The Analysis of Permutations

by B.R. Dansie

A thesis submitted for the degree of Doctor of Philosophy in the

Department of Statistics, the University of Adelaide,

February 1988.

Awarded: June 15, 1988.

Contents

Abstract

The objective of this thesis is to present various parametric models for the analysis of permutations.

Chapter 1 is a summary of the models presented in the literature for the analysis of data which consists of the $r!$ frequencies corresponding to the rankings of r items by a group of individuals. The main models covered are those of Plackett [32] and the transportation theory work of Daganzo [7] and McFadden [27].

In Chapter 2 we look closely at Plackett's series of models and show that we can represent them in the framework of the so called order statistics models, that is models which involve the use of underlying variables such as those discussed in Chapter 1. Chapter 3 contains a simplistic analysis of variance approach for specifying dependence in the ranking probabilities. An algorithm for fitting the model using iterative least squares is presented.

In Chapter 4 we look at the normal distribution as a basis for these models. Some consideration is given to specification and identification problems which arise and an example is discussed from a paper by Horowitz et al. [20]. A useful extension of Henery's [17] Taylor series approximation is also given and is used to generate approximate parameter estimates using the least squares procedure derived in Chapter 3.

In Chapter 5 we derive the choice probabilities for a number of logistic models, in particular those for which it is possible to write down explicit expressions. Some work on multivariate logistic distributions is also covered in this chapter. In Chapter 6 we look at some data sets with the intention of demonstrating some of the numerical properties of the models previously considered. Comparisons between the different models are also made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due reference is given in the text of the thesis.

I consent to the thesis being made available for photocopying and loan if it is accepted for the award of the degree.

Brenton Dansie

Feb 1st, 1988

Acknowledgement

I would like to gratefully acknowledge the assistance and guidance given to me during the preparation of this thesis by my supervisor Dr. G.M. Tallis of the Statistics Department of the University of Adelaide. I would also like to thank colleagues in both the Statistics Department of the University of Adelaide and the School of Mathematics and Computer Studies at the South Australian Institute of Technology, for the many useful discussions that took place regarding the contents of this thesis. In addition, I would also like to acknowledge the assistance given to me by the editors and referees of the journals in which material from this thesis has been published.

Dedication

I would like to dedicate this thesis to my wife Dianne and my Mum and Dad for without their support, encouragement, understanding and patience over the years this work would not have been completed.

Chapter 1

Introduction and Review

1.1 Outline of Problem

The main focus of this thesis is the specification of mathematical models for the parametric analysis of permutation data. In general this data consists of a set of $r!$ frequencies corresponding to the rankings of r objects, made according to the individual preferences of a group of n people. The field of market research is one area where problems of this type arise. The intent of this thesis is to investigate models which allow association in the way individuals perceive the value or attraction of various subsets of the r items.

The approach taken involves the fitting of parametric models and therefore it is usually the case that n is much larger than $r!$. I will concentrate on the application to 3 and 4 objects being ranked, although in most cases the procedures are, in theory at least, easily applied to larger values of r. From a practical point of view this is not a great restriction, especially in such experimental areas as consumer preference studies, where the ability of a subject to differentiate between, and give a ranking of, more than 4 objects may be regarded as being doubtful.

I will in general use the following notation. Let $ijk\ldots l$ denote a permutation of $1, 2, \ldots r$ with probability $p_{ijk\ldots l}$ and corresponding frequency of occurrence $f_{ijk\ldots l}$ in a sample of size n. Partial sums are defined by

$$p_i = \sum_{j,k,\ldots,l} p_{ijk\ldots l}$$

where the sum is taken over all permutations of $1, 2, \ldots, r$ excluding i, f_i being defined similarly.

1.2 Background material

The literature on the analysis of permutations is reasonably small. For the case $r = 3$ in earlier years there were two main models suggested. Bradley and Terry[2] suggested the model

$$p_{ijk} = \frac{\pi_i \pi_j}{(\pi_i + \pi_j + \pi_k)(\pi_j + \pi_k)}$$

where π_i is a measure of the individual merit of object i. This model was a direct extension of their paired comparisons model which we shall consider in more depth shortly.

Pendergrass and Bradley[30] considered the model

$$p_{ijk} = \frac{\pi_i^2 \pi_j}{\pi_i^2(\pi_j + \pi_k) + \pi_j^2(\pi_i + \pi_k) + \pi_k^2(\pi_i + \pi_j)}$$

This model arises by considering a ranking of 3 as being made up of three pairwise comparisons.

In the recent literature there have been several discussions on the analysis of permutations. Plackett[32] discussed the problem of specifying models for permutations and proposed a hierarchical system of models which contain the Bradley-Terry model as the first stage. The associated estimation theory and an example of voting patterns in Local Government Council elections in England and Wales due to Upton and Brook[40] and Brook and Upton[5] are also discussed. This series of models will be considered in more detail in Chapter 2.

In an attempt to produce a model with fewer parameters, Schulman[37] considers a model based on the definition of a modal permutation. By counting inversions from the modal permutation, he introduces a model with 2 parameters, p_m the probability of the modal permutation and ρ the factor controlling the decline in probability as we move away from the mode.

Henery[17] considers Plackett's first order model and discusses the common features of permutation models at a general level. His discussion centres on the concept of a natural ordering amongst the permutations, akin to Schulmann's modal permutation. He also gives a Taylor series expansion for $p_{ijk...l}$ for the independent normal variates model which we shall look at in more detail later. Henery[18] also considers the approximate calculation of place probabilities in the independent normal statistics model and the calculation of permutation probabilities from independent gamma variables [19]. In a similar vein, Pettitt[31] calculates an approximation to the rank order probability in the independent normal case.

There are two other bodies of theory which have developed largely independently, although somewhat similarly, which also provide useful background material. These are the theory of paired comparisons and disaggregate demand modelling. While it is true that both of these pieces of theory are centered mainly on individuals making a single choice rather than a ranking, it is clear that models used in these situations are suitable for extension to the permutation problem.

Let us firstly briefly consider the development of paired comparison theory. As the name suggests this work deals mainly with the situation where a choice is made from two objects presented to a subject, such as consumer goods. A variety of other situations including the results of tennis matches have also been analysed using these methods. A comprehensive review of this theory is given by Bradley[4] and an accompanying bibliography compiled by Davidson and Farquhar[13].

The models considered by Thurstone[39] and Mosteller[28] and Bradley and Terry[2] form the basis for a great deal of the literature on this subject. The usual hypothesis is that an individual makes a random but unobservable evaluation X_i of object or treatment i. The probability that treatment i is preferred to treatment j is given by

$$p_{ij} = Pr(X_i > X_j) = Pr(Y_{ij} = X_i - X_j > 0)$$

Thurstone and Mosteller assume that the distribution function of X_i and therefore also of Y_{ij} is normal. Thurstone assumed that the X_i's were independent with mean μ_i and variance σ^2. Mosteller showed that the same probabilities arose when the independence condition was replaced by an assumption that the variables were equally correlated, i.e. $\text{corr}(X_i, X_j) = \rho$.

The Bradley-Terry model has a simpler computational form, because unlike the normal model, the choice probabilities can be written down explicitly. The probabilities have the form

$$p_{ij} = \frac{\pi_i}{(\pi_i + \pi_j)}$$

where π_i is a measure of preference for object i. Bradley[3] noted that this model arises if we assume that the distribution of the differences Y_{ij} is logistic with location parameter $\log \pi_i - \log \pi_j$, and thus the comparison of the two models is similar to the one made between the logit and probit models in biological assay. Davidson[12] pointed out that if the X_i's were independently distributed with the extreme value distribution

$$F(x, \theta) = \exp(-e^{-(x-\theta)}) \qquad -\infty < x < \infty, \qquad -\infty < \theta < \infty$$

then $Y_{ij} = X_i - X_j$ has the logistic distribution with location parameter $\theta_i - \theta_j$ and this, as we have seen, results in the Bradley-Terry model. There has been a great deal written on extensions to this model, for example to allow for ties and order effects. The reader is directed to Bradley[4] for a further discussion of these.

The other major area of work that is useful as preparation is the theory of disaggregate demand modelling, most commonly used in mathematical economics. The main application of this work is the situation in which a subject is confronted with a choice from r alternatives. To each alternative is assigned a random utility U_i, which summarises the attraction or worth of alternative i to the decision maker. The utility is often written as

$$U_i = V_i + \epsilon_i$$

where $V_i = V(\mathbf{t_i})$, the deterministic element of utility, is usually a linear function of a vector of covariates and ϵ_i is the random component. It is more convenient to talk about U_i as a random variable with mean V_i and a given error structure. The two most useful models in this area, not surprisingly, are analogous to those of the paired comparisons model.

If the variables U_i are assumed independent and distributed with the extreme value distribution given earlier with location parameter V_i, then the selection probability of alternative i is given by

$$p_i = \frac{e^{V_i}}{\sum_{j=1}^{r} e^{V_j}}$$

This model is commonly referred to as the multinomial logit model. A non-constructive proof of the existence of such a model was given by Block and Marshack[1]; a simpler constructive proof was given in Luce and Suppes[23], attributed to Holman and Marley. McFadden[25] showed that under very mild conditions the extreme value distribution is unique in giving the multinomial logit model. A survey of developments in this field is given by McFadden[26]. Yellot[41] and Luce[24] are also useful review articles in this area. In the main development recently, McFadden[27] has introduced a family of generalised extreme value distributions which permit limited departures from the assumption of independence. We shall consider these distributions at a later stage.

The other natural choice for the distribution function is the normal distribution. The basic multinomial probit model is obtained when the X_i's are independent N(V_i, 1) variables. The extension to the situation in which the X_i's have a general covariance matrix Σ is discussed by Daganzo[7]. Conceptually this is a very favourable model. It allows a method for specifying the structure of dependencies between pairs of alternatives and it allows for different variances to be attached to the different alternatives. There are computational difficulties with the evaluation of the choice probabilities, but recently developed numerical techniques go some way in solving this problem. We shall consider this model in greater detail at a later stage.

1.3 Summary

This thesis is aimed at presenting various models for the analysis of permutations. In Chapter 2 we look closely at Plackett's series of models and show that we can represent them in the framework of the so called order statistics models, that is models which involve the use of underlying variables such as those just discussed. Chapter 3 contains a simplistic analysis of variance approach for specifying dependence in the ranking probabilities. In Chapter 4 we look at the normal theory and study some interesting features of the specification of these models. A useful extension of Henery's[17] Taylor series approximation is also given.

In Chapter 5 we look at the logistic models, in particular those for which it is possible to write down explicit expressions for the choice probabilities. In Chapter 6 we look at some data sets with the intention of demonstrating some of the models previously considered, and for making comparisons between them.

Chapter 2

An Interpretation of Plackett's Model

2.1 Introduction

Material from this chapter has been published in Dansie[9].

We saw in the review of previous work that the paper of Plackett[32] contains one of the most recent system of models for the analysis of permutations. In this chapter we begin by giving a brief summary of the models. We will then show how we can interpret this system from an underlying variable viewpoint, and so place it in the class of models which includes the logit and probit models already introduced. The chapter concludes with consideration of the practical questions relating to the application of these models as an introduction to chapter 3.

