \) to give

\[
L_{\text{max}}(\lambda) = -\frac{1}{2} N \log \hat{\sigma}^2(\lambda; z)
\]

where \( \hat{\sigma}^2(\lambda; z) = S(\lambda; z)/N = (z(\lambda) - \hat{z}(\lambda))T(z(\lambda) - \hat{z}(\lambda))/N. \)

The MLE, \( \hat{\lambda} \), may be read from the plot of \( L_{\text{max}}(\lambda) \) against trial values of \( \lambda \).
On the other hand, from a Bayesian point of view, it is found that the contribution of the observations to the posterior distribution of $\lambda$ is 
$$
\left(\hat{\sigma}^2(\lambda;z)\right)^{-\frac{1}{2}(N-k)} \propto \left(\frac{S(\lambda;z)}{(N-k)}\right)^{-\frac{1}{2}(N-k)}
$$
as a factor, where $k$ is the number of parameters of the fitted model. On a log scale, it is
$$
L_b = -\frac{1}{2}(N-k) \log \hat{\sigma}^2(\lambda;z).
$$
$L_{\text{max}}$ and $L_b$ differ by the substitution of degrees of freedom for the residual, $N-k$ for $N$. Both of their maxima occur when $S(\lambda;z)$ is a minimum.

Therefore, in practice, when the prior density of $\lambda$ is taken to be uniform, the posterior distribution may be obtained by plotting 
$$
p(\lambda|y) = K x S(\lambda;z)^{-\frac{1}{2}(N-k)}
$$
where $K$ is the normalizing constant.

Maximum Likelihood Predictor (MLP) and Bayesian Predictor (BP) estimates can be found in the following way:

**Step 1:** Choose a set $A$ of trial values of $\lambda$

**Step 2:** Apply the normalized Box-Cox transformation on the observed values to obtain $z^{(\lambda)}$ for $\lambda$ chosen amongst elements of $A$

**Step 3:** Find the least squares estimate $\hat{z}^{(\lambda)}$ of $z^{(\lambda)}$

**Step 4:** Find $S(\lambda;z)$ to obtain $p(\lambda|y)$.

Repeat the procedure by going to **Step 2** until $A$ is exhausted.

**Step 5:** Determine

a) MLP, by

$$
\hat{y} = \begin{cases} 
\lambda_0 \frac{\lambda_0}{y} \left( y_{o}^{\lambda_0} + 1 \right)^{1/\lambda_0} \quad & \lambda_0 \neq 0 \\
\exp \left\{ \frac{1}{y} \hat{z}^{(\lambda_0)} \right\} \quad & \lambda_0 = 0
\end{cases}
$$

for $\lambda_0$ which maximizes $L_{\text{max}}(\lambda)$ (maximizes $p(\lambda|y)$).
b) BP, by

\[ \hat{y} = \sum_{\lambda \in \Lambda} (\lambda \hat{z}(\lambda) \hat{y}^{\lambda-1} + 1)^{1/\lambda} p(\lambda | \hat{y}) + \exp \left( \frac{1}{\hat{y}} \hat{z}(0) \right) p(\lambda=0 | \hat{y}). \]

4.3 Suggestions to Overcome Impracticalities of Box-Cox Transformations

In practice, some problems may arise working with the Box-Cox transformation. First of all, since the transformation is from \( \mathbb{R}^+ \) into \( \mathbb{R} \), it is not always possible to apply the inverse transformation. For example, after finding \( \hat{z}(\lambda) \) as in Step 3 one might obtain a negative value for \( (\lambda \hat{z}(\lambda) \hat{y}^{\lambda-1} + 1) \) which makes the inverse transformation impossible for some values of \( \lambda \). Therefore it is not always possible to choose the set \( \Lambda \) as large as we wish, as it is restricted by non-feasible values of \( \hat{y} \) for some \( \lambda \).

Secondly, even if the value is in the range of the transformation, it is possible to obtain very poor estimates. For example, a positive but very small value of the term might cause a poor prediction when Step 5 is applied for some \( \lambda \). In such a case MPL may be and BP will be affected seriously since for BP the predictions for all \( \lambda \in \Lambda \) are considered.

One possible way to overcome the first type of problem, is to apply a different transformation - e.g. folded power transformation - which approximates Box-Cox transformation.
Consider the family of transformations, from the interval $[0,1]$ onto $\mathbb{R}$, introduced by Tukey,

$$P(p) = \frac{\lambda_1 - (1-p)^{\lambda_1}}{\lambda_1}$$  \hspace{1cm} (4.2)

for $\lambda_1 \neq 0$. We can also define

$$P(p) = \log \frac{p}{1-p} \text{ for } \lambda_1 = 0$$  \hspace{1cm} (4.3)

Now consider a transformation from $\mathbb{R}^+$ onto $(0,1)$, which is defined by

$$Q(y) = p = \frac{y}{a+y}, \quad a > 0$$  \hspace{1cm} (4.4)

Define the Folded Power transformation $y_{FP}^{(\delta)}$ from $\mathbb{R}^+$ onto $\mathbb{R}$ $PoQ$ where $\lambda_1 = -\delta$ in (4.2) and (4.3), that is

$$y_{FP}^{(\delta)} = \begin{cases} \frac{(a+y)^{\delta} - (a+y)^\delta}{a} & \delta \neq 0 \\ \log y - \log a & \delta = 0 \end{cases}$$  \hspace{1cm} (4.5)

The 'new' family of transformations has the following features:

i) $y_{FP}^{(\delta)}$ is defined on $\mathbb{R}^+$ and is onto $\mathbb{R}$ since $y_{FP}^{(\delta)} = g = PoQ$, $P$ and $Q$ defined by (4.2), (4.3) and (4.4) are onto and the domain of $P$ is the range of $Q$. 
\[
\lim_{\delta \to 0} \delta \left( \frac{(a+y)^\delta}{a} - \frac{(a+y)^\delta}{y} \right) = \log y - \log a, \text{ hence it is continuous at } \delta = 0.
\]

iii) Box-Cox transformation family can be approximated by Folded Power Family with suitably chosen \( a \), i.e.

a) \( y_{FP}^{(\delta)} \) approximates to \( y^{(\lambda)} \) for \( a \to 0 \) where \( \lambda = \delta \)

b) \( y_{FP}^{(\delta)} \) approximates to \( y^{(\lambda)} \) for \( a \to \infty \) where \( \lambda = -\delta \)

c) \( y_{FP}^{(\delta)} = y^{(\lambda)} \) for \( a = 1 \) where \( \lambda = \delta = 0 \).

Therefore, if a set \( \lambda_A \) of elements \( (\delta, a) \) corresponding to values of \( \lambda \in A \) is chosen according to (iii) and a Folded Power transformation is applied, the corresponding approximate transformed Box-Cox variables can be obtained. The normalized transformation can also be considered similarly, such as

\[
Z_{FP}^{(\delta)} = \frac{y_{FP}^{(\delta)}}{J_{FP}^{1/N}}
\]

where the Jacobian of the transformation is

\[
J_{FP} = \prod_{i=1}^{N} (a+y_i)^{\delta-1} \left( \left( \frac{1}{a} \right)^{\delta} + a \left( \frac{1}{y_i} \right)^{\delta+1} \right)
\]

for \( \delta \neq 0 \) and \( J_{FP} = \frac{1}{y^N} \) for \( \delta = 0 \). Modifications to obtain MLP and BP should be made for Step 5, i.e.
Step 5'. Determine

a) FMLP by

\[ \hat{y} = \begin{cases} \frac{a_0 \hat{p}}{1-p} & (\delta_0, a_0) \neq (0,1) \\ \exp\left(\frac{1}{\hat{y}} Z(\delta_0)\right) & (\delta_0, a_0) = (0,1) \end{cases} \]

for \((\delta_0, a_0)\) which maximizes \(L_{\text{max}} (\delta_0, a_0)\).

b) FBP by

\[ \hat{y} = \sum_{(\delta, a) \in \Lambda} \frac{a \hat{p}}{1-p} p((\delta, a)|y) + \exp\left(\frac{1}{\hat{y}} Z_{FP}(0)\right) p((\delta, a)=(0,1)|y) \]

where \(\hat{p}\) is solved numerically from \(\hat{p}^{-\delta} - (1-\hat{p})^{-\delta} + \delta \frac{1}{\hat{y}} Z(\delta)_{FP} = 0\).

Applying a Folded Power transformation solves the problem of inversion, however, one can still obtain some poor estimates since the 'goodness' of the estimates is measured in the transformed observations metric. To overcome this problem, the particular set, \(\Lambda^*\), of values \(\lambda^*\) of the parameter, which cause poor estimates can be dropped from \(\Lambda\) and estimator can be based on the set \(\Lambda_1 = \Lambda - \Lambda^*\).
4.4 Simulation Study

4.4.1 Introduction

In this section PRPT is compared with some non-parametric techniques used in Chapter 3, namely MNIS, MDIS, RTIC, and parametric techniques of Sections 4.2 and 4.3 (BP/FBP, MLP/FMLP) and Correct Transformation Predictor (CTP) which is obtained by applying the Box-Cox transformations with the parameter $\lambda = \lambda_0$, that is the correct value of $\lambda$. CTP is considered as the optimal predictor. $d_3$ is chosen to represent PRPT since it is obtained by taking the median of the predictive probabilities hence the most 'non-parametric' in nature. Results of Chapter 3 also show that $d_3$ is a 'good' estimate even when the assumptions of PRPT are not fully satisfied.

The simulation study is based on the observations which are obtained by applying inverse Box-Cox transformations to a standard linear model so that all the assumptions of PRPT are satisfied i.e. the equation (4.1) provides the correct model. The Mean Squared Errors (MSE) of the predictors are used as the measure of 'goodness'. Comparisons are also made when inverse Box-Cox transformations are applied to a non-standard linear model where the distribution of the heteroscedastic errors is double exponential.

4.4.2 Details of Calculation

Similar to the experiment of Chapter 3, one independent observation at each of nine observation points $x_j = 0.1, ..., 0.9$ are generated by $Y_j = 4x_j + e_j$. The errors are generated by using:
a) NAG (1977) subroutine G05DDF for standard normal \( e_j \)'s;

b) technique described in subsection 3.4.2 for double exponential \( e_j \)'s where Step (3) is modified by using \( 0.316, \ldots, 0.949 \) instead of the constant so that variance is \( 2x_j \).

The observed data sets are assumed to consist of \((Y_j, x_j)\) \( j=1, \ldots, 9 \) where \( Y_j \)'s are derived from the \( Y_j' \)'s by four separate monotone increasing transformations. These are:

i) \( Y = Y' \)

ii) \( Y = (Y' + c)^2 \)

iii) \( Y = \exp(Y') \)

iv) \( Y = (Y' + c)^{4/3} \)

In the first experiment, the data sets \((Y_j, x_j)\) for each case give the standard linear model after a Box-Cox transformation for the correct value of \( \lambda \) which is 1 for (i), 0.5 for (ii), 0 for (iii) and 0.75 for (iv) where the errors are generated according to (a). In (ii) and (iv), constant \( a \) is added to \( Y' \) to obtain a positive value for \( Y' + c \). Case (i) is also taken as \( Y = Y' + c \) to be able to employ parametric techniques since the transformations are defined in \( \mathbb{R}^+ \). This change does not make any difference to the results of other estimators/predictors. Thus \( c \) is taken as 4 to ensure that \( Y' + c \) is positive with probability close to 1.

The cases (i), (ii) and (iv) are considered for the second experiment where errors are generated by (b). Case (iii) is omitted since the derived observations are too widely spread for any method to handle.
The set of trial values \( \lambda \), is chosen as \( \Lambda = \{-1,-0.5,0,0.5,1\} \) for BP and MLP since it is not always possible to predict the response value if \(|\lambda|>1\), as explained in Section 4.3. The larger set of values \( \Lambda' = \Lambda \cup \{-1.5,1.5\} \) is used for FBP and FMLP. \( a \) in (4.5) is taken as \( 10\hat{y} \) and \( \frac{\hat{y}}{10} \) for 'large' and 'small' values respectively where \( \hat{y} = \left( \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^{2/3} \).

The MSE's of the various estimators were calculated using the median of the distribution of \( Y \). This measure is also used by Carrell and Ruppert (1981) in their theory. Since the median of \( Y_j \) is \( 4x_j \), the median of \( Y_j \) is the relevant transformation of \( 4x_j \). Of course, this is in agreement with \( d_3 \), which uses the median of the rank predictive distribution and with MDIS. Any great differences arising from considering the median of \( Y \) rather than its mean should be reflected by great differences between the two isotonic estimators, one uses the median and the other the arithmetic mean as the 'average'.

4.5 Results

To compare the efficiency of the predictor \( d_3 \), ratios of MSE's of CTP to RTIC, MNIS, MDIS and \( d_3 \) - which do not use any information, and BP/FBP, MLP/FMLP those use some information about the correct transformation, are calculated for the first experiment and tabulated in Table 4.1. Results of Bayesian and Maximum Likelihood Predictors using Box-Cox and Folded Power Transformation are given under one heading since the ratios for FBP and FMLP are slightly smaller than ratios for BP and MLP. Table 4.1 reveals that \( d_3 \) is clearly superior
to MNIS, MDIS and RTIC. Two exceptions are observed for \( x = 0.1 \) and \( x = 0.2 \) for case (iii) where MNIS and/or MDIS perform better than \( d_3 \). However for \( x \geq 0.5 \) these techniques are far inferior to \( d_3 \). It should also be noted that \( d_3 \) is much better than the 'optimal' estimate CTP at \( x = 0.9 \) for case (iii).

On the other hand ratios of MSE's show that BP and MLP are better than \( d_3 \) almost uniformly, with the value of MSE(CTP)/MSE(BP or MLP) being much closer to 1 than MSE(CTP)/MSE(\( d_3 \)) mostly, there being some exceptions, Table 1b, \( x = 0.9 \), Table 1c, \( x = 0.8, 0.9 \) and Table 1.4, \( x = 0.9 \). The parametric estimates are quite close to optimal estimates and the loss in the efficiency is less for BP than for MLP as expected. However unreasonably large MSE's for both BP and MLP can occur at some observation points (Table 4.1c, \( x = 0.9 \)). If we examine a sample from case (iii) the generated observations are \((0.78, 1.13, 4.67, 1.07, 5.68, 68, 13.72, 64.96, 22.43, 88.50)^T\) at \( x = 0.1, ..., 0.9 \) and if we consider the predictions which are found by applying Box-Cox transformations when \( \lambda = -0.5 \), which is close to the true value. \( \lambda = 0 \), we obtain \((0.92, 1.21, 1.64, 2.36, 3.69, 6.53, 14.60, 57.36, 172722.06)^T\).

This happens as frequently as once in twenty repetitions and sometimes estimates are even poorer such as 1,124,845,056 for a value (155.17) at \( x = 0.9 \). BP is obviously affected more than MLP, since BP takes estimate for every \( \lambda \) into account. The ratio in parentheses for \( x = 0.9 \) in Table 4.1c is found by ignoring some values of the parameter which cause such poor estimates.
The effect of mis-specified errors in the generation of data is investigated in the second simulation study (case (b)). Ratios of MSE are tabulated in Table 4.2 as in Table 4.1. Results reveal that even if the errors are mis-specified $d_3$ is superior to the other non-parametric predictors and not as good as parametric ones. It should also be noted that the ratio of MSE of CTP to BP and/or MLP is either very close to or greater than one at almost all observation points.

4.6 Conclusion

It is concluded that the median estimator $d_3$ based on predictive probabilities is superior to the other non-parametric estimates since the ratio of MSE of the optimal estimate CTP to $d_3$ is greater than the ratio of MSE of CTP to any other non-parametric estimate at almost all observation points and efficiency is about 70 percent. However, the efficiency of the estimate RTIC is more than 80 percent of the efficiency of $d_3$ for all cases except case (iii), $x=0.9$ and therefore it is not really poor compared to $d_3$. It is not strange to obtain better results for $d_3$ than the 'optimal' estimate CTP at some observation points since the 'transformed best estimate' is not always 'best' in the transformed metric.

Parametric techniques provide better estimates than $d_3$ almost everywhere, even when the transformation required to obtain the standard linear model is mis-specified but the $\lambda$ required for the correct transformation is in the range of the trial set $\Lambda$. The values of $\lambda$ are restricted to a discrete parameter space. Carroll (1980) has used
a restricted parameter space in Box-Cox transformations and he found very little difference between predictions based on restricted and unrestricted parameter spaces.

Results of parametric techniques are contaminated by unreasonably large MSE's detected at $x = 0.9$ for case (iii). This shows how unreliable BP and MLP can be for some cases especially when the observations are increasingly rapidly.

Comparing Tables 4.1 and 4.2 shows that double exponential heteroscedastic errors do affect some non-parametric predictors. In fact poor performance of MNIS shows that it is affected more than CTP except at $x = 0.1$. MDIS is affected more than CTP but less than MNIS. RTIC is influenced less than CTP at most of the points whereas $d_3$ shows its best performance which proves that departures from normality of distribution of errors for $d_3$ is far less important than for CTP.

BP and MLP perform at their best except at the extreme points $x = 0.9$.

At many observation points in Table 4.2, ratio values of parametric techniques are greater than one and it shows that BP and MLP are better predictors than CTP when the assumptions of the standard linear model are not fully satisfied.
4.7 Illustration

As an illustration of the various techniques we consider some data given by Child (1973) in the investigation of the organization of companies. In Figure 4.1 we have plotted the number of affirmative replies to 90 questions against the logarithm (base 10) of the size (number of employees) of various companies. Theory suggests a monotone increasing relationship between the number of yes's and company size. It is not reasonable to model the number of yes's by a binomial variable since companies are more likely to answer in the affirmative to some questions than others and replies to questions are certainly related. From the plot of the observations, there appears to be increasing scatter as size increases. It is therefore difficult to specify a parametric model which would adequately describe the data.

We can however use some of the techniques described here to smooth the data and this has been done. From Figure 4.1, MDIS appears unsatisfactory; it is a step function and appears to be biased upwards for most of the size values. On the other hand BP, $d_3$ and RTIC give similar results. BP is a smooth curve being a mixture of smooth parametric curves, $d_3$ and RTIC tend to waggle about BP, with the greatest difference between $d_3$ and RTIC occurring for large values of $\log (\text{size})$, where RTIC gives smaller values. In the type of exercise with BP, $d_3$ and RTIC giving similar results one might have greater confidence in the parametric method BP being robust against model specification. Of course, one can give confidence intervals for predicted values using $d_3$ and BP which is not possible with RTIC nor MDIS.
TABLE 4.1: Comparison of RTIC, MNIS, MDIS, $d_3$, BP, MLP when a transformation is applied on response variables for observations $y'_j = 4x_j + e_j$ where errors have normal distribution with zero mean and variance one. Results are based on 600 repetitions.

a) $y = y' (y = y' + 4$ for BP and MLP)

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\text{var(CTP)}/\text{MSE(RTIC)}$</th>
<th>$\text{var(CTP)}/\text{MSE(MNIS)}$</th>
<th>$\text{var(CTP)}/\text{MSE(MDIS)}$</th>
<th>$\text{var(CTP)}/\text{MSE(d}_3\text{)}$</th>
<th>$\text{var(CTP)}/\text{MSE(BP)}$</th>
<th>$\text{var(CTP)}/\text{MSE(MLP)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.61</td>
<td>.58</td>
<td>.57</td>
<td>.75</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.2</td>
<td>.63</td>
<td>.57</td>
<td>.56</td>
<td>.76</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.3</td>
<td>.59</td>
<td>.48</td>
<td>.45</td>
<td>.68</td>
<td>.99</td>
<td>.97</td>
</tr>
<tr>
<td>.4</td>
<td>.57</td>
<td>.36</td>
<td>.34</td>
<td>.55</td>
<td>.94</td>
<td>.90</td>
</tr>
<tr>
<td>.5</td>
<td>.47</td>
<td>.31</td>
<td>.29</td>
<td>.53</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>.6</td>
<td>.60</td>
<td>.38</td>
<td>.37</td>
<td>.63</td>
<td>.93</td>
<td>.88</td>
</tr>
<tr>
<td>.7</td>
<td>.65</td>
<td>.50</td>
<td>.46</td>
<td>.71</td>
<td>.96</td>
<td>.93</td>
</tr>
<tr>
<td>.8</td>
<td>.66</td>
<td>.61</td>
<td>.57</td>
<td>.78</td>
<td>.95</td>
<td>.93</td>
</tr>
<tr>
<td>.9</td>
<td>.66</td>
<td>.59</td>
<td>.58</td>
<td>.82</td>
<td>.90</td>
<td>.88</td>
</tr>
</tbody>
</table>

b) $y = (y' + 4)^2$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\text{MSE(CT)}/\text{MSE(RTIC)}$</th>
<th>$\text{MSE(CT)}/\text{MSE(MNIS)}$</th>
<th>$\text{MSE(CT)}/\text{MSE(MDIS)}$</th>
<th>$\text{MSE(CT)}/\text{MSE(d}_3\text{)}$</th>
<th>$\text{MSE(CT)}/\text{MSE(BP)}$</th>
<th>$\text{MSE(CT)}/\text{MSE(MLF)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.58</td>
<td>.67</td>
<td>.68</td>
<td>.69</td>
<td>1</td>
<td>.94</td>
</tr>
<tr>
<td>.2</td>
<td>.58</td>
<td>.61</td>
<td>.61</td>
<td>.70</td>
<td>1.04</td>
<td>.96</td>
</tr>
<tr>
<td>.3</td>
<td>.53</td>
<td>.46</td>
<td>.44</td>
<td>.62</td>
<td>1.02</td>
<td>.93</td>
</tr>
<tr>
<td>.4</td>
<td>.50</td>
<td>.35</td>
<td>.33</td>
<td>.57</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td>.5</td>
<td>.47</td>
<td>.29</td>
<td>.28</td>
<td>.52</td>
<td>.90</td>
<td>.79</td>
</tr>
<tr>
<td>.6</td>
<td>.60</td>
<td>.36</td>
<td>.34</td>
<td>.63</td>
<td>.90</td>
<td>.83</td>
</tr>
<tr>
<td>.7</td>
<td>.67</td>
<td>.48</td>
<td>.44</td>
<td>.73</td>
<td>.95</td>
<td>.92</td>
</tr>
<tr>
<td>.8</td>
<td>.72</td>
<td>.61</td>
<td>.56</td>
<td>.84</td>
<td>.94</td>
<td>.96</td>
</tr>
<tr>
<td>.9</td>
<td>.71</td>
<td>.53</td>
<td>.52</td>
<td>.90</td>
<td>.72</td>
<td>.89</td>
</tr>
</tbody>
</table>
c) \[ y = e^{y'} \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>MSE(CTP)/MSE(RTIC)</th>
<th>MSE(CTP)/MSE(MNIS)</th>
<th>MSE(CTP)/MSE(MDIS)</th>
<th>MSE(CTP)/MSE(( d_3 ))</th>
<th>MSE(CTP)/MSE(BP)</th>
<th>MSE(CTP)/MSE(MLP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.38</td>
<td>.61</td>
<td>.76</td>
<td>.46</td>
<td>.80</td>
<td>.76</td>
</tr>
<tr>
<td>.2</td>
<td>.35</td>
<td>.39</td>
<td>.57</td>
<td>.46</td>
<td>.75</td>
<td>.73</td>
</tr>
<tr>
<td>.3</td>
<td>.32</td>
<td>.22</td>
<td>.29</td>
<td>.40</td>
<td>.70</td>
<td>.68</td>
</tr>
<tr>
<td>.4</td>
<td>.31</td>
<td>.18</td>
<td>.22</td>
<td>.34</td>
<td>.64</td>
<td>.61</td>
</tr>
<tr>
<td>.5</td>
<td>.38</td>
<td>.11</td>
<td>.16</td>
<td>.40</td>
<td>.64</td>
<td>.62</td>
</tr>
<tr>
<td>.6</td>
<td>.56</td>
<td>.14</td>
<td>.19</td>
<td>.57</td>
<td>.75</td>
<td>.74</td>
</tr>
<tr>
<td>.7</td>
<td>.71</td>
<td>.24</td>
<td>.24</td>
<td>.72</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>.8</td>
<td>.76</td>
<td>.24</td>
<td>.25</td>
<td>.93</td>
<td>.93</td>
<td>.96</td>
</tr>
<tr>
<td>.9</td>
<td>.53</td>
<td>.17</td>
<td>.17</td>
<td>1.14</td>
<td>.61x10^{-5}(.46)</td>
<td>.86x10^{-4}</td>
</tr>
</tbody>
</table>

d) \[ y = (y' + 4)^{4/3} \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>MSE(CTP)/MSE(RTIC)</th>
<th>MSE(CTP)/MSE(MNIS)</th>
<th>MSE(CTP)/MSE(MDIS)</th>
<th>MSE(CTP)/MSE(( d_3 ))</th>
<th>MSE(CTP)/MSE(BP)</th>
<th>MSE(CTP)/MSE(MLP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.62</td>
<td>.58</td>
<td>.59</td>
<td>.75</td>
<td>.95</td>
<td>.99</td>
</tr>
<tr>
<td>.2</td>
<td>.63</td>
<td>.60</td>
<td>.59</td>
<td>.75</td>
<td>.96</td>
<td>.98</td>
</tr>
<tr>
<td>.3</td>
<td>.57</td>
<td>.48</td>
<td>.45</td>
<td>.66</td>
<td>.99</td>
<td>.98</td>
</tr>
<tr>
<td>.4</td>
<td>.50</td>
<td>.36</td>
<td>.34</td>
<td>.53</td>
<td>.92</td>
<td>.75</td>
</tr>
<tr>
<td>.5</td>
<td>.47</td>
<td>.31</td>
<td>.29</td>
<td>.53</td>
<td>.89</td>
<td>.85</td>
</tr>
<tr>
<td>.6</td>
<td>.59</td>
<td>.37</td>
<td>.35</td>
<td>.62</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>.7</td>
<td>.65</td>
<td>.49</td>
<td>.45</td>
<td>.71</td>
<td>.95</td>
<td>.93</td>
</tr>
<tr>
<td>.8</td>
<td>.68</td>
<td>.62</td>
<td>.57</td>
<td>.81</td>
<td>.93</td>
<td>.95</td>
</tr>
<tr>
<td>.9</td>
<td>.70</td>
<td>.57</td>
<td>.56</td>
<td>.87</td>
<td>.83</td>
<td>.92</td>
</tr>
</tbody>
</table>
TABLE 4.2: Comparison of RTIC, MNIS, MDIS, $d_3$, BP, MLP when a transformation is applied on response variables for observations $y_j = 4x_j + e_j$ where errors have double exponential distribution with zero mean and variance $2x_j$. Results are based on 400 repetitions.