2.2 The Plackett System

2.2.1 Review

This system is designed as a hierarchical set of models for the analysis of permutations. Parameters are added in stages, the final stage gives the saturated model of

$r! - 1$ independent parameters. Parameters in stage i are used only when certain relationships are satisfied amongst the parameters of stage $i + 1$.

For stage 1 define

$$\kappa = \log p_r \quad \lambda_a = \log(p_a/p_r) \quad a = 1, 2, \ldots, r-1$$

The relation $\lambda_a = 0$ for $a = 1, 2, \ldots, r-1$ gives the uniform probability model, i.e. $p_i = 1/r$ for $i = 1, 2, \ldots, r$.

For the second stage of the model put

$$s = \max(1, 2, \ldots, r \text{ excluding } i),$$

and define

$$\lambda_{ib} = \log\left(\frac{p_{ib}p_s}{p_{is}p_b}\right) \qquad b = 1, 2, \ldots, r \text{ excluding } i, s$$

A model which uses only first order parameters is obtained when $\lambda_{ib} = 0$ i.e. when $p_{ib}/p_{is} = p_b/p_s$.

This process is continued in a similar way until the saturated model of stage $r-1$. It is clear that the system is hierarchical, since parameters of stage i only appear when all of the parameters of stage $i+1$ are set to zero. The cumulative number of parameters used at stage i is $r(r-1)\ldots(r-i+1)-1$, where we see that the parameter κ is determined by the values of λ_a and the requirement that $\sum p_i = 1$. For $r = 3$ then, the first stage contains 2 parameters and the second stage is the saturated model of 5 parameters. For $r = 4$ the three stages contain 3,11 and 23 parameters respectively.

The zero order model obtained when all of the parameters $\{\lambda_a\}, \{\lambda_{ib}\}$ etc. are zero is the uniform probability model. The first order model is equivalent to the Bradley-Terry model mentioned in chapter 1. For example consider the case of $r = 4$. Having set the second and third order parameters equal to zero it is fairly easy to show that the expression for p_{1234} using a first order model is

$$p_{1234} = \frac{p_1 p_2 p_3 p_4}{(1-p_1)(1-p_1-p_2)(1-p_1-p_2-p_3)}$$

which we can recognize as the Bradley-Terry model if we rewrite it as

$$\frac{p_1 p_2 p_3}{(p_2+p_3+p_4)(p_3+p_4)}$$

The basis for this model is that object 1 is chosen first with probability p_1. Object 2 is then chosen from the remaining three objects with probability $p_2/(p_2+p_3+p_4)$ and finally object 3 is chosen ahead of object 4 with probability $p_3/(p_3+p_4)$.

To give an example of the structure of a second order model Plackett gives

$$p_{1234} = \frac{p_{12} p_{23}}{(p_{23}+p_{24})}$$

We can, if we wish, easily write this expression in terms of the original parameters $\{\lambda_a\}, \{\lambda_{ib}\}$.

The interpretation of these models is reasonably clear. The first order model allocates different preferences to the individual objects being ranked, this preference is unaltered by any other choices already made. The second order model allows association between pairs of objects to determine the probability of the selection of a particular permutation. Higher order models clearly allow for association between larger groups of objects.

2.2.2 Some preliminary results

We will show that we can construct Plackett's series of models using a system of underlying variables. Suppose that $X_1, \ldots, X_r$ are independently distributed exponential random variables with parameters $\alpha_1, \ldots, \alpha_r$. The density function is of the form

$$\alpha_i \exp(-\alpha_i x_i) \qquad x_i > 0 \quad \alpha_i > 0$$

with distribution function

$$F_{X_i}(x_i) = 1 - \exp(-\alpha_i x_i).$$

To simplify the construction we first give the following four results.

Result 1 Put $Y = X_{(1)}$ the first order statistic amongst the set of r independent random variables $\{X_i\}$. Then $Y \sim \exp(\alpha)$ where $\alpha = \alpha_1 + \ldots + \alpha_r$.

Proof $F_Y(y)= \Pr(Y \leq y)$

$$= 1 - \Pr(Y > y)$$

$$= 1 - \Pr(X_1 > y, \ldots, X_r > y)$$

$$= 1 - (1 - F_{X_1}(y)) \ldots (1 - F_{X_r}(y))$$

$$= 1 - \exp(-\sum \alpha_i y)$$

Result 2 $$\Pr(X_1 < X_2 < \ldots < X_r) = \prod_{i=1}^{r} \alpha_i / \prod_{i=1}^{r}(\sum_{j=1}^{i} \alpha_{r+1-j})$$

The proof of this result is notationally tedious but logically simple in the general case. As an example for $r = 3$ we see:

$$\begin{aligned}
\Pr(X_1 < X_2 < X_3) &= \int_0^\infty \int_{x_1}^\infty \int_{x_2}^\infty \prod_{i=1}^{3} \alpha_i e^{(-\sum \alpha_i x_i)} \, dx_3 dx_2 dx_1 \\
&= \frac{\prod \alpha_i}{\alpha_3} \int_0^\infty e^{-\alpha_1 x_1} \int_{x_1}^\infty e^{-(\alpha_2+\alpha_3)x_2} \, dx_2 dx_1 \\
&= \frac{\prod \alpha_i}{\alpha_3(\alpha_2+\alpha_3)} \int_0^\infty e^{-(\alpha_1+\alpha_2+\alpha_3)x_1} \, dx_1 \\
&= \frac{\prod \alpha_i}{\alpha_3(\alpha_3+\alpha_2)(\alpha_3+\alpha_2+\alpha_1)} \\
&= \prod_{i=1}^{3} \alpha_i \Big/ \prod_{i=1}^{3} \Big(\sum_{j=1}^{i} \alpha_{4-j}\Big)
\end{aligned}$$

Result 3 $\Pr(X_i < X_j \quad \forall j \neq i) = \dfrac{\alpha_i}{\sum_{j=1}^r \alpha_j}$

Proof

$$\begin{aligned}
\Pr(X_i < X_j \quad \forall j \neq i) &= \int_0^\infty \alpha_i e^{-\alpha_i x} \prod_{j \neq i} (1 - F_{x_i}(x)) \, dx \\
&= \int_0^\infty \alpha_i e^{-\alpha_i x} \prod_{j \neq i} e^{-\alpha_j x} \, dx \\
&= \int_0^\infty \alpha_i e^{-\sum_{j=1}^r \alpha_j x} \, dx \\
&= \frac{\alpha_i}{\sum_{j=1}^r \alpha_j}
\end{aligned}$$

Result 4 Let F_1 be a subset of $\{2, \ldots, r\}$ with $k \geq 2$ elements and suppose that $a \in F_1$. The events defined by

$$A(X_1 < X_i \quad i = 2, \ldots, r) \qquad B(X_a < X_j \quad \forall j, \quad j \in F_1, \quad j \neq a)$$

are independent. The result is easily extended to two or more disjoint subsets of $\{2, \ldots, r\}$ with at least two elements in each.

Proof $\Pr(A) = \dfrac{\alpha_1}{\sum_{j=1}^r \alpha_j}$ by Result 3.

$\Pr(B) = \sum \Pr(\, X_a < X_{i_k})$

where the sum is taken over all permutations fixing a as the first element and permuting the remaining elements $i_2, \ldots i_k$ of F_1. Let $\mathbf{d} = d_2, \ldots, d_r$ be a permutation of $2, \ldots, r$ in which the element a appears before any element of F_1. It is clear that Pr(B) is the sum of all such permutation probabilities, which by result 2 is

$$\Pr(B) = \sum_{\mathbf{d}} \frac{\prod_{i=2}^{r} \alpha_{d_i}}{\prod_{i=1}^{r-1} \sum_{j=1}^{i} \alpha_{d_{r+1-j}}}$$

The joint probability of events A and B is the sum of the probabilities of all permutations with 1 as the first element and a occurring before any other of the elements of F_1. Let $d_1, \ldots, d_r$ be one such permutation where $d_1 = 1$. By result 2 we have

$$\Pr(X_{d_1} < X_{d_2} < \ldots < X_{d_r}) = \frac{\prod_{i=1}^{r} \alpha_{d_i}}{\prod_{i=1}^{r} \sum_{j=1}^{i} \alpha_{d_{r+1-j}}}$$

The only term in the denominator involving $\alpha_{d_1} = \alpha_1$ is the term $i = r$ and so extracting this term we can factorise this as

$$\Pr(X_{d_1} < X_{d_2} < \ldots < X_{d_r}) = \frac{\alpha_1}{\sum_{j=1}^{r} \alpha_j} \left(\frac{\prod_{i=2}^{r} \alpha_{d_i}}{\prod_{i=1}^{r-1} \sum_{j=1}^{i} \alpha_{d_{r+1-j}}} \right)$$

The result follows by summing the bracketed term over all such permutations; we see that $\Pr(A \cap B) = \Pr(A)\Pr(B)$ and therefore conclude that the two events are independent.

2.2.3 The Underlying Variable Equivalence

We use the following notation. Let $i_1 i_2 \ldots i_r$ be a permutation of $1, 2, \ldots r$. Define $r!$ random variables $X_{i_1 i_2 \ldots i_r}$. These variables are independently exponentially distributed with parameters $\alpha_{i_1 i_2 \ldots i_r}$, where the density function is of the form

$\alpha_i e^{-\alpha_i x}$. We use α_i instead of $1/\alpha_i$ because it can be directly related to p_i the probability that i is placed first.

Adopting notation analogous to that of Plackett we put

$$\alpha_{i_1} = \sum_{i_2,\ldots,i_r} \alpha_{i_1 i_2 \ldots i_r}$$

where the sum is taken over all permutations of $1, \ldots, r$ excluding the element i_1. Similarly

$$\alpha_{i_1 i_2} = \sum_{i_3,\ldots,i_r} \alpha_{i_1 i_2 \ldots i_r}$$

For uniqueness we define

$$\sum_{i=1}^{r} \alpha_i = 1$$

In order to note the general equivalence with Plackett and to show the nature of his set of models, consider the construction of the jth stage, for $1 \leq j < r$. Define the random variables $X_{i_1 i_2 \ldots i_j}$ derived from the original variables as follows:

$$X_{i_1 i_2 \ldots i_j} = \min_{k_1 k_2 \ldots k_{r-j} \neq i_1 i_2 \ldots i_j} \{X_{i_1 i_2 \ldots i_j k_1 k_2 \ldots k_{r-j}}\}.$$

Since this minimisation is over disjoint sets of the original variables, the $X_{i_1 i_2 \ldots i_j}$ are independent of each other and, by result 1, are distributed as $\exp(\alpha_{i_1 i_2 \ldots i_j})$. The probability of the permutation $i_1 i_2 \ldots i_r$ is defined via the order statistics of the jth stage X's whose suffixes occur as consecutive runs of length j in $i_1 i_2 \ldots i_r$ as follows. The first j positions $p_1 p_2 \ldots p_j$ are chosen by selecting the first order statistic from the j stage X's:

$$X_{p_1 p_2 \ldots p_j} = \min_{i'_1 i'_2 \ldots i'_j} X_{i'_1 i'_2 \ldots i'_j}.$$

Subsequent positions are filled by a procedure which is conditional on the preceeding positions having been fixed. Thus if the first k positions are filled by $p_1 p_2 \ldots p_k$ we choose p_{k+1} by

$$X_{p_{k-j+2} \cdots p_k p_{k+1}} = \min_{i'_1 \neq p_1 p_2 \cdots p_k} X_{p_{k-j+2} \cdots p_k i'_1}.$$

For instance when $r = 4$ and $j = 2$ there are ${}^4\mathrm{P}_2 = 12$ independent variables of type $X_{i_1 i_2}$. The permutation 1234 is chosen if X_{12} is the smallest of these 12 variables and also X_{23} is smaller than X_{24}.