a) $y = y' \ (y = y' + 4 \text{ for BP and MLP})$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\text{MSE(CTP)}/\text{MSE(RTIC)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE(MNIS)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE(MDIS)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE}(d_3)$</th>
<th>$\text{MSE(CTP)}/\text{MSE(BP)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE(MLP)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.82</td>
<td>1.16</td>
<td>1.14</td>
<td>.85</td>
<td>1.21</td>
<td>1.15</td>
</tr>
<tr>
<td>.2</td>
<td>.72</td>
<td>.64</td>
<td>.64</td>
<td>.79</td>
<td>1.15</td>
<td>1.13</td>
</tr>
<tr>
<td>.3</td>
<td>.56</td>
<td>.38</td>
<td>.40</td>
<td>.64</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.4</td>
<td>.50</td>
<td>.29</td>
<td>.30</td>
<td>.54</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td>.5</td>
<td>.47</td>
<td>.30</td>
<td>.32</td>
<td>.60</td>
<td>.93</td>
<td>.93</td>
</tr>
<tr>
<td>.6</td>
<td>.60</td>
<td>.36</td>
<td>.38</td>
<td>.68</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>.7</td>
<td>.70</td>
<td>.48</td>
<td>.52</td>
<td>.82</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.8</td>
<td>.77</td>
<td>.61</td>
<td>.64</td>
<td>.93</td>
<td>.90</td>
<td>.93</td>
</tr>
<tr>
<td>.9</td>
<td>.61</td>
<td>.47</td>
<td>.48</td>
<td>.92</td>
<td>.76</td>
<td>.76</td>
</tr>
</tbody>
</table>

b) $y = (y' + 4)^2$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\text{MSE(CTP)}/\text{MSE(RTIC)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE(MNIS)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE(MDIS)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE}(d_3)$</th>
<th>$\text{MSE(CTP)}/\text{MSE(BP)}$</th>
<th>$\text{MSE(CTP)}/\text{MSE(MLP)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.75</td>
<td>1.38</td>
<td>1.36</td>
<td>.75</td>
<td>1.09</td>
<td>1.03</td>
</tr>
<tr>
<td>.2</td>
<td>.65</td>
<td>.66</td>
<td>.68</td>
<td>.74</td>
<td>1.13</td>
<td>1.08</td>
</tr>
<tr>
<td>.3</td>
<td>.53</td>
<td>.38</td>
<td>.40</td>
<td>.62</td>
<td>1.06</td>
<td>.98</td>
</tr>
<tr>
<td>.4</td>
<td>.50</td>
<td>.29</td>
<td>.30</td>
<td>.65</td>
<td>.99</td>
<td>.88</td>
</tr>
<tr>
<td>.5</td>
<td>.50</td>
<td>.30</td>
<td>.31</td>
<td>.63</td>
<td>1.00</td>
<td>.90</td>
</tr>
<tr>
<td>.6</td>
<td>.65</td>
<td>.35</td>
<td>.38</td>
<td>.75</td>
<td>1.08</td>
<td>1.01</td>
</tr>
<tr>
<td>.7</td>
<td>.77</td>
<td>.47</td>
<td>.51</td>
<td>.91</td>
<td>1.08</td>
<td>1.07</td>
</tr>
<tr>
<td>.8</td>
<td>.85</td>
<td>.55</td>
<td>.56</td>
<td>1.02</td>
<td>.94</td>
<td>.96</td>
</tr>
<tr>
<td>.9</td>
<td>.58</td>
<td>.37</td>
<td>.37</td>
<td>1.00</td>
<td>.57</td>
<td>.67</td>
</tr>
</tbody>
</table>
c) \( y = (y' + 4)^{4/3} \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \frac{\text{MSE(CTP)}}{\text{MSE(RTIC)}} )</th>
<th>( \frac{\text{MSE(CTP)}}{\text{MSE(MNIS)}} )</th>
<th>( \frac{\text{MSE(CTP)}}{\text{MSE(MDIS)}} )</th>
<th>( \frac{\text{MSE(CTP)}}{\text{MSE(d_{3})}} )</th>
<th>( \frac{\text{MSE(CTP)}}{\text{MSE(BP)}} )</th>
<th>( \frac{\text{MSE(CTP)}}{\text{MSE(MLP)}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.80</td>
<td>1.24</td>
<td>1.21</td>
<td>.82</td>
<td>1.17</td>
<td>1.11</td>
</tr>
<tr>
<td>.2</td>
<td>.70</td>
<td>.64</td>
<td>.66</td>
<td>.79</td>
<td>1.15</td>
<td>1.11</td>
</tr>
<tr>
<td>.3</td>
<td>.56</td>
<td>.38</td>
<td>.41</td>
<td>.64</td>
<td>1.03</td>
<td>1.01</td>
</tr>
<tr>
<td>.4</td>
<td>.50</td>
<td>.29</td>
<td>.30</td>
<td>.55</td>
<td>.94</td>
<td>.89</td>
</tr>
<tr>
<td>.5</td>
<td>.49</td>
<td>.30</td>
<td>.32</td>
<td>.60</td>
<td>.96</td>
<td>.90</td>
</tr>
<tr>
<td>.6</td>
<td>.62</td>
<td>.36</td>
<td>.38</td>
<td>.71</td>
<td>1.03</td>
<td>.98</td>
</tr>
<tr>
<td>.7</td>
<td>.73</td>
<td>.48</td>
<td>.52</td>
<td>.85</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>.8</td>
<td>.80</td>
<td>.60</td>
<td>.62</td>
<td>.95</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>.9</td>
<td>.61</td>
<td>.44</td>
<td>.44</td>
<td>.95</td>
<td>.73</td>
<td>.73</td>
</tr>
</tbody>
</table>
FIGURE 4.1: Number of affirmative replies to 90 questions against size of the companies

Key: Step function MDIS, Solid Curve: BP, Dashes: RTIC, Dots and dashes = $d_3$
CHAPTER 5

DISCRIMINANT ANALYSIS BY USING RANKS AND A CASE STUDY

5.1 Introduction

The objective of the classical discriminant analysis problem is to assess to which of the possible populations an individual \( Z^* \) belongs, given the \( k \)-variate observation vector \( z^* \) value of \( Z^* \), the population parameters and the prior probabilities \( q_i \) that \( Z^* \) belongs to population \( \Pi_i \). In this Chapter the discrimination problem where individuals belong to one of two populations \( \Pi_1 \) and \( \Pi_2 \) is considered where the associated costs are taken to be equal.

The probability of being a member of population \( \Pi_i \) can be computed from Bayes formula as:

\[
P_i = \frac{q_i \ p(z^*|\theta_i)}{\sum_{i=1}^{2} q_i \ p(z^*|\theta_i)} \quad i = 1, 2
\]

(5.1)

where \( p(z^*|\theta_i) \) is the \( i \)-th population density function of \( z^* \) given the vectors of population parameters \( \theta_i \). In practice \( \theta_i \) is not known but it is possible to have some information about it by using the data from the set of observations of confirmed cases. The rule which minimizes the probability of misclassification is to classify \( Z^* \) into \( \Pi_1 \) if \( P_1 > P_2 \).
Inference for $P_i$ has been either 'estimative' or 'predictive'.

An estimative estimate of $P_i$ can be obtained by replacing $p(z^*|\theta_i)$ in (5.1) by $\hat{p}(z^*|\hat{\theta}_i)$, the assumed $i$-th population density of $Z^*$ where $\hat{\theta}_i$ is the vector of an estimate of the population parameters.

A predictive estimate of $P_i$ can be obtained by replacing $p(z^*|\theta_i)$ in (5.1) by predictive densities

$$p(z^*|\text{data}) = \int p(z^*|\theta_i) p(\theta_i|\text{data}) d\theta_i$$

(5.2)

where $p(\theta_i|\text{data})$ is the posterior distribution of $\theta_i$ based on a prior $p(\theta_i)$ and data.

An alternative estimate of $P_i$ can be obtained by treating $\Pr(Z^* \in \Pi_i|z^*,\theta_i)$ as a random variable, a function of the parameter $\theta_i$ and considering its distribution, using the posterior distribution of $\theta_i$ so that some estimates of $P_i$ and confidence intervals can be found.

Aitchison et al (1977) points out the differences between estimative and predictive methods and their importance in medical work. They also conducted a simulation study and it was concluded that the predictive method is superior to the estimative method. Discrimination techniques are widely used in medical work, mostly to diagnose the type of illness, but the aim of statistical diagnosis is to determine the probabilities $P_i$ of possible types rather than choosing 'the possible type'.
Although much has been published regarding the parametric techniques used in discriminant analysis (Anderson (1958), Aitchison and Dunsmore (1980)), a few statisticians have employed non-parametric techniques to determine an estimate of $P_i$ (Randles et al. (1978), Iman and Conover (1980)). In this Chapter some semi-parametric techniques, based on the ranks of the observations, are proposed for the two population discrimination problem. Since the ordering of multivariate observations is not well-defined, a discriminant function $D(\cdot): \mathbb{R}^k \rightarrow \mathbb{R}$ is used to project $k$-variate observations into the real line. The predictive distribution of $Y^* = D(Z^*)$ given the rank of $Y^*$ in the projected combined sample is used in two different ways to find the predictive estimate of the posterior probability of $Y^*$ being a member of the projected populations $\Pi_i$. This probability is taken as an estimate of the probability of $Z^*$ being a member of $\Pi_i$.

5.2 Approximation to the Predictive Estimate of $P_i$ by Using Ranks

One way to obtain an approximation to the predictive estimate of $P_i$ by using ranks of the observations is to replace $p(z^* | \theta_i)$ in equation (5.1) by an approximation $p(\text{rank of } z^* | \text{ranks of data})$ to the predictive density $p(z^* | \text{data})$.

Let $Z_{11}, \ldots, Z_{1n_1}$ and $Z_{21}, \ldots, Z_{2n_2}$ be the confirmed cases from $k$-variate populations $\Pi_1$ and $\Pi_2$ respectively. It is not possible to find the ranks of the multivariate observations unless corresponding univariate observations are obtained by using a function, therefore a well-known
discriminant function $D(.) : \mathbb{R}^{k} \rightarrow \mathbb{R}$ is chosen to apply, to obtain the projection into the real line in the correct order, i.e. $Y_{ij} = D(Z_{ij})$ $i = 1, 2, j = 1, \ldots, n_i$. Assume that after an unknown monotone transformation, the $Y$'s have an approximate normal distribution with different location parameters $\beta x_j$, where $\beta$ is the unknown parameter, variance one and are approximately independent.

Let the dummy variable $x_{ij}$ corresponding to $Y_{ij}$ be

$$x_{ij} = \begin{cases} 0 & \text{if } i = 1 \\ 1 & \text{if } i = 2 \end{cases}$$

then, after an arbitrary (unknown) monotone transformation, it is assumed that $Y_{ij_1}$'s and $Y_{2j_2}$'s are approximately distributed as $N(0,1)$ and $N(\beta,1)$ respectively.

Let $Y^*$ be the projection $D(Z^*)$ of a new case $Z^*$ to be classified and $\Pi^*_1$ be the projected form of population $\Pi_1$. Consider $r(j)$, the vector of ranks of $Y^*_1, \ldots, Y^*_n$, $Y^*_2, \ldots, Y^*_2n_2$, where the rank of $Y^*$, $r^* = j$.

The $i^{th}$ population predictive density of rank of $Y^*$ in the combined sample from $\Pi_1^*$ and $\Pi_2$ given the data $r$ (i.e. given ranks of $Y^*_1, \ldots, Y^*_2n_2$ amongst themselves) is given by

$$p_i(r(j)|r) = \int p_i(r(j)|\beta) p(\beta|\pi) \, d\beta$$

(5.3)

For the two sample problem, the approximation (2.4) to the rank likelihood $p_i(r(j)|\beta)$ in equation (5.3) is simplified to be

$$\pi_i(r(j)|\beta) = \frac{1}{M(ij)} \exp\left\{ \frac{1}{2} \frac{m^2(ij)}{M(ij)} \right\} \exp\left\{ -\frac{1}{2} \frac{(\beta - m(ij))^2}{M(ij)} \right\}$$

(5.4)
where $N = n_1 + n_2 + 1$, $m_{ij} = \sum_{k=1}^{N} (x^*(i))^k (a(j))^k M(ij)$,

\[
M(ij) = \left( \sum_{k=1}^{N} (x^*(i))^2 - \sum_{k=1}^{N} \sum_{\ell=1}^{N} (x^*(i))^k (x^*(i))^\ell (A(j))^\ell \right)^{-1}
\]

$(a(j))^k = (\varepsilon)(r(j))^k (A(j))^k, k, \ell = 1, \ldots, N$

\[
x^* = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ x^*(i) \end{bmatrix}
\]

with $x^*(i) = \begin{cases} 0 & \text{if } i = 1 \\ 1 & \text{if } i = 2 \end{cases}$ for $Y^* \in \Pi_1$, $i = 1, 2$.

$\xi$ is the vector of expected values and $\Sigma$ is the covariance matrix of normal order statistics of a sample of size $N$. The predictive probability of $r^* = j$ amongst $Y_{11}, \ldots, Y_{2n_2}$, $Y^*$ given $Y^* \in \Pi_1$ and $r$ (i.e. equation 5.3) is

\[
\rho_{ij} = \frac{\exp \left[ \frac{1}{2} \frac{m^2(ij)}{M(ij)} \right] \sqrt{M(ij)}}{\sum_{j=1}^{N} \exp \left[ \frac{1}{2} \frac{m^2(ij)}{M(ij)} \right] \sqrt{M(ij)}}
\]
if the approximation \( \pi_i(r(j)|\beta) \) (equation 5.4) is used for \( p_i(r(j)|\beta) \) and the prior distribution of \( \beta \) is taken to be locally uniform (Acar and Pettitt, 1984).

Therefore a predictive estimate of \( P_i \) is,

\[
\hat{P}_i = \Pr(Y \in \Pi_i | \mathbf{r}_* = j, \mathbf{r})
\]

\[
= \frac{q_i \Pr(r_0 = j | Y \in \Pi_i, \mathbf{r})}{\Pr(r_0 = j | \mathbf{r})}
\]

\[
= \frac{q_i \rho_{ij}}{\sum_{i=1}^{2} q_i \rho_{ij}} \quad (5.6)
\]

This technique is illustrated in Section 5.5.

5.3 Approximation to the Predictive Estimate of \( P_i \) by Using its Distribution

An alternative way to obtain an estimate of \( P_i \), by using ranks of the observations, is to find the approximate distribution of \( P_i \), where \( p_i \) is taken as a function of random variables.

Suppose that the approximation \( \pi_i(r(j)|\beta) \) to the rank likelihood be treated as a random variable and denoted by \( \pi_i \) for simplicity. Then \( P_i \) can be approximated by
\[ P_i = \text{pr}(Y^* \in \Pi_i | r^* = j, \beta) \]

\[ = \frac{q_i \text{pr}(r^* = j | Y^* \in \Pi_i, \beta)}{\text{pr}(r^* = j | \beta)} \]

\[ = \frac{q_i \pi_i}{q_1 \pi_1 + q_2 \pi_2} \quad (5.7) \]

If \( P_i \) is treated as a random variable, a function of \( \beta \), then estimates of \( P_i \) and confidence interval for \( P_i \) can be obtained from the distribution of \( P_i \). It is convenient to obtain the distribution of \( W = \log \frac{\pi_1}{\pi_2} \) which can be used to determine the distribution of \( P_i \) since it is possible to derive \( P_i \) by a monotone transformation of \( W \)

\[ i.e. \quad P_i = g(W) = \left\{ 1 + \left( \frac{q_1}{q_2} e^W \right)^{(-1)^i} \right\}^{-1} \quad (5.8) \]

A rank predictive approximation to \( W \) can be obtained by (5.4) as

\[ W = \log \pi_1 - \log \pi_2 = \frac{1}{2} \left[ \frac{m^2(1j)}{M(1j)} - \frac{m^2(2j)}{M(2j)} + \frac{(\beta-m(2j))^2}{M(2j)} - \frac{(\beta-m(1j))^2}{M(1j)} \right] \quad (5.9) \]

If the algebraic simplification,
\[
\frac{(x-a)^2}{A} - \frac{(x-b)^2}{B} = \frac{(x-c)^2}{C} + \frac{(a-b)^2}{A-B} \text{ is used}
\]

where \( c = -\frac{aB+bA}{A-B} \) and \( C = -\frac{AB}{A-B} \), (5.9) can be written as

\[
W = \frac{1}{2M^*} [(b-m^*)^2 - m^*2] = \frac{1}{2M^*} (b^2 - 2b m^*)
\]

(5.10)

where \( m^* = \frac{m(1j)M(2j) - m(2j)M(1j)}{M(2j) - M(1j)} \) and \( M^* = \frac{M(1j) M(2j)}{M(2j) - M(1j)} \).

If \( \beta \) has uniform vague prior, then an approximate posterior distribution of \( \beta \) given data \( r \) is normal with mean \( m \) and variance \( M \), where \( m \) and \( M \) are similar to \( m(1j) \) and \( M(1j) \), equation (5.4), but they are obtained by taking the \( n_1+n_2 \) dimensional vectors \( X \) as \( (n_1 \ 0's \ followed \ by \ n_2 \ 1's) \) and \( r \), as the ranks of \( Y_{11}, \ldots, Y_{2n_2} \) amongst themselves rather than the \( n_1+n_2+1 \) dimensional vectors \( X^*(i), r(j) \), ranks of \( Y_{11}, \ldots, Y_{2n_2}, Y^* \) where \( r^* = j \). Therefore, the probability distribution function of \( W \), is approximated by

\[
F(W) = \Phi \left( \frac{(2M^*^2 + m^*2)^{\frac{1}{2}} + m^* - m}{\sqrt{M}} \right) - \Phi \left( \frac{-(2M^*^2 + m^*2)^{\frac{1}{2}} + m^* - m}{\sqrt{M}} \right)
\]

(5.11)

for \( M^* > 0 \), and
\[ F_W(\omega) = 1 + \frac{\left(\frac{m^* \omega + m^* z}{\sqrt{M}}\right)^{\frac{1}{2}}}{M} - \frac{\left(\frac{2M^* \omega + m^* z}{\sqrt{M}}\right)^{\frac{1}{2}}}{M} \] (5.12)

for \( M^* < 0 \) where \( \Phi(\cdot) \) is the standard normal distribution function.

Differentiating (5.11) and (5.12) gives

\[
f_W(\omega) = \frac{M^*}{[2\pi M(2M^* \omega + m^* z)]^{\frac{1}{2}}} \exp \left\{ - \frac{1}{2} \left( \frac{(m^* - m)^2 + (2M^* \omega + m^* z)}{M} \right) \right\}
\]

\[
\times 2 \cosh \left\{ \frac{(m^* - m)(2M^* \omega + m^* z)^{\frac{1}{2}}}{M} \right\} \times \text{sgn}(M^*)
\] (5.13)

where \( \text{sgn}(M^*) = \begin{cases} 1 & \text{if } M^* > 0 \\ -1 & \text{if } M^* < 0 \end{cases} \)

Let \( \bar{P}_i \) denote the rank predictive approximation of random variable \( P_i \), equation (5.7), then an estimate of the probability of being a member of population \( \pi_i E_{\beta|\bar{P}_i}(\bar{P}_i) \), can be approximated by a Taylor series approximation to \( E_{\beta|\bar{P}_i}(g(W)) \). Taking expectation of the expansion of \( g(W) \) by Taylor series about \( g(E_{\beta|\bar{P}_i}(W)) \) up to second order gives

\[
E_{\beta|\bar{P}_i}(g(W)) = g(E_{\beta|\bar{P}_i}(W)) + \frac{1}{2} g''(E_{\beta|\bar{P}_i}(W)) \sigma^2_{\beta|\bar{P}_i}(W)
\] (5.14)

The expected value and variance of \( W \) in (5.14) can be easily obtained by finding

\[
E_{\beta|\bar{P}_i}(W^k) = \int \frac{1}{(2M^*)^k} (\beta^2 - 2\beta m^*)^k \frac{1}{\sqrt{2\pi M}} e^{-\frac{(\beta - M)^2}{2M}} d\beta
\] (5.15)
for \( k = 1 \) and \( 2 \) since the approximate posterior distribution of \( \beta \) is normal with mean \( m \) and variance \( M \). Therefore, (5.14) can be written as

\[
E_{\beta|\tau} (P_1) = \frac{1}{1 + k_1e^0} \{1 + \frac{1}{2} \frac{k_1e^0(k_1e^0 - 1)}{(1 + k_1e^0)^2} S\} \tag{5.16}
\]

where \( k_1 = \frac{q_2}{q_1} \), \( k_2 = \frac{1}{k_1} \)

\[
Q = (-1)^j \{ (-1)^{j/2} (\mu_2^j - 2m^* \mu_1^j) \}
\]

\[
S = \frac{1}{4m^{*2}} \{ \mu_4^j - \mu_2^j \mu_2^j + 4m^* (\mu_2^j \mu_1^j - \mu_3^j) + 4m^{*2} (\mu_2^j - \mu_1^{j/2}) \}
\]

\( Q \) and \( S \) are \((-1)^j \) x mean and variance of \( W \) respectively, where \( \mu_j \) stands for, \( \mu_j(\beta) \), the raw moments of \( \beta \). The central moments of \( \beta \) are given by

\[
\mu_j(\beta) = E((\beta - m)^j) = \begin{cases} 
0 & \text{if } j \text{ is odd} \\
\frac{j!}{(j/2)!} M^{j/2} & \text{if } j \text{ is even}
\end{cases}
\tag{5.17}
\]

and the relation

\[
\mu_k^j(\beta) = E(\beta^k) = \sum_{j=0}^{\infty} \binom{k}{j} \mu_j(\beta)(\mu_j^*)^{k-j} \tag{5.18}
\]
for \( k = 2,3,4 \) where \( \mu_k'(\beta) = 1 \) and \( \mu_k'(\beta) = m \) gives

\[
\begin{align*}
\mu'_2 &= m^2 + M \\
\mu'_3 &= m(m^2 + 3M) \\
\mu'_4 &= m^4 + 6m^2M + 3M^2
\end{align*}
\] (5.19)

to insert in (5.16).

The distribution function (5.11) or (5.12) for \( W \) can be used to obtain median and other percentiles for \( W \). Therefore, another estimate, the median of \( P_i \) is easier to find because

\[
\text{med}(\tilde{P}_i) = \text{med}(g(W)) = g(\text{med}(W))
\] (5.20)

since \( \tilde{P}_i \) is a monotone function of \( W \). Similarly, confidence intervals of \( \tilde{P}_i \) can be found by using percentiles for \( W \).

5.4 Use of Pearson Curves

The distribution function of \( W \) is obtained in Section 5.3 equations (5.11), (5.12)), however a very practical way to find estimates of \( \tilde{P}_i \) is to approximate the distribution function of \( W \) by Pearson Curves. It has been found that Pearson distributions are sometimes useful in approximating to theoretical distributions from their known first four moments (Kendall and Stuart, 1963). The mean of \( W \) is given in (5.16) as \( Q \). The second, third and fourth moments of \( W \) are
\[ E_{\beta}r(W^2) = \frac{1}{(2M*)^2} \left( \mu'_4 - 4m^4 \mu'_3 + 4m^2 \mu^2_2 \right) \quad (5.21) \]

\[ E_{\beta}r(W^3) = \frac{1}{(2M*)^3} \left( \mu'_6 - 6m^5 \mu'_5 + 12m^2 \mu^2_4 - 8m^3 \mu^3_3 \right) \quad (5.22) \]

\[ E_{\beta}r(W^4) = \frac{1}{(2M*)^4} \left( \mu'_8 - 8m^7 \mu'_7 + 24m^2 \mu^2_6 - 32m^3 \mu^3_5 + 16m^4 \mu^4_4 \right) \quad (5.23) \]

where \( \mu'_j \) is the \( j \)th raw moment of \( \beta \). Addition to those given in (5.19)

\[ \mu'_5 = m(m^4 + 10m^2M + 15M^2) \]

\[ \mu'_6 = m^6 + 15m^4M + 45m^2M^2 + 15M^3 \]

\[ \mu'_7 = m(m^6 + 21m^4M + 105m^2M^2 + 105M^3) \]

\[ \mu'_8 = m^8 + 28m^6M + 210m^4M^2 + 420m^2M^3 + 105M^4 \quad (5.24) \]

are obtained to substitute in (5.21), (5.22) and (5.23) where \( m \) and \( M \) are the mean and variance of \( \beta \) respectively, \( m^* \) and \( M^* \) are as given in (5.10).

The first four moments of \( W \) can therefore be used to calculate skewness, kurtosis of the distribution of \( W \) and criterion to determine the type to which the distribution belongs. A computer program written in FORTRAN by Stephens & David (1983) which provides the median and some other percentage
points of the appropriate distribution given the first four moments
is also used in this work.

5.5 Illustrations

A sample of size 12 is chosen in the following illustrations, to summa-
ritize how the techniques, introduced in Sections 5.2 and 5.3 to predict
$P_i$ by using ranks, are applied.

Let $Z_{11}, Z_{12}, \ldots, Z_{16}$ and $Z_{21}, Z_{22}, \ldots, Z_{26}$ be the $k$-variate observations
from $\Pi_1$ and $\Pi_2$ respectively and let $Y_{11}, \ldots, Y_{16}$ and $Y_{21}, \ldots, Y_{26}$ be the
observations, projected into real line by a discriminant function $D(\cdot)$,
which forms $\Pi'_1$ and $\Pi'_2$. Suppose that the ranks of the $Y_{1j}$, $j=1, \ldots, 6$
amongst the combined sample of $Y_{1j}$'s and $Y_{2j}$'s are 1, 2, 4, 5, 8, 10. Let
$r$ be the vector of ranks of the $Y_{1j}$'s followed by those of $Y_{2j}$'s.

It is assumed that after a monotone transformation the $Y_{1j}$'s and
$Y_{2j}$'s are approximately distributed as $N(0,1)$ and $N(\beta,1)$ where $\beta$
estimated by $m$, based on the ranks of 12 observations.

\[
m = \left( \sum_{k=7}^{12} (\xi_k) r_k \right) \left( 6 - \sum_{k=7}^{12} \sum_{i=7}^{12} (\xi) r_k r_i \right) \text{where } \xi \text{ is the vector}
\]
of expected values and $\Xi$ is the covariance matrix of normal order
statistics of a sample of size 12.
5.5.1 Illustration I

In this section the technique to estimate $P_1$ introduced in Section 5.2, which replaces $p(z^*|\theta_i)$ of equation (5.1) by rank predictive approximation of $p(z^*|\text{data})$, is illustrated.

To consider the problem of which population the new observation $Z^*$ belongs to, first suppose that the projected new observation $Y^* = D(Z^*)$ belongs to $\Pi_1^i$, where the rank of $Y^*$ is known in the combined sample. The predictive probability of $r^*=j$ amongst $Y_{11}, \ldots, Y_{26}, Y^*$ given $Y^* \in \Pi_1^i$ and data $r$ (i.e. $p(r(j)|Y^* \in \Pi_1^i, r)$) is approximated by $p_{ij}$, equation (5.5) where

\[
(r)_{k} = \begin{cases} 
(r)_{k} & \text{if } j > r_k \\
(r)_{k+1} & \text{if } j \leq r_k 
\end{cases}
\]

$k = 1, \ldots, 12$ and $(r(j))_{13} = j$, e.g. for $Y^*$ having a value between 8th and 9th ranked observations we obtain $r(9) = (1,2,4,5,8,11,3,6,7,10,12,13,9)$. The required values of $m(lj)$ and $M(lj)$ are found as in (5.4) by using $X^*(1) = (0,0,0,0,0,0,1,1,1,1,1,1,0)^T$ and rearranged $\xi(a)$ and $\Xi(A)$ based on $r(j)$, so $p_{ij}$, $j=1, \ldots, 13$ are obtained. Similarly $Y^* \in \Pi_2$ is assumed and the procedure is repeated by using $X^*(2) = (0,0,0,0,0,1,1,1,1,1,1,1,1)^T$ to find $p_{2j}$, $j=1, \ldots, 13$. $p_{ij}$, $i=1,2$, $j=1, \ldots, 13$ are listed in Table 5.1. Therefore the predictive probability of $Y^* \in \Pi_1^i (Z^* \in \Pi_1)$ can be found by substituting $p$-values in equation (5.6). In the case of $Y^*$ having a value between the 8th and the 9th in the combined ranked observations, we obtained $\hat{P}_1 = 0.466$, if the prior probabilities are given as $q_i = 0.5$, $i = 1,2$. 
5.5.2 Illustration II

In Section 5.3 estimates of $P_i$ are obtained by treating $P_i$ as a random variable and observing its distribution given the rank of a new case $Y^*$. In this section the example used to demonstrate in Illustration I is considered again.

When the ranks of $Y_{lj}$, $j=1,\ldots,6$; members of population $\Pi_1$ are 1,2,4,5,8,10 amongst the combined sample of 12 observations, $m$ and $M$ based on ranks of the observations are found as 0.930 and 0.373 respectively. Furthermore $m(i,j)$ and $M(ij)$, $i=1,2$, $j=1,\ldots,13$ are found by using $r(j)$, $X^*(1)$ and $X^*(2)$ as in previous subsection and hence $m^*'$s and $M^*'$s are found as in equation (5.10) so that percentiles for $W = \log \frac{\pi_1}{\pi_2}$ where $\pi_1$ is the probability of $r^*=j$ amongst $Y_{ll},\ldots,Y_{26}$, $Y^* \in \Pi_1^i$, $i=1,2$ and $\beta$ can be obtained either by using approximate distribution function equations (5.11),(5.12), or by substituting the first four moments, $-Q$ in equation (5.16), (5.21) - (5.23) to fit a Pearson Curve. Therefore the percentiles for (say) $\tilde{P}_1$ can be obtained by the relation

$$P_1 = \frac{1}{1+e^{-W}} \text{ for } q_1 = q_2.$$ 

In Table (5.2) the median of $W$, $\tilde{P}_1$ and, mean of $P_1$ found by equation (5.16), are listed for 13 possible ranks of a new case $Y^*$. If, for example, the new case has rank $r^*=9$, the median of $W$ found as -0.390 gives $\text{med}(\tilde{P}_1) = 0.404$, where mean (of $\tilde{P}_1$) is found as 0.402 by using $k_1=1$, mean of $W = -0.403 (-Q)$ and variance of $W = 0.076(5)$. The mean and median of $\tilde{P}_1$ are very close to each other since the distribution of $\tilde{P}_1$ is nearly symmetric and the variance of $W$ is very small. The percentiles of $W$ and $\tilde{P}_1$ are also given in Table 5.2 for $r^*=9$. The
first four moments of $W$, -0.403, 0.238, -0.163 and 0.127 respectively are obtained and used to find percentiles of the fitted approximate distribution. Therefore 80% confidence interval is $(0.318, 0.485)$.