Using results 1-4 the probability of any permutation is easily evaluated if it is realised that the filling of subsequent positions in the permutation uses disjoint sets of the j stage X's and hence gives independent events.

In the example of r=4 we see that the probability of the permutation 1234 is given by

$$\begin{aligned} \mathrm{P}(1234) &= \mathrm{P}[\{X_{12} < X_{ij} \quad \forall ij \neq 12\} \text{ and } \{X_{23} < X_{24}\}] \\ &= \alpha_{12} \times \alpha_{23}/(\alpha_{23} + \alpha_{24}) \end{aligned}$$

This expression is identical to the one given by Plackett's model.

It is interesting to note that the sequence of random variables used in the decision process does not form a set of ordered variables; although it is true that $X_{p_1 p_2 \cdots p_j}$ is the smallest of the j stage X's it is not true in general that

$$X_{p_2 p_3 \cdots p_{j+1}} < X_{p_3 p_4 \cdots p_{j+2}} < \cdots < X_{p_{r-j+1} \cdots p_r}$$

since these variables are chosen independently of each other.

2.3 Practical considerations

We have seen that the exponential distribution is the central part of the order statistic interpretation of Plackett's model. As an error distribution in this situation the exponential distribution is not as intuitively as plausible as the normal or the extreme value distributions. In addition we note that Result 3 is the basis of the multinomial logit model presented in the theory of disaggregate demand. We can in fact show that there is a very simple relationship between the exponential and multinomial logit models.

Looking at the definition of the probability of a permutation in the exponential case e.g.

$$p_{12\ldots r} = \Pr(X_1 < X_2 < \cdots < X_r)$$

we see that if ϕ is a monotonic increasing function then we can write

$$p_{12\ldots r} = \Pr(\phi(X_1) < \phi(X_2) < \cdots < \phi(X_r))$$

and likewise if ϕ is a monotonic decreasing function then

$$p_{12\ldots r} = \mathrm{P}(\phi(X_1) > \phi(X_2) > \cdots > \phi(X_r))$$

It is a well known result (see for example Johnson and Kotz[21]) that if $X \sim \exp(\alpha)$ and $Y = -\log X$ then Y has an extreme value distribution with location parameter $\log \alpha$. Thus the exponential and extreme value distributions lie in a class of distributions which generate the same probability models for the analysis of permutations.

The main practical problem with the Plackett system as pointed out by Schulmann[37], is that the number of parameters used rises quite quickly with the number of stages. The system lacks a little flexibility because of it's hierarchical nature. It is not possible to mix parameters of different stages, nor is it possible to have, for example, second order association limited to a subset of the items being ranked. For $r = 3$ the first stage has 2 parameters and the second stage is the saturated model; there is no in-between stage of association. For $r = 4$ the first stage has 3 parameters, the second stage has a total of 11 parameters, which is quite large considering that we are dealing with 23 independent frequencies.

Our main objective then is to provide a flexible series of models in which there is greater freedom to eliminate non-significant parameters, thus providing fairly simple models with a limited number of parameters. The next chapter considers a simple analysis of variance type approach and the subsequent chapters investigate the order statistics approach.

Chapter 3

An Alternative Approach to the Analysis of Permutations

3.1 Introduction

Material from this chapter has been published in Tallis and Dansie[38].

As I pointed out at the conclusion of the previous chapter, the two main practical problems with the implementation of the Plackett series of models are the inflexibility within a stage and the rapid addition of parameters in higher order stages.

The model that we put forward in this chapter as an alternative is analogous to the standard design models used in analysis of variance. It is a relatively simple model intended mainly for use as an exploratory technique, rather than as a definitive model of the process under study. Unlike other models, parameters in this model can be estimated using standard analytic maximum likelihood techniques without recourse to numerical maximisation of the likelihood function.

3.2 The Model

We shall use notation similar to that of Plackett[32]. Consider first the case of r=3. We write the model as

$$p_{ijk} = \frac{1}{6} + \frac{\alpha_i}{2} + \beta_{ij} \quad i \neq j$$

For identifiability of the parameters we impose the constraints

$$\beta_{ij} + \beta_{ik} = 0 \quad j \neq k \neq i \quad i = 1, 2, 3$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 0$$

We have used 9 parameters in the specification and a system with a total of 4 constraints which gives us the required number of independent parameters for the saturated model. For clarity we can write the probabilities as

$$\begin{array}{lclclcl} p_{123} & = & \frac{1}{6} + \frac{\alpha_1}{2} + \beta_{12} & p_{132} & = & \frac{1}{6} + \frac{\alpha_1}{2} - \beta_{12} \\ p_{213} & = & \frac{1}{6} + \frac{\alpha_2}{2} + \beta_{21} & p_{231} & = & \frac{1}{6} + \frac{\alpha_2}{2} - \beta_{21} \\ p_{312} & = & \frac{1}{6} + \frac{\alpha_3}{2} + \beta_{31} & p_{321} & = & \frac{1}{6} + \frac{\alpha_3}{2} - \beta_{31} \end{array}$$

subject only to the constraint that

$$\alpha_1 + \alpha_2 + \alpha_3 = 0$$

The interpretation of the two types of parameters is reasonably clear. The parameters $\{\alpha_i,\ i = 1, 2, 3\}$ are related to the probability that object i is ranked first. The second order parameters $\{\beta_{ij}\}$ represent the association between the rankings of elements i and j.

Two submodels of the saturated model are worthy of note. When we have $\beta_{ij} = 0 \quad \forall j \neq i, \quad i = 1, 2, 3$ we see that

$$p_{ijk} = p_{ikj} = \frac{1}{6} + \frac{\alpha_i}{2}$$

In terms of the data this model says that preference for a given ordering is determined by the preference for the first object i in the ordering and that any further preference is equally distributed between the orderings jk and kj. In addition if $\alpha_i = 0$ we have the uniform probability model viz.

$$p_{ijk} = \frac{1}{6}$$

The generalisation of the model to $r > 3$ objects is straightforward. We put

$$p_{ijk\ldots lmn} = \frac{1}{r!} + \frac{\alpha_i}{(r-1)!} + \frac{\beta_{ij}}{(r-2)!} + \frac{\gamma_{ijk}}{(r-3)!} + \ldots + \delta_{ijk\ldots lm}$$

The system of constraints is

$$\sum_m \delta_{ijk\ldots lm} = \cdots = \sum_k \gamma_{ijk} = \sum_j \beta_{ij} = \sum_i \alpha_i = 0$$

The specification consists of $r + r(r-1) + r(r-1)(r-2) + \ldots + r!$ parameters and $1 + r + \ldots + r(r-1)\ldots 3$ constraints which gives r!-1 independent parameters as required. The reason for including the factorial terms in the model is that it simplifies the estimation procedure.

In the analysis of permutation data, especially in relation to such areas as consumer preference studies, we may be interested in other aspects of the data apart from the ordering effects presented in the model. Take for example the parameters β_{12} and β_{21}. These parameters convey an idea of the way in which elements 1 and 2 are ordered together in the data, i.e the association between the rankings of objects 1 and 2. In many circumstances we may be interested in the hypothesis $\beta_{12} = \beta_{21}$, that is we are looking at the preference for objects 1 and 2 as a subgroup rather than at the ordering within this subgroup.

Consider an artificial example. Three brands of bread are tasted by a large number of people and their preferences are recorded. Brands 1 and 2 are brown breads, brand 3 is white. We may be interested in detecting preferences for each individual brand, for example the preference for brand 1 over brand 2. Alternatively we might look for a preference for brown bread over white bread. We may find for example that of those people who prefer brown bread, regardless of which brand they select first they will have a common tendency to place the remaining brown bread as their second choice. This would indicate a preference for the group of brown breads as opposed to the white bread and would be conveniently summarised by putting $\beta_{12} = \beta_{21}$ in the model.

3.3 Inference

The main simplicity of the model is that maximum likelihood estimates of the parameters are analytically obtainable from the usual estimates of the probabilities $p_{ijk...l}$. Consider the case r=3 and the saturated model. In this situation we have the same number of parameters as independent multinomial probabilities. The maximum likelihood estimate of p_{ijk} is f_{ijk}/n. We can write each parameter as a function of a set of the probabilities and by the invariance property of maximum likelihood estimation the required estimates are calculated by substituting f_{ijk}/n for p_{ijk} in the appropriate expressions, e.g.

$$\begin{array}{rclrcll} \alpha_i & = & p_i - \frac{1}{3} & \beta_{ij} & = & (p_{ij} - p_{ik})/2 & k \neq j \neq i \\ \hat{\alpha}_i & = & (f_i/n) - \frac{1}{3} & \hat{\beta}_{ij} & = & \{(f_{ij}/n) - (f_{ik}/n)\}/2 & \end{array}$$

It is quite clear that we can also estimate the variance matrix of our estimates, from the known variance matrix of a multinomial variable. This calculation is a special case of the procedure to be explained in more detail shortly. Following Plackett we can construct a table of Z^2, the square of the parameter estimates divided by their respective estimated variances. This table is useful in suggesting which parameters may be omitted from the saturated model. By setting a subset of the parameters to zero, such as all of the highest order terms, in such a way as to obtain a model which has the same number of independent probabilities as independent parameters, it is clear that the estimates of the remaining parameters are unchanged. We can proceed very simply in finding a model which gives good accord with the data and which will give some basic understanding of the process by which the data has been generated.

As we have mentioned, a part of the modelling process may involve looking for certain grouped preferences, either suggested a priori or by the analysis of the data. As an example consider the model with $r = 3$ and $\beta_{12} = \beta_{21} = \beta, \beta_{31} = 0$. This gives the six probabilities as

$$
\begin{aligned}
p_{123} &= \tfrac{1}{6} + \tfrac{\alpha_1}{2} + \beta & p_{132} &= \tfrac{1}{6} + \tfrac{\alpha_1}{2} - \beta \\
p_{213} &= \tfrac{1}{6} + \tfrac{\alpha_2}{2} + \beta & p_{231} &= \tfrac{1}{6} + \tfrac{\alpha_2}{2} - \beta \\
p_{312} &= \tfrac{1}{6} + \tfrac{\alpha_3}{2} = p_{321}
\end{aligned}
$$

We now have a situation of 4 independent probabilities and 3 independent parameters. Denote by $\mathbf{Y}(r!\text{x}1)$ the vector of relative frequencies $\{f_{ijk\ldots r}/n\}$. We can write the general model as $\mathrm{E}(\mathbf{Y})=\mathbf{P}(\theta)$ where $\mathbf{P}(r!\text{x}1) = (P_1(\theta),\ldots,P_{r!}(\theta))$ subject to the constraint $\sum Y_i = 1$; $\mathrm{Var}(\mathbf{Y})= V$ where V is the variance matrix of a multinomial variable. Parameter estimates can be obtained by the principle of least squares viz

$$
\text{minimise} \mathcal{Q}(\theta) = (\mathbf{Y} - \mathbf{P}(\theta))' V^g (\mathbf{Y} - \mathbf{P}(\theta))
$$

where V^g indicates a generalised inverse of V. We can in fact show that the least squares approach is equivalent to the method of maximum likelihood in the following way. Put

$$
\Lambda^{-1} = \mathrm{diag}(P_1^{1/2}(\theta),\ldots,P_{r!}^{1/2}(\theta)) \qquad \phi = \Lambda^{-1}\mathbf{1}
$$

where $\mathbf{1}$ represents a vector of 1's of length $r!$. We can then represent V, the multinomial variance matrix as

$$
V = \Lambda^{-1}(I - \phi\phi')\Lambda^{-1}
$$

It is known then that

$$
\Lambda^2 = \mathrm{diag}(P_i^{-1}(\theta))
$$

is a generalised inverse of V since

$$
\begin{aligned}
V\Lambda^2 V &= \Lambda^{-1}(I - \phi\phi')\Lambda^{-1}\Lambda^2\Lambda^{-1}(I - \phi\phi')\Lambda^{-1} \\
&= \Lambda^{-1}(I - \phi\phi' - \phi\phi' + \phi\phi'\phi\phi')\Lambda^{-1} \\
&= V
\end{aligned}
$$

because $\phi'\phi = \mathbf{1}'\mathbf{P} = 1$. Since minimisation of $\mathcal{Q}(\theta)$ is independent of the choice of V^g we can write

$$\mathcal{Q}(\theta) = (\mathbf{Y} - \mathbf{P}(\theta))'\Lambda^2(\mathbf{Y} - \mathbf{P}(\theta))$$

Assuming Λ^2 known we can find the least squares equations by finding the partial derivatives of $\mathcal{Q}$ with respect to θ_i. This gives

$$\sum_j \frac{(Y_j - P_j(\theta))}{P_j(\theta)} \frac{\partial P_j(\theta)}{\partial \theta_i} = 0.$$

and since

$$\sum_j P_j = 1 \quad \Rightarrow \quad \sum_j \frac{\partial P_j}{\partial \theta_i} = 0$$

then the least squares equations are

$$\sum_j \frac{Y_j}{P_j(\theta)} \frac{\partial P_j(\theta)}{\partial \theta_i} = 0.$$

Writing this in the form

$$\sum_j Y_j \frac{\partial \ln(P_j(\theta))}{\partial \theta_i} = 0$$

we see that these are simply the ordinary maximum likelihood equations based on the likelihood function

$$\mathcal{L}(\theta) = \prod_j P_j(\theta)^{nY_j}$$

We can write $\mathbf{P}(\theta) = X\theta$ where X is the appropriate design matrix; standard theory gives the least squares solution as

$$\hat{\theta} = (X'V^gX)^gX'V^g\mathbf{Y}$$

It can be shown that this solution is asymptotically equivalent to minimising $\mathcal{Q}(\theta)$ by assuming that Λ^{-1} is not known. Since V^g is a function of θ the least squares solution must be obtained by iteration. A suitable calculating procedure may be developed in the following way.

Write the model in the form

$$\mathbf{Y} = \gamma + X\beta + \epsilon \quad \mathbf{Y}(r!\text{x}1), \quad \beta(q\text{x}1), \quad X(r!\text{x}q), \quad \text{rank}(X) = q$$

$$\text{E}(\epsilon) = 0, \qquad \text{Var}(\epsilon) = \text{Var}(\mathbf{Y}) = V.$$

Let $H'\mathbf{Y} = \Gamma$ specify the constraints amongst the components of $\mathbf{Y}, \gamma$ is such that $H'\gamma = \Gamma$. H' is nxk and of rank k. In the model presented in this chapter the constraint is clearly that $\sum Y_j = 1 = \Gamma$. In this case H is a vector of 1's (of rank 1) and $\gamma = \mathbf{1}/6$. We note that

$$H'\gamma = \mathbf{1}'\mathbf{1}/6 = 1 = \Gamma$$

Put $\mathbf{Z} = \mathbf{Y} - \gamma$. Partition $\mathbf{Z}$ as $(\mathbf{Z}_1'\mathbf{Z}_2')'$ where $\mathbf{Z}_1$ is $(r! - k)$x1. Similarly partition X and V so that X_1 is $(r! - k)$xq and $V_{11}[(r! - k)\text{x}(r! - k)]$ is the variance matrix of $\mathbf{Z}_1$. Since

$$H'\mathbf{Z} \quad = \quad H'\mathbf{Y} - H'\gamma \quad = \quad 0$$

it is clear that there are k constraints amongst the components of $\mathbf{Z}$. For convenience write the last k components of $\mathbf{Z}$ in terms of the first $n - k$. Thus there exists K of rank $n - k$ such that

$$\mathbf{Z} = K\mathbf{Z}_1 \qquad \text{and} \qquad V = KV_{11}K'$$

We also note that

$$\mathrm{E}(\mathbf{Z}) = X\beta = KX_1\beta$$

We can now rewrite

$$\begin{aligned}
\mathcal{Q}(\beta) &= (\mathbf{Y} - \gamma - X\beta)'V^g(\mathbf{Y} - \gamma - X\beta) \\
&= (\mathbf{Z} - X\beta)'V^g(\mathbf{Z} - X\beta) \\
&= (\mathbf{Z_1} - X_1\beta)'K'V^gK(\mathbf{Z_1} - X_1\beta)
\end{aligned}$$

We can also show that $K'V^gK = V_{11}^{-1}$ This follows by noting that the matrix $V_{11}^{1/2}K'(KV_{11}K')^gKV_{11}^{1/2}$ is idempotent and of rank $n-k$ and is therefore equal to I_{n-k}. This implies that

$$\begin{aligned}
K'V^gK &= K'(KV_{11}K')^gK \\
&= V_{11}^{-1}
\end{aligned}$$

We can therefore write

$$\mathcal{Q}(\beta) = (\mathbf{Z_1} - X_1\beta)'V_{11}^{-1}(\mathbf{Z_1} - X_1\beta)$$

from which it follows that

$$\hat{\beta} = (X_{11}'V_{11}^{-1}X_1)^{-1}X_1'V_{11}^{-1}Z_1$$

Estimates may be obtained by the process of iterative least squares although using the relative frequencies to consistently estimate V_{11} the solution obtained on the first cycle is usually of sufficient accuracy.

3.4 Example calculations

As a very simple example to demonstrate the calculations involved, consider the following set of data extracted from a vehicle testing trial. The full set of data involves the rankings of 8 test vehicles by a group of 800 people. I am grateful to the Melbourne office of Data Sciences Pty. Ltd. for this data set and their general assistance.

The data used here refers to the preferences expressed for the first three cars. Unfortunately I do not have much background knowledge of the vehicles, apart from the information that cars 1 and 3 are new experimental models and car 2 is a currently available make. The observed frequencies of the 6 permutations are given in the first row of Table 3.1. We see for example that 135 of the 800 people selected car 1 first, and then chose car 2 ahead of car 3. To fit the saturated model

Table 3.1: Rankings of three cars

Order	123	132	213	231	312	321	Total
Observed Frequency	135	98	152	139	126	150	800
Expected Frequency (A)	135	98	145.5	145.5	138	138	800
Expected Frequency (B)	137.4	99.8	148.1	148.1	133.3	133.3	800

we calculate

$$\hat{p}_1 = \frac{233}{800} \qquad \hat{p}_2 = \frac{291}{800} \qquad \hat{p}_3 = \frac{276}{800}$$

and using $\hat{\alpha}_i = \hat{p}_i - 1/3$ we have

$$\hat{\alpha}_1 = -0.042 \qquad \hat{\alpha}_2 = 0.030 \qquad \hat{\alpha}_3 = 0.012$$

We can calculate the estimates of the interaction parameters in the following way:

$$\hat{\beta}_{12} = \left(\frac{f_{12}}{800} - \frac{f_{13}}{800}\right)/2 = \left(\frac{135-98}{800}\right)/2 = 0.023$$

Similarly we have

$$\hat{\beta_{21}} = 0.008 \qquad \hat{\beta_{31}} = -0.015$$

To calculate the variance of these estimates we have for example:

$$\mathrm{Var}(\hat{\alpha_1}) = \mathrm{Var}(\hat{p_1}) \simeq \frac{\hat{p_1}(1-\hat{p_1})}{800} = 0.00026$$

Similarly $\mathrm{Var}(\hat{\alpha_2}) \simeq 0.00029$ and $\mathrm{Var}(\hat{\alpha_3}) \simeq 0.00028$. For the interaction parameters we have for example:

$$\begin{aligned}
\mathrm{Var}(\hat{\beta_{12}}) &= \mathrm{Var}\left(\frac{\hat{p_{12}} - \hat{p_{13}}}{2}\right) \\
&\simeq \frac{1}{4}\left(\frac{\hat{p_{12}}(1-\hat{p_{12}})}{800} + \frac{\hat{p_{13}}(1-\hat{p_{13}})}{800} + \frac{2\hat{p_{12}}\hat{p_{13}}}{800}\right) \\
&= \frac{1}{4}\left(\frac{135}{800}(1-135/800)/800 + \frac{98}{800}(1-98/800)/800 + \frac{2\times 135\times 98}{800^3}\right) \\
&= 0.000090
\end{aligned}$$

Similarly $\mathrm{Var}(\hat{\beta_{21}}) \simeq 0.000114$ and $\mathrm{Var}(\hat{\beta_{31}}) \simeq 0.000108$.

An analysis of the saturated model as previously outlined in this chapter is given in Table 3.2. We have included α_3 in this table for convenience, it must be understood that since $\alpha_1 + \alpha_2 + \alpha_3 = 0$ there are only 2 degrees of freedom for these three parameters.

It is reasonable to deduce from Table 3.2 that the parameters β_{21} and β_{31} are non significant. The expected frequencies of the model obtained by leaving these parameters out are shown in Table 3.1 (labelled A), the estimates of the remaining parameters are unchanged. This model gives a χ^2 value of 2.67 on 2 degrees of freedom.