5.6 Difference Between $\hat{P}_i$ and Mean of $\tilde{P}_i$

The rank predictive estimate $\hat{P}_i$, equation (5.6), of $P_i$ is clearly different than $E_{\beta|_{\tilde{r}}}(\tilde{P}_i)$ equation (5.16) since $\hat{P}_i$ is obtained by

$$\frac{q_1 E_{\beta|_{\tilde{r}}}(\pi_1(\beta))}{q_1 E_{\beta|_{\tilde{r}}}(\pi_1(\beta)) + q_2 E_{\beta|_{\tilde{r}}}(\pi_2(\beta))}$$

whereas

$$E_{\beta|_{\tilde{r}}}(\tilde{P}_i) = E_{\beta|_{\tilde{r}}} \frac{q_1 \pi_1(\beta)}{q_1 \pi_1(\beta) + q_2 \pi_2(\beta)}$$

Rigby (1982) considered an approximation to the distribution of logarithm of the likelihood ratio for one particular choice of the $i^{th}$ population density of $Y^*$ given population parameters and showed the difference between a predictive estimate of $P_i$ and the mean of $P_i$.

The difference between rank predictive estimate $\hat{P}_i$ and the mean of $\tilde{P}_i$ can be shown similarly as follows:
\[
\hat{p}_i = \frac{q_ip(r^*=j|Y^* \in \Pi_i, r)}{p(r^*=j|r)}
\]

\[
= \int \frac{q_ip(r^*=j|Y^* \in \Pi_i, \beta)p(\beta|r) d\beta}{p(r^*=j|r)}
\]

\[
= \int \frac{(q_ip(r^*=j|Y^* \in \Pi_i, \beta)p(r^*=j, \beta|r) d\beta}{p(r^*=j|r) p(r^*=j|\beta, r)}
\]

\[
= \int \hat{p}_i p(\beta|r^*=j, r) d\beta \quad (5.25)
\]

whereas
\[
E_{\beta|r}(\hat{p}_i) = \int \hat{p}_i p(\beta|r) d\beta \quad (5.26)
\]

5.7 A Similarity of Rank and Parametric Techniques

Discriminant analysis using ranks resembles the classical discriminant analysis in the sense that, when the observations belong to the region where two populations overlap, i.e. the most uncertain cases to classify to one of the populations, then the knowledge about the location parameter does not contribute to the solution of the problem as it ought to do. We give a simple example.

Let \( \Pi_1 \) and \( \Pi_2 \) be two univariate normal populations with mean \(-\delta, +\delta\) and variance 1. The classical discriminant function, the logarithm of
the ratio of density functions of the populations is \( W = \log \frac{f_1}{f_2} = -2y^*\delta \) for a new case \( Y^* \), that is, the decision is made by using a multiple of \( \delta \). If \( y^* \) is in a small neighbourhood of zero i.e. \( y^* \) is one of the most uncertain cases, a very small multiple of \( \delta \) is taken. For \( y^* = 0 \), no matter how much we know about \( \delta \), the test cannot indicate to which population the individual belongs.

A very similar feature can be observed in the ranks technique. Note that the logarithm of the approximated rank likelihood ratio \( W \) can be written as:

\[
\log \frac{\pi_1}{\pi_2} = - \frac{1}{2} \beta^2 \left( \frac{1}{M(1j)} - \frac{1}{M(2j)} \right) + \beta \left( \frac{m(1j)}{M(1j)} - \frac{m(2j)}{M(2j)} \right)
\]

where \( j \) is the rank of \( Y^* \) amongst the combined observations, and \( \beta = 2\delta \). If \( Y^* \) is larger than most of the members of \( \Pi_1 \) and smaller than most of the members of \( \Pi_2 \) then \( m(1j) \to m(2j) \) and \( M(1j) \to M(2j) \). Therefore the posterior distribution of \( W \) involves \( \beta \) amplified by a small amount and so it has little information.

5.8 Error Rate Estimations

The error rate \( e_1(e_2) \) associated with a sample discrimination rule is the probability of allocating the new case \( Z^* \) into population \( \Pi_2(\Pi_1) \) where in fact \( Z^* \) is a member of population \( \Pi_1(\Pi_2) \). In practice widely used 'estimative' techniques provide an answer to which population the case \( Z^* \) belongs to. In this case, the obvious estimate of the error
rate is

\[ \hat{e}_i = \frac{\text{No. of } Z_{ij}'s \text{ which classified as member of } \Pi_{3-i}, \ i=1,2, \ j=1,...,n_i}{n_i} \]

(5.27)

where it is given that \( Z_{i1},...,Z_{in_i} \) are members of population \( \Pi_i \).

\( \hat{e}_i \) tends to underestimate the true error rates.

Efron (1979) suggested the 'bootstrap' estimate of the bias correction \( E(e_i - \hat{e}_i) \). In summary, he obtained the new bootstrap random samples, associated discriminant functions and the difference \( d_i \) of error rates (5.26) when the function is applied to the original sample and to new bootstrap samples. The mean of differences is obtained after repeated realizations and taken as bias correction. He also showed that bootstrap method outperforms the commonly used 'leave-one-out' technique.

Error rates can also be estimated by using the approximation to the predictive probability of rank of the observation \( Y^* \), \( r^*=j \), given which population \( Y^* \) belongs to. Obviously for k-variate observations, \( Z_{11},...,Z_{1n_1} \) and \( Z_{21},...,Z_{2n_2} \) of population \( \Pi_1 \) and \( \Pi_2 \) respectively need to be projected into real line by using a function \( \text{D}(.) : \mathbb{R}^k \rightarrow \mathbb{R} \) to obtain \( Y_{11},...,Y_{1n_1} \) and \( Y_{21},...,Y_{2n_2} \) of populations \( \Pi_1' \) and \( \Pi_2' \) respectively and consequently to enable us to apply rank techniques. Therefore the first error rate \( e_1 \) can be written as
Let $C$ be the set of ranks of the observations which can be classified as a member of $\pi_2$ according to the technique proposed in Section 5.3, in other words

$$C = \{j: \hat{P}_2 > \hat{P}_1 \text{ for the observation } Y^* \text{ whose rank is } j\}$$

i.e.

$$C = \{j: q_2 p_{2j} > q_1 p_{1j}\}$$

where $p_{ij}$ is the probability of rank of $Y^*$, $r^*=j$ amongst $Y_{11}, \ldots, Y_{2n_2}, Y^*$ given $Y^* \in \pi_i$. Hence $e_1$ can be estimated by the probability of $r^*$ being equal to $j$ for all possible $j \in C$.

i.e.

$$\hat{e}_1 = \sum_{j \in C} \text{pr}(r^*=j | Y^* \in \pi_1) = \sum_{j \in C} p_{1j} \tag{5.30}$$

Similarly $e_2$ can be defined as $\text{pr}(Y^* \in \pi_1' | Y^* \in \pi_2')$ and estimated by

$$\hat{e}_2 = \sum_{j \notin C} \text{pr}(r^*=j | Y^* \in \pi_2) = \sum_{j \notin C} p_{2j} = 1 - \sum_{j \in C} p_{2j} \tag{5.31}$$

since $\sum_{j=1}^{n_1+n_2+1} p_{2j} = 1$. 

$$\text{is classified as } e_1 = \text{pr}(Z^* \in \pi_2 | Z^* \in \pi_1) = \text{pr}(Y^* \in \pi_2' | Y^* \in \pi_1')$$

(5.28)
The performance of the discrimination techniques can also be measured by calculating residual sum of squares. Let the observed value of \( x_j(i), j=1, \ldots, n_1+n_2 \) be either 0 or 1 according to \( i=1 \) or 2 where the confirmed membership of the case \( Y^* \) for \( \Pi_i \). The expected value of estimate of \( \hat{x}_j(i) \) is,

\[
E(\hat{x}_j(i)) = \sum_{i=1}^{2} x_j(i) \text{pr}(\hat{x}_j(i) = x_j(i)|\text{data})
\]

where the probability in the sum is \( \text{pr}(Y^*e\Pi_i|\text{data}) \), that is \( E(\hat{x}_j(i)) \) can be taken as various estimates of \( P_2 \) for \( Y_j, j=1, \ldots, n_1+n_2 \) which is treated as a new case \( Y^* \). Therefore different techniques can be compared and be called 'better' for smaller residual sum of squares

\[
RSS = \sum_{j=1}^{n_1+n_2} (\text{observed value of } x_j(i) - E(\hat{x}_j(i)))^2 \quad (5.32)
\]

5.9 Application of Predictive Techniques to Medical Data

Approximations to the predictive estimate of \( P_i \), the probability of being a member of population \( \Pi_i \), by using ranks of the observations, described in this Chapter, are applied to published data in this Section.

The example is taken from Habbema et al (1974) in the context of genetic counselling. The problem is discriminating between normal women and haemophilia A carriers with the two variables \( z_1 = \log_{10} \text{(AHF-activity)} \) and \( z_2 = \log_{10} \text{(AHF-like antigen)} \). Reference data of 30 non-carriers
and 45 obligatory carriers were available. More data on \( z_1 \) and \( z_2 \) were given for a test group of 23 possible carriers. The prior probability on carriership equals the genetic chance on carriership and this chance is known from the history for each of the possible carriers. The population of non-carriers is taken as \( \Pi_1 \) and of carriers as \( \Pi_2 \).

Habbema et al. (1974) concluded from an examination of the reference data that the assumption of bivariate normal distributions with equal covariance matrices was a very reasonable assumption. Therefore an 'estimative' technique Fisher's Linear Discriminant Function, defined by \( (z^* - \frac{1}{2} (\hat{\mu}^{(1)} + \hat{\mu}^{(2)}))^T \Sigma^{-1} (\hat{\mu}^{(1)} - \hat{\mu}^{(2)}) \) (LDF), where \( \hat{\mu}^{(1)} \) and \( \hat{\mu}^{(2)} \) are means and \( \Sigma \) is the covariance matrix of the populations, to obtain an estimate of \( P_i \), is only used as a tool in rank techniques to obtain a predictive estimate of \( P_i \). Hence LDF is applied to \( \Pi_1 \), \( \Pi_2 \) and undiagnosed observations where

\[
\hat{\mu}^{(1)} = \begin{bmatrix} -0.135 \\ -0.078 \end{bmatrix}, \quad \hat{\mu}^{(2)} = \begin{bmatrix} -0.308 \\ -0.006 \end{bmatrix}, \quad S = \begin{bmatrix} 0.023 & 0.015 \\ 0.015 & 0.022 \end{bmatrix}
\]

are sample means and pooled covariance matrix of the populations and the obtained univariate values are considered as populations \( \Pi_1', \Pi_2' \) and projected undiagnosed observations. The sample covariance matrices,
are also used to find alternative groups of observations by applying, Quadratic Discriminant Function, defined by
\[
(z^* - \mu_2(2))^T \Sigma_2^{-1} (z^* - \mu_2(2)) - (z^* - \mu_1(1))^T \Sigma_1^{-1} (z^* - \mu_1(1)) + \log \frac{|\Sigma_2|}{|\Sigma_1|},
\]
(QDF) to form \(\pi_1^1, \pi_2^2\) and the set of undiagnosed observations.

5.9.1 Application of Predictive Techniques to the Reference Data

Rank estimates of the probability of being a non-carrier, \(P_1\), namely \(\hat{P}_{\text{lin}}, \hat{P}_{\text{quad}}\), are obtained by treating this reference data as if it were an unclassified sample consisting of 75 observations where LDF and QDF are used to rank the observations before applying equation (5.6). The data is also treated in the same way to obtain the mean and median of the approximate distribution of \(P_1\), namely \(\text{mean}(\bar{P}), \text{med}(\bar{P})\) (equations (5.16),(5.20)) where the projected observations are obtained by LDF. A parametric predictive estimate \(\hat{P}_{\text{par}}\), proposed by Aitchison and Dunsmore (1975) is also considered; that is replacing \(p(z*|\theta_i)\) by
\[
p(z*|\text{data}) = \frac{\Gamma(\frac{1}{2}(m+1))}{\{\pi^k \Gamma(\frac{1}{2}(m-k+1))|\mu U_1|^\frac{1}{2}(1+(z^* - z_1)^T(\mu U_1)^{-1}(z^* - z_1))^{\frac{1}{2}(m+1)}\}
\]
where \(m = n_1 + n_2 - 2,\quad U_1 = (1 + \frac{1}{n_i})S\)
in equation (5.1). Prior probability of being a member of \(\pi_1\) is taken as \(\frac{n_1}{n_1 + n_2}\), i.e. \(q_1 = 0.4, q_2 = 0.6.\)
After allocating the observations to $\pi_1$ if the estimate of $P_1 > 0.5$ and $\pi_2$ otherwise, it is observed that the rank techniques misallocated 9 (5+4) observations if LDF is employed to obtain univariate samples whereas the parametric predictive technique misallocates 11 (4+7) observations (Table 5.3). The crude sample error rate estimates of the rank techniques $\hat{e}_1 = 0.16$ $\hat{e}_2 = 0.09$ where the corresponding bootstrap estimates of bias are 0.007 and 0.09 whereas rank estimate of both error rates are nearly 0.17.

In this case, QDF is not a better choice of the function $D: \mathbb{R}^2 \to \mathbb{R}$ than LDF, to obtain the univariate observations and 'correct ranking', but rank techniques are applied to such univariate samples and only 11 (6+5) observations are misallocated. However the importance of correct ranking will be witnessed in detail in Section 5.9.2 where the probabilities are observed more closely. It can also be observed when the performance of the techniques is compared by measuring residual sum of squares (RSS), equation (5.32).

RSS of $\hat{P}_{\text{lin}}$, $\hat{P}_{\text{quad}}$, mean($\tilde{P}$) and med($\tilde{P}$) are calculated for the set of 75 observations and found as 6.59, 8.51, 6.57, 6.56 respectively whereas that for the parametric estimate $P_{\text{par}}$ is 6.77. Therefore it can be claimed that, if LDF is used for univariate samples which is believed to give correct rankings in this case; rank based methods perform better than the parametric technique.
An important feature of rank techniques, being robust against outliers, plays a considerable role in getting better results than parametric techniques when they are applied to solve discrimination problems. If the sample(s) contain(s) some outliers, better results should be obtained by taking robust estimates of means and covariance matrices.

5.9.2 Comparison of Predictive Estimates of \( P_1 \) for Further Observations

Rank estimates of \( P_1 \); \( \hat{P}_{\text{lin}} \), \( \hat{P}_{\text{quad}} \), mean(\( \hat{P} \)), med(\( \hat{P} \)) are also obtained for 23 possible carriers and listed in Table 5.4 along with the parametric predictive estimate \( \hat{P}_{\text{par}} \).

It is observed that the mean and median of \( \hat{P}_1 \) are very close to each other since \( \text{var} \left( \log \frac{\pi_1}{\pi_2} \right) \) is small for each new observation. They are also very close to the rank estimate \( \hat{P}_{\text{lin}} \) which shows that the rank of \( Y^* \) does not give further information about the parameter.

Table 5.4 also reveals that the rank estimate \( \hat{P}_{\text{lin}} \), med(\( \hat{P} \)), mean(\( \hat{P} \)) are reasonably close to the parametric predictive estimate \( \hat{P}_{\text{par}} \) although it is slightly less extreme than \( \hat{P}_{\text{par}} \). They are in close agreement for the smaller and larger indexed observations for which the \( P \) values are close to 0 or 1 (observations 1-8, 15-23). The difference between \( \hat{P}_{\text{par}} \) and rank estimates is relatively large for observations 9-14 and it appears to be more serious for observations 13 and 14. \( \hat{P}_{\text{par}} \) classifies 13 and 14 to the population of non-carriers while rank estimates classify
them to the population of carriers. However, as mentioned in Section 5.1, the purpose of finding these probabilities is to observe how close a patient is to being a non-carrier rather than deciding that she is a non-carrier or a carrier. Therefore assuming that the parametric predictive estimate is nearest to the correct probability, we can say that the rank techniques give reliable results by choosing cases 13 and 14 as most doubtful individuals. Taking the function QDF to obtain univariate samples in this case study gives misleading results for some critical cases, for example, case 14 is given a probability of being a non-carrier as small as 0.247. However this result is not very surprising since the 'incorrect' ranks, obtained by applying QDF, are used while the assumption of having common covariance matrix is reasonable. Success of the rank methods depend upon having ranks of the univariate observations which represent multivariate populations correctly. It should also be noted that the discrimination techniques based on ranks do not perform very well for small sized populations ($n_1+n_2 < 20$).

5.10 Conclusions

The techniques presented here require more computation than standard parametric techniques but have the advantage of being robust against outlying or extreme observations. Their application depends upon the size of the populations and the application of some transformation to reduce multivariate observations to observations on the line. The weak assumption of normality after some unspecified transformation is required. Central limit theory applied to linear discriminants and
other functions suggests the normality assumption to be reasonable. The example considered in Section 5.9 suggests that when parametric normal assumptions hold, the rank based techniques perform remarkably efficiently.
TABLE 5.1:
Predicted posterior probabilities that $Y^*$ has rank $j$ amongst the combined observations given $Y^*$ belong to $\pi_1$ (columns I and II), and predictive probability of $Y^*$ belonging to $\pi_1$ given the rank of $Y^*$ (column III)

<table>
<thead>
<tr>
<th>Rank of $Y^*$</th>
<th>(I) $\rho_{1j}$</th>
<th>(II) $\rho_{2j}$</th>
<th>(III) $\hat{P}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$j$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.089</td>
<td>.053</td>
<td>.626</td>
</tr>
<tr>
<td>2</td>
<td>.089</td>
<td>.060</td>
<td>.595</td>
</tr>
<tr>
<td>3</td>
<td>.089</td>
<td>.067</td>
<td>.570</td>
</tr>
<tr>
<td>4</td>
<td>.087</td>
<td>.067</td>
<td>.565</td>
</tr>
<tr>
<td>5</td>
<td>.087</td>
<td>.071</td>
<td>.550</td>
</tr>
<tr>
<td>6</td>
<td>.087</td>
<td>.077</td>
<td>.531</td>
</tr>
<tr>
<td>7</td>
<td>.080</td>
<td>.077</td>
<td>.508</td>
</tr>
<tr>
<td>8</td>
<td>.073</td>
<td>.077</td>
<td>.487</td>
</tr>
<tr>
<td>9</td>
<td>.073</td>
<td>.084</td>
<td>.466</td>
</tr>
<tr>
<td>10</td>
<td>.067</td>
<td>.084</td>
<td>.444</td>
</tr>
<tr>
<td>11</td>
<td>.067</td>
<td>.094</td>
<td>.416</td>
</tr>
<tr>
<td>12</td>
<td>.059</td>
<td>.094</td>
<td>.388</td>
</tr>
<tr>
<td>13</td>
<td>.051</td>
<td>.094</td>
<td>.352</td>
</tr>
</tbody>
</table>

where $\rho_{ij}$ of columns I and II are calculated as in equation (5.5) and $\hat{P}_1$ as in equation (5.6).
TABLE 5.2:
The median, mean and variance of the variable $W = \log \frac{\pi_1(\beta)}{\pi_2(\beta)}$ where $\pi_i$ is the probability that $Y^*$ has rank $j$ amongst the combined observations given $Y^*$ belongs to $\Pi_i$, $i=1,2$ (Columns I,II) and predictive probability of $Y^*$ belonging to $\Pi_1$ given the rank of $Y^*$ (Columns III,IV).

<table>
<thead>
<tr>
<th>Rank of $Y^*$</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>med($W$)</td>
<td>E($W$)</td>
<td>var($W$)</td>
<td>med($\hat{p}_1$)</td>
</tr>
<tr>
<td>1</td>
<td>1.642</td>
<td>1.679</td>
<td>1.290</td>
<td>0.836</td>
</tr>
<tr>
<td>2</td>
<td>1.170</td>
<td>1.209</td>
<td>0.687</td>
<td>0.763</td>
</tr>
<tr>
<td>3</td>
<td>0.853</td>
<td>0.881</td>
<td>0.365</td>
<td>0.701</td>
</tr>
<tr>
<td>4</td>
<td>0.603</td>
<td>0.622</td>
<td>0.181</td>
<td>0.646</td>
</tr>
<tr>
<td>5</td>
<td>0.394</td>
<td>0.409</td>
<td>0.080</td>
<td>0.597</td>
</tr>
<tr>
<td>6</td>
<td>0.188</td>
<td>0.193</td>
<td>0.017</td>
<td>0.547</td>
</tr>
<tr>
<td>7</td>
<td>-0.010</td>
<td>-0.015</td>
<td>0.0002</td>
<td>0.497</td>
</tr>
<tr>
<td>8</td>
<td>-0.195</td>
<td>-0.204</td>
<td>0.020</td>
<td>0.451</td>
</tr>
<tr>
<td>9</td>
<td>-0.390</td>
<td>-0.403</td>
<td>0.076</td>
<td>0.404</td>
</tr>
<tr>
<td>10</td>
<td>-0.599</td>
<td>-0.617</td>
<td>0.176</td>
<td>0.354</td>
</tr>
<tr>
<td>11</td>
<td>-0.850</td>
<td>-0.877</td>
<td>0.359</td>
<td>0.299</td>
</tr>
<tr>
<td>12</td>
<td>-1.167</td>
<td>-1.204</td>
<td>0.681</td>
<td>0.237</td>
</tr>
<tr>
<td>13</td>
<td>-1.639</td>
<td>-1.675</td>
<td>1.284</td>
<td>0.163</td>
</tr>
</tbody>
</table>

Equations (5.11), (5.12), (5.16) and $\text{med}(g(W)) = g(\text{med}(W))$ equation (5.20), where $g(W) = \frac{1}{1+e^{-W}}$ are used to calculate the values.
Table 5.2 ... continued

Pearson Curves - some percentiles of \( W \) and \( \tilde{P}_1 \) for \( r^* = 9 \) \( E(W) = -0.403 \), \( E(W^2) = 0.238 \), \( E(W^3) = -0.163 \), \( E(W^4) = 0.127 \).

Mean = -0.403, Variance = 0.076, Skewness = 0.275, Kurtosis = 3.1

<table>
<thead>
<tr>
<th>Level ( \alpha )</th>
<th>( \omega_\alpha )</th>
<th>( (\tilde{P}<em>1)</em>\alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0025</td>
<td>-1.26</td>
<td>0.221</td>
</tr>
<tr>
<td>0.005</td>
<td>-1.18</td>
<td>0.235</td>
</tr>
<tr>
<td>0.01</td>
<td>-1.10</td>
<td>0.250</td>
</tr>
<tr>
<td>0.025</td>
<td>-0.98</td>
<td>0.273</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.88</td>
<td>0.293</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.76</td>
<td>0.319</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.58</td>
<td>0.359</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.39</td>
<td>0.404</td>
</tr>
<tr>
<td>0.75</td>
<td>-0.21</td>
<td>0.45</td>
</tr>
<tr>
<td>0.90</td>
<td>-0.06</td>
<td>0.485</td>
</tr>
<tr>
<td>0.95</td>
<td>0.03</td>
<td>0.507</td>
</tr>
<tr>
<td>0.975</td>
<td>0.10</td>
<td>0.525</td>
</tr>
<tr>
<td>0.99</td>
<td>0.18</td>
<td>0.545</td>
</tr>
<tr>
<td>0.995</td>
<td>0.24</td>
<td>0.560</td>
</tr>
<tr>
<td>0.9975</td>
<td>0.28</td>
<td>0.570</td>
</tr>
</tbody>
</table>

\[ \text{pr}(\omega \leq \omega_\alpha) = \text{pr}(P_1 \leq \tilde{P}_1\alpha) = \alpha \]
TABLE 5.3:
Predictive probabilities of being a 'non-carrier' for reference data for 30 haemophilia A non-carriers and 45 obligatory carriers