Table 3.2: Analysis of the Saturated Model

Parameter	Estimate	Z^2
α_1	-0.042	6.86
α_2	0.030	3.20
α_3	0.012	0.48
β_{12}	0.023	5.92
β_{21}	0.008	0.58
β_{31}	-0.015	2.09

As an example of the least squares calculating procedure, consider the model obtained by deleting the parameter α_3. Taking into consideration the constraints between the parameters the model can be written as

$$\begin{aligned} p_{123} &= \tfrac{1}{6} + \alpha_1/2 + \beta_{12} & p_{132} &= \tfrac{1}{6} + \alpha_1/2 - \beta_{12} \\ p_{213} &= p_{231} = \tfrac{1}{6} - \alpha_1/2 & p_{312} &= p_{321} = \tfrac{1}{6} \end{aligned}$$

For the iterative procedure we have the following:

$$X_1 = \begin{bmatrix} 1/2 & 1 \\ 1/2 & -1 \\ -1/2 & 0 \\ -1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \beta = \begin{bmatrix} \alpha_1 \\ \beta_{12} \end{bmatrix}$$

$$Z_1 = \begin{pmatrix} \frac{135}{800} - \frac{1}{6} \\ \vdots \\ \frac{126}{800} - \frac{1}{6} \end{pmatrix} \qquad \hat{P}^{-1} = \begin{bmatrix} \frac{800}{135} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \frac{800}{126} \end{bmatrix} \qquad \hat{p}_6 = \frac{150}{800}$$

We can calculate $V_{11}^{-1} = 800(\hat{P}^{-1} + \mathbf{11}'/\hat{p}_6)$ as

$$\begin{bmatrix} 9007 & 4267 & 4267 & 4267 & 4267 \\ & 10797 & 4267 & 4267 & 4267 \\ & & 8477 & 4267 & 4267 \\ & & & 8871 & 4267 \\ & & & & 9346 \end{bmatrix}$$

We calculate the initial estimate of β as

$$\hat{\beta} = (X_1'V_{11}^{-1}X_1)^{-1}X_1'V_{11}^{-1}Z_1 = \begin{pmatrix} -0.0366 \\ 0.0236 \end{pmatrix}$$

By calculating $P = \mathbf{1}/6 + X\beta$ and then recalculating V_{11}^{-1} we can iterate to find the least squares solutions which are $\hat{\alpha_1} = -0.0369$ and $\hat{\beta_{12}} = 0.0235$. We can see that the initial estimates are quite close to the final ones, in fact only one iteration was required to produce convergence to 4 decimal places. The expected frequencies obtained are shown in Table 3.1 (labelled B), and the corresponding χ^2 value is 3.22 on 3 degrees of freedom. The interpretation of this model is reasonably simple. There is a decrease equal to the magnitude of α_1 in the probability that a person will choose car 1 first in preference to car 3, and a corresponding increase in the probability of selecting car 2. We note that this indicates a preference for the current make. Of the people who indicate a preference for car 1 first, there is a preference, measured by the parameter β_{12}, for choosing car 2 over car 3 for the second position. When either car 2 or 3 is chosen first there is no preference between the remaining choices for the second position.

Chapter 4

Normal Order Statistics as Permutation Models

4.1 Introduction

Material from this chapter has been published in Dansie[10] and Dansie[11].

In considering distributions which allow us to model association between objects via correlated random variables, that is the so called order statistic models, the normal family is an obvious choice. The advantage of being able to choose a general correlation structure amongst pairs of variables seemingly gives us a reasonable deal of flexibility in modelling. The main disadvantage, as with many techniques based on the normal distribution, is that models are usually algebraically intractable and hence a fair degree of numerical work is involved. It is clear from the type of integrals we have already seen in specifying models for permutations that this is the case here.

4.2 The Normal Model Specification

4.2.1 Specification problems

We will begin by discussing the case $r = 3$ objects being ranked, for the simplicity of this situation enables the model specification and its associated problems to be clearly seen. The generalisation to $r = 4$ and beyond is reasonably straightforward.

As we have seen before in both the paired comparisons work and the disaggregate demand theory, the basis for our modelling is the use of underlying variables X_1, X_2, X_3. Values for these variables are unobserved, we model the probability of a permutation ijk as

$$p_{ijk} = \Pr(X_i > X_j > X_k)$$

We regard X_i as a random utility or perceived attraction for object i to a subject, and we postulate that the subject will select a ranking by maximising the perceived attraction for the 3 objects. In the case of the normal distribution we could equivalently write

$$p_{ijk} = \Pr(X_i^* < X_j^* < X_k^*)$$

since this is simply taking a linear transformation of the original variables and we thus remain within the normal family. The main difference would be the sign of the means, we would be interpreting X_i^* as a negative attraction or dislike for the object being ranked, in which case the subject is minimising dislike to form a ranking.

Returning to our original definition of perceived attraction, suppose that the variable $\mathbf{X} = (X_1, X_2, X_3)'$ has a general trivariate normal distribution with mean vector μ and variance matrix V, i.e. $\mathrm{E}(X_i) = \mu_i$, $\mathrm{Var}(X_i) = V_i^2$ and $\mathrm{cov}(X_i, X_j) = V_{ij}$. We can easily show that

$$p_{ijk} = \Pr(Z_1 < z_{ij}, Z_2 < z_{jk})$$

where $Z_1 = X_j - X_i, Z_2 = X_k - X_j$ have a standard (zero means, unit variances) bivariate normal distribution with correlation

$$\rho = \frac{V_{ij} + V_{jk} - V_{ik} - V_j^2}{\sqrt{(V_i^2 + V_j^2 - 2V_{ij})(V_j^2 + V_k^2 - 2V_{jk})}}$$

and

$$z_{ij} = \frac{\mu_i - \mu_j}{\sqrt{(V_i^2 + V_j^2 - 2V_{ij})}}$$

There are two problems with this specification for the normal model, the first of which is an identification problem. It is clear that one of the mean parameters is redundant, since we could translate the distribution by adding a constant c to each variable without altering the probabilities. We therefore impose a constraint amongst the means, it is useful for interpretation to set one of the means to zero. It is convenient to choose to set $\mu_3 = 0$. It is also clear that the same change of scale on each variable leaves the probabilities invariant, we could therefore set one of the variances to 1. There is however a deeper problem with the choice of variance parameters in the model which is highlighted by the following two examples.

Consider two seemingly different specifications. In model A set $V_{ij} = 0 \; \forall i, j \; i \neq j$ and $V_3^2 = 1$, so that the model consists of the four parameters $\{\mu_1', \mu_2', V_1^2, V_2^2\}$. For Model B set $V_i^2 = 1, i = 1, 2, 3$ and $V_{23} = 0$ giving also a four parameter model viz. $\{\mu_1, \mu_2, V_{12}, V_{13}\}$. Using these two models on a set of data we notice that the fitted values are identical. There are many other examples of this problem. In a similar way we find that the maximum number of unique parameters is 4 and thus it is, in general, impossible to fit a saturated model for the case $r = 3$ since there are five independent frequencies.

The second type of problem can also arise in a variety of ways. In the model specification it is implicit that V should be positive definite, since it is a variance matrix. This requirement, of course, places constraints on the parameters of the matrix. It is reasonably simple to construct a series of values for these parameters which give perfectly valid probabilities, in the sense that they sum to 1 and are non-negative, and yet the variance matrix formed is not positive definite. As an example consider the specification of model B. If we set $V_{12} = -3/2, V_{13} = 0$ then the derived probabilities using arbitrary values for the mean parameters are valid. With unit variances V_{12} is a correlation for which a value of -3/2 is non-sensical.

The resolution of these two problems lies in the realisation that constructing a model on the r-dimensional distribution of $\{X_1, X_2, \cdots, X_r\}$ results in a redundant specification in the general case. It is also true that this specification is not identical to a model constructed on the differences. In our case define

$$Y_1 = X_1 - X_3 \qquad Y_2 = X_2 - X_3$$

For the model on the differences put $\mathbf{Y} = (Y_1, Y_2)'$. Suppose that $\mathbf{Y}$ has a bivariate normal distribution with mean $\boldsymbol{\theta}$ and variance matrix Σ, where $\mathrm{E}(Y_i)=\theta_i$, $\mathrm{Var}(Y_i) = \sigma_i^2$ and $\mathrm{Cov}(Y_1, Y_2) = \sigma_{12}$. We can write all of the six probabilities in terms of Y_1 and Y_2 e.g.

$$p_{123} \quad = \quad \Pr(X_1 > X_2 > X_3) \quad = \quad \Pr(Y_1 - Y_2 > 0, Y_2 > 0)$$

No conflict arises between the two methods of specification in the case of independence and equal variances in the three dimensional model. The only parameters involved then are the means and it is fairly clear that the models are interchangeable. However in the case of a general variance matrix V in the three dimensional model, the two specifications are not identifiable, that is to say that there does not exist a unique 1-1 relationship between them.

It is firstly clear that one of the variance parameters of the distribution of $\mathbf{Y}$ is redundant, for example we can write the above expression for p_{123} as

$$p_{123} \quad = \quad \Pr((Y_1 - Y_2)/\sigma > 0, Y_2/\sigma > 0)$$

We can thus, without loss of generality, set $\sigma_2^2 = 1$. We will refer to the model with $\text{Var}(Y_2)=1$ as the standardised or scaled model. This model is achieved simply by dividing the original variables by $\sqrt{\text{Var}(Y_2)}$. We are then left with a four parameter model viz. $\{\theta_1, \theta_2, \sigma_1^2, \sigma_{12}\}$. We shall investigate the interpretation of these parameters later in the chapter. Let us firstly consider the two specification problems mentioned previously.

We can write down the equivalence relations between the parameters of the two and three dimensional models. We see that in terms of the original parameters of the three dimensional model i.e. mean vector μ and variance matrix V the bivariate model is

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} V_1^2 + V_3^2 - 2V_{13} & V_{12} + V_3^2 - V_{23} - V_{13} \\ & V_2^2 + V_3^2 - 2V_{23} \end{pmatrix}\right]$$

By considering the bivariate specifications of the relevant models we can clarify the two problems mentioned previously. In model A we had set $V_{12}, V_{13}, V_{23} = 0$ and $V_3^2 = 1$. We were therefore left with the four parameters $\{\mu_1', \mu_2', V_1^2, V_2^3\}$ The equivalent bivariate specification for this model is

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1' \\ \mu_2' \end{pmatrix}, \begin{pmatrix} 1 + V_1^2 & 1 \\ & 1 + V_2^2 \end{pmatrix}\right]$$

which we can rescale to give the standardised bivariate model:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1'/\sqrt{1+V_2^2} \\ \mu_2'/\sqrt{1+V_2^2} \end{pmatrix}, \begin{pmatrix} (1+V_1^2)/(1+V_2^2) & 1/(1+V_2^2) \\ & 1 \end{pmatrix}\right]$$

For model B with $V_i^2 = 1$, $i = 1,2,3$, and $V_{23} = 0$, the standardised bivariate specification arrived at in the same way is

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1/\sqrt{2} \\ \mu_2/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1 - V_{13} & (1 - V_{13} + V_{12})/2 \\ & 1 \end{pmatrix}\right]$$

The two models are clearly equivalent. It is a simple matter to write down the relationships between the parameters of Models A and B as

$$\mu_i = \frac{\sqrt{2}}{\sqrt{1+V_2^2}}\,\mu_i' \quad i = 1,2 \qquad V_{12} = \frac{1 - V_1^2}{1 + V_2^2} \qquad V_{13} = \frac{V_2^2 - V_1^2}{1 + V_2^2}$$

It is also useful to have the relationship between the trivariate specification of Model B using the correlation parameters and the standardised bivariate specification with 4 parameters $\theta_1, \theta_2, \sigma_1^2, \sigma_{12}$. These relationships are

$$\theta_i = \mu_i/\sqrt{2} \quad \sigma_1^2 = 1 - V_{13} \quad \sigma_{12} = \frac{1 + V_{12} - V_{13}}{2}$$

It is clear from looking at the standard bivariate specification of model B that the requirement of a positive definite variance matrix in the bivariate case gives different constraints on the values of the variance parameters from the original specification. As an example consider the problem raised earlier when for model B we have the clearly invalid model $V_{12} = -3/2, V_{13} = 0$. The equivalent bivariate specification is

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1/\sqrt{2} \\ \mu_2/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1 & -1/4 \\ & 1 \end{pmatrix}\right]$$

This variance matrix is positive definite and so provides a valid bivariate model.

The practical implication of these observations is that we should use the bivariate form of specification for $r = 3$ and in general the $r-1$ dimensional specification for the case of r objects being ranked. The interpretation of parameters and the symmetry of the r dimensional model however make it an attractive and easily interpreted form to use. In most cases in practice, the problem of seemingly invalid parameter values does not arise because the values which occur for the correlations in the r dimensional model are usually fairly close to zero. It is then likely that the r dimensional model is valid and hence no problem exists. It is easy enough to find the equivalent $r-1$ dimensional model if, on occasion, the need arises.

The problems posed by the parameterisation however can be slightly more serious. The example considered showed that two seemingly different four parameter models were equivalent. In general we have noted that we have two mean and two variance parameters to use in our modelling. The choice of which two variance parameters to use is not an arbitrary one; we in fact run a risk of introducing a redundant parameter if we do not have regard for the bivariate specification. To illustrate this point consider the following example from a paper of Horowitz et al.[20]

4.2.2 An example of a Redundant Specification

The aim of the paper of Horowitz et al. [20] is to investigate the use of an approximation due to Clark[6] for calculating probabilities using a multinomial probit model. Adopting, for this section only, the notation of the paper of Horowitz et al. we write the utility of alternative j expressed as a function of an attribute vector $\mathbf{X}_j$ as

$$U_j = \mathbf{X}'_j\theta + \epsilon_j$$

We regard $\mathbf{U}$, the vector of utilities, as a trivariate normal random variable with mean vector $\mathbf{V}$, which has components $V_j = \mathbf{X}'_j\theta$.

The method used to investigate the accuracy of the approximation was to firstly select a parametrisation for the model. Two independent sets of experiments were conducted. The particular model of interest is the one used in the Berkeley experiments, we shall refer to it as Model A . The particular details of the model are a mean structure given as

$$V_1 = \theta_1 + \theta_3 x_1 \quad V_2 = \theta_2 + \theta_3 x_1 \quad V_3 = 0$$

and the covariance matrix denoted by σ_A given as

$$\sigma_A = \begin{bmatrix} 1 & \sigma_{12} & 0 \\ \sigma_{12} & 1 & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}$$

One hundred values for the explanatory variables x_1 were chosen randomly. Using known values for the parameters, a set of choices were simulated and then an estimation procedure based on the Clark approximation was implemented. The experiments were repeated with different values chosen for the parameters. These values were selected to represent a wide range of possible situations likely to be encountered in practice. The analysis of the experiment consisted of a comparison of the actual values used and the estimated values given by the approximation.

The problem with the specification of model A is that it contains a redundant parameter. Horowitz et al. note that there are only two non-redundant variance parameters available for modelling, but as we observed previously the choice of these parameters in the three dimensional model is not an arbitrary one. To see clearly the nature of the redundancy consider two submodels of Model A that have the same mean structure but covariance matrices σ_B and σ_C respectively, given by

$$\sigma_B = \begin{bmatrix} 1 & \sigma_{12} & 0 \\ \sigma_{12} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \sigma_C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}$$

The estimation problem that arises is that these three models are equivalent. To see this consider the equivalent standardised bivariate variance matrices formed from the three models. These matrices are

$$\begin{array}{ccc} \text{ModelA} & \text{ModelB} & \text{ModelC} \\ \begin{bmatrix} 1 & \frac{\sigma_{12}+\sigma_{33}}{1+\sigma_{33}} \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & \frac{1+\sigma_{12}}{2} \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & \frac{\sigma_{33}}{1+\sigma_{33}} \\ & 1 \end{bmatrix} \end{array}$$

Each model has the same bivariate parameter in it and therefore is equivalent to the other two. Considering model A as Horowitz et al. have, we see that one of the two parameters in the variance matrix is redundant. This causes a problem in the estimation procedure because any pair of estimates of $(\sigma_{12}, \sigma_{13})$ that produces the same value of the bivariate covariance parameter

$$\rho = \frac{\sigma_{12} + \sigma_{33}}{1 + \sigma_{33}}$$

are valid estimates for the two parameters. This of course makes comparison of the estimated values with the true values very difficult.

As an example consider one of the set of values, labelled 15(B), listed in the results of the experiments. Table 4.1 contains the original values given by Horowitz et al. A comparison of the actual values used for the parameters and those estimated

Table 4.1: Comparison of redundant and scaled parameter values

		θ_1	θ_2	θ_3	σ_{12}	σ_{33}	ρ
ORIGINAL	ACTUAL	-5.00	5.00	10.00	-0.8	25.00	*
	ESTIMATED	-4.07	4.45	9.27	1.00	25.40	*
SCALED	ACTUAL	-0.98	0.98	1.96	*	*	0.93
	ESTIMATED	-0.79	0.87	1.80	*	*	1.00

from the generated data shows a peculiar value for the correlation parameter σ_{12}. Having generated data with a value of -0.8 for this correlation we notice that the estimated value is given as 1.0. I believe that the redundant parameterisation has had a significant effect on this comparison. To illustrate this we can scale both the initial and estimated values to their non-redundant bivariate equivalents. The relevant scalings are

$$\theta_i' = \frac{\theta_i}{(1 + \sigma_{33})^{1/2}} \qquad \rho = \frac{\sigma_{12} + \sigma_{33}}{1 + \sigma_{33}}$$

The scaled values obtained by these transformations, that is the values of the means and correlation of the equivalent standardised bivariate model are given in table 4.1. On first inspection it appears that the comparison of these parameter values would suggest that the approximation was reasonable in this case. A deeper analysis is required to draw this conclusion however, since it is known that the performance of the Clark approximation is affected by the normalisation chosen. It is not possible to easily separate out the effects of the redundancy and the a-posterior scaling.

4.2.3 General model specification

One of the desirable aspects of the normal distribution which makes it attractive for the modelling of permutation probabilities is the notion that we can have correlations between variables, which in addition can have unequal variances. As we saw in the introduction the users of the multinomial probit model certainly had this in mind. The generalised extreme value distributions introduced by McFadden[27], which we shall consider in detail in the next chapter, do not allow this unequal variance in their main form and hence are seen as being inflexible in this sense. Horowitz et al.[20] for example, is one of many sources which draw attention to this claimed advantage of the normal model.

We have seen in the equivalence of the two models A and B in the first section of this chapter, that for the case $r = 3$ all of our models can be written with a parametrisation of the variance matrix which uses either two variance parameters or two correlation parameters. The example of the redundant specification of the previous section would suggest that if models are to be specified in the three dimensional form then redundancies are most easily avoided if we do not attempt to use both variance and correlation parameters, but rather restrict our usage to one or the other.

In the case of $r = 4$ there are not enough variance parameters in the four dimensional model for us to express any model using just variance parameters. However it is still the case that we can generate any model if we restrict our set of parameters just to the set of correlation parameters. To see this define

$$Y_1 = X_1 - X_4 \qquad Y_2 = X_2 - X_4 \qquad Y_3 = X_3 - X_4$$

Using notation analogous to the that originally introduced we can see that the joint distribution of $\mathbf{Y} = (Y_1, Y_2, Y_3)'$ is

$$\mathrm{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} V_1^2 + V_4^2 - 2V_{14} & V_{12} + V_4^2 - V_{14} - V_{24} & V_{13} + V_4^2 - V_{14} - V_{34} \\ & V_2^2 + V_4^2 - 2V_{24} & V_{23} + V_4^2 - V_{24} - V_{34} \\ & & V_3^2 + V_4^2 - 2V_{34} \end{pmatrix}\right]$$

where we have assumed that the mean of X_4 is 0. It is once again clear that we can standardise the scale of one of the variates without altering the probabilities and so without loss of generality divide the variates by the standard deviation of Y_3. We are thus left with five parameters in this variance matrix and three mean parameters, a total of 8 parameters in all. This should be compared to the 11 parameters required for the second stage of the Plackett system and thus it it reasonable to regard these normal models as being of second order. We are considerably short of the 23 parameters required for the saturated model but in practice this is of little concern.

By inspection of the previous expression for the standardised variance matrix two points become apparent. It is not possible to generate any full 5 parameter variance model using only the four variance parameters $\{V_1^2, V_2^2, V_3^2, V_4^2\}$, since there are simply not enough parameters. It is also clear that any model could be written in terms of any five of the six parameters $\{V_{12}, V_{13}, V_{14}, V_{23}, V_{24}, V_{34}\}$. It is simplest to set $V_{34} = 0$ leaving the other five parameters to be our modelling set. Setting $V_i^2 = 1 \quad i = 1, \ldots, 4$ we see that the reduced trivariate specification (allowing for scaling $\mathrm{Var}(Y_3) = 1$) is

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1/\sqrt{2} \\ \mu_2/\sqrt{2} \\ \mu_3/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1 - V_{14} & \frac{1+V_{12}-V_{14}-V_{24}}{2} & \frac{1+V_{13}-V_{14}}{2} \\ & 1 - V_{24} & \frac{1+V_{23}-V_{24}}{2} \\ & & 1 \end{pmatrix}\right]$$

Each of the five terms in the variance matrix which are not equal to one contain a unique member of the set of five parameters mentioned previously and therefore this set does indeed generate any model we require. To illustrate how we can deduce equivalent specifications between different four variate models, consider for example a model with parameters $\{\mu_1', \mu_2', \mu_3', V_1^2, V_3^2\}$. The equivalent scaled trivariate specification is given by

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_1'/\sqrt{1+V_3^2} \\ \mu_2'/\sqrt{1+V_3^2} \\ \mu_3'/\sqrt{1+V_3^2} \end{pmatrix}, \begin{pmatrix} \frac{1+V_1^2}{1+V_3^2} & \frac{1}{1+V_3^2} & \frac{1}{1+V_3^2} \\ & \frac{2}{1+V_3^2} & \frac{1}{1+V_3^2} \\ & & 1 \end{pmatrix}\right]$$

Comparing this with the previous expression we see that we can deduce the equivalent model using the five correlation parameters given by

$$\mu_i = \frac{\sqrt{2}}{\sqrt{1+V_3^2}}$$

$$1 - V_{24} = \frac{2}{1 + V_3^2} \quad \rightarrow \quad V_{24} = \frac{V_3^2 - 1}{1 + V_3^2}$$

$$1 - V_{14} = \frac{1 + V_1^2}{1 + V_3^2} \quad \rightarrow \quad V_{14} = \frac{V_3^2 - V_1^2}{1 + V_3^2}$$

$$V_{12} = V_{14} \qquad V_{13} = \frac{1 - V_1^2}{1 + V_3^2} = V_{14} - V_{24} \qquad V_{23} = 0$$

It is clear that the original specification of the normal model in terms of the variance of the original variables can easily be respecified in terms of a set of correlation parameters. The form of the four variate variance matrix using two parameters $\alpha = V_{14}$ and $\beta = V_{24}$ would be

$$\begin{pmatrix} 1 & \alpha & \alpha - \beta & \alpha \\ & 1 & 0 & \beta \\ & & 1 & 0 \\ & & & 1 \end{pmatrix}$$

The same model expressed in the trivariate form would have variance matrix given by

$$\begin{pmatrix} \gamma & \delta & \delta \\ & 2\delta & \delta \\ & & 1 \end{pmatrix}$$

where $\gamma = (1 + V_1^2)/(1 + V_3^2)$ and $\delta = 1/(1 + V_3^2)$.