<table>
<thead>
<tr>
<th>Obsn. No</th>
<th>$Z_1$</th>
<th>$Z_2$</th>
<th>$r_1$(LDF)</th>
<th>$r_1$(QDF)</th>
<th>$\hat{\beta}_{\text{par}}$</th>
<th>$\hat{\beta}_{\text{lin}}$</th>
<th>$\hat{\beta}_{\text{quad}}$</th>
<th>$\gamma_{\text{med}(P)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.006</td>
<td>-0.166</td>
<td>75</td>
<td>73</td>
<td>.994</td>
<td>.992</td>
<td>.979</td>
<td>.995</td>
</tr>
<tr>
<td>2</td>
<td>-0.170</td>
<td>-0.159</td>
<td>67</td>
<td>65</td>
<td>.919</td>
<td>.928</td>
<td>.904</td>
<td>.937</td>
</tr>
<tr>
<td>3</td>
<td>-0.347</td>
<td>-0.188</td>
<td>42</td>
<td>42</td>
<td>.417*</td>
<td>.378*</td>
<td>.378*</td>
<td>.385*</td>
</tr>
<tr>
<td>4</td>
<td>-0.089</td>
<td>0.006</td>
<td>55</td>
<td>56</td>
<td>.779</td>
<td>.715</td>
<td>.740</td>
<td>.725</td>
</tr>
<tr>
<td>5</td>
<td>-0.168</td>
<td>0.071</td>
<td>29</td>
<td>28</td>
<td>.218*</td>
<td>.147*</td>
<td>.135*</td>
<td>.147*</td>
</tr>
<tr>
<td>6</td>
<td>-0.084</td>
<td>0.011</td>
<td>57</td>
<td>57</td>
<td>.781</td>
<td>.763</td>
<td>.763</td>
<td>.773</td>
</tr>
<tr>
<td>7</td>
<td>-0.198</td>
<td>-0.001</td>
<td>36</td>
<td>36</td>
<td>.342*</td>
<td>.252*</td>
<td>.252*</td>
<td>.256*</td>
</tr>
<tr>
<td>8</td>
<td>-0.076</td>
<td>0.039</td>
<td>50</td>
<td>51</td>
<td>.723</td>
<td>.586</td>
<td>.613</td>
<td>.595</td>
</tr>
<tr>
<td>9</td>
<td>-0.191</td>
<td>-0.212</td>
<td>70</td>
<td>68</td>
<td>.946</td>
<td>.957</td>
<td>.938</td>
<td>.965</td>
</tr>
<tr>
<td>10</td>
<td>-0.109</td>
<td>-0.119</td>
<td>69</td>
<td>70</td>
<td>.947</td>
<td>.948</td>
<td>.957</td>
<td>.957</td>
</tr>
<tr>
<td>11</td>
<td>-0.527</td>
<td>-0.477</td>
<td>54</td>
<td>46</td>
<td>.726</td>
<td>.691</td>
<td>.477*</td>
<td>.700</td>
</tr>
<tr>
<td>12</td>
<td>-0.084</td>
<td>0.025</td>
<td>51</td>
<td>52</td>
<td>.739</td>
<td>.613</td>
<td>.639</td>
<td>.622</td>
</tr>
<tr>
<td>13</td>
<td>-0.023</td>
<td>-0.058</td>
<td>72</td>
<td>72</td>
<td>.968</td>
<td>.972</td>
<td>.972</td>
<td>.979</td>
</tr>
<tr>
<td>14</td>
<td>0.008</td>
<td>0.078</td>
<td>63</td>
<td>63</td>
<td>.863</td>
<td>.876</td>
<td>.876</td>
<td>.886</td>
</tr>
<tr>
<td>15</td>
<td>-0.183</td>
<td>-0.114</td>
<td>59</td>
<td>59</td>
<td>.816</td>
<td>.806</td>
<td>.806</td>
<td>.816</td>
</tr>
<tr>
<td>16</td>
<td>0.124</td>
<td>0.214</td>
<td>62</td>
<td>61</td>
<td>.845</td>
<td>.860</td>
<td>.843</td>
<td>.870</td>
</tr>
<tr>
<td>17</td>
<td>-0.470</td>
<td>-0.310</td>
<td>39</td>
<td>38</td>
<td>.356*</td>
<td>.311*</td>
<td>.290*</td>
<td>.316*</td>
</tr>
<tr>
<td>18</td>
<td>-0.152</td>
<td>-0.069</td>
<td>58</td>
<td>58</td>
<td>.791</td>
<td>.785</td>
<td>.785</td>
<td>.795</td>
</tr>
<tr>
<td>19</td>
<td>0.001</td>
<td>-0.115</td>
<td>74</td>
<td>74</td>
<td>.990</td>
<td>.985</td>
<td>.985</td>
<td>.990</td>
</tr>
<tr>
<td>20</td>
<td>-0.202</td>
<td>-0.050</td>
<td>43</td>
<td>45</td>
<td>.522</td>
<td>.402*</td>
<td>.451*</td>
<td>.409*</td>
</tr>
<tr>
<td>21</td>
<td>-0.193</td>
<td>-0.229</td>
<td>71</td>
<td>69</td>
<td>.956</td>
<td>.965</td>
<td>.948</td>
<td>.972</td>
</tr>
<tr>
<td>22</td>
<td>0.151</td>
<td>0.093</td>
<td>73</td>
<td>75</td>
<td>.981</td>
<td>.979</td>
<td>.992</td>
<td>.985</td>
</tr>
<tr>
<td>23</td>
<td>-0.126</td>
<td>-0.067</td>
<td>61</td>
<td>62</td>
<td>.854</td>
<td>.843</td>
<td>.860</td>
<td>.853</td>
</tr>
<tr>
<td>24</td>
<td>-0.155</td>
<td>-0.125</td>
<td>64</td>
<td>64</td>
<td>.895</td>
<td>.891</td>
<td>.891</td>
<td>.900</td>
</tr>
<tr>
<td>25</td>
<td>-0.195</td>
<td>-0.101</td>
<td>52</td>
<td>53</td>
<td>.744</td>
<td>.639</td>
<td>.665</td>
<td>.648</td>
</tr>
<tr>
<td>26</td>
<td>0.029</td>
<td>0.044</td>
<td>68</td>
<td>71</td>
<td>.939</td>
<td>.938</td>
<td>.965</td>
<td>.947</td>
</tr>
<tr>
<td>27</td>
<td>-0.223</td>
<td>-0.171</td>
<td>60</td>
<td>60</td>
<td>.842</td>
<td>.825</td>
<td>.825</td>
<td>.835</td>
</tr>
<tr>
<td>28</td>
<td>-0.100</td>
<td>-0.073</td>
<td>66</td>
<td>67</td>
<td>.911</td>
<td>.916</td>
<td>.928</td>
<td>.926</td>
</tr>
<tr>
<td>29</td>
<td>-0.197</td>
<td>-0.061</td>
<td>47</td>
<td>50</td>
<td>.590</td>
<td>.504</td>
<td>.586</td>
<td>.512</td>
</tr>
<tr>
<td>30</td>
<td>-0.087</td>
<td>-0.056</td>
<td>65</td>
<td>66</td>
<td>.908</td>
<td>.904</td>
<td>.916</td>
<td>.914</td>
</tr>
<tr>
<td>Obsn No</td>
<td>( Z_1 )</td>
<td>( Z_2 )</td>
<td>( r_i ) (LDF)</td>
<td>( r_i ) (QDF)</td>
<td>( \hat{p}_{\text{par}} )</td>
<td>( \hat{p}_{\text{lin}} )</td>
<td>( \hat{p}_{\text{quad}} )</td>
<td>( \gamma_{\text{med}(p)} )</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>31</td>
<td>-0.348</td>
<td>0.115</td>
<td>7</td>
<td>7</td>
<td>0.007</td>
<td>0.015</td>
<td>0.015</td>
<td>0.013</td>
</tr>
<tr>
<td>32</td>
<td>-0.362</td>
<td>-0.301</td>
<td>41</td>
<td>41</td>
<td>0.402</td>
<td>0.355</td>
<td>0.355</td>
<td>0.361</td>
</tr>
<tr>
<td>33</td>
<td>-0.499</td>
<td>-0.086</td>
<td>8</td>
<td>11</td>
<td>0.010</td>
<td>0.017</td>
<td>0.026</td>
<td>0.015</td>
</tr>
<tr>
<td>34</td>
<td>-0.501</td>
<td>-0.298</td>
<td>28</td>
<td>29</td>
<td>0.209</td>
<td>0.135</td>
<td>0.147</td>
<td>0.135</td>
</tr>
<tr>
<td>35</td>
<td>-0.133</td>
<td>0.010</td>
<td>48</td>
<td>49</td>
<td>0.594*</td>
<td>0.530*</td>
<td>0.558*</td>
<td>0.539*</td>
</tr>
<tr>
<td>36</td>
<td>-0.691</td>
<td>-0.339</td>
<td>12</td>
<td>17</td>
<td>0.022</td>
<td>0.029</td>
<td>0.050</td>
<td>0.027</td>
</tr>
<tr>
<td>37</td>
<td>-0.361</td>
<td>0.124</td>
<td>6</td>
<td>5</td>
<td>0.005</td>
<td>0.013</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td>38</td>
<td>-0.453</td>
<td>-0.168</td>
<td>22</td>
<td>24</td>
<td>0.070</td>
<td>0.081</td>
<td>0.096</td>
<td>0.078</td>
</tr>
<tr>
<td>39</td>
<td>-0.348</td>
<td>-0.172</td>
<td>37</td>
<td>39</td>
<td>0.350</td>
<td>0.271</td>
<td>0.311</td>
<td>0.275</td>
</tr>
<tr>
<td>40</td>
<td>-0.354</td>
<td>0.072</td>
<td>9</td>
<td>9</td>
<td>0.011</td>
<td>0.020</td>
<td>0.020</td>
<td>0.018</td>
</tr>
<tr>
<td>41</td>
<td>-0.472</td>
<td>-0.108</td>
<td>14</td>
<td>15</td>
<td>0.021</td>
<td>0.037</td>
<td>0.041</td>
<td>0.034</td>
</tr>
<tr>
<td>42</td>
<td>-0.361</td>
<td>-0.040</td>
<td>21</td>
<td>21</td>
<td>0.047</td>
<td>0.074</td>
<td>0.074</td>
<td>0.071</td>
</tr>
<tr>
<td>43</td>
<td>-0.323</td>
<td>0.167</td>
<td>5</td>
<td>4</td>
<td>0.005</td>
<td>0.010</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>44</td>
<td>-0.432</td>
<td>-0.069</td>
<td>15</td>
<td>16</td>
<td>0.023</td>
<td>0.041</td>
<td>0.043</td>
<td>0.038</td>
</tr>
<tr>
<td>45</td>
<td>-0.273</td>
<td>-0.002</td>
<td>26</td>
<td>26</td>
<td>0.116</td>
<td>0.114</td>
<td>0.114</td>
<td>0.113</td>
</tr>
<tr>
<td>46</td>
<td>-0.362</td>
<td>-0.201</td>
<td>1</td>
<td>1</td>
<td>0.001</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>47</td>
<td>-0.376</td>
<td>-0.187</td>
<td>32</td>
<td>34</td>
<td>0.292</td>
<td>0.186</td>
<td>0.217</td>
<td>0.188</td>
</tr>
<tr>
<td>48</td>
<td>-0.495</td>
<td>-0.015</td>
<td>4</td>
<td>6</td>
<td>0.004</td>
<td>0.008</td>
<td>0.013</td>
<td>0.007</td>
</tr>
<tr>
<td>49</td>
<td>-0.511</td>
<td>-0.248</td>
<td>24</td>
<td>25</td>
<td>0.090</td>
<td>0.096</td>
<td>0.105</td>
<td>0.095</td>
</tr>
<tr>
<td>50</td>
<td>-0.165</td>
<td>0.213</td>
<td>17</td>
<td>13</td>
<td>0.032</td>
<td>0.050</td>
<td>0.033</td>
<td>0.048</td>
</tr>
<tr>
<td>51</td>
<td>-0.245</td>
<td>-0.041</td>
<td>33</td>
<td>33</td>
<td>0.295</td>
<td>0.201</td>
<td>0.201</td>
<td>0.203</td>
</tr>
<tr>
<td>52</td>
<td>-0.423</td>
<td>-0.100</td>
<td>20</td>
<td>20</td>
<td>0.042</td>
<td>0.067</td>
<td>0.067</td>
<td>0.065</td>
</tr>
<tr>
<td>53</td>
<td>-0.238</td>
<td>0.288</td>
<td>3</td>
<td>2</td>
<td>0.004</td>
<td>0.007</td>
<td>0.005</td>
<td>0.006</td>
</tr>
<tr>
<td>54</td>
<td>-0.220</td>
<td>0.005</td>
<td>30</td>
<td>30</td>
<td>0.238</td>
<td>0.160</td>
<td>0.160</td>
<td>0.160</td>
</tr>
<tr>
<td>55</td>
<td>-0.215</td>
<td>-0.022</td>
<td>38</td>
<td>37</td>
<td>0.349</td>
<td>0.290</td>
<td>0.271</td>
<td>0.295</td>
</tr>
<tr>
<td>56</td>
<td>-0.345</td>
<td>0.010</td>
<td>16</td>
<td>18</td>
<td>0.030</td>
<td>0.045</td>
<td>0.055</td>
<td>0.043</td>
</tr>
<tr>
<td>57</td>
<td>-0.254</td>
<td>-0.057</td>
<td>34</td>
<td>35</td>
<td>0.316</td>
<td>0.217</td>
<td>0.234</td>
<td>0.220</td>
</tr>
<tr>
<td>58</td>
<td>-0.378</td>
<td>-0.268</td>
<td>49</td>
<td>48</td>
<td>0.598*</td>
<td>0.558*</td>
<td>0.530*</td>
<td>0.567*</td>
</tr>
<tr>
<td>59</td>
<td>-0.405</td>
<td>-0.116</td>
<td>23</td>
<td>23</td>
<td>0.071</td>
<td>0.088</td>
<td>0.088</td>
<td>0.086</td>
</tr>
<tr>
<td>60</td>
<td>-0.064</td>
<td>-0.157</td>
<td>35</td>
<td>31</td>
<td>0.321</td>
<td>0.234</td>
<td>0.172</td>
<td>0.237</td>
</tr>
<tr>
<td>Obsn No</td>
<td>$z_1$</td>
<td>$z_2$</td>
<td>$r_i$(LDF)</td>
<td>$r_i$(QDF)</td>
<td>$\hat{P}_{\text{par}}$</td>
<td>$\hat{P}_{\text{lin}}$</td>
<td>$\hat{P}_{\text{quad}}$</td>
<td>$\text{med}(\tilde{P})$</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>------------</td>
<td>------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>61</td>
<td>-0.335</td>
<td>-0.137</td>
<td>31</td>
<td>32</td>
<td>.278</td>
<td>.172</td>
<td>.186</td>
<td>.173</td>
</tr>
<tr>
<td>62</td>
<td>-0.015</td>
<td>0.154</td>
<td>45</td>
<td>44</td>
<td>.549*</td>
<td>.451</td>
<td>.426</td>
<td>.459</td>
</tr>
<tr>
<td>63</td>
<td>-0.031</td>
<td>0.140</td>
<td>44</td>
<td>43</td>
<td>.533*</td>
<td>.426</td>
<td>.402</td>
<td>.433</td>
</tr>
<tr>
<td>64</td>
<td>-0.174</td>
<td>-0.078</td>
<td>53</td>
<td>54</td>
<td>.745*</td>
<td>.675*</td>
<td>.691*</td>
<td>.674*</td>
</tr>
<tr>
<td>65</td>
<td>-0.142</td>
<td>0.164</td>
<td>25</td>
<td>22</td>
<td>.093</td>
<td>.105</td>
<td>.081</td>
<td>.104</td>
</tr>
<tr>
<td>66</td>
<td>-0.151</td>
<td>0.114</td>
<td>27</td>
<td>27</td>
<td>.160</td>
<td>.124</td>
<td>.124</td>
<td>.124</td>
</tr>
<tr>
<td>67</td>
<td>-0.096</td>
<td>0.053</td>
<td>46</td>
<td>47</td>
<td>.589*</td>
<td>.477</td>
<td>.504*</td>
<td>.485</td>
</tr>
<tr>
<td>68</td>
<td>-0.264</td>
<td>0.087</td>
<td>19</td>
<td>19</td>
<td>.037</td>
<td>.061</td>
<td>.061</td>
<td>.059</td>
</tr>
<tr>
<td>69</td>
<td>-0.023</td>
<td>0.080</td>
<td>56</td>
<td>55</td>
<td>.778*</td>
<td>.740*</td>
<td>.715*</td>
<td>.749*</td>
</tr>
<tr>
<td>70</td>
<td>-0.335</td>
<td>0.088</td>
<td>10</td>
<td>10</td>
<td>.012</td>
<td>.023</td>
<td>.023</td>
<td>.021</td>
</tr>
<tr>
<td>71</td>
<td>-0.188</td>
<td>0.251</td>
<td>11</td>
<td>8</td>
<td>.013</td>
<td>.025</td>
<td>.017</td>
<td>.024</td>
</tr>
<tr>
<td>72</td>
<td>-0.174</td>
<td>0.189</td>
<td>18</td>
<td>14</td>
<td>.039</td>
<td>.055</td>
<td>.037</td>
<td>.053</td>
</tr>
<tr>
<td>73</td>
<td>-0.405</td>
<td>-0.242</td>
<td>40</td>
<td>40</td>
<td>.374</td>
<td>.333</td>
<td>.333</td>
<td>.338</td>
</tr>
<tr>
<td>74</td>
<td>-0.244</td>
<td>0.161</td>
<td>13</td>
<td>12</td>
<td>.018</td>
<td>.033</td>
<td>.029</td>
<td>.030</td>
</tr>
<tr>
<td>75</td>
<td>-0.478</td>
<td>0.028</td>
<td>2</td>
<td>3</td>
<td>.003</td>
<td>.005</td>
<td>.007</td>
<td>.004</td>
</tr>
</tbody>
</table>


$z_1 - \log_{10}$ (AHF activity)

$z_2 - \log_{10}$ (AHF-like antigen)

$r_i$(LDF), $r_i$(QDF) - rank of $Y_i$ where LDF and QDF are used to obtain univariate observations

$\hat{P}_{\text{par}}$ - predictive probability found by using the parametric method of Aitchison and Dunsmore (1975)

$\hat{P}_{\text{lin}}, \hat{P}_{\text{quad}}$ - LDF and QDF are used to obtain univariate observations probabilities found by equation (5.6)

$\text{med}(\tilde{P})$ - median of $\tilde{P}$ found by using approximate distribution of $P_1$
TABLE 5.4:
Predicted probabilities of being a 'non-carrier' for 23 possible carriers

<table>
<thead>
<tr>
<th>Obsn</th>
<th>$\hat{p}_{\text{par}}$</th>
<th>$\hat{p}_{\text{lin}}$</th>
<th>$\hat{p}_{\text{quad}}$</th>
<th>$\text{med}(\hat{p})$</th>
<th>$\text{mean}(\hat{p})$</th>
<th>90% CI for $\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.996</td>
<td>.992</td>
<td>.973</td>
<td>.994</td>
<td>.992</td>
<td>(.978, .998)</td>
</tr>
<tr>
<td>2</td>
<td>.990</td>
<td>.991</td>
<td>.993</td>
<td>.991</td>
<td>.991</td>
<td>(.979, .997)</td>
</tr>
<tr>
<td>3</td>
<td>.988</td>
<td>.983</td>
<td>.988</td>
<td>.986</td>
<td>.983</td>
<td>(.957, .996)</td>
</tr>
<tr>
<td>4</td>
<td>.986</td>
<td>.987</td>
<td>.989</td>
<td>.987</td>
<td>.987</td>
<td>(.969, .996)</td>
</tr>
<tr>
<td>5</td>
<td>.984</td>
<td>.983</td>
<td>.978</td>
<td>.986</td>
<td>.983</td>
<td>(.957, .996)</td>
</tr>
<tr>
<td>6</td>
<td>.972</td>
<td>.973</td>
<td>.973</td>
<td>.973</td>
<td>.973</td>
<td>(.940, .991)</td>
</tr>
<tr>
<td>7</td>
<td>.944</td>
<td>.950</td>
<td>.949</td>
<td>.953</td>
<td>.947</td>
<td>(.888, .983)</td>
</tr>
<tr>
<td>8</td>
<td>.936</td>
<td>.929</td>
<td>.938</td>
<td>.931</td>
<td>.927</td>
<td>(.874, .965)</td>
</tr>
<tr>
<td>9</td>
<td>.841</td>
<td>.802</td>
<td>.819</td>
<td>.797</td>
<td>.796</td>
<td>(.775, .815)</td>
</tr>
<tr>
<td>10</td>
<td>.803</td>
<td>.691</td>
<td>.714</td>
<td>.688</td>
<td>.687</td>
<td>(.652, .720)</td>
</tr>
<tr>
<td>11</td>
<td>.780</td>
<td>.742</td>
<td>.741</td>
<td>.739</td>
<td>.738</td>
<td>(.717, .754)</td>
</tr>
<tr>
<td>12</td>
<td>.721</td>
<td>.646</td>
<td>.535</td>
<td>.636</td>
<td>.634</td>
<td>(.567, .704)</td>
</tr>
<tr>
<td>13</td>
<td>.586</td>
<td>.490</td>
<td>.368</td>
<td>.484</td>
<td>.482</td>
<td>(.456, .505)</td>
</tr>
<tr>
<td>14</td>
<td>.507</td>
<td>.440</td>
<td>.247</td>
<td>.433</td>
<td>.432</td>
<td>(.391, .467)</td>
</tr>
<tr>
<td>15</td>
<td>.183</td>
<td>.168</td>
<td>.184</td>
<td>.161</td>
<td>.164</td>
<td>(.095, .243)</td>
</tr>
<tr>
<td>16</td>
<td>.048</td>
<td>.060</td>
<td>.050</td>
<td>.054</td>
<td>.058</td>
<td>(.026, .099)</td>
</tr>
<tr>
<td>17</td>
<td>.035</td>
<td>.055</td>
<td>.027</td>
<td>.049</td>
<td>.053</td>
<td>(.023, .093)</td>
</tr>
<tr>
<td>18</td>
<td>.027</td>
<td>.055</td>
<td>.050</td>
<td>.049</td>
<td>.053</td>
<td>(.023, .093)</td>
</tr>
<tr>
<td>19</td>
<td>.015</td>
<td>.041</td>
<td>.028</td>
<td>.035</td>
<td>.040</td>
<td>(.012, .085)</td>
</tr>
<tr>
<td>20</td>
<td>.011</td>
<td>.027</td>
<td>.024</td>
<td>.023</td>
<td>.026</td>
<td>(.008, .054)</td>
</tr>
<tr>
<td>21</td>
<td>.006</td>
<td>.020</td>
<td>.024</td>
<td>.016</td>
<td>.020</td>
<td>(.005, .049)</td>
</tr>
<tr>
<td>22</td>
<td>.006</td>
<td>.020</td>
<td>.028</td>
<td>.016</td>
<td>.020</td>
<td>(.005, .049)</td>
</tr>
<tr>
<td>23</td>
<td>.003</td>
<td>.005</td>
<td>.011</td>
<td>.004</td>
<td>.005</td>
<td>(.0008, .012)</td>
</tr>
</tbody>
</table>

$\hat{p}_{\text{par}}$ - predictive probability found by using the parametric method of Aitchison and Dunsmore (1975)

$\hat{p}_{\text{lin}}$, $\hat{p}_{\text{quad}}$ - LDF and QDF are used to obtain univariate observations, probabilities found by (5.6)

$\text{med} \ P$, $\text{mean} \ P$, CI - median, mean, confidence interval found by using approximate distribution of $P_1$
TABLE 5.4 ... continued
Test group of POSSIBLE CARRIERS

<table>
<thead>
<tr>
<th>Observation No.</th>
<th>$z_1$</th>
<th>$z_2$</th>
<th>$q_1$</th>
<th>$r_1$(LDF)</th>
<th>$r_1$(QDF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.112</td>
<td>-0.279</td>
<td>0.50</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>2</td>
<td>-0.059</td>
<td>-0.068</td>
<td>0.75</td>
<td>71</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>0.064</td>
<td>0.012</td>
<td>0.50</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>-0.043</td>
<td>-0.052</td>
<td>0.67</td>
<td>71</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>-0.050</td>
<td>-0.058</td>
<td>0.50</td>
<td>73</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>-0.094</td>
<td>-0.113</td>
<td>0.50</td>
<td>71</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>-0.123</td>
<td>-0.143</td>
<td>0.34</td>
<td>71</td>
<td>71</td>
</tr>
<tr>
<td>8</td>
<td>-0.011</td>
<td>0.037</td>
<td>0.50</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>9</td>
<td>-0.210</td>
<td>-0.090</td>
<td>0.67</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>-0.126</td>
<td>-0.019</td>
<td>0.50</td>
<td>51</td>
<td>52</td>
</tr>
<tr>
<td>11</td>
<td>-0.210</td>
<td>-0.044</td>
<td>0.75</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>12</td>
<td>0.069</td>
<td>0.192</td>
<td>0.34</td>
<td>55</td>
<td>51</td>
</tr>
<tr>
<td>13</td>
<td>0.030</td>
<td>0.224</td>
<td>0.50</td>
<td>43</td>
<td>38</td>
</tr>
<tr>
<td>14</td>
<td>-0.002</td>
<td>0.206</td>
<td>0.50</td>
<td>41</td>
<td>32</td>
</tr>
<tr>
<td>15</td>
<td>-0.371</td>
<td>-0.133</td>
<td>0.50</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>16</td>
<td>-0.162</td>
<td>0.162</td>
<td>0.34</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>17</td>
<td>-0.097</td>
<td>0.263</td>
<td>0.34</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>-0.250</td>
<td>0.095</td>
<td>0.34</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>19</td>
<td>-0.287</td>
<td>0.137</td>
<td>0.50</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>-0.357</td>
<td>0.031</td>
<td>0.34</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>21</td>
<td>-0.461</td>
<td>-0.003</td>
<td>0.50</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>22</td>
<td>-0.590</td>
<td>-0.139</td>
<td>0.50</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>23</td>
<td>-0.561</td>
<td>-0.060</td>
<td>0.50</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

$z_1 - \log_{10}(AHF$ activity;  $z_2 - \log_{10}(AHF$-like antigen)
$q_1$ - prior probability of non-carriership
CHAPTER 6
QUARTIC APPROXIMATION TO THE RANK
LIKELIHOOD FOR LOGISTIC DENSITY

6.1 Introduction
The marginal likelihood of ranks \( f(\mathbf{r}|\theta) \) which is given by

\[
\prod_{y=1}^{N} f(y_j - \theta_j) \, dy
\]

(6.1)

where \( \Omega = \{(y_{a_1}, \ldots, y_{a_N}) : y_{a_1} < \ldots < y_{a_N} \} \) can well be approximated by using the Taylor expansion of \( \log f \) when \( f(.) \) is a normal density. Consequently various techniques have been developed and applied successfully to tackle statistical problems by using this approximation to the rank marginal likelihood for example Pettitt (1982b, 1983a,b), Acar and Pettitt (1984, 1985a). When \( f(.) \) is proportional to a normal density then second order Taylor series approximation, becomes exact, since the higher order derivatives of \( \log f \) are zero. For a non-normal density \( f(.) \), one should include higher order terms to obtain more accurate approximations.

In this Chapter \( f(.) \) is taken as logistic density and an approximation is obtained by including fourth order terms of Taylor approximation.

6.2 Quartic Approximation to the Rank Likelihood
The rank likelihood \( pr(Y_{a_1} < Y_{a_2} < \ldots < Y_{a_N} | \theta) \) given by (6.1) can be obtained numerically as an N-type integral. To extend the results
given in Pettitt (1982a), consider the approximation to
\[
\log f(y_j - \theta_j)
\]
\[
= \log (y_j - \theta_j) = \log(y_j) + g(y_j)\theta_j - \frac{1}{2} g'(y_j)\theta_j^2 + \frac{1}{6} g''(y_j)\theta_j^3
\]
\[
- \frac{1}{24} g'''(y_j)\theta_j^4
\]

where \( g(y) = -\frac{f'(y)}{f(y)} \).

Substituting \( \exp(x) \) in (6.1) gives
\[
\text{pr} (y_1 < \ldots < y_N | \theta) = \int \exp \left\{ \sum g(y_j)\theta_j - \frac{1}{2} \sum g'(y_j)\theta_j^2 + \frac{1}{6} \sum g''(y_j)\theta_j^3 \right. \\
- \frac{1}{24} \sum g'''(y_j)\theta_j^4 \} \times \{ \Pi f(y_j)dy_j \}
\]
\[
= \frac{1}{N!} E \left( \exp \left( \sum g(Z_{r_j})\theta_j - \frac{1}{2} \sum g'(Z_{r_j})\theta_j^2 + \frac{1}{6} \sum g''(Z_{r_j})\theta_j^3 - \frac{1}{24} \sum g'''(Z_{r_j})\theta_j^4 \right) \right)
\]
(6.3)

where the summations and product are over \( j=1, \ldots, N \) and \( Z_1 < \ldots < Z_N \) are the order statistics of a sample of size \( N \) from the distribution \( f(.) \) and \( r_i \) is the rank of \( y_i, i=1, \ldots, N \). Consider the approximation of (6.3) by a function of the form
\[
\frac{1}{N!} \exp \left\{ \sum a_i \theta_i - \frac{1}{2} \sum b_{ij} \theta_i \theta_j + \frac{1}{6} \sum c_{ijk} \theta_i \theta_j \theta_k - \frac{1}{24} \sum d_{ijkl} \theta_i \theta_j \theta_k \theta_l \right\}
\]
(6.4)
Equating the coefficients of (6.3) and (6.4) gives

\[ a_i = \mathbb{E} \{ g(Z_{r_i}) \} \]

\[ b_{ij} = a_i a_j - \mathbb{E}(g(Z_{r_i})g(Z_{r_j})) - \delta_{ij} g'(Z_{r_i}) \]

\[ c_{ijk} = \mathbb{E}(g(Z_{r_i})g(Z_{r_j})g(Z_{r_k})) - \delta_{ij} g'(Z_{r_i}) g(Z_{r_j}) - \delta_{ik} g'(Z_{r_i}) g(Z_{r_k}) - \delta_{jk} g'(Z_{r_j}) g(Z_{r_k}) - \delta_{ik} \delta_{jk} g''(Z_{r_i}) - a_1 a_j a_k + b_{ij} a_k + b_{ik} a_j + b_{jk} a_i \]

\[ d_{ijk\ell} = a_i a_j a_k a_\ell - b_{ij} a_k a_\ell - b_{ik} a_j a_\ell - b_{jk} a_i a_\ell + c_{i\ell k} a_j + c_{i\ell k} a_j + c_{i\ell k} a_j + c_{i\ell k} a_j + c_{i\ell k} a_j + c_{i\ell k} a_j \]

\[ - \mathbb{E}(g(Z_{r_i})g(Z_{r_j})g(Z_{r_k})g(Z_{r_\ell})) - \delta_{ij} g'(Z_{r_i}) g(Z_{r_j}) g(Z_{r_k}) g(Z_{r_\ell}) \]

\[ - \delta_{ik} g'(Z_{r_i}) g(Z_{r_j}) g(Z_{r_k}) g(Z_{r_\ell}) \]

\[ - \delta_{i\ell} g'(Z_{r_i}) g(Z_{r_j}) g(Z_{r_k}) g(Z_{r_\ell}) - \delta_{jk} g'(Z_{r_j}) g(Z_{r_i}) g(Z_{r_k}) g(Z_{r_\ell}) \]

\[ - \delta_{k\ell} g'(Z_{r_k}) g(Z_{r_i}) g(Z_{r_j}) g(Z_{r_\ell}) \]
where $\delta_{ij}$ is Kronecker delta function.

Therefore an approximation to (6.1) can be obtained by substituting (6.5), (6.6), (6.7) and (6.8) into (6.4).

6.3 Logistic Distribution Case

For any density function $f(.)$ in (6.1), the coefficients $a_i, b_{ij}, c_{ijk}$ and $d_{ijkl}$ obtained in Section 6.2 can be written by using central moments of combinations of $g(.), g'(.), g''(.)$ and $g'''(.)$. Let $g(Z_r)$ be denoted by $g_i$, and the central moments be denoted as follows:

$\mu(U) = E(U)$

$\sigma(U,V) = E((U-E(U))(V-E(V)))$
\[ \kappa (U, V, W) = E((U-E(U))(V-E(V))(W-E(W))) \]

\[ \tau(U, V, W, T) = E((U-E(U))(V-E(V))(W-E(W))(T-E(T))) \]

for variables U, V, W and T. Therefore, (6.5), (6.6), (6.7) and (6.8) can be re-written as

\[ a_i = \mu(g_i) \quad (6.9) \]

\[ b_{ij} = -\sigma(g_i, g_j) + \delta_{ij} \mu(g_i) \quad (6.10) \]

\[ c_{ijk} = \kappa(g_i, g_j, g_k) - \delta_{ij} \sigma(g_i, g_k) - \delta_{jk} \sigma(g_j, g_i) - \delta_{ik} \sigma(g_i, g_j) \]

\[ + \delta_{ij} \delta_{jk} \mu(g_i) \quad (6.11) \]

\[ d_{ijk\ell} = -\tau(g_i, g_j, g_k, g_\ell) + \delta_{ij} \kappa(g_i, g_k, g_\ell) + \delta_{ik} \kappa(g_i, g_j, g_\ell) \]

\[ + \delta_{i\ell} \kappa(g_i, g_j, g_k) + \delta_{jk} \kappa(g_j, g_i, g_\ell) + \delta_{ik} \kappa(g_k, g_i, g_j) \]

\[ + \delta_{j\ell} \kappa(g_j, g_i, g_k) - \delta_{ij} \delta_{k\ell} \sigma(g_i, g_k) - \delta_{ik} \delta_{j\ell} \sigma(g_i, g_j) \]

\[ - \delta_{i\ell} \delta_{jk} \sigma(g_i, g_j) - \delta_{ij} \delta_{k\ell} \sigma(g_i, g_\ell) - \delta_{ik} \delta_{j\ell} \sigma(g_k, g_j) \]

\[ - \delta_{ik} \delta_{j\ell} \sigma(g_j, g_\ell) - \delta_{jk} \delta_{i\ell} \sigma(g_j, g_i) + \sigma(g_i, g_j, g_\ell, g_\ell) + \sigma(g_i, g_j, g_k, g_\ell) \]

\[ + \sigma(g_i, g_\ell) \sigma(g_j, g_k) + \sigma(g_i, g_k) \sigma(g_j, g_\ell) \]

\[ + \delta_{ij} \delta_{jk} \delta_{k\ell} \mu(g_i, g_j, g_k) \quad (6.12) \]
When the density function \( f(.) \) in (6.1) is taken to be logistic i.e.

\[
f(y) = \frac{e^{-y}}{(1 + e^{-y})^2} = F(y) (1 - F(y))
\]

where \( F(y) = \frac{1}{1 + e^{-y}} \)

\[g(y) = 2F(y) - 1\]

\( g_i (g(Z_{r_i})) \) and its derivatives can be written as

\[
g_i = 2U_{r_i} - 1
\]

\[
g'_i = 2U_{r_i} (1 - U_{r_i})
\]

\[
g''_i = 2U_{r_i} (1 - U_{r_i})(1 - 2U_{r_i})
\]

\[
g'''_i = 2U_{r_i} (1 - U_{r_i})(2U_{r_i} - 1 + \frac{\sqrt{3}}{3})(3U_{r_i} - \frac{3}{2} - \frac{\sqrt{3}}{2})
\]

where \( 0 < U_1 < U_2 < \ldots < U_N \) are the \( N \) order statistics of a sample of size \( N \) from the uniform distribution. Therefore, to obtain the coefficients \( a_i, b_{ij}, c_{ijk} \) and \( d_{ijkl} \) moments of the combinations of \( g_i \) and its derivatives, in terms of moments of uniform order statistics.
are required. The abbreviations \( \nu_i \), \( \sigma_{ijk} \), \( \kappa_{ijkl} \) and \( \tau_{ijkl} \) are used to represent the first four central moments of uniform order statistics. The typical moments are calculated as:

\[
\mu(g_i) = 2\nu_{r_i} - 1 \tag{6.13}
\]

\[
\mu(g_i') = -2\left(\sigma_{r_i r_i} - \nu_{r_i} (1 - \nu_{r_i})\right) \tag{6.14}
\]

\[
\mu(g_i'') = 2(2\kappa_{r_i r_i} r_i - 3\sigma_{r_i r_i} (1 - 2\nu_{r_i}) + \nu_{r_i} (1 - \nu_{r_i})(1 - 2\nu_{r_i})) \tag{6.15}
\]

\[
\mu(g_i''') = 2(-6\nu_{r_i} r_i + 12\kappa_{r_i r_i} r_i (1 - 2\nu_{r_i}) + \sigma_{r_i r_i} (-36\nu_{r_i}^2 + 36\nu_{r_i} - 7)
+ \nu_{r_i} (1 - \nu_{r_i})(6\nu_{r_i}^2 - 6\nu_{r_i} + 1)) \tag{6.16}
\]

\[
\sigma(g_i, g_j) = 4\sigma_{r_i r_j} \tag{6.17}
\]

\[
\sigma(g_i', g_j) = 4(-\kappa_{r_i r_i} r_i + \sigma_{r_i r_j} (1 - 2\nu_{r_i})) \tag{6.18}
\]

\[
\sigma(g_i'', g_j) = 4(2\nu_{r_i} r_i r_j + 3\kappa_{r_i r_i} r_j (2\nu_{r_i} - 1) + \sigma_{r_i r_j} (6\nu_{r_i}^2 - 6\nu_{r_i} + 1)) \tag{6.19}
\]

\[
\sigma(g_i', g_j') = 4(\nu_{r_i} r_i r_j + \kappa_{r_i r_i} r_j (2\nu_{r_i} - 1) + \kappa_{r_i r_j} r_j (2\nu_{r_j} - 1)
+ \sigma_{r_i r_j} (2\nu_{r_i} - 1)(2\nu_{r_j} - 1) - \sigma_{r_i r_i} \sigma_{r_j r_j}) \tag{6.20}
\]

\[
k(g_i, g_j, g_k) = 8\kappa_{r_i r_j r_k} \tag{6.21}
\]
\[ \kappa(g_1', g_2', g_3', g_4') = 8\{ -r_1 r_2 r_3 r_4 + \kappa_{12} r_1 r_2 r_3 r_4 (1-2u_1) + \sigma_{12} r_1 r_2 r_3 r_4 \} \] (6.22)

\[ \tau(g_1', g_2', g_3', g_4') = 16r_1 r_2 r_3 r_4 \] (6.23)

where \( u_1, \sigma_{ij}, \kappa_{ijk}, \tau_{ijkl} \) are the first four central moments of \( U_1, U_2, U_3, U_4 \) from uniform order statistics \( U_1 < U_2 < \ldots < U_N \).

Hence, the coefficients can be calculated by substituting equations (6.13)-(6.23) into (6.9)-(6.12) where the central moments of uniform order statistics can be found in the following Section.

6.4 Moments of Uniform Order Statistics

In this Section we will be concerned with the moments of uniform order statistics. A random sample of size \( N, U_1, U_2, \ldots, U_N \) from a population with continuous cumulative distribution function \( F_U(u) \), the marginal distribution of \( U(i)(U_{a_i}), i=1, \ldots, N \) is given by

\[ f_{U(i)}(u) = N \binom{N-1}{i-1} (F_U(u))^{i-1} (1-F_U(u))^{N-i} f_U(u) \] (6.24)

(Gibbons, 1971). For \( f_U(.) \) uniform in \((0,1)\) mean of the \( i \)th order statistics is easily calculated by

\[ \mu_i = E(U(i)) = N \binom{N-1}{i-1} \int_0^1 u^i (1-u)^{N-i} \, du \]

\[ = \frac{i}{N+1} \] (6.25)
The product moments to obtain $\sigma_{ij}$, $\kappa_{ijk}$ and $\tau_{ijk\ell}$ can be found by considering the marginal distribution of $(U_{(i)}, U_{(j)})$, $(U_{(i)}, U_{(j)}, U_{(k)})$ and $(U_{(i)}, U_{(j)}, U_{(k)}, U_{(\ell)})$ respectively. For example

$$f_{U_{(i)}U_{(j)}U_{(k)}U_{(\ell)}}(u,v,w,t) = K u^{i-1} v^{j-i-1} w^{k-j-1} t^{\ell-k-1} (1-t)^{N-\ell}$$

where $K = \frac{N!}{(i-1)!(j-i-1)!(k-j-1)!(\ell-k-1)!(N-\ell)!}$ with $i<j<k<\ell$.

The new variables

$$u_1 = \frac{u}{v}, \quad v_1 = \frac{v}{w}, \quad w_1 = \frac{w}{t} \quad \text{and} \quad t_1 = t$$

can be defined and substituted in (6.26) to obtain a simpler form

$$f(u_1,v_1,w_1,t_1) = K u_1^{i-1} (1-u_1)^{j-i-1} v_1^{j-1} (1-v_1)^{k-j-1} w_1^{k-1} (1-w_1)^{\ell-k-1} t_1^{\ell-1} (1-t_1)^{N-\ell}$$

which leads to the product moment

$$E(U_{(i)}U_{(j)}U_{(k)}U_{(\ell)}) = \frac{i(j+1)(k+2)(\ell+3)}{(N+1)(N+2)(N+3)(N+4)}$$

(6.27)

Therefore,

$$\sigma_{ij} = \frac{i(N-j+1)}{(N+1)^2(N+2)}$$

(6.28)
\[
\kappa_{ijk} = \frac{2i(N-2j+1)k}{(N+1)^3(N+2)(N+3)}
\]

\[
\tau_{ijk\ell} = \frac{i(j+1)(k+2)(\ell+3)}{(N+1)(N+2)(N+3)(N+4)} - \frac{1}{(N+1)^2} \frac{4ij(N-2k+1)(N-\ell+1)}{(N+2)(N+3)}
\]

\[
+ \frac{3ijk(N-\ell+1)+2ij\ell(N-k+1)+ik\ell(N-j+1)}{N+2}
\]

\[
+2i(N-2j+1)(k(N-2\ell+1)-\ell(N+1))+ijk\ell}
\]

6.5 Calculation of Scores

The programs are written in FORTRAN to find the coefficients \(a_i, b_{ij}, c_{ijk}\) and \(d_{ijk\ell}\); those will be used to obtain the approximation (6.4) to \(pr(Y_{a_1} \leq \ldots \leq Y_{a_N})\) when \(f(.)\) is logistic, and for any \(N \geq 2\). They are calculated and tabulated in Table 6.1 for \(N=2,3,4,5\) and \(r_i=i, i=1,\ldots,N\). Scores are listed for only \(i \leq j \leq k \leq \ell\), i.e. \(n, a_i's; n(n+1)/2, b_{ij}'s; n(n+1)(n+2)/6, c_{ijk}'s\) and \(n(n+1)(n+2)(n+3)/24, d_{ijk\ell}'s\). The ones which are not listed can be found by using:

i) \(b_{ij} = b_{i',j'}\), where \(i' = N+1-i, j' = N+1-j\)

ii) \(c_{ijk} = -c_{i',j',k'}\), where \(i' = N+1-i, j' = N+1-j, k' = N+1-k\)

iii) \(d_{ijk\ell} = d_{i',j',k',\ell'}\), where \(i' = N+1-i, j' = N+1-j, k' = N+1-k, \ell' = N+1-\ell\).
Also:

I) \[ \sum_{i=1}^{N} a_i = 0 \]

II) \[ \sum_{j=1}^{N} b_{i,j} = 0 \text{ for all } i=1, \ldots, N \]

III) \[ \sum_{k=1}^{N} c_{i,j,k} = 0 \text{ for all } i,j=1, \ldots, N \]

IV) \[ \sum_{\ell=1}^{N} d_{i,j,k,\ell} = 0 \text{ for all } i,j,k,\ell=1, \ldots, N \] \hspace{1cm} (6.32)

Despite the complicated looking equations (6.5)-(6.8) or (6.9)-(6.12), a very simple program is written to obtain the logistic scores, since some of the terms are only included when some (or all) of \( i, j, k, \ell \) are equal to each other. These terms are included by using a few carefully chosen 'IF' statements of FORTRAN programming language. First, the scores \( b_{ij}, c_{ijk} \) and \( d_{ijk\ell} \) are taken, as if none of the subscripts are equal to each other, i.e. \( i<j<k<\ell \). Then the cases of equality are considered. It is straightforward for \( b_{ij} \), since there is only one equality case \( i=j \) to include terms with \( \delta_{ij} \). For a more complicated situation for example in calculating \( d_{ijk\ell}, i<j<k<\ell \), cases are considered in the following order where the first three statements are also used for \( c_{ijk} \):

1. 'i=j' to include the terms with \( \delta_{ij} \)
2. 'i=k' to include the terms with \( \delta_{ij}\delta_{jk} \) and \( \delta_{ik} \)
3. 'j=k' to include the terms with \( \delta_{jk} \)
4. 'j=\ell' to include the terms with \( \delta_{jk}\delta_{k\ell} \) and \( \delta_{j\ell} \)
5. 'k=\ell' to include the terms with $\delta_{k\ell}$

6. 'i=\ell' to include the terms with $\delta_{i\ell}$, $\delta_{i\ell j} \delta_{j\ell k}$, $\delta_{i\ell j} \delta_{j\ell k}$, $\delta_{i\ell k} \delta_{k\ell j}$, $\delta_{i\ell k} \delta_{k\ell i}$, $\delta_{i\ell j} \delta_{j\ell k}$

7. 'i=j' and 'k=\ell' to include the terms with $\delta_{i\ell j}$

Calculation of the approximation (6.4), therefore, can be considerably reduced by using the aforementioned algorithm and following formulae:

\[
\sum_{i,j=1}^{N} b_{ij} \theta_i \theta_j = \sum_{i=1}^{N} b_{ii} \theta_i^2 + 2 \sum_{i<j}^{N} b_{ij} \theta_i \theta_j
\]

\[
\sum_{i,j,k=1}^{N} c_{ijk} \theta_i \theta_j \theta_k = \sum_{i=1}^{N} c_{iii} \theta_i^3 + 3 \sum_{i<j}^{N} (c_{iij} \theta_i^2 \theta_j + c_{ijj} \theta_i \theta_j^2) + 6 \sum_{i<j<k}^{N} c_{ijk} \theta_i \theta_j \theta_k
\]

\[
\sum_{i,j,k,l=1}^{N} d_{ijkl} \theta_i \theta_j \theta_k \theta_l = \sum_{i=1}^{N} d_{i\iiii} \theta_i^4 + 4 \sum_{i<j}^{N} (d_{iiij} \theta_i^3 \theta_j + d_{iijj} \theta_i \theta_j^3) + 6 \sum_{i<j}^{N} d_{iijj} \theta_i \theta_j^2
\]

\[
+ 12 \sum_{i<j<k}^{N} (d_{iijk} \theta_i^2 \theta_j \theta_k + d_{iijk} \theta_i \theta_j^2 \theta_k + d_{iikk} \theta_i \theta_j \theta_k^2)
\]

\[
+ 24 \sum_{i<j<k<l}^{N} d_{ijkl} \theta_i \theta_j \theta_k \theta_l
\]
6.6 Approximation to Asymptotic Likelihood

When the distribution of $Y_j$'s, after a monotone (increasing) transformation, are assumed to be logistic with mean $\theta_j$ and variance $\sigma_j^2/3$, the quadratic, cubic and quartic approximation to likelihood for $N<7$ can now be obtained by using the $a_i, b_{ij}, c_{ijk}, d_{ijkl}$ values in Table 6.1 and the approximation

$$f(r|\theta) = \frac{1}{N} \exp \left( \sum_{i=1}^{N} a_i \theta_i - \frac{1}{2} \sum_{i=1}^{N} b_{ij} \theta_i \theta_j + \frac{1}{6} \sum_{i=1}^{N} c_{ijk} \theta_i \theta_j \theta_k \right)$$

$$- \frac{1}{24} \sum_{i=1}^{N} d_{ijkl} \theta_i \theta_j \theta_k \theta_l \right)$$

(6.34)

Terms up to (including) the summation consisting of $b_{ij}$ are taken for quadratic, $c_{ijk}$ for cubic and $d_{ijkl}$ for quartic approximation. Scores for $N>7$ can be found by using the same program.

Listings of coefficients in Table 6.1 reveal that, when $N=3$, $c_{ijk}=0$ for all $i,j,k$, that is the cubic approximation is exactly the same as the quadratic one. When $N=4$, the $c_{ijk}$ for the logistic distribution are of opposite sign to the $c_{ijk}$ for the normal distribution for most $ijk$'s. Therefore, the cubic approximation improvements for the normal distribution may not be observed for the logistic distribution (Pettitt, 1982b). This is an interesting result, since more accurate approximations should be expected when the approximation is taken to higher order terms. The asymptotic likelihood is considered for investigating how appropriate
a Taylor approximation of the form (6.34) to the rank marginal likelihood when the underlying distribution is logistic.

Let \( r_{h_1}, \ldots, r_{h_K} \) be the vectors of every possible rankings where \( K=N! \) and \( r_{h_i} = (r_{h_1}, \ldots, r_{h_N}) \) is the rank vector where ranks are assigned to \( N \) objects in order of preference. The likelihood for set of \( n \) rankings \( r_n \) is given by

\[
f(r_n | \theta) = \prod_{m=1}^{K} f(r_{h_m} | \theta)^{a_m}
\]  

(6.35)

where \( a_m \) is the frequency of the ranking \( r_{h_m} \), \( \sum_{m=1}^{K} a_m = n \).

Each \( f(r_{h_m} | \theta) \) in (6.35) can be approximated by equation (6.34) by replacing \( a_i, b_{ij}, c_{ijk} \) and \( d_{ijk} \) with \( a_{h_i}, b_{h_i h_j}, c_{h_i h_j h_k} \) and \( d_{h_i h_j h_k} \) respectively where the expectations in the definitions of the coefficients are taken according to the rank vector \( r_{h_m} \), \( m=1, \ldots, n \).

The asymptotic likelihood is given by

\[
\lim_{n \to \infty} \left[ f(r_n | \theta) \right]^{1/n} = \prod_{m=1}^{K} f(r_{h_m} | \theta)^{q_m}
\]  

(6.36)

where \( q_m \) is the probability of having rank vectors \( r_{h_m} \) given the value of \( \theta \) generating \( r_{h_m} \) since \( a_m/n \to q_m \). A numerical integration technique, Milton's Algorithm (1970) can be used to obtain \( q_m \): for \( \theta = 0 \)

\[ q_m = \frac{1}{K} \]

Therefore (quadratic, cubic and) quartic approximation to asymptotic likelihood can be written as
where the summations are taken over h=1,...,K, and i,j,k,l=1,...,N.

In fact, \( \mathbf{\theta}^T = (\theta_1,...,\theta_N) \) can be redefined as a vector of dimension N-1, such that \( \mathbf{\theta}^T = (\theta_1-\theta_N, \theta_2-\theta_N,...,\theta_{N-1}-\theta_N,0) \) since the probability \( \text{pr}(Y_{\alpha_1} < ... < Y_{\alpha_N}) \) is invariant under the addition of a constant to the \( \theta_j \)'s. Therefore, summations in (6.37) can be taken over i,j,k,l=1,...,N-1.

6.7 Application to a Two-Sample Population and a Trend Model

Assume that after an unknown monotone transformation, \( Y_j \)'s have an approximate logistic distribution with means \( \theta_1 \) or \( \theta_2 \) according to \( Y_j \)'s belonging to Sample I or Sample II respectively. For \( \theta \) related by a simple linear model, the \( \theta_j \)'s in equation (6.37) should be replaced by \( x_j \beta \) where in two sample case, the dummy variable \( x \) can be taken as \( n_1 \), the size of the Sample I, 1's followed by \( n_2 \) 0's where \( n_1+n_2=N \). Then asymptotic likelihood (6.37) can be found by

\[
\frac{1}{N!} \exp \left\{ \sum_{h=1}^{K} q_h a_h \sum_{i=1}^{n_1} a_{h_i} - \frac{1}{2} \sum_{h,i,j} q_h b_{h_i} b_{h_j} \right\} \sum_{i,j=1}^{n_1} b_{h_i} b_{h_j} + 2 \sum_{i,j=1}^{n_1} b_{h_i} b_{h_j} + 2 \sum_{i,j=1}^{n_1} b_{h_i} b_{h_j} + 6 \sum_{i,j=1}^{n_1} b_{h_i} b_{h_j} + 6 \sum_{i,j=1}^{n_1} b_{h_i} b_{h_j}
\]

\[
= 1 + \frac{1}{6} \beta^2 ( \sum_{i=1}^{n_1} c_{h_i} + c_{h_1} h) + 3 \sum_{i<j} (c_{h_i} h_j + c_{h_1} h_j) + 6 \sum_{i<j<k} c_{h_i} h_j h_k
\]
An alternative approximation; "a quadratic approximation to the quartic approximation to the asymptotic log likelihood" can also be considered with

\[ z_A(\beta) = \text{constant} - \frac{1}{2} (\beta - \hat{\beta})^2 \left. \frac{d^2 \lambda(\beta)}{d\beta^2} \right|_{\beta = \hat{\beta}} \quad (6.39) \]

where \( \lambda(\beta) \) is \(-\log(N!) + \) exponent of equation (6.38) and \( \hat{\beta} \) is the value which maximizes the quartic approximate, equation (6.38). A numerical method can be used to solve the equation \( \frac{d\lambda(\beta)}{d\beta} = 0 \), to obtain \( \hat{\beta} \) in a small neighbourhood of quadratic solution of the equation.

For sample sizes \( N=4,5,6 \) where \( Y_1 \) and \( Y_2 \) are taken to be members of the first sample and the rest of the observations are taken to be members of the second sample for \( \beta_0 \) values 0.5, 1, 1.5, 2. The quadratic, cubic and quartic approximations, equation (6.38) to the asymptotic log likelihood are found over a range of \( \beta \in (\beta_0 - 5, \beta_0 + 5) \) with 0.1 increments. A quadratic approximation to the quartic approximation to the asymptotic...
log likelihood is also obtained over the same range together with the 'exact' values. Graphical displays of approximated log likelihood are made against varying $\beta$ values for comparison. Generally, the quartic approximation seems quite satisfactory for smaller values of $\beta_0$ for all $N$ considered. Its values are almost the same as other approximations and exact values in the close neighbourhood of $\beta_0$. Quartic approximation values are closer (but on the opposite side) to the exact values than quadratic, cubic and $\varepsilon_A(\beta)$ approximations - those are almost identical - in a wider neighbourhood of $\beta_0$. However, for larger values of $\beta_0$, the quartic approximation becomes impractical. Examples can be seen in Diagrams 6.1a,b, where the asymptotic log likelihood values are plotted against $\beta$ values where $\beta_0 = 0.5$ in (a) and $\beta_0 = 2$ in (b), where $N=4$. Symmetric $\varepsilon_A$ gives a better approximation than the others for $\beta>\beta_0$, but far worse than the rest when $\beta<\beta_0$. All other approximations give almost the same as exact values for $\beta<\beta_0$. For $\beta>\beta_0$, quadratic and cubic approximations are close to each other but not close to the exact values. The quartic approximation is very close to the exact values in a small neighbourhood of $\beta_0$. However features like giving far greater values than it should be for some $\beta>\beta_0$ and also being non-unimodel make it a poorer approximation than earlier ones.

An ad hoc solution to the impracticality of quartic approximation is to take the average of cubic, and quartic approximations which is equivalent to dividing $d_{ijk\ell}$ by 2 for all $i,j,k,\ell$. This suggestion appears to give the best approximation of all for every $\beta_0$ and $N$ values tested, however it does not have any theoretical basis.
A trend model is also considered with $\theta_j = j\theta_1$ for $j=1, \ldots, N$, $\theta_1 = \beta$, the approximations are observed for $N=4, 5, 6$ for $\beta_0$ values 0.1, 0.5, 1. In Diagrams 6.2a and b, very similar results to the two sample population results, can be observed where $\beta_0 = 0.5$ and $\beta_0 = 1$ respectively for $N=5$.

6.8 Conclusions

It has been indicated that an approximation of the form

$$\frac{1}{N^t} \exp \left( \frac{a}{4} i_{i} \theta_1 - \frac{1}{2} \sum_{i} b_{ij} \theta_i \theta_j + \frac{1}{6} \sum_{i,j,k} c_{ijk} \theta_i \theta_j \theta_k - \frac{1}{24} \sum_{i,j,k,l} d_{ijkl} \theta_i \theta_j \theta_k \theta_l \right)$$

to the rank marginal likelihood when the density of the transformed $Y_j$'s is logistic may not be appropriate. This point is verified by observing the quadratic, cubic and quartic approximations closely for two sample populations and trend models. The quartic approximation is a good approximation only in the range $(-\infty, \beta_0)$ and should not be used for making inference for $\beta$ since it is not unimodal. It should be noted that the aforementioned approximation is an approximation to

$$\frac{1}{N!} \mathbb{E}(\exp(\log(Z_{r_j}) \theta_j) - \sum_{k=2}^{4} \frac{1}{k!} \mathbb{E} \left( \frac{d}{d\theta_j} \log(Z_{r_j}) \theta_j \right) - \sum_{k=2}^{4} \frac{1}{k!} \mathbb{E} \left( \frac{d^2}{d\theta_j^2} \log(Z_{r_j}) \theta_j \right) + \ldots)$$

which is exact for the normal distribution when truncated to the second order in $\theta_j$ so that the results for the normal distribution involve only one approximation whereas for any other distribution results involve two approximations.
TABLE 6.1:
Logistic scores $a_i$, $b_{ij}$, $c_{ijk}$ and $d_{ijkl}$ for $N = 2, 3, 4, 5$ when $r_i = i$
i = 1, ..., $N$

<table>
<thead>
<tr>
<th>SAMPLE SIZE</th>
<th>2</th>
<th>SAMPLE SIZE</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A SCOPES</td>
<td></td>
<td>A SCOPES</td>
<td></td>
</tr>
<tr>
<td>-0.3333</td>
<td>0.3333</td>
<td>-0.5000</td>
<td>0.0000  0.5000</td>
</tr>
<tr>
<td>B SCOPES</td>
<td></td>
<td>B SCOPES</td>
<td></td>
</tr>
<tr>
<td>0.1111</td>
<td>-0.1111</td>
<td>0.1111</td>
<td></td>
</tr>
<tr>
<td>C SCOPES</td>
<td></td>
<td>C SCOPES</td>
<td></td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>-0.74074194E-02</td>
<td>1 1 1 1</td>
<td>0.52154064E-07</td>
</tr>
<tr>
<td>1 1 2 2</td>
<td>0.74074045E-02</td>
<td>1 1 1 2</td>
<td>0.00000000E+00</td>
</tr>
<tr>
<td>1 2 2 2</td>
<td>-0.74074157E-02</td>
<td>1 2 2 2</td>
<td>0.00000000E+00</td>
</tr>
<tr>
<td>2 2 2 2</td>
<td>0.74074445E-02</td>
<td>2 2 2 2</td>
<td>0.00000000E+00</td>
</tr>
<tr>
<td>D SCOPES</td>
<td></td>
<td>D SCOPES</td>
<td></td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>-0.014817</td>
<td>1 1 1 2</td>
<td>-0.37252930E-08</td>
</tr>
<tr>
<td>1 1 2 2</td>
<td>0.014814</td>
<td>2 2 2 2</td>
<td>-0.44763484E-07</td>
</tr>
<tr>
<td>1 2 2 2</td>
<td>-0.014816</td>
<td>2 2 2 2</td>
<td>-0.0023572</td>
</tr>
<tr>
<td>2 2 2 2</td>
<td>0.014824</td>
<td>2 2 2 2</td>
<td>0.015714</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>-0.023572</td>
<td>2 2 2 2</td>
<td>-0.014286</td>
</tr>
<tr>
<td>1 1 2 2</td>
<td>0.015714</td>
<td>2 2 2 2</td>
<td>0.007857</td>
</tr>
<tr>
<td>1 1 1 2</td>
<td>-0.014286</td>
<td>1 1 2 3</td>
<td>0.015714</td>
</tr>
<tr>
<td>1 1 2 3</td>
<td>0.014286</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>1 2 2 3</td>
<td>-0.0023572</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>2 2 2 3</td>
<td>0.007857</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>2 2 2 3</td>
<td>-0.014286</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>2 2 2 3</td>
<td>0.007857</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>2 2 2 3</td>
<td>-0.014286</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>2 2 2 3</td>
<td>0.007857</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
<tr>
<td>2 2 2 3</td>
<td>-0.014286</td>
<td>2 2 2 3</td>
<td>-0.00429</td>
</tr>
</tbody>
</table>
TABLE 6.1 ... continued