As a result of this small example we can see that models in general will be much simpler to build and interpret in the fully dimensional case. We have seen that for $r = 4$ we can avoid redundancies by selecting only from the correlation parameters which we shall see in the next section have a natural interpretation of association. For general r this is also the case. The $r - 1$ variate specification of the model of r objects contains $r(r-1)/2 - 1$ independent variance matrix parameters. The r dimensional model has $r(r-1)/2$ covariance parameters and by the same construction used for the previous case it is clear that we can state any model using all but one of these parameters. It is also clear that we will avoid any redundancies by following this rule. This is the suggested method for constructing such models.

We shall now turn our attention to the extension of Henery's[17] Taylor series approximation to these probabilities

4.3 Taylor series approximation

The aim of this section is to develop the first order Taylor series approximation to p_{ijk}, extending the result of Henery[17]. The resultant expression is a useful approximation model in it's own right, and it also serves as a good approximation to the computationally more difficult normal model. As we suggested at the end of the previous section, the r variate distribution is more convenient for modelling purposes. This is also the case with the argument to follow and so we will develop the approximation for the r dimensional model.

Henery[17] obtained a Taylor series expansion for p_{ijk} in the case of independence as follows. Suppose that $X_1, X_2, \ldots, X_r$ are independent normal random variables with means $\mu_1, \mu_2, \ldots, \mu_r$. Denote by $P(\pi)$ the probability that $X_1 > X_2 > \ldots > X_r$. We can write $P(\pi)$ as

$$P(\pi) = \int_{-\infty}^{\infty} \phi(x_r - \mu_r) \int_{x_r}^{\infty} \phi(x_{r-1} - \mu_{r-1}) \ldots \int_{x_2}^{\infty} \phi(x_1 - \mu_1) dx_1 \ldots dx_r$$

where $\phi(z)$ is the standard normal density function. The terms required in the Taylor series for $P(\pi)$ are the evaluation of $P(\pi)$ and its first derivatives at $\mu_1 = \mu_2 = \ldots = \mu_r = 0$. $P(\pi)$ evaluated at $\mu_i = 0, \;\; i = 1, \ldots, r$ is clearly 1/r!. A typical derivative is

$$\frac{\partial P(\pi)}{\partial \mu_i} = \int_{-\infty}^{\infty} \phi(x_r) \ldots \int_{x_{i+1}}^{\infty} x_i \phi(x_i) \int_{x_i}^{\infty} \phi(x_{i-1}) \ldots \int_{x_2}^{\infty} \phi(x_1) dx_1 \ldots dx_r$$

This expression is equal to $\theta_{i:r}/r!$ where $\theta_{i:r}$ is the expected value of the ith order statistic in a random sample of r N(0, 1) random variables arranged in decreasing order. For small values of μ_i Henery obtains the approximation:

$$P(\pi) = \frac{1}{r!} + \frac{\sum \mu_i \theta_{i:r}}{r!}$$

Note that this summary is slightly different to the derivation given by Henery since he defined the permutation probability using the smallest order statistic as the one chosen first so that

$$P(\pi) = \Pr(X_1^* < X_2^* < \ldots < X_r^*)$$

We have previously noted the equivalence of these two approaches in the Normal case.

As an example of the non-independent case consider $r = 3$. For small values of r, expressions are available for such constants as $\theta_{i:r}$ (see Jones[22] and Godwin[14]). Consulting these tables gives us $\theta_{1:3} = 3/(2\sqrt{\pi}) = a, \theta_{2:3} = 0$ and $\theta_{3:3} = -\theta_{1:3} = -a$ We can thus write an approximation for p_{123} in the independence case as

$$p_{123} = \frac{1}{6} + \frac{a\mu_1}{6} - \frac{a\mu_3}{6}$$

In the general case for $r = 3$ suppose that $\mathbf{X} = (X_1, X_2, X_3)'$ has a trivariate Normal distribution with unit variances. We can express the probability density function in the following useful form (see for example Johnson and Kotz[21]).

$$f(x_1, x_2, x_3) = (2\pi)^{-3/2}\delta^{-1/2} \exp\left\{-\delta^{-1}\sum_{i=1}^{3}\sum_{j=i}^{3} A_{ij}(x_i - \mu_i)(x_j - \mu_j)\right\}$$

where

$$\begin{aligned}
\delta &= 1 - V_{23}^2 - V_{12}^2 - V_{13}^2 + 2V_{23}V_{13}V_{12} \\
A_{11} &= (1 - V_{23}^2)/2 \quad A_{22} = (1 - V_{13}^2)/2 \quad A_{33} = (1 - V_{12}^2)/2 \\
A_{12} &= V_{12}V_{23} - V_{12} \quad A_{13} = V_{12}V_{23} - V_{13} \\
A_{23} &= V_{12}V_{13} - V_{23}
\end{aligned}$$

Consider the probability p_{123}:

$$p_{123} = \Pr(X_1 > X_2 > X_3) = \int_{-\infty}^{\infty}\int_{x_3}^{\infty}\int_{x_2}^{\infty} f(x_1, x_2, x_3)dx_1dx_2dx_3$$

where for our purposes we set μ_3 and V_{23} =0 leaving the four parameter model that we require. As suggested by Henery[17] we find the Taylor series in the parameters $\beta = (\mu_1, \mu_2, V_{12}, V_{13})'$ as far as the linear terms. We require p_{123} evaluated at β=0, which as in Henery is clearly 1/3!. In addition we must evaluate the following at β=0:

$$\frac{\partial p_{123}}{\partial \mu_i} \quad i = 1, 2 \quad \text{and} \quad \frac{\partial p_{123}}{\partial V_{1j}} \quad j = 2, 3$$

We will begin by showing that the first of these derivatives gives the same expression as given by Henery[17] i.e.

$$\frac{\partial p_{123}}{\partial \mu_i} = \theta_{i:3} \quad i = 1, 2$$

With $f(x_1, x_2, x_3)$ written in the above form consider the example of the case $i = 1$. We can express this derivative as

$$\frac{\partial f(x_1, x_2, x_3)}{\partial \mu_1} = (2\pi)^{-3/2}\delta^{-1/2}\exp(\delta^{-1}T)\frac{\partial T}{\partial \mu_1}$$

where

$$T = -\sum_{i=1}^{3}\sum_{j=i}^{3} A_{ij}(x_i - \mu_i)(x_j - \mu_j)$$

and

$$\frac{\partial T}{\partial \mu_1} = 2A_{11}(x_1 - \mu_1) + A_{12}(x_2 - \mu_2) + A_{13}(x_3 - \mu_3)$$

Evaluating this expression at β=0, we see that $\delta = 1, A_{11} = A_{22} = A_{33} = 1/2$ and $A_{ij} = 0 \quad \forall i \neq j$. This gives

$$\frac{\partial f(x_1, x_2, x_3)}{\partial \mu_1} = (2\pi)^{-3/2} x_1 \exp\left(-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)\right)$$

which we recognise as the product of x_1 and the three independent standard normal densities $\phi(x_i)$. We can therefore write

$$\frac{\partial p_{123}}{\partial \mu_1} = \int_{-\infty}^{\infty} \phi(x_3) \int_{x_3}^{\infty} \phi(x_2) \int_{x_2}^{\infty} x_1 \phi(x_1) dx_1 dx_2 dx_3$$

This is equivalent to the expression given by Henery(1981) and is therefore equal to $\theta_{1:3}/6$ where, as before, $\theta_{1:3}$ is the expected value of the first order statistic in a sample of size three from a standard normal distribution. It is also clear, by analogy, that

$$\frac{\partial p_{123}}{\partial \mu_2} = \frac{\theta_{2:3}}{3!}$$

For the second type of derivative using V_{12} as an example, we see that

$$\begin{aligned}
\frac{\partial f(x_1, x_2, x_3)}{\partial V_{12}} &= (2\pi)^{-3/2}\left[\delta^{-1/2}\exp(\delta^{-1}T)\frac{\partial(\delta^{-1}T)}{\partial V_{12}} - \exp(\delta^{-1}T)\frac{1}{2}\delta^{-3/2}\frac{\partial \delta}{\partial V_{12}}\right] \\
&= (2\pi)^{-3/2}\delta^{-1/2}\exp(\delta^{-1}T)\left[\frac{\partial(\delta^{-1}T)}{\partial V_{12}} - \frac{1}{2}\delta^{-1}\frac{\partial \delta}{\partial V_{12}}\right]
\end{aligned}$$

Now

$$\frac{\partial(\delta^{-1}T)}{\partial V_{12}} = -\left[\sum_{i=1}^{3}\sum_{j=i}^{3} A_{ij}(x_i-\mu_i)(x_j-\mu_j)(-\delta^{-2}\frac{\partial\delta}{\partial V_{12}})\right.$$

$$\left.+\delta^{-1}\sum_{i=1}^{3}\sum_{j=i}^{3}\frac{\partial A_{ij}}{\partial V_{12}}(x_i-\mu_i)(x_j-\mu_j)\right]$$

Since

$$\frac{\partial\delta}{\partial V_{12}} = -2V_{12}+2V_{23}V_{13}$$

which when evaluated at $\beta = \mathbf{0}$ is 0, the first term in this expression is 0. In addition, all of the partial derivatives of A_{ij} with respect to V_{12} when evaluated at $\beta = \mathbf{0}$ are also 0 except for

$$\frac{\partial A_{12}}{\partial V_{12}} = -1$$

Since $\delta = 1$ when $\beta = \mathbf{0}$ we have

$$\frac{\partial(\delta^{-1}T)}{\partial V_{12}} = x_1x_2$$

and hence

$$\frac{\partial f(x_1,x_2,x_3)}{\partial V_{12}} = (2\pi)^{-3/2}x_1x_2\exp\left(-\frac{1}{2}(x_1^2+x_2^2+x_3^2)\right)$$

i.e. the product of x_1x_2 and the three independent normal densities. We thus have

$$\frac{\partial p_{123}}{\partial V_{12}} = \int_{-\infty}^{\infty}\int_{x_3}^{\infty}\int_{x_2}^{\infty} x_1x_2\phi(x_1)\phi(x_2)\phi(x_3)dx_1dx_2dx_3$$

This expression is equal to $\theta_{12:3}/6$ where $\theta_{12:3}$ is the expected value of the product of the first two order statistics in an independent sample of size three from the standard normal distribution. By symmetry we can see that at $\beta = 0$ we have

$$\frac{\partial p_{123}}{\partial V_{13}} = \frac{\theta_{13:3}}{3!}$$

For small values of the parameters $\mu_1, \mu_2, V_{12}, V_{13}$ we can write approximately

$$p_{123} = \frac{1}{6} + \frac{\mu_1 \theta_{1:3}}{6} + \frac{\mu_2 \theta_{2:3}}{6} + \frac{V_{12}\theta_{12:3}}{6} + \frac{V_{13}\theta_{13:3}}{6}$$

The values of these constants are available from the tables mentioned previously.