<table>
<thead>
<tr>
<th>SAMPLE SIZE</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>D SCORES</td>
<td></td>
</tr>
<tr>
<td>-0.6000</td>
<td>-0.2000</td>
</tr>
<tr>
<td>B SCORES</td>
<td></td>
</tr>
<tr>
<td>0.1600</td>
<td></td>
</tr>
<tr>
<td>-0.0800</td>
<td>0.2400</td>
</tr>
<tr>
<td>-0.0333</td>
<td>-0.1067</td>
</tr>
<tr>
<td>-0.0267</td>
<td>-0.0533</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C SCORES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D SCORES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
### TABLE 6.1 ... continued

<table>
<thead>
<tr>
<th>SAMPLE SIZE</th>
<th>B SCORES</th>
<th>C SCORES</th>
<th>D SCORES</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.6667</td>
<td>-0.9333</td>
<td>0.5000</td>
</tr>
<tr>
<td></td>
<td>0.8888</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td></td>
<td>-0.9333</td>
<td>0.8888</td>
<td>0.5000</td>
</tr>
<tr>
<td></td>
<td>0.6667</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td></td>
<td>-0.625</td>
<td>0.2540</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.317</td>
<td>-0.8952</td>
<td>0.2857</td>
</tr>
<tr>
<td></td>
<td>-0.0159</td>
<td>-0.2166</td>
<td>0.1587</td>
</tr>
<tr>
<td></td>
<td>0.0967</td>
<td>0.0967</td>
<td>0.0967</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B SCORES</th>
<th>0.1587</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0625</td>
<td>0.2540</td>
</tr>
<tr>
<td>-0.0476</td>
<td>-0.0852</td>
</tr>
<tr>
<td>-0.0317</td>
<td>-0.0952</td>
</tr>
<tr>
<td>-0.0159</td>
<td>-0.0476</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C SCORES</th>
<th>1 2 4 4 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 0</td>
<td>0.2645015E-01</td>
</tr>
<tr>
<td>1 1 2 0</td>
<td>-1.0352604E-01</td>
</tr>
<tr>
<td>1 2 2 0</td>
<td>-0.5291005E-02</td>
</tr>
<tr>
<td>1 3 2 0</td>
<td>-0.8216442E-01</td>
</tr>
<tr>
<td>1 4 2 0</td>
<td>-0.7026514E-02</td>
</tr>
<tr>
<td>2 1 2 0</td>
<td>0.0000006E+00</td>
</tr>
<tr>
<td>2 2 2 0</td>
<td>0.0000006E+00</td>
</tr>
<tr>
<td>2 3 2 0</td>
<td>0.0000006E+00</td>
</tr>
<tr>
<td>2 4 2 0</td>
<td>0.0000006E+00</td>
</tr>
<tr>
<td>3 1 2 0</td>
<td>-0.5291005E-02</td>
</tr>
<tr>
<td>3 2 2 0</td>
<td>-0.5291005E-02</td>
</tr>
<tr>
<td>3 3 2 0</td>
<td>-0.5291005E-02</td>
</tr>
<tr>
<td>3 4 2 0</td>
<td>-0.5291005E-02</td>
</tr>
<tr>
<td>4 1 2 0</td>
<td>-0.2166442E-01</td>
</tr>
<tr>
<td>4 2 2 0</td>
<td>-0.2166442E-01</td>
</tr>
<tr>
<td>4 3 2 0</td>
<td>-0.2166442E-01</td>
</tr>
<tr>
<td>4 4 2 0</td>
<td>-0.2166442E-01</td>
</tr>
<tr>
<td>5 1 2 0</td>
<td>-0.2645015E-02</td>
</tr>
<tr>
<td>5 2 2 0</td>
<td>-0.2645015E-02</td>
</tr>
<tr>
<td>5 3 2 0</td>
<td>-0.2645015E-02</td>
</tr>
<tr>
<td>5 4 2 0</td>
<td>-0.2645015E-02</td>
</tr>
<tr>
<td>5 5 2 0</td>
<td>-0.2645015E-02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D SCORES</th>
<th>3 3 3 3 1</th>
<th>-0.001134</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 2 3</td>
<td>0.000000</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4</td>
<td>0.001479</td>
<td></td>
</tr>
<tr>
<td>1 3 4 5</td>
<td>-0.005569</td>
<td></td>
</tr>
<tr>
<td>1 4 5 6</td>
<td>-0.001134</td>
<td></td>
</tr>
<tr>
<td>2 1 2 3</td>
<td>-0.012472</td>
<td></td>
</tr>
<tr>
<td>2 2 3 4</td>
<td>0.006083</td>
<td></td>
</tr>
<tr>
<td>2 3 4 5</td>
<td>0.013605</td>
<td></td>
</tr>
<tr>
<td>2 4 5 6</td>
<td>0.048016</td>
<td></td>
</tr>
<tr>
<td>3 1 2 3</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>3 2 3 4</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>3 3 4 5</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>3 4 5 6</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>4 1 2 3</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>4 2 3 4</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>4 3 4 5</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>4 4 5 6</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>5 1 2 3</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>5 2 3 4</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>5 3 4 5</td>
<td>0.001147</td>
<td></td>
</tr>
<tr>
<td>5 4 5 6</td>
<td>0.001147</td>
<td></td>
</tr>
</tbody>
</table>

| 1 1 2 3  | -0.001134 |
| 1 2 3 4  | 0.000000 |
| 1 3 4 5  | 0.001479 |
| 1 4 5 6  | -0.005569|
| 2 1 2 3  | -0.012472|
| 2 2 3 4  | 0.006083 |
| 2 3 4 5  | 0.013605 |
| 2 4 5 6  | 0.048016 |
| 3 1 2 3  | 0.001147 |
| 3 2 3 4  | 0.001147 |
| 3 3 4 5  | 0.001147 |
| 3 4 5 6  | 0.001147 |
| 4 1 2 3  | 0.001147 |
| 4 2 3 4  | 0.001147 |
| 4 3 4 5  | 0.001147 |
| 4 4 5 6  | 0.001147 |
| 5 1 2 3  | 0.001147 |
| 5 2 3 4  | 0.001147 |
| 5 3 4 5  | 0.001147 |
| 5 4 5 6  | 0.001147 |

| 1 1 2 3  | -0.001134 |
| 1 2 3 4  | 0.000000 |
| 1 3 4 5  | 0.001479 |
| 1 4 5 6  | -0.005569|
| 2 1 2 3  | -0.012472|
| 2 2 3 4  | 0.006083 |
| 2 3 4 5  | 0.013605 |
| 2 4 5 6  | 0.048016 |
| 3 1 2 3  | 0.001147 |
| 3 2 3 4  | 0.001147 |
| 3 3 4 5  | 0.001147 |
| 3 4 5 6  | 0.001147 |
| 4 1 2 3  | 0.001147 |
| 4 2 3 4  | 0.001147 |
| 4 3 4 5  | 0.001147 |
| 4 4 5 6  | 0.001147 |
| 5 1 2 3  | 0.001147 |
| 5 2 3 4  | 0.001147 |
| 5 3 4 5  | 0.001147 |
| 5 4 5 6  | 0.001147 |

| 1 1 2 3  | -0.001134 |
| 1 2 3 4  | 0.000000 |
| 1 3 4 5  | 0.001479 |
| 1 4 5 6  | -0.005569|
| 2 1 2 3  | -0.012472|
| 2 2 3 4  | 0.006083 |
| 2 3 4 5  | 0.013605 |
| 2 4 5 6  | 0.048016 |
| 3 1 2 3  | 0.001147 |
| 3 2 3 4  | 0.001147 |
| 3 3 4 5  | 0.001147 |
| 3 4 5 6  | 0.001147 |
| 4 1 2 3  | 0.001147 |
| 4 2 3 4  | 0.001147 |
| 4 3 4 5  | 0.001147 |
| 4 4 5 6  | 0.001147 |
| 5 1 2 3  | 0.001147 |
| 5 2 3 4  | 0.001147 |
| 5 3 4 5  | 0.001147 |
| 5 4 5 6  | 0.001147 |
DIAGRAM 6.1a: Comparison of quadratic (2), cubic (3), quartic (4), quadratic approximation to quartic (A) and average of cubic and quartic (AVG) approximations with exact values (E) of log likelihood for two sample population when $\beta_0 = 0.5$, $N=4$
Diagram 6.1b: Comparison of quadratic (2), cubic (3), quartic (4), quadratic approximation to quartic (A) and average of cubic and quartic (AVG) approximations with exact values (E) of log likelihood for two sample population when $\beta_0 = 2$, $N=4$. 
DIAGRAM 6.2a: Comparison of quadratic (2), cubic (3), quartic (4), quadratic approximation to quartic (A) and average of cubic and quartic (AVG) approximations with exact values (E) of log likelihood for a trend model when $\beta_0 = 0.5$, $N=5$. 
DIAGRAM 6.2b: Comparison of quadratic (2), cubic (3), quartic (4), quadratic approximation to quartic (A) and average of cubic and quartic (AVG) approximations with exact values (E) of log likelihood for a trend model when $\beta=1, N=5$. 
CHAPTER 7

APPROXIMATION TO THE RANK MARGINAL LIKELIHOOD
UNDER FIRST ORDER AUTOREGRESSIVE DEPENDENCE

7.1 Introduction

The assumption of independence is essential for most inferential procedures concerning population characteristics. However in many situations the observations are collected sequentially so that dependency between successive observations may be suspected. Non-parametric techniques, and in particular rank based techniques, are robust against model mis-specifications. Yet they are very sensitive to dependence between observations, hence they are not robust against model mis-specification regarding the independence of the observations. The effect of serial correlation on the significance level of the Mann-Whitney statistic is illustrated in Box (1976).

Recently various test statistics based on ranks for detecting serial dependence have been investigated, for example Bartels (1982), Knoke (1977), Gupta and Govindarajulu (1980). Cox (1966) and Bartels (1977) have shown that the first order serial correlation coefficient may be badly affected when the underlying distribution is not normal. Rank version of statistics for testing independence is shown to have good power properties in Bartels (1982).

In this Chapter, the possibility of making inferences for \( \rho \), the parameter in a first order serial correlation model is investigated by considering an approximation to a marginal likelihood based on ranks. Ranks of the observations are invariant under monotone (increasing)
transformations. Therefore, the serial correlation model need not apply to the original observations but to transformed observations where the assumption of normality is reasonable.

7.2 An Approximation to Marginal Likelihood for Dependent Variables

Suppose \( Y_1, \ldots, Y_N \) are random variables with joint density \( f(y) \); where \( r_j \) is the rank of \( Y_j \) and \( \alpha_j \) is the anti-rank of \( Y_j \) as defined in Chapter 2. Recall that the marginal likelihood of ranks is given by

\[
pr(Y_{\alpha_1} < \ldots < Y_{\alpha_N}) = \int_{\Omega} f(y) \, dy \quad (7.1)
\]

where \( \Omega = \{(y_{\alpha_1}, \ldots, y_{\alpha_N}) : -\infty < y_{\alpha_1} < \ldots < y_{\alpha_N} < \infty\} \).

In this Section we will consider the case where the joint distribution of the \( Y \)'s is given by a multivariate normal distribution with zero mean and covariance matrix \( \Sigma_p \), depending on a single parameter \( p \).

Therefore the probability in (7.1) is given by

\[
p(r|\rho) = \int_{\Omega} \frac{1}{(2\pi)^{N/2} |\Sigma_p|^{1/2}} \exp \left\{-\frac{1}{2} y^T \Sigma_p^{-1} y\right\} \, dy \quad (7.2)
\]

The integral needs to be computed numerically or approximated. Consider the Taylor expansion approximation to \( \varepsilon(\rho) = \log p(r|\rho) \):
\[ \ell(p) = \ell(0) + p \ell'(0) + \frac{1}{2} p^2 \ell''(0) \]  

(7.3)

where \( \ell'(p) = \frac{d}{dp} \ell(p) \)

\[ \ell''(p) = \frac{d^2}{dp^2} \ell(p). \]

Also consider the expansion of the \( N \times N \) covariance matrix \( \Sigma_p \) in powers of \( p \) to give

\[ \Sigma_p^{-1} = I + pA + p^2B + p^3C + \ldots \]  

(7.4)

where \( \Sigma_p=0 = \Sigma_p^{-1} = I \), the identity matrix, that is the \( Y \)'s constitute a random sample from the standard normal distribution when \( p = 0 \).

Let \( Z_1 < Z_2 < \ldots < Z_N \) be the order statistics of a random sample of size \( N \) from the standard normal distribution, and \( \xi^T = (Z_{r_1}, \ldots, Z_{r_N}) \), then the first derivative of

\[ \text{Int} = \int_{\Omega} \frac{1}{(2\pi)^{N/2}} \exp\left( -\frac{1}{2} \xi^T (I + pA + p^2B + \ldots) \xi \right) \text{d}\xi \]  

(7.5)

with respect to \( p \) at \( p = 0 \) gives
\[ \frac{d \text{Int}}{d \rho} \bigg|_{\rho=0} = \frac{1}{N!} E \left( -\frac{1}{2} Z^T AZ \right) \quad (7.6) \]

and similarly, the second derivative of (7.5) gives

\[ \frac{d^2 \text{Int}}{d \rho^2} \bigg|_{\rho=0} = \frac{1}{N!} E \left( -Z^T BZ + \frac{1}{4} (Z^T AZ)^2 \right) \quad (7.7) \]

Therefore,

\[ \kappa'(\rho) \bigg|_{\rho=0} = \left( \frac{d}{d \rho} \left( -\frac{1}{2} \log |\Sigma_\rho| \right) \right)_{\rho=0} + \frac{d}{d \rho} \left( \frac{\text{Int}}{\text{Int}} \right) \bigg|_{\rho=0} \]

\[ = d_1 - \frac{1}{2} E (Z^T AZ) \quad (7.8) \]

and

\[ \kappa''(\rho) \bigg|_{\rho=0} = \left\{ \frac{d^2}{d \rho^2} \left( -\frac{1}{2} \log |\Sigma_\rho| \right) \right\}_{\rho=0} + \left\{ \frac{d^2}{d \rho^2} \left( \frac{\text{Int}}{\text{Int}} \right) \right\}_{\rho=0} - \left( \frac{d}{d \rho} \left( \frac{\text{Int}}{\text{Int}} \right) \right)^2_{\rho=0} \]

\[ = d_2 + E(-Z^T BZ + \frac{1}{4} (Z^T AZ)^2) - \frac{1}{4}[E(Z^T AZ)]^2 \]

\[ = d_2 + \frac{1}{4} \text{var} (Z^T AZ) - E(Z^T BZ) \quad (7.9) \]
If $\log|\Sigma_\rho|$ can easily be found as a function of $\rho$, then there is no need to approximate it.

If $\xi$ is the vector of expected values and $\Xi$ is the covariance matrix of normal order statistics in a random sample of size $N$, then let $\bar{m}$ and $\bar{V}$ denote the mean vector and covariance matrix of $Z$ respectively, i.e. $m^T = (\xi_1, \ldots, \xi_N)$ and $(V)_{i,j} = (\Xi)_{r_i,r_j}$. Hence (7.8) can be written as

$$\lambda'(\rho)_{\rho=0} = d_1 - \frac{1}{2} m^T A m - \frac{1}{2} \text{tr}(AV)$$

(7.10)

The expression for $\lambda''(\rho)_{\rho=0}$ involves the variance of a quadratic form in the order statistics $Z_1, \ldots, Z_N$ which itself involves third and fourth order cumulants of the $Z$'s. Asymptotically the distribution of non-extreme order statistics tends to normality. Therefore third and higher order cumulants are assumed to be zero and an approximation to variance can be taken as

$$\text{var} \left( Z^T A Z \right) = 2 \text{tr}(AVAV) + 4 \bar{m}^T AV \bar{m}$$

That is, $\lambda''(\rho)_{\rho=0}$ can be approximated by
The approximation (7.3) can now be computed by substituting (7.10) and (7.11) for a given set of ranks and \( \lambda(0) = -\log N! \). Approximate inference for \( \rho \) can be made by likelihood ratio test of \( H_0 : \rho = 0 \) against \( H_1 : \rho \neq 0 \) by using the approximate log likelihood, \( \lambda(\rho) \), eq. (7.3). The approximation also allows \( \rho \) to be estimated as explained in Section 7.4.

7.3 Serial Correlation Model

A common model to describe temporal dependence is the serial correlation model which can be formulated as follows. Let \( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N \) be independently distributed standard normal random variables, and consider the model

\[
Y_j = \rho Y_{j-1} + \varepsilon_j \quad j = 2, \ldots, N
\]

and

\[
Y_1 = \frac{\varepsilon_1}{\sqrt{1 - \rho^2}} \quad ; \quad |\rho| < 1
\]  

(7.12)

giving a stationary distribution for \( Y \)'s. Following King (1980) the vector of errors \( \varepsilon \) is
\( e = T Y \)

where

\[
T = \begin{bmatrix}
\sqrt{1-\rho^2} & 0 & \cdots & 0 & 0 \\
-\rho & 1 & \cdots & 0 & 0 \\
0 & -\rho & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & -\rho & 1
\end{bmatrix}
\]

Therefore the inverse of covariance matrix of \( Y \)'s is

\[
\Sigma_p^{-1} = (T^{-1} I (T^{-1})^T)^{-1} = T^T T,
\]

i.e.

\[
\Sigma_p^{-1} = \begin{bmatrix}
1 & -\rho & 0 & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\
0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \cdots & -\rho & 1+\rho^2 \\
0 & 0 & \cdots & 0 & -\rho & 1
\end{bmatrix}
\] (7.13)

Equation (7.13) can be written as

\[
\Sigma_p^{-1} = I + \rho A + \rho^2 B \text{ with }
\]
\[
A = \begin{bmatrix}
    0 & -1 & 0 & 0 & \ldots & 0 \\
    -1 & 0 & -1 & 0 & \ldots & 0 \\
    0 & -1 & 0 & -1 & \ldots & 0 \\
    \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & \ldots & \ldots & \ldots & -1 & 0 & -1 \\
    0 & \ldots & \ldots & \ldots & -1 & 0 & \ldots
\end{bmatrix}
\]

(7.14)

and

\[
B = \begin{bmatrix}
    0 & 0 & \ldots & 0 \\
    0 & 1 & \ldots & 0 \\
    0 & 0 & 1 & \ldots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & \ldots & 1 & 0 \\
    0 & \ldots & 0 & 0 & \ldots
\end{bmatrix}
\]

(7.15)

Note that the approximation (7.4) to \( \Sigma_p^{-1} \) is exact with simple A and B. The determinant of \( \Sigma_p \) is

\[
|\Sigma_p| = \frac{1}{|\Sigma_p^{-1}|} = \frac{1}{|T|^2} = \frac{1}{1 - \rho^2}
\]

(7.16)

In view of the simplicity of \( |\Sigma_p| \) there is no need to approximate its logarithm, therefore the approximation (7.3) to logarithm of marginal likelihood of ranks can be written as
\[ \mathcal{L}(\rho) = \mathcal{L}(0) + \frac{1}{2} \log (1 - \rho^2) + a_1 \rho - \frac{1}{2} \delta_2 \rho^2 \]  

(7.17)

where \( a_1 \) and \( a_2 \) are obtained by using (7.10), (7.11) (by omitting the terms \( d_1 \) and \( d_2 \)), (7.14) and (7.15). That is,

\[
a_1 = \mathcal{L}'(\rho) \bigg|_{\rho=0} - d_1 = - \frac{1}{2} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j A_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} V_{ji} \right) \\
= \sum_{i=1}^{N-1} \xi_{r_i} \xi_{r_i+1} + \sum_{i=1}^{N-1} (\Xi) r_i + r_{i+1} 
\]

(7.18)

and

\[
a_2 = - \mathcal{L}''(\rho) \bigg|_{\rho=0} + d_2 = - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} m_{im} A_{ik} A_{jn} V_{jk} V_{ni} \\
- \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} m_{im} A_{ik} \alpha_{jk} \alpha_{ni} V_{kl} \\
+ \sum_{i=1}^{N} \sum_{j=1}^{N} m_{ij} B_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} V_{ji} \\
= - \frac{1}{2} k_1 - k_2 + k_3
\]

(7.19)
where

\[ k_1 = (V_{2,1})^2 + (V_{N-1,N})^2 + 2V_{2,N}V_{N-1,1} + \sum_{i=2}^{N-1} \{2(V_{i-1,1}V_{2,i} + V_{i-1,N}V_{N-1,i}) + V_{i+1,1}V_{2,i} + V_{i+1,N}V_{N-1,i}\} \]

\[ + \sum_{k=2}^{N-1} (V_{i-1,k} + V_{i+1,k})(V_{k-1,i} + V_{k+1,i}) \}

\[ k_2 = m_1(m_2V_{2,1} + m_{N-1}V_{2,N}) + m_N(m_2V_{N-1,1} + m_{N-1}V_{N-1,N}) \]

\[ + \sum_{i=2}^{N-1} \{(m_{i-1} + m_{i+1})(m_1V_{2,i} + m_NV_{N-1,i}) \}

\[ + m_1(m_2(V_{i-1,1} + V_{i+1,1}) + m_{N-1}(V_{i-1,N} + V_{i+1,N}) \}

\[ + \sum_{k=2}^{N-1} (m_{k-1} + m_{k+1}) \sum_{i=2}^{N-1} (V_{i-1,k} + V_{i+1,k}) \}

and

\[ k_3 = \sum_{i=2}^{N-1} (m_i^2 + V_{i,i}) \]
7.4 Inference for Serial Correlation Parameter $\rho$

Various statistics have been developed to test the hypothesis $H_0: \rho = 0$ against $H_1: \rho \neq 0$, based on ranks of the observations, for example Knoke (1977), Gupta and Govindarajulu (1980), Bartels (1982). The approximation (7.3) of Section 7.2 to $t(\rho)$ provides much more extensive inference for $\rho$ to be made than other statistics. An estimate of $\rho$ can be obtained as well as a statistic for testing whether $\rho$ is zero or not.

Note that equation (7.3) can also be written as

$$t(\rho) = \text{constant} - \frac{1}{2} \frac{(\rho - u)^2}{U}$$

(7.20)

where $u = Uz'(0)$, $U = \frac{1}{\tilde{z}''(0)}$ and constant $= \log \left( \frac{1}{N!} \right) - \frac{1}{2} \frac{(\tilde{z}'(0))^2}{\tilde{z}''(0)}$.

Therefore the probability (7.2) can be approximated by

$$\Pr(Y_{a_1} < \ldots < Y_{a_N} | \rho) = \text{constant} \times \exp\left( -\frac{1}{2} \frac{(\rho - u)^2}{U} \right)$$

(7.21)

which is a normal distribution with mean $u$ and variance $U$.

Therefore $\rho$ can be estimated by

$$\hat{\rho}_A = -\frac{z'(0)}{\tilde{z}''(0)} \quad \text{with variance} \quad \sigma^2 = -\frac{1}{\tilde{z}''(0)}$$

(7.22)
and in particular, for model (7.12),

\[ \hat{\rho}_A = \frac{a_1}{a_2 + 1} \]

with variance \( \frac{1}{a_2 + 1} \)

since \( d_1 = 0, d_2 = -1 \).

Now consider the likelihood ratio test of \( H_0 : \rho = 0 \) against \( H_1 : \rho \neq 0 \) by using the approximation to the marginal likelihood of ranks \( p(\rho) \)

\[ -2 \log \left( \frac{p(0)}{p(\hat{\rho}_A)} \right) = -2 \left( -\rho \frac{\lambda'}{2} \rho^2 \kappa''(0) \right) \]

\[ = \frac{(\lambda'(0))^2}{-\kappa''(0)} \]

(7.23)

Therefore the rank likelihood ratio test is to accept \( H_0 \) if

\[ \frac{(\lambda'(0))^2}{-\kappa''(0)} \leq \chi^2_{\alpha} \]

where \( \chi^2_{\alpha} \) is the upper \( \alpha \) point of the \( \chi^2 \) distribution with one degree of freedom.

Therefore the equivalent test for model (7.13) can be taken as

\[ \frac{a_1^2}{a_2 + 1} \leq \chi^2_{\alpha} \]
7.5 Approximation to Marginal Likelihood for Regression Model with Correlated Errors

In this section the approximation to rank marginal likelihood is obtained when the situation is more complicated than Section 7.2, such as after an arbitrary monotone transformation, the observations $Y$'s are assumed to be dependent and normally distributed with mean $\theta^T = (\theta_1, \ldots, \theta_N)$ rather than $0$, and covariance matrix $\Sigma_\rho$. That is, considering the model

$$Y_j = \theta_j + \delta_j \quad j = 1, \ldots, N$$  \hfill (7.24)

where the error term $\delta_j$ follows the model

$$\delta_j = \rho \delta_{j-1} + \epsilon_j \quad j = 2, \ldots, N$$  \hfill (7.25)

$$\delta_1 = \frac{\epsilon_1}{\sqrt{1-\rho^2}} \quad ; \; |\rho| < 1$$

and $\epsilon_j$'s are independently distributed standard normal random variables.

The marginal likelihood of ranks given by (7.1) is then

$$p(r|\tilde{s},\rho) = \int \frac{1}{(2\pi)^{N/2} \mid \Sigma_\rho \mid^{1/2}} \exp \left\{ -\frac{1}{2} (Y - \theta)^T \Sigma_{\rho}^{-1} (Y - \theta) \right\} dy$$  \hfill (7.26)
Let the vector of parameters be \( \phi = (\theta_1, \ldots, \theta_N, \rho) \). Then the Taylor expansion approximation \( I(\phi) \) to \( \log p(r|\theta, \rho) \) is

\[
I(\phi) = I(0) + \phi^T I'(0) + \frac{1}{2} \phi^T I''(0) \phi
\]

(7.27)

where \( I'(\phi) = \begin{bmatrix}
\frac{\partial}{\partial \theta_1} I(\phi) \\
\vdots \\
\frac{\partial}{\partial \theta_N} I(\phi) \\
\frac{\partial}{\partial \rho} I(\phi)
\end{bmatrix} \) and \( -I'' \) is the observed information matrix

\[
I''(\phi) = \begin{bmatrix}
\frac{\partial^2}{\partial \theta_1^2} I(\phi) & \cdots & \frac{\partial^2}{\partial \theta_1 \partial \theta_N} I(\phi) & \frac{\partial^2}{\partial \theta_1 \partial \rho} I(\phi) \\
\vdots & & \vdots & \vdots \\
\frac{\partial^2}{\partial \theta_N \partial \theta_1} I(\phi) & \cdots & \frac{\partial^2}{\partial \theta_N \partial \theta_N} I(\phi) & \frac{\partial^2}{\partial \theta_N \partial \rho} I(\phi) \\
\frac{\partial^2}{\partial \rho \partial \theta_1} I(\phi) & \cdots & \frac{\partial^2}{\partial \rho \partial \theta_N} I(\phi) & \frac{\partial^2}{\partial \rho^2} I(\phi)
\end{bmatrix}
\]

(7.28)

Assume that the inverse of covariance matrix can be written in powers of \( \rho \) as in (7.4). Consider the integral (7.26) without the term \( |\Sigma_\rho|^{-1/2} \) as \( \text{Int}_1 \). Differentials of \( \text{Int}_1 \) with respect to \( \theta \) give
\[ \frac{\partial}{\partial \phi} \text{Int}_1 \bigg|_{\phi=0} = \frac{1}{N} E(Z) \]

\[ \frac{\partial^2}{\partial \theta \partial \phi} \text{Int}_1 \bigg|_{\phi=0} = \frac{1}{N} E(Z \bar{Z}^T - I) \]

whereas differentials of \( \text{Int}_1 \) with respect to \( \rho \) (at \( \phi = 0 \)) are as in (7.6) and (7.7). Second derivatives of \( \text{Int}_1 \) with respect to \( \theta \) and \( \rho \) at \( \phi = 0 \) gives

\[ \frac{\partial^2}{\partial \theta \partial \rho} \text{Int}_1 \bigg|_{\phi=0} = \frac{1}{N} E(AZ - \frac{1}{2} \bar{Z} \bar{Z}^T AZ) \]

Therefore, following the steps similar to Section 8.2

\[ \varepsilon'(\phi) \bigg|_{\phi=0} = \begin{bmatrix} \frac{\partial}{\partial \phi} \varepsilon(\phi) \end{bmatrix}_{N \times 1} = \begin{bmatrix} E(Z) \end{bmatrix} \]

That is, if \( \mu \) and \( V \) are the mean vector and covariance matrix as in Section 7.2,

\[ \varepsilon'(\phi) \bigg|_{\phi=0} = \begin{bmatrix} \mu \\ d_1 - \frac{1}{2} \bar{Z}^T A \bar{Z} - \frac{1}{2} \text{tr}(AV) \end{bmatrix} \]
where \( d_1 = \left\{ \frac{\partial}{\partial \phi} \left(- \frac{1}{2} \log |\Sigma_{\phi}| \right) \right\}_{\phi=0} \)

The matrix \( \mathcal{L}''(\phi) \bigg|_{\phi=0} \) is found as

\[
\frac{\partial^2}{\partial \phi \partial \phi'} \mathcal{L}(\phi) \bigg|_{\phi=0} = E(\mathbf{Z} \mathbf{Z}^T - I) - E(\mathbf{Z})E(\mathbf{Z}^T) = \text{var}(\mathbf{Z}) - I
\]

\[
\frac{\partial^2}{\partial \phi \partial \phi'} \mathcal{L}(\phi) \bigg|_{\phi=0} = E(\mathbf{A} \mathbf{Z} - \frac{1}{2} \mathbf{Z} \mathbf{Z}^T \mathbf{A} \mathbf{Z}) + \frac{1}{2} E(\mathbf{Z}) E(\mathbf{Z}^T \mathbf{A} \mathbf{Z}) = E(\mathbf{A} \mathbf{Z}) - \frac{1}{2} \text{cov}(\mathbf{Z}, \mathbf{Z}^T \mathbf{A} \mathbf{Z})
\]

The approximation

\[
\text{cov}(\mathbf{Z}, \mathbf{Z}^T \mathbf{A} \mathbf{Z}) = 2 \mathbf{V} \mathbf{A} \mathbf{m}
\]

can be substituted in \( \frac{\partial^2}{\partial \phi \partial \phi'} \mathcal{L}(\phi) \bigg|_{\phi=0} \). This is equivalent to assuming that distribution of the order statistics is normal.

Therefore

\[
\mathcal{L}''(\phi) \bigg|_{\phi=0} \propto \begin{bmatrix}
(V-I)_{N \times N} & (I-V)\mathbf{A} \mathbf{m} \\
0 & d_2 + \frac{1}{2} \text{tr}(\mathbf{A} \mathbf{V} \mathbf{A}^T) + \mathbf{m}^T \mathbf{A} \mathbf{V} \mathbf{m}^T - \text{tr}(\mathbf{B} \mathbf{V})
\end{bmatrix} \tag{7.30}
\]
where \( d_2 = \left\{ \frac{2}{\sigma^2} \left( -\frac{1}{2} \log |\Sigma| \right) \right\}_{\psi = 0} \)

Note that the approximation (7.27) can be written as

\[
\mathcal{L}(\varphi) = \text{constant} - \frac{1}{2} (\varphi - \mu)^T U^{-1} (\varphi - \mu) \tag{7.31}
\]

where \( \mu = U \varphi'(0) \) and \( U^{-1} = -\varphi''(0) \)

which is approximate normal with mean \( \mu \) and covariance matrix \( U \).

Therefore inference for \( \theta \) can be made by considering the marginal distribution of \( \theta \) that can be found by taking the appropriate components of \( \mu \) and \( U \). The simple regression model is considered to illustrate it in the following section.

### 7.6 Test of Location Parameter for Simple Regression Model with Correlated Errors

When the model (7.24) is considered for a simple regression model, i.e. \( \theta = X\beta \), the vector of parameters can be taken as \( \psi = (\beta) \). In this particular case \( \varphi'(\varphi) \) and \( \varphi''(\varphi) \) become

\[
\mathcal{L}'(\varphi) \bigg|_{\psi = 0} = \begin{bmatrix} X^T m \\ \frac{1}{2} m^T Am - \frac{1}{2} \text{tr}(AV) \end{bmatrix} \tag{7.32}
\]
The statistic \(\frac{\partial \xi}{\partial \beta}\) is the locally most powerful rank statistic to test \(\beta = 0\) for independent random variables (Hajek and Sidak, 1967). Similarly \(\frac{\partial \xi}{\partial p}\) is the locally most powerful rank statistic to test \(p = 0\) for serially correlated variables (Gupta and Govindarajulu, 1980).

Let the entries of \(\xi' (\phi)\) and \(-\xi'' (\phi)\) be denoted by \(\xi'_{i}\) and \(\xi''_{ij}\) for \(i,j = 1,2\). Then the distribution of \(\phi\) can be approximated by multivariate normal distribution (eq. 7.31) with mean.

\[
\begin{pmatrix}
\frac{1}{2} \\
\frac{1}{2} \\
\end{pmatrix}
\]

and covariance matrix

\[
\begin{pmatrix}
\xi_{22} & \xi_{12} \\
\xi_{12} & \xi_{11} \\
\end{pmatrix}
\]
Therefore the approximate marginal distribution of $\hat{\beta}$ is

$$\hat{\beta} \sim N \left( \frac{\ell_{12}^n}{\ell_{11}^n \ell_{22}^n - (\ell_{12}^n)^2}, \frac{\ell_{22}^n}{\ell_{11}^n \ell_{22}^n - (\ell_{12}^n)^2} \right)$$  \hspace{1cm} (7.34)$$

Under the hypothesis $H_0: \beta = 0$ the distribution of the test

$$T_R = \frac{E(\beta)}{\sqrt{\sigma_\beta^2}} = \frac{\ell_{12}^n}{\sqrt{(\ell_{11}^n \ell_{22}^n - (\ell_{12}^n)^2)}} \ell_{22}^n$$  \hspace{1cm} (7.35)$$

is approximately $N(0,1)$. The performance of the approximation to the rank marginal likelihood when the $Y$'s are related by regression model with correlated errors will be investigated by undertaking various simulation experiments in Chapter 8.
CHAPTER 8
APPLICATION OF THE APPROXIMATION TO THE
RANK MARGINAL LIKELIHOOD UNDER DEPENDENCE

1.1 Introduction
In this Chapter, the performance of the approximation to rank marginal likelihood for dependent variables is investigated. The investigation is made by comparing:

a) The values of the approximated likelihood with calculated values by using numerical techniques

b) The results of the tests based upon approximated likelihood with various tests based on ranks to make inference for serial correlation parameter \( \rho \)

c) Curvature values with suggested values for a 'good approximation'.

Furthermore, various designs are considered to investigate the 'goodness' of the approximation to find the null distribution of the test of location parameter. Some modifications are suggested and implemented to overcome the difficulties faced during the applications.

The multiple integral,

\[
pr(Y_{a_1} < \ldots < Y_{a_N}) = \int_{\Omega} f(y)dy
\]

where \( \Omega = \{(y_{a_1}, y_{a_2}, \ldots, y_{a_N}) = -\infty < y_{a_1} < \ldots < y_{a_N} < \infty\} \) and \( f(y) \) is the joint density of the random variables \( Y_1, \ldots, Y_N \), is complex and
numerical methods do not exist for the choice of density \( f(y) \) as multivariate normal with zero mean and covariance matrix \( \Sigma \). A computational method is developed for evaluating the integral (8.1).

### 8.2 Algorithm to Compute Marginal Likelihood of Ranks

The technique to find \( \text{pr}(Y_{a_1} < ... < Y_{a_N}) \) is based upon the idea of representing any definite integral as an expected value of some random variable. The Sample Mean Monte Carlo method is considered by Rubinstein (1981) for a single integral. This idea is to compute

\[
E\left(\frac{h(x)}{g_X(x)}\right) = \int_a^b \frac{h(x)}{g_X(x)} g_X(x) \, dx
\]  

(8.2)

to find \( \int_a^b h(x) \, dx \), where \( g_X(x) \) is any probability density function.

In the case of taking

\[
g_X(x) = \begin{cases} 
  \frac{1}{b-a} & \text{if } a < x < b \\
  0 & \text{otherwise}
\end{cases}
\]

The integral is estimated by

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{h(x_i)}{g_X(x_i)} = (b-a) \frac{1}{N} \sum_{i=1}^{N} h(x_i)
\]
The same idea is now implemented to find the multiple integral (8.1).

Let the region \( \Omega' \) be a subset of \( \mathbb{R}^N \) such that \( a < y_{a1} < \ldots < y_{aN} < b \).

Then the integral \( \int_{\Omega'} f(y) \, dy \) can be estimated by

\[
\text{Vol.} \times \frac{1}{N_0} \sum_{i=1}^{N_0} f(y_i)
\]

where Vol denotes the volume of the region \( \Omega' \).

The following algorithm is used to evaluate

\[
\int_{\Omega'} \exp \left\{ -\frac{1}{2} y^T \Sigma^{-1} y \right\} \, dy \quad (8.3)
\]

1. Generate a sequence of random real numbers \( U_1, \ldots, U_N \) between 0 and 1.
2. Set \( y_i = (b-a) U_i + a, \quad i = 1, \ldots, N \)
3. Sort \( y_i \)'s so that \( y(1) < \ldots < y(N) \)
4. Calculate \( f_j(y) \) where \( f_j \) is the integrand of (8.3) and \( \Sigma = (y(r_1), \ldots, y(r_N)) \)
5. Repeat steps 1-4 \( N_0 \) times giving \( N_0 f_j(y) \) values
6. Estimate the integral (8.3) by

\[
\frac{1}{N_0} (b-a)^N \times \frac{1}{N_0} \sum_{j=1}^{N_0} f_j(y)
\]
where \( \frac{1}{N!} (b-a)^N \) is the volume of the region \( y_{a_1} < ... < y_{a_N} \).

Therefore the integral (8.1) for serial correlation model
\[ Y_j = \rho Y_{j-1} + e_j, j=2,\ldots,N \]
is computed as
\[ \frac{1}{N!} ((b-a) \times 0.3989)^N \sqrt{1-\rho^2} \frac{1}{N_0} \sum_{j=1}^{N_0} f_j(y) \]  
(8.4)

8.3 Details of Computation

In this experiment random points in the N dimensional Cartesian product of \([a,b]\) has been generated by using G05CAF subroutine of NAG (1977) library. \(a\) and \(b\) were taken as \(b = -a = 4\) since the probability of \(|y_j| > 4\) is very small for a normally distributed \(y_j\), with mean zero and variance \(1 + \rho^2 (|\rho| < 1)\). \(N_0 = 400,000\) replications have been made to obtain reasonably small standard errors.

For a given ranking \(p(r|\rho)\) was found using the same random number for \(\rho = -0.9, -0.6, -0.3, (0), 0.3, 0.6\) and 0.9 to save CPU times. However application of the numerical algorithm still needed a long computation time, therefore \(N_0\) would not be chosen greater than 400,000. To obtain some more information, the value of the integral was computed for some additional \(\rho\) values in the interval that the maximum was indicated to occur. For example if there is enough evidence that \(p(r|\rho)\) reached a maximum value in \((-0.3,0)\) then \(p(r|\rho)\) was also computed for \(-0.20, -0.15, -0.10\).
These estimates were then smoothed by using the formula

$$\text{smooth } p(r|\rho) = p(r|\rho=0) \sqrt{1-\rho^2} \exp\{b_1\rho - \frac{1}{2} b_2 \rho^2 + b_3 \rho^3\} \quad (8.5)$$

so that values of $b_1$, $b_2$ are directly comparable with the coefficients $a_1$, $a_2$ of the approximation of the form

$$\varepsilon(\rho) = \varepsilon(0) + \frac{1}{2} \log (1-\rho^2) + a_1\rho - \frac{1}{2} a_2\rho^2 \quad (8.6)$$

8.4 Results

In Table 8.1, values of the integral found by using (i) approximation, (ii) numerical algorithm and (iii) smoothing formula are listed for some rankings where the sample size is 6 or 8. The smoothed values of $p(r|\rho)$ were found by using least squares technique (MINITAB) and were well within a Monte Carlo standard error of the estimated values (about 5 per cent of the estimated probability).

In Table 8.1 some values of the coefficients $a_1, a_2$ from formula of approximation (8.6) and $b_1, b_2, b_3$ from the smoothed Monte Carlo estimates (8.5) are also given. The numerical results are very limited in their extent, but the results of some typical cases do show that there is reasonable agreement between $a_1$ and $b_1$, for both values of $N$ and reasonable agreement between $a_2$ and $b_2$ for $N=8$, although the values of $b_3$ tend to be quite large showing that the quadratic approximation is not that adequate at larger values of $|\rho|$.
Comparing the Monte Carlo values of the probability with the approximation shows that the approximation is within 10 per cent (two standard errors) of the Monte Carlo values for cases listed in Table 8.1 (which are very similar to the cases considered but not listed) at least for $|\rho| \leq 0.3$. In fact it is observed that the range of $\rho$ values for which the approximation within 10 per cent of the numerical solution of the integral is wider if $\hat{\rho}_A$ is positive. For example, the ranges are $(-0.4, 0.4)$, $(-0.5, 0.4)$, $(-0.9, 0.5)$ for the first three examples where $N = 8$, where $\hat{\rho}_A$ is $-0.44$, $-0.09$ and $0.62$ respectively, i.e. the approximation appears better for $\hat{\rho}_A > 0$ than for $\hat{\rho}_A < 0$.

8.5 Observed Information

The approximation

$$\mathcal{I}_A(\rho)\bigg|_{\rho=0} = d_2 + \frac{1}{2} \text{tr}(AVAV) + m^T AVAm - m^T Bm - \text{tr}(BV)$$

to

$$\mathcal{I}(\rho)\bigg|_{\rho=0} = d_2 + \frac{1}{4} \text{var}(Z^T AZ) - E(Z^T BZ)$$

has also been investigated numerically to a limited extent.

Theoretically,

$$E(-\mathcal{I}''(\rho)\bigg|_{\rho=0}) = \text{var}(\mathcal{I}'(\rho)\bigg|_{\rho=0}) \quad (8.7)$$
where the expectations are over the randomization distribution of the ranks when \( p = 0 \). Therefore, if the approximation is reasonable, \( E(-z_A^n(p))_{p=0} \) should be close to \( \text{var}(z'(p))_{p=0} \).

\[ \text{var}(z') \text{ and } E(-z_A^n) \text{ are listed in Table 8.2 for selected sample sizes.} \]

The possible \( N! \) arrangements of the \( N \) ranks are used for \( N = 5,6,7 \)
whereas a random selection of 4000 arrangements are used for \( N = 10-15 \)
because of computation time limits. The difference, \( E(z_A^n(p))_{p=0} - \text{var}(z'(p))_{p=0} \) becomes smaller when the sample size becomes larger,
e.g. difference is 0.2 for \( N=5 \), whereas 0.09 for \( N=15 \). Therefore,
it is concluded that \( z_A^n \) is a reasonable approximation to \( z^n \) for larger \( N \).

### 8.6 An Illustrative Example and Comparison with Some Non-Parametric Techniques

Inference for the serial correlation parameter \( p \) can be made by using ranks as explained in Section 7.4 when observations are available on \( Y \)'s satisfying the model
\[ Y_j = \rho Y_{j-1} + \epsilon_j \text{ for } j=2,\ldots,N; \quad Y_1 = \frac{\epsilon}{\sqrt{1-\rho^2}} \]
where \(|\rho| < 1\) and \( \epsilon_j \)'s are independently distributed standard normal random variables. If this were the case, inference for \( p \) can be made by using observations themselves and the (parametric) statistic
\[ S_{\text{par}} = \frac{\sum_{i=1}^{N-1} Y_i Y_{i+1} - n \overline{Y}^2}{\sum_{i=1}^{N} (Y_i - \overline{Y})^2} \quad (8.8) \]
studied by Anderson (1942). However, if the model above is assumed to be satisfied after an arbitrary monotone (increasing) transformation then rank techniques are useful to make inference for \( \rho \). We expect inference for \( \rho \) made by using ranks to be reasonably invariant to the choice of error distribution in the serial correlation model of Section 7.4. Thus, if the \( e_j \)'s were taken to have some other distribution than the normal, then inference for \( \rho \) based on the ranks of the observations would remain almost unchanged, as experienced in the study of other models using ranks (Acar and Pettitt 1984).

Consider a sample size \( N = 20 \) which gives the sequence of ranks

\[
\mathbf{r} = (6, 8, 14, 20, 17, 18, 11, 19, 16, 3, 12, 4, 10, 15, 2, 9, 5, 1, 13, 7)^T
\]

in serial order. This sample of ranks was obtained by generating errors \( e_j \) with the standard normal distribution by using the NAG (1977) subroutine G05DDF and \( \rho = 0.5 \).

For these data, the values of \( a_1 \) and \( a_2 \) are found by using equations (7.18) and (7.19) as 4.77 and 14.09 respectively. The NAG (1977) subroutine E04ABF is used to minimize resulting log likelihood function, i.e. maximize \( \ell(\rho) \) of equation (7.18), to obtain \( \hat{\rho}_m = 0.31 \). When \( \rho \) is estimated by equation (7.22) i.e. \( \hat{\rho}_A = \frac{\hat{a}_1}{\hat{a}_2 + 1} \), a very close value is found, \( \hat{\rho}_A = 0.32 \). The approximate likelihood ratio test statistics is \( 2 \ell_A(\hat{\rho}_A) - 2\ell(0) = 1.50 \), thus the hypothesis \( \rho = 0 \) is well supported by the data on referring 1.50 to the \( \chi^2 \) distribution with one degree of freedom.
Gupta and Govindarajulu (1980) proposed the statistics based on ranks

\[ T_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N-1} \left( Z_{r_i} - Z_{r_{i+1}} \right) \tag{8.9} \]

for first order autoregressive model where \( Z_1 < \ldots < Z_N \) are the order statistics of a random sample of size \( N \) and \( r_i \) is the rank of the observation \( Y_i \). They also conducted a simulation study to show the statistic \( T_N \) is more powerful than \( S_{\text{par}} \). In the same paper, the mean and variance of \( T_N \) under the hypothesis of randomness, \( \rho = 0 \) are given by zero and

\[ \sigma^2(T_N) = \frac{1}{N^2} \sum_{i,j=1}^{N} (E(Z_iZ_j)^2 - \frac{N+1}{N^2(N-1)} \sum_{i=1}^{N} (E(Z_i^2))^2 + \frac{2}{N(N-1)} \tag{8.10} \]

respectively.

It should be noted that \( \sqrt{N} x T_N \) is equivalent to the rank statistics

\[ a_1 = \lambda'(\rho)\big|_{\rho=0} = \sum_{i=1}^{N-1} \xi_{r_i} \xi_{r_{i+1}} + \sum_{i=1}^{N-1} (\xi)_{r_i} r_{i+1} \]

hence, under the hypothesis \( H_0: \rho = 0 \), \( a_1 \) has an approximate normal distribution with mean of zero and variance equal to 14.17 for \( N=20 \).

We also note that in the approximation of the form (8.6), the observed information at \( \rho = 0 \) is given by \( a_2 + 1 \) and the value of expected information, \( E(-\lambda''(\rho))\big|_{\rho=0} = 14.17 \) is close to the value of the observed information 15.09.
Using expected information instead of observed information gives $\hat{\rho}_E = 0.34$ and the $\chi^2$ value equal to 1.61.

An alternative, estimate of $\rho$ can be derived from Bartels (1982) who considered a rank version of Von Neumann's test statistics which is given by

$$RVN = \frac{\sum_{i=1}^{N-1} (r_i - r_{i+1})^2}{\sum_{i=1}^{N} (r_i - \bar{r})^2}$$  \hspace{1cm} (8.11)

where $r_i$ is the rank of the observation $Y_i$.

Critical values for RVN are also tabulated and an estimate of circular rank correlation parameter $\rho'$ is given by

$$\hat{\rho}_{RVN} = 1 - \frac{1}{2} \frac{\sum_{i=1}^{N} (r_i - r_{i+1})^2}{N(N^2-1)/12}$$

where $r_{N+1}$ is taken to be equal to $r_1$, in the same paper.

For the data above, RVN is found to be equal to 1.52 and corresponds to an estimate of $\hat{\rho}_{RVN} = 0.24$. A one-sided $P$-value corresponding to 1.52 is in excess of 10 per cent.

Several examples like this were considered and it was observed that the method based on rank marginal likelihood was in agreement with
other non-parametric techniques. All the comparisons give similar results of testing hypothesis $H_0: \rho = 0$.

It was also observed that generally the rank version of Von Neumann statistics gives estimates $\hat{\rho}_{\text{RVN}}$ whose value are slightly closer to the real value of $\rho$ than $\hat{\rho}_A$. A small Monte Carlo experiment was undertaken to confirm this point. Mean Squared Error (MSE) and Mean Absolute Errors (MAE) were found for varying $\rho$ values, $\pm 0.6$, $\pm 0.3$, $\pm 0.2$, $\pm 0.1$ and $0$, where the model (7.13) was considered with errors generated by NAG (1977) subroutine G05DDF for $N=20$. Results of 100 iterations to compare RVN and the rank approximation with observed and expected information, show that, regardless of $\rho$ value, MSE of $\hat{\rho}_{\text{RVN}}$ is around 0.05 whereas MSE of $\hat{\rho}_A$ varies in the range 0.07 to 0.10. $\hat{\rho}_E$ gives very similar results to $\hat{\rho}_{\text{RVN}}$ with MSE equal to 0.06 everywhere except at the end points $\rho = \pm 0.6$, where the MSE is 0.04. The MAE's follow a very similar pattern.

8.7 Curvature

An alternative way of investigating the approximation to the rank marginal log likelihood

$$
\ell(\rho) = \ell(0) + \rho \ell'(\rho) + \frac{1}{2} \rho^2 \ell''(\rho)
$$

(8.12)

is to consider the curvature $\gamma$ of the likelihood function of ranks

$$
p(r|\rho) = \int_{\Omega} f(y) \, dy
$$
Analytically, the curvature of a curve at $X$ is given by

$$
\gamma_X = \left( \frac{Y''}{1 + (Y')^2} \right)^{\frac{1}{2}}
$$

where $Y = Y(X)$ defines $\mathcal{Y}$, $Y'$ and $Y''$ are first and second derivatives with respect to $X$ which are assumed to exist continuously in a neighbourhood of the value $X$ where the curvature is being evaluated and interpreted as the rate of change of direction of $\mathcal{Y}$ with respect to arc length along the curve.

Statistical curvature is defined in Efron (1975) by

$$
\gamma(\rho) = \frac{1}{i(\rho)} \left[ \text{var}(\mathcal{Z}''(\rho)) - \frac{\text{cov}(\mathcal{Z}'(\rho), \mathcal{Z}''(\rho))}{i(\rho)} \right]^{\frac{3}{2}} \quad (8.13)
$$

where $i_\rho = E(-\mathcal{Z}''(\rho)) = E(\mathcal{Z}'(\rho)^2)$ and is interpreted as a measure of how well a distribution can be approximated, near a given parameter value by expanding log likelihood in a Taylor series.

He also suggested in the same paper that if $\gamma^2 \geq \frac{1}{8}$ then the curvature is 'seriously large'.

The curvature at $\rho=0$ can be approximated for the serial correlation model of Section 7.3. For the likelihood based on ranks, $\mathcal{Z}'(\rho)]_{\rho=0}$ can be found exactly as in equation (7.8), however, $\mathcal{Z}''(\rho)]_{\rho=0}$, equation
(7.9) has to be approximated by \( z_A''(\rho) \) \( \rho=0 \), equation (7.11). The experiment is undertaken to find the expectations of \( z'(\rho) \) \( \rho=0 \) and \( z_A''(\rho) \) \( \rho=0 \) required for the curvature at \( \rho=0 \). Table 8.2 gives details of numerical results and estimates of the curvature.

We note that \( \gamma^2 \) is estimated to be less than \( 1/8 \) for \( N>11 \). Using the exact value of \( \text{var}(z'(\rho)\mid \rho=0) \) instead of \( E(-z''_A(\rho)\mid \rho=0) \) does not make much difference since the two values are close to each other for larger \( N \).

Incidentally, Efron (1975) gave the curvature for the serial correlation model with normal observations. He found, for \( \rho=0 \), \( \gamma^2 = \frac{8N+2}{(N+1)} \), which is substantially larger than the curvature based on the ranks model, for example \( N=15 \) gave \( \gamma^2 = 0.227 \) compared with 0.061 (or 0.060) for the ranks model when \( \text{var}(z') \) (or \( E(-z''_A) \)) was used in equation (8.13).

Thus from this limited study, the indications are that the curvature of the marginal likelihood based on ranks is reasonably small for increasing \( N \), and thus to make the approximate inference for \( \rho \) based on equation (8.12) is adequate for larger values of \( N \). The effect of marginalization over ranks of the serial correlation model appears to reduce the curvature of the resulting likelihood function. However, this is not a general effect because it does not occur with a marginal likelihood of ranks based on the normal location shift model where the original observations likelihood has zero curvature and rank likelihood has non-zero curvature (see Reid 1984).
8.8 Application to a Two-Sample Problem

According to the theory developed in Sections 7.5 and 7.6, the approximation to the rank marginal likelihood for the linear model, should allow the effect of serial correlation in a simple regression problem or the two-sample problem to be investigated if the approximation is reasonably good. When the model $Y_j = x_j \beta + \delta_j$, where the error term $\delta_j$ is taken as in Section 7.5 (eq. 7.25) is considered the marginal distribution of $\beta$ is given in 7.34. Under the null hypothesis $H_0: \beta = 0$ the distribution of $T_{R\phi}$, (eq. 7.35), which can be written as

$$T_{R\phi} = \frac{\lambda_1' - \frac{\lambda_{12}'}{\lambda_{22}} \lambda_2'}{(\lambda_{11} - \frac{(\lambda_{12}')^2}{\lambda_{22}})^{\frac{1}{2}}}$$

should be standard normal where $\lambda_1$ and $\lambda_1''$ denote the entries of $\lambda'(\phi)|_{\phi=0}$ and $-\lambda''_A(\phi)|_{\phi=0}$ respectively. This criterion can be used to assess the goodness of the approximation to the rank marginal likelihood for the linear model with correlated errors. For two-sample problem, with dummy variable $X$ and location parameter $x_j \beta$

i.e.

$$x_j = \begin{cases} 
1 & \text{for observations from sample I} \\
0 & \text{for observations from sample II}
\end{cases}$$

where $n_1 + n_2 = N$
\( n_i \) is the size of the \( i^{th} \) sample, the entries \( r_i^* \), \( r_{ij}^* \) in \( T_{R,\phi} \) become

\[
\begin{align*}
\xi_{11} &= \sum_{i=1}^{N-1} r_i \xi_i \\
\xi_{21} &= \sum_{i=1}^{N-1} \xi_i r_{i+1} + \sum_{i=1}^{N-1} (\xi_i^*) r_i r_{i+1} \\
\xi_{12} &= \sum_{j=2}^{N-1} \sum_{i=1}^{N-1} (\delta_{ij} - (\xi_i^*) r_i r_j)(\xi_j r_{j-1} + \xi_j r_{j+1}) \\
\xi_{22} &= -\frac{1}{2} k_1 - k_2 + k_3 \text{ where } k_1, k_2, k_3 \text{ are as in (7.19)}.
\end{align*}
\]

An A simulation study was undertaken to find the distribution of \( T_{R,\phi} \) for various designs for the two-sample set-up when \( \beta = 0 \).

Independent normal errors \( \xi \) are generated by using NAG (1977) subroutine G05DDF and \( \rho \) is given values \( 0, \pm 0.15, \pm 0.30, \pm 0.45 \) and \( \pm 0.60 \) to
obtain serially correlated errors \( \delta \) as in model (7.25). The program is written in FORTRAN to allocate the observations in either sample \( I, Y_j = \theta_j + \delta_j \) or in sample \( II, Y_j = \delta_j \); randomly and in the following sequences:

a) I I II II I I II II ...

b) I II I II ...

c) I I ... I II II ... II

The mean and variance of \( T_{R\phi} \) is calculated for \( N=20 \), based on 3000 replications for each \( \rho \). The mean of \( T_{R\phi} \) is always close to zero however, the variance of it is far greater than one for allocation c.

The results are summarized in Table 8.3 for allocation a, as expected \( T_{R\phi} \) had (near) zero mean and variance close to one and \( T_{R\phi} \) outperforms the test statistics \( S_{R\phi} = \frac{2^{1/2}}{\sqrt{n}} \) which is based on the approximation to the rank marginal likelihood for independent variables.

For allocation b \( \text{var}(T_{R\phi}) \) is affected by varying \( \rho \) however it is less affected than \( \text{var}(S_{R\phi}) \).