In general for p_{ijk} the coefficient of μ_1 is $\theta_{i':3}/6$ where i' is the position in the permutation occupied by object 1. The coefficient of V_{12} is $\theta_{i'j':3}$ where i' and j' are the positions occupied by objects 1 and 2 respectively in the ranking. For example

$$p_{213} = \frac{1}{6} + \frac{\mu_1 \theta_{2:3}}{6} + \frac{\mu_2 \theta_{1:3}}{6} + \frac{V_{12}\theta_{21:3}}{6} + \frac{V_{13}\theta_{23:3}}{6}$$

As a consequence of these results we can approximate the vector of probabilities $\mathbf{p}$ as a linear function of the parameters β,viz.

$$\mathbf{p} = \frac{1}{6} + X\beta \qquad \mathbf{p}(6 \times 1) \qquad X(6 \times 4) \qquad \beta(4 \times 1)$$

In its own right this expression represents a model for permutation probabilities. The two main advantages of this model are its simplicity and ease of interpretation. In addition, its parameter estimates are very easy to calculate, by the adaptation given below of the iterative least squares method, as outlined in chapter 3. Following the notation of that chapter denote by $\mathbf{Y}$ the (6×1) vector of relative frequencies. Our model is

$$\mathrm{E}(\mathbf{Y}) = \mathbf{1}/6 + X\beta, \qquad \mathrm{V}(\mathbf{Y}) = V$$

where V is the variance matrix of a set of multinomial probabilities and $\mathbf{1}$ denotes the unit vector of length 6. We have one constraint, $\mathbf{1}'\mathbf{Y} = 1$ and so we have

$\gamma = 1/6$. Construct $\mathbf{Z} = \mathbf{Y} - \gamma$ and partition $\mathbf{Z}$ as $(\mathbf{Z}'_1, \mathbf{Z}'_2)'$ where $\mathbf{Z}_1$ is (5×1). Similarly partition X and V so that X_1 is (5×4) and V_{11} is (5×5). The least squares estimates are given by

$$\hat{\beta} = (X_1' V_{11}^{-1} X_1)^{-1} X_1' V_{11}^{-1} \mathbf{Z}_1.$$

Estimates are obtained by iterative least squares, although using the relative frequencies as consistent estimates in V_{11}, the solution on the first cycle is often of sufficient accuracy. The variance matrix of the estimates is given by $(X_1' V_{11}^{-1} X_1)^{-1}$. It is a simple matter to fit the 4 parameter model and, using the variance matrix, eliminate non significant parameters. As we shall see in the practical example to follow, this approximation is a good one, in particular of course for small values of the parameters. It is a useful technique for obtaining very good initial estimates, for use in the fitting of the normal probabilities model. A suggested procedure is to find a minimal model using the Taylor series approximation, and then to calculate the estimates of the normal model by numerical maximisation of the likelihood function using, for example, search algorithms such as the method outlined by O'Neill[29]. The calculation of the normal probabilities requires the evaluation of an $r - 1$ dimensional multivariate integral. For the most common uses of these models, viz $r = 3$ or 4 this is a well established technique. (See for example Daley[8]). For more recent work and methods suitable for higher dimensions the reader is referred to Schervish[35],[36] and Russell et al.[33].

4.4 Interpretation of the model

4.4.1 The $r = 3$ case

Making use of the Taylor series approximation let us briefly consider the interpretation of the parameters of the three dimensional model. Table 4.2 contains the coefficient matrix X from the Taylor series expansion $\mathbf{p} = 1/6 + X\beta$. The

Table 4.2: Parameter coefficients in the $r = 3$ Taylor series approximation

	Parameter			
Permutation	μ_1	μ_2	V_{12}	V_{13}
123	a	0	b	-2b
132	a	-a	-2b	b
213	0	a	b	b
231	-a	a	-2b	b
312	0	-a	b	b
321	-a	0	b	-2b

$$a = (4\sqrt{\pi})^{-1} \quad b = (4\pi\sqrt{3})^{-1}$$

interpretation of the mean parameter μ_i is clear, it is a measure of the attraction for object i, relative to a zero attraction for object 3. For example, if $\mu_1 > 0$ then there is a preference for object 1 over object 3, and thus permutations beginning with 1, i.e. 123,132 have a higher probability than those ending with 1. The other two parameters, V_{12} and V_{13}, reflect patterns of association in the rankings of pairs of objects. It is clear that V_{12} is a measure of the association in the perception of the value of objects 1 and 2, since in those permutations where objects 1 and 2 are together, its coefficient is positive, and for those where 1 and 2 are apart, namely

132 and 231, the coefficient is negative. If $V_{12} > 0$ therefore, there is an increase in the probability of those permutations which have 1 and 2 together, i.e. a positive association between the rankings of objects 1 and 2. The parameter V_{13} has a similar interpretation in terms of objects 1 and 3; if $V_{13} > 0$ there is a positive association between the ranking of 1 and 3. Clearly the value of a probability will depend on the relative magnitude of the individual parameters.

In order to see the expansion of these probabilities in terms of the parameters of the standardised bivariate model we make use of the relationships given for the parameters of the two models. These relationships are

$$\theta_i = \mu_i/\sqrt{2} \quad \sigma_1^2 = 1 - V_{13} \quad \sigma_{12} = \frac{1 + V_{12} - V_{13}}{2}$$

The values chosen to calculate the Taylor expansion around i.e. $\mu_i = V_{ij} = 0$ are of course the values for the zero means independence model. Using the given relations these values correspond in the bivariate case to $\theta_i = 0, \sigma_1^2 = 1, \sigma_{12} = 1/2$. If we put $\phi = (\theta_1, \theta_2, \sigma_1^2 - 1, \sigma_{12} - 1/2)'$ we deduce from the given equivalence relations that we can write ϕ as a linear function of the original vector of parameters β. We write $\phi = A\beta$ where A is given by

$$\begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1/2 & -1/2 \end{bmatrix}$$

A is clearly non-singular and hence invertible. It is clear that from the original expansion $\mathbf{p} = 1/6 + X\beta$ we can write $\mathbf{p} = 1/6 + X^*\phi$ where $X^* = XA^{-1}$. This construction can be derived from considerations of the bivariate model alone, but the symmetry of the trivariate specification produces a simple derivation which can be easily generalised. The coefficient matrix X^* is given in Table 4.3. From

Table 4.3: Coefficient matrix in bivariate Taylor series approximations

	Parameter			
Permutation	θ_1	θ_2	$\sigma_1^2 - 1$	$\sigma_{12} - 1/2$
123	$\sqrt{2}$a	0	b	2b
132	$\sqrt{2}$a	-$\sqrt{2}$a	b	-4b
213	0	$\sqrt{2}$a	-2b	2b
231	-$\sqrt{2}$a	$\sqrt{2}$a	b	-4b
312	0	-$\sqrt{2}$a	-2b	2b
321	-$\sqrt{2}$a	0	b	2b

$$a = (4\sqrt{\pi})^{-1} \quad b = (4\pi\sqrt{3})^{-1}$$

this table the interpretation of the parameters of the bivariate model follows fairly simply. Relative to a zero attraction for object 3, the parameters θ_1 and θ_2 measure the preference for objects 1 and 2 respectively. It is clear that σ_1^2 is a measure of the association between objects 2 and 3, since in those permutations that have objects 2 and 3 together the coefficient of (σ_1^2) is positive and negative otherwise. If $\sigma_1^2 > 1$ therefore, there is an increase in the probability of those rankings which have objects 2 and 3 adjacent. The parameter σ_{12} has a similar interpretation in terms of objects 1 and 2. If $\sigma_{12} > 1/2$ there is a positive association between objects 1 and 2.

There are a few submodels of the bivariate model that should be noted at this stage. The bivariate model with $\sigma_1^2 = 1$ and $\sigma_{12} = 1/2$ of course corresponds to the independence model. To this basic model in the trivariate system we can add a single correlation parameter in 3 different ways. There are three corresponding models in the bivariate specification. An obvious model is to set $\sigma_{12} = 1/2$ leaving σ_1^2 as the third parameter. This corresponds to a trivariate model with the parameter V_{23}. Alternatively we can set $\sigma_1^2 = 1$ leaving the parameter σ_{12} free, this corresponds to the trivariate model with the parameter V_{12}. The corresponding model to the trivariate model with parameter V_{13} is constructed by putting $\sigma_1^2 = \delta$ and $\sigma_{12} = \delta/2$. If we look at Table 4.3 we will see that the coefficient of delta in the expansion would be the coefficient of $(\sigma_1^2 - 1)$ plus a half of the coefficient of $(\sigma_{12} - 1/2)$. Performing this calculation it is clear that δ is a measure of the association between objects 1 and 3 with $\delta > 1$ representing a negative association because of the way we have defined Y_1 and Y_2.

4.4.2 The $r = 4$ model

The expansion of probabilities in the case $r > 3$ proceeds by direct analogy. For $r = 4$ the matrix X in the expansion $\mathbf{p} = \mathbf{1}/24 + X\beta$ is given in Table 4.4

We can find the expansion in terms of the trivariate difference distribution in a similar way to the method employed previously. The relationships between the two sets of parameters are

$$
\begin{aligned}
\theta_i &= \mu_i/\sqrt{2} \quad i = 1, 2, 3 \\
\sigma_1^2 - 1 &= -V_{14} \\
\sigma_2^2 - 1 &= -V_{24} \\
\sigma_{12} - 1/2 &= (V_{12} - V_{14} - V_{24})/2
\end{aligned}
$$

Table 4.4: Parameter coefficients in Taylor series approximations to the permutation probabilities for $r = 4$

	Parameter							
Permutation	μ_1	μ_2	μ_3	V_{12}	V_{13}	V_{14}	V_{23}	V_{24}
1234	a	b	-b	c	d	e	-d	d
1243	a	b	-a	c	e	d	d	-d
1342	a	-b	b	d	c	e	-d	c
1342	a	-a	b	e	c	d	d	c
1423	a	-b	-a	d	e	c	c	-d
1432	a	-a	-b	e	d	c	c	d
2134	b	a	-b	c	-d	d	d	e
2143	b	a	-a	c	d	-d	e	d
2314	-b	a	b	d	-d	c	c	e
2341	-a	a	b	e	d	c	c	d
2413	-b	a	-a	d	c	-d	e	c
2431	-a	a	-b	e	c	d	d	c
3124	b	-b	a	-d	c	d	d	c
3142	b	-a	a	d	c	-d	e	c
3214	-b	b	a	-d	d	c	c	d
3241	-a	b	a	d	e	c	c	-d
3412	-b	-a	a	c	d	-d	e	d
3421	-a	-b	a	c	e	d	d	-d
4123	b	-b	-a	-d	d	c	c	d
4132	b	-a	-b	d	-d	c	c	e
4213	-b	b	-a	-d	c	d	d	c
4231	-a	b	-b	d	c	e	-d	c
4312	-b	-a	b	c	-d	d	d	e
4321	-a	-b	b	c	d	e	-d	d

$$\mathrm{a} = \frac{\sqrt{\pi}}{24}\left[\frac{2}{5}\frac{15}{4\pi} + \frac{2}{5}\frac{15}{2\pi^2}\arcsin(\frac{1}{3})\right]$$

$$\mathrm{b} = \frac{\sqrt{\pi}}{24}\left[\frac{2}{5}\frac{15}{4\pi} - \frac{6}{5}\frac{15}{2\pi^2}\arcsin(\frac{1}{3})\right]$$

$$\mathrm{c} = \frac{\sqrt{3}}{24\pi} \qquad d = \frac{-(2\sqrt{3}-3)}{24\pi} \qquad e = \frac{-3}{24\pi}$$

$$\sigma_{13} - 1/2 = (V_{13} - V_{14})/2$$

$$\sigma_{23} - 1/2 = (V_{23} - V_{24})/2$$

Put $\phi = (\theta_1, \theta_2, \theta_3, \sigma_1^2 - 1, \sigma_2^2 - 1, \sigma_{12} - 1/2, \sigma_{13} - 