8.9 Conclusions

Various extensive and limited Monte Carlo studies in this Chapter, to compare the values of the approximated rank marginal likelihood with exact values of the integral, found by using a computational algorithm, the test based on this approximation with existing rank
based tests; and small curvature values for $N > 11$, show that the approximation proposed in Chapter 7 for serially correlated random variables is a good one for $|\rho| < 0.3$. The test statistics for the location parameter based on the approximation to the rank marginal likelihood under dependence is more robust against serially correlated errors than similar test statistics based on the standard linear model for some designs of a two-sample population.
Some examples of the marginal rank likelihood (integral eq. 7.2) found by using approximation $\xi_A(p) = \xi(0) + \frac{1}{2} \log (1-p^2) + a_1 p - \frac{1}{2} a_2 p^2$
with numerical solution $\xi_N(p)$ of the integral for serially correlated random variables. The model $Y_j = \rho Y_{j-1} + \varepsilon_j$, $j=2,...,N$ and
$Y_1 = \frac{\varepsilon_1}{\sqrt{1-\rho^2}}$, $|\rho| < 1$ is considered where $\varepsilon_j$'s are independently distributed standard normal random variables. The numerical results are smoothed by using $\xi_{NS}(p) = \xi(0) + \frac{1}{2} \log (1-p^2) + b_1 p - \frac{1}{2} b_2 p^2 + b_3 p^3$.

Example 1:
$N = 8$ ranks = 5, 8, 1, 3, 4, 2, 7, 6

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\xi_N(p) \times 10^5$</th>
<th>se $\times 10^6$</th>
<th>$\xi_{NS}(p) \times 10^5$</th>
<th>$\xi_A(p) \times 10^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.199</td>
<td>0.168</td>
<td>0.188</td>
<td>0.273</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.999</td>
<td>0.529</td>
<td>1.107</td>
<td>1.188</td>
</tr>
<tr>
<td>-0.3</td>
<td>2.128</td>
<td>0.861</td>
<td>2.246</td>
<td>2.242</td>
</tr>
<tr>
<td>0</td>
<td>2.398</td>
<td>0.933</td>
<td>2.480</td>
<td>2.480</td>
</tr>
<tr>
<td>0.3</td>
<td>1.667</td>
<td>0.744</td>
<td>1.806</td>
<td>1.664</td>
</tr>
<tr>
<td>0.6</td>
<td>0.878</td>
<td>0.502</td>
<td>0.980</td>
<td>0.654</td>
</tr>
<tr>
<td>0.9</td>
<td>0.364</td>
<td>0.263</td>
<td>0.342</td>
<td>0.111</td>
</tr>
</tbody>
</table>

$a_1 = -0.50$  \quad a_2 = 4.51
$b_1 = -0.45$  \quad b_2 = 3.59  \quad b_3 = 0.97
TABLE 8.1... continued

Example 2:

\( N = 8 \) ranks = 6, 3, 7, 1, 4, 2, 9, 5

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \xi_N(\rho) \times 10^5 )</th>
<th>se ( \times 10^6 )</th>
<th>( \xi_{NS}(\rho) \times 10^5 )</th>
<th>( \xi_A(\rho) \times 10^5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.671</td>
<td>0.323</td>
<td>0.641</td>
<td>1.561</td>
</tr>
<tr>
<td>-0.6</td>
<td>2.550</td>
<td>0.880</td>
<td>2.643</td>
<td>3.418</td>
</tr>
<tr>
<td>-0.3</td>
<td>3.423</td>
<td>1.116</td>
<td>3.478</td>
<td>3.605</td>
</tr>
<tr>
<td>0</td>
<td>2.398</td>
<td>0.933</td>
<td>2.480</td>
<td>2.480</td>
</tr>
<tr>
<td>0.3</td>
<td>1.154</td>
<td>0.613</td>
<td>1.244</td>
<td>1.151</td>
</tr>
<tr>
<td>0.6</td>
<td>0.490</td>
<td>0.392</td>
<td>0.532</td>
<td>0.349</td>
</tr>
<tr>
<td>0.9</td>
<td>0.188</td>
<td>0.205</td>
<td>0.180</td>
<td>0.051</td>
</tr>
</tbody>
</table>

\( a_1 = -1.90 \) \( b_1 = -1.84 \) \( a_2 = 3.32 \) \( b_2 = 2.86 \) \( b_3 = 1.40 \)

Example 3:

\( N = 8 \) ranks = 6, 7, 3, 1, 2, 4, 8, 5

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \xi_N(\rho) \times 10^5 )</th>
<th>se ( \times 10^6 )</th>
<th>( \xi_{NS}(\rho) \times 10^5 )</th>
<th>( \xi_A(\rho) \times 10^5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.030</td>
<td>0.054</td>
<td>0.029</td>
<td>0.019</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.204</td>
<td>0.215</td>
<td>0.219</td>
<td>0.187</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.826</td>
<td>0.510</td>
<td>0.889</td>
<td>0.853</td>
</tr>
<tr>
<td>0</td>
<td>2.398</td>
<td>0.933</td>
<td>2.480</td>
<td>2.480</td>
</tr>
<tr>
<td>0.3</td>
<td>4.588</td>
<td>1.287</td>
<td>4.660</td>
<td>4.754</td>
</tr>
<tr>
<td>0.6</td>
<td>5.209</td>
<td>1.727</td>
<td>5.409</td>
<td>5.802</td>
</tr>
<tr>
<td>0.9</td>
<td>2.921</td>
<td>0.756</td>
<td>2.720</td>
<td>3.332</td>
</tr>
</tbody>
</table>

\( a_1 = 2.86 \) \( a_2 = 3.59 \) \( b_1 = 2.79 \) \( b_2 = 3.34 \) \( b_3 = 0.32 \)
**Example 4:**

\[ N = 6 \text{ ranks} = 1, 3, 6, 4, 2, 5 \]

<table>
<thead>
<tr>
<th>( \varepsilon )</th>
<th>( \varepsilon_N(\rho) \times 10^3 )</th>
<th>( se \times 10^4 )</th>
<th>( \varepsilon_{NS}(\rho) \times 10^3 )</th>
<th>( \varepsilon_A(\rho) \times 10^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.243</td>
<td>0.175</td>
<td>0.244</td>
<td>0.237</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.712</td>
<td>0.392</td>
<td>0.712</td>
<td>0.684</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.192</td>
<td>0.546</td>
<td>1.125</td>
<td>1.115</td>
</tr>
<tr>
<td>0</td>
<td>1.150</td>
<td>0.623</td>
<td>1.389</td>
<td>1.389</td>
</tr>
<tr>
<td>0.3</td>
<td>1.152</td>
<td>0.614</td>
<td>1.481</td>
<td>1.368</td>
</tr>
<tr>
<td>0.6</td>
<td>1.339</td>
<td>0.530</td>
<td>1.407</td>
<td>1.030</td>
</tr>
<tr>
<td>0.9</td>
<td>0.951</td>
<td>0.350</td>
<td>0.941</td>
<td>0.438</td>
</tr>
</tbody>
</table>

\( a_1 = 0.34 \)

\( b_1 = 0.42 \)

<table>
<thead>
<tr>
<th>( \varepsilon )</th>
<th>( \varepsilon_N(\rho) \times 10^3 )</th>
<th>( se \times 10^4 )</th>
<th>( \varepsilon_{NS}(\rho) \times 10^3 )</th>
<th>( \varepsilon_A(\rho) \times 10^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.226</td>
<td>0.166</td>
<td>0.230</td>
<td>0.259</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.699</td>
<td>0.382</td>
<td>0.683</td>
<td>0.695</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.182</td>
<td>0.538</td>
<td>1.097</td>
<td>1.099</td>
</tr>
<tr>
<td>0</td>
<td>1.496</td>
<td>0.624</td>
<td>1.389</td>
<td>1.389</td>
</tr>
<tr>
<td>0.3</td>
<td>1.596</td>
<td>0.633</td>
<td>1.545</td>
<td>1.451</td>
</tr>
<tr>
<td>0.6</td>
<td>1.538</td>
<td>0.570</td>
<td>1.573</td>
<td>1.211</td>
</tr>
<tr>
<td>0.9</td>
<td>1.164</td>
<td>0.387</td>
<td>1.166</td>
<td>0.597</td>
</tr>
</tbody>
</table>

\( a_1 = 0.46 \)

\( b_1 = 0.53 \)

**Example 5:**

\[ N = 6 \text{ ranks} = 1, 2, 6, 3, 5, 4 \]

<table>
<thead>
<tr>
<th>( \varepsilon )</th>
<th>( \varepsilon_N(\rho) \times 10^3 )</th>
<th>( se \times 10^4 )</th>
<th>( \varepsilon_{NS}(\rho) \times 10^3 )</th>
<th>( \varepsilon_A(\rho) \times 10^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.226</td>
<td>0.166</td>
<td>0.230</td>
<td>0.259</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.699</td>
<td>0.382</td>
<td>0.683</td>
<td>0.695</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.182</td>
<td>0.538</td>
<td>1.097</td>
<td>1.099</td>
</tr>
<tr>
<td>0</td>
<td>1.496</td>
<td>0.624</td>
<td>1.389</td>
<td>1.389</td>
</tr>
<tr>
<td>0.3</td>
<td>1.596</td>
<td>0.633</td>
<td>1.545</td>
<td>1.451</td>
</tr>
<tr>
<td>0.6</td>
<td>1.538</td>
<td>0.570</td>
<td>1.573</td>
<td>1.211</td>
</tr>
<tr>
<td>0.9</td>
<td>1.164</td>
<td>0.387</td>
<td>1.166</td>
<td>0.597</td>
</tr>
</tbody>
</table>

\( a_1 = 0.46 \)

\( b_1 = 0.53 \)

<table>
<thead>
<tr>
<th>( \varepsilon )</th>
<th>( \varepsilon_N(\rho) \times 10^3 )</th>
<th>( se \times 10^4 )</th>
<th>( \varepsilon_{NS}(\rho) \times 10^3 )</th>
<th>( \varepsilon_A(\rho) \times 10^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.9</td>
<td>0.226</td>
<td>0.166</td>
<td>0.230</td>
<td>0.259</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.699</td>
<td>0.382</td>
<td>0.683</td>
<td>0.695</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.182</td>
<td>0.538</td>
<td>1.097</td>
<td>1.099</td>
</tr>
<tr>
<td>0</td>
<td>1.496</td>
<td>0.624</td>
<td>1.389</td>
<td>1.389</td>
</tr>
<tr>
<td>0.3</td>
<td>1.596</td>
<td>0.633</td>
<td>1.545</td>
<td>1.451</td>
</tr>
<tr>
<td>0.6</td>
<td>1.538</td>
<td>0.570</td>
<td>1.573</td>
<td>1.211</td>
</tr>
<tr>
<td>0.9</td>
<td>1.164</td>
<td>0.387</td>
<td>1.166</td>
<td>0.597</td>
</tr>
</tbody>
</table>

\( a_1 = 0.46 \)

\( b_1 = 0.53 \)
TABLE 8.2:

Values of estimates concerned with the curvature of the ranks serial correlation model. For $N = 5, 6, 7$ values given are exact expectations; for $N = 10-15$ values are based on random sampling.

$\gamma_{Ef}^2$ is $\frac{8N+2}{(N+1)^2}$ as suggested by Efron (1975). $\gamma_1^2$ and $\gamma_2^2$ are found by equation (8.13) where $\text{var}(\xi')$ and $\text{E}(-\xi''_A)$ are substituted for $i_x$ respectively.

<table>
<thead>
<tr>
<th>$N$</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>var ($\xi'$)</td>
<td>1.18</td>
<td>1.86</td>
<td>2.61</td>
<td>5.06</td>
<td>5.92</td>
<td>6.80</td>
<td>7.70</td>
<td>8.60</td>
<td>9.51</td>
</tr>
<tr>
<td>$\text{E}(-\xi''_A)$</td>
<td>1.40</td>
<td>2.06</td>
<td>2.80</td>
<td>5.29</td>
<td>6.14</td>
<td>6.94</td>
<td>7.94</td>
<td>8.66</td>
<td>9.59</td>
</tr>
<tr>
<td>var($\xi''_A$)</td>
<td>1.10</td>
<td>1.64</td>
<td>2.15</td>
<td>3.69</td>
<td>4.16</td>
<td>4.79</td>
<td>5.20</td>
<td>5.26</td>
<td></td>
</tr>
<tr>
<td>cov($\xi',\xi''_A$)</td>
<td>-0.51</td>
<td>-0.74</td>
<td>-0.93</td>
<td>-1.16</td>
<td>-1.42</td>
<td>-1.85</td>
<td>-2.22</td>
<td>-2.09</td>
<td>-2.11</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>1.10</td>
<td>0.76</td>
<td>0.37</td>
<td>0.15</td>
<td>0.115</td>
<td>0.101</td>
<td>0.086</td>
<td>0.074</td>
<td>0.061</td>
</tr>
<tr>
<td>$\gamma_2^2$</td>
<td>0.75</td>
<td>0.47</td>
<td>0.32</td>
<td>0.14</td>
<td>0.107</td>
<td>0.097</td>
<td>0.080</td>
<td>0.073</td>
<td>0.060</td>
</tr>
<tr>
<td>$\gamma_{Ef}^2$</td>
<td>1.36</td>
<td>1.04</td>
<td>0.82</td>
<td>0.46</td>
<td>0.391</td>
<td>0.336</td>
<td>0.292</td>
<td>0.257</td>
<td>0.227</td>
</tr>
</tbody>
</table>
TABLE 8.3:  
Distribution of $T_R = \frac{x_1'' - \frac{x_{12}'}{x_{22}''}}{x_{11}'' - \frac{(x_{12}')^2}{x_{22}''}}$ and $S_{R_B} = \frac{x_1'}{\sqrt{x_{11}''}}$

for various designs of two sample where $x_i'$ and $x_{ij}'$, $i,j=1,2$ are as in eq. (8.14). Results based on 3000 replications when $N=20$.

<table>
<thead>
<tr>
<th>Design</th>
<th>I</th>
<th>I</th>
<th>II</th>
<th>I</th>
<th>II</th>
<th>II</th>
<th>II</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>E($T_R$)</td>
<td>0.03</td>
<td>var($T_R$)</td>
<td>0.98</td>
<td>E($S_{R_B}$)</td>
<td>-0.02</td>
<td>var($S_{R_B}$)</td>
<td>0.53</td>
</tr>
<tr>
<td>-0.3</td>
<td></td>
<td>0.00</td>
<td></td>
<td>1.13</td>
<td></td>
<td>0.00</td>
<td></td>
<td>0.86</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.05</td>
<td></td>
<td>1.16</td>
<td></td>
<td>0.05</td>
<td></td>
<td>1.15</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>0.00</td>
<td></td>
<td>1.11</td>
<td></td>
<td>0.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.02</td>
<td>0.94</td>
<td>-0.02</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>I</th>
<th>I</th>
<th>II</th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>E($T_R$)</td>
<td>0.03</td>
<td>var($T_R$)</td>
<td>0.65</td>
<td>E($S_{R_B}$)</td>
</tr>
<tr>
<td>-0.3</td>
<td></td>
<td>0.02</td>
<td></td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0.03</td>
<td></td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>0.01</td>
<td></td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.03</td>
<td>0.63</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

For random design (regardless of $\rho$ values) results based on 10,000 replications

$E(T_R) = 0.03 \quad \text{var}(T_R) = 1.11 \quad E(S_{R_B}) = 0.01 \quad \text{var}(S_{R_B}) = 1.10$
In this thesis, some semiparametric techniques based on approximations to the rank marginal likelihood (ARML) are proposed and applied to various problems. These techniques require an assumption about the joint distribution of random variables $Y_1, \ldots, Y_N$; therefore they are not distribution-free. However, they use the ranks of the observations, thus the assumption about the distribution of $g(Y_1), \ldots, g(Y_N)$ would be sufficient where $g(.)$ is an arbitrary monotone transformation since the ranks of $g(Y_1), \ldots, g(Y_N)$ are the same as ranks of $Y_1, \ldots, Y_N$. The actual values of the observations are only required to obtain ranks of the observations or to provide an answer in terms of actual values.

The introduced techniques have the advantages of the rank techniques like being robust against error mis-specification and outliers or extreme observations, and being applicable when the data consist of ranks of the observations as in judges problem, consumer preferences etc. They require longer computations but usually give more extensive inference than non-parametric techniques. Suggested techniques are useful in analysing data from the social and medical sciences where the assumption of the population distribution for the original observations might not be tenable but one for some transformation of the observations might be reasonable.

For the problem of predicting a future response, when the data can be transformed to a standard linear model, i.e.
\[ g(Y) = X\beta + e \]

where \( X \) is regressor variable, \( \beta \) is unknown parameter and the distribution of \( e \) is standard normal, estimators are proposed. These estimators are based on the predictive probabilities for a future response \( Y^* \) to have a specified ranks amongst the data. Proposed estimators are assessed by comparing them with each other and with non-parametric and parametric techniques. Simulation experiments are undertaken in various different circumstances mainly at three levels: when the assumption of predictive rank probabilities technique (PRPT) is not satisfied, partly satisfied and fully satisfied.

It is concluded that PRPT provides very good estimates even if the explanatory variables and the error distribution are mis-specified. Rank based techniques cannot follow any rapid increase in observations, whereas isotonic estimators are better for rapid increasing and poorer for slowly increasing observations. If the regressor variables can be transformed, to be correctly specified, or at least not to be badly mis-specified, a substantial improvement is seen in the performance of PRPT. These transformations do not affect other rank based or isotonic techniques. The superiority of PRPT to least squares estimates, with both assuming the same parametric model, shows the robustness of rank estimates to mis-specification of the error distribution.

PRPT is superior to its non-parametric competitors when the transformation of the response variables are considered and the assumptions of the standard linear model are fully satisfied on the transformed
metric. Parametric techniques as applied to transformed observations perform better than PRPT almost uniformly when the prediction of back-transformed response is considered for the former. On the other hand, PRPT has the advantage of being applicable in the untransformed metric hence poor results caused by back transformations are avoided. Mis-specification of the error distribution by taking heteroscedastic errors in the simulation study affect some non-parametric estimators, however PRPT seems to be least affected amongst them.

In summary, PRPT provides better estimators than Mean Isotonic, Median Isotonic and Rank Transform techniques and performs reasonably well when it is compared with Bayesian and Maximum Likelihood Predictors. Although the studies considered relationships between a response variable and an explanatory variable, PRPT can be used with vector explanatory variables. It can also give confidence intervals for future responses.

ARML is applied to a two sample population problem and rank techniques for (two sample) discriminant analysis problem are suggested. Various estimators for probability $P_i$, for a new case $Z^*$, of being a member of possible two populations, are obtained; either by replacing classical predictive probabilities $p(z^*|\text{data})$ by PRP, or by substituting ARML for $p(z^*|\beta)$ and treating $P_i$ as a random variable and finding the approximate distribution of it. A case study of a medical data set suggests that new estimators perform remarkably efficiently when the assumptions of parametric model hold. Rank estimators of error rates also perform well in the same example. A very important factor in having successful results when using rank techniques is the correct
rankings of the multivariate observations. This may be a disadvan-
tage of rank techniques since the way to obtain a projection of
multivariate observations into a real line is not unique. If the
univariate observations are obtained by using an unsuitable function,
the number of misallocated observations does not change dramatically.
However, when the discrimination techniques are used in medical
diagnosis, the aim is not to allocate the cases for deciding whether
or not they suffer from a particular disease, but it is to find 'the
probability of having the disease'. In this respect, misleading results
can be obtained by choosing incorrect rankings. Similar to the classi-
cal parametric techniques, information about the location parameter does
not contribute much to the solution, for the most uncertain cases.

If the distribution of the random variables is taken to be logistic
with variance $\pi^2/3$ and mean $\theta = X\beta$ a quartic ARML of the form

$$
\frac{1}{N^n} \exp \left( \sum_{i=1}^{N} a_i \theta_i - \frac{1}{2} \sum_{i,j} b_{ij} \theta_i \theta_j + \frac{1}{6} \sum_{i,j,k} c_{ijk} \theta_i \theta_j \theta_k \right)
$$

$$
- \frac{1}{24} \sum_{i,j,k} d_{ijk\ell} \theta_i \theta_j \theta_k \theta_\ell
$$

may not be appropriate. Two sample and trend models are considered to
demonstrate that quartic approximation to the asymptotic likelihood may
not be unimodel and should not be used to make inferences. Results of
the same experiment also suggest that the average of cubic and quartic
approximations may provide a good approximation, however it is merely an ad-hoc solution.

When the underlying joint distribution of serially correlated variables

\[ Y_j = \rho Y_{j-1} + e_j \]

is multivariate normal with zero mean and covariance matrix \( \Sigma_\rho \), depending on a single parameter \(|\rho|<1\), where \( e_j \)'s are independently distributed standard normal random variables, an ARML is obtained. ARML is found successful for serial correlation \(|\rho|<0.3\) by comparing the results of developed numerical algorithm to evaluate the multiple integral of RML with approximated values. The approximation provides a test of hypothesis of independence, \( H_0 : \rho = 0 \), as well as an estimate of serial correlation coefficient. The performance of the test statistics can be improved if the observed information is replaced by the expected information. Its results are in agreement with results of other rank based techniques, although a rank version of the Von Neumann estimator provides slightly better estimates. Unlike the ARML based on the normal location shift model, the effect of marginalization over ranks of the serially correlated variables appears to reduce the curvature of the likelihood function. Curvature of the approximation is smaller for increasing sample sizes, thus, proposed estimator of serial correlation coefficient is more appropriate for sample sizes \( N \gg 1 \). ARML is extended for the model

\[ Y_j = \theta_j + \delta_j \]
where

\[ \delta_j = \rho \delta_{j-1} + e_j \]

A test of locations parameter \( H_0: \beta = 0 \) and an estimator of \( \beta \) is suggested for a simple regression model, \( \theta = X\beta \). Simulation studies show that this test statistic is more robust against serially correlated errors than the similar test of location parameter, based on the standard linear model, for some designs of two sample, however the expected robustness may not be obtained for every design.

**Current Work Using Ranks by Some Other Authors**

Recently Clayton and Cuzick (1985) considered another semiparametric method, using ranks, to find an approximate maximum likelihood estimate of \( \beta \) for the model

\[ g(Y_j) = x_j^T \beta + e_j \]

where \( e_j \) has the density \( f_e \). They consider the log likelihood equations, assuming \( g(Y_j)'s \) are known,

\[ \frac{3}{\beta} \sum_{j=1}^{N} \log f_e (g(Y_j) - x_j^T \beta) = 0 \]

Given the value of the current iterate \( \beta^* \), they consider replacing the \( g(Y_j) \) by approximations to
and solve equation (9.1) to obtain the next iterate $\beta^{**}$. They also provide computing formulae for some choice of $f_e$.

Pettitt also reports (unpublished) that Monte Carlo techniques can be used to find $\hat{\beta}$ for the normal error model. He uses an idea by Doksum called the 'ordered bootstrap' i.e. approximating the rank likelihood by

$$\text{lik}(\beta) = \frac{1}{N} \sum_{j=1}^{N} \frac{\phi(Z_{r_j} - x^T_j \beta)}{\phi(Z_{r_j})}$$

(9.2)

where $Z_1 < ... < Z_N$ are the order statistics of a sample from standard normal distribution and expectation in equation (9.2) is estimated by

$$\frac{1}{M} \sum_{i=1}^{M} \prod_{j=1}^{N} \left\{ \frac{\phi(Z_{r_{j,i}} - x^T_j \beta)}{\phi(Z_{r_{j,i}})} \right\}$$

where $Z_{1,i} < ... < Z_{N,i}$ are independent realizations of standard normal order statistics, $i=1, ..., M$. 

$$E(g(Y_j)\mid r, \beta=\beta^*)$$
REFERENCES

Prediction using nonparametric techniques: some comparisons  

ACAR, B.S. and PETTITT, A.N. (1985a)  
Using ranks in discriminant analysis.  

ACAR, B.S. and PETTITT, A.N. (1985b)  
Inference for the serial correlation model using an approximation to  
the rank likelihood. 
(Submitted to Biometrika).

Statistical Prediction Analysis  
London: Cambridge University Press.

A critical comparison of two methods of statistical discrimination  

Distribution of the serial correlation coefficient  

An Introduction to Multivariate Statistical Analysis  
New York: John Wiley and Sons.

BARTELS, R. (1977)  
Estimation in a first order autoregressive scheme with non-normal stable  
disturbances  
The rank version of Von Neumann's ratio test for randomness.

BICKEL, P.J. (1965)
On some robust estimates of location
Annals of Math. Stats., 36, 847-858

Science and statistics

An analysis of transformations

BRADLEY, J.V. (1968)
Distribution-Free Statistical Tests
New Jersey: Prentice-Hall.

BROOKS, R.J. (1978)
Bayesian analysis of a two sample problem based on the rank order
statistics

Estimation of isotonic regression
Non-Parametric Techniques in Statistical Inference
Cambridge University Press.

CARROLL, R.J. (1980)
A robust method for testing transformations to achieve approximate
normality
On prediction and the power transformation family
*Biometrika*, 63, 609-615.

CHILD, J. (1973)
Predicting and understanding organization structure

CLAYTON, D. and CUZICK, J. (1985)
Multivariate generalization of the proportional hazards model
(To appear in *J.R. Statist. Soc. A*).

CONOVER, W.J. and IMAN, R.L. (1976)
On some alternative procedures using ranks for the analysis of experimental designs

Theoretical Statistics
London: Chapman and Hall.

CRYER, J.D., ROBERTSON, T., WRIGHT, F.T. and CASADY, R.J. (1972)
Monotone median regression

AS 128. Approximating the covariance matrix of normal order statistics
*Appl. Statist.*, 27, 206-212.

Proportional hazards, transformation models, partial likelihood, the order bootstrap and adaptive reference, II Report 3/84 Dept. Maths.
University of Trondheim, Norway.
EFRON, B. (1975)
Defining the curvature of a statistical problem (with applications to second order efficiency)

EFRON, B. (1979)
Bootstrap methods: another look at the jackknife

GIBBONS, J.D. (1971)
Nonparametric Statistical Inference

Nonparametric tests of randomness against autocorrelated normal alternatives
Biometrika, 67, 375-379.

A stepwise discriminant analysis program using density estimation

HÁJEK, J. and SIDÁK, Z.S. (1967)
Theory of Rank Tests

HENERY, R.T. (1981)
Permutation probabilities as models for horse races

A power study of a rank transform for the two-way classification model when interaction may be present
IMAN, R.L. and CONOVER, W.J. (1979)
The use of the rank transform in regression
Technometrics, 21, 499-509.

IMAN, R.L. and CONOVER, W.J. (1980)
The rank transformation as a method of discrimination with some examples

KALBFLEISCH, J.D. (1978)
Likelihood methods and nonparametric tests

KENDAL, M.G. and STUART, A. (1963)
The Advanced Theory of Statistics

Robust tests for spherical symmetry and their application to least squares regression
The Annals of Statist., 8, 1265-1271.

KNOKE, J.D. (1977)
Testing for randomness against autocorrelation: alternative tests
Biometrika, 64, 523-529.

LEHMAN, E.L. (1975)
Nonparametrics: Statistical Methods Based on Ranks
San Francisco: Holden-Day.

MILTON, R.C. (1970)
Rank Order Probabilities: Two-sample Normal Shift Alternatives
New York: Wiley

MINITAB (1982)
Statistical Computing System, Pennsylvania, USA.
MORRISON, D.F. (1976)
*Multivariate Statistical Methods*

NAG (1977)
Numerical Algorithms Group, UK.

NG, T.G. and RANDLES, R.H. (1983)
Rank procedures in many population forced discrimination problems

PEARSON, E.S. and HARTLEY, H.O. (1972)
*Biometrika Tables for Statisticians,* Vol 2
London: Cambridge University Press.

Inference for the linear model using a likelihood based on ranks

PETTITT, A.N. (1982b)
Parametric tests for agreement amongst groups of judges

PETTITT, A.N. (1983a)
An approximate analysis of incomplete ranked data

PETTITT, A.N. (1983b)
Approximate methods using ranks for regression with censored data
*Biom. J.,* 70, 121-132.

PLACKETT, R.L. (1975)
The analysis of permutations
RANDLES, R.H., BROFFITT, J.D., RAMBERG, J.S. and HOGG, R.V. (1978a) Discriminant analysis based on ranks

RANDLES, R.H., BROFFITT, J.D., RAMBERG, J.S. and HOGG, R.V. (1978b) Generalized linear and quadratic discriminant functions using robust estimates
*Jour. Amer. Statist. Assoc.*, 73, 564-568.

(To appear in *J.R. Statist. Soc. B*).

RIGBY, R.A. (1982) A credibility interval for the probability that a new observation belongs to one of two multivariate normal populations

*Appl. Statist.*, 31, 161-165.

New York: John Wiley and Sons.

*Appl. Statist.*, 30, 158-162.

An alternative approach to the analysis of permutations

WILCOXON, F. (1945)
Individual comparisons by ranking methods
Biometrics Bull., 1, 80-83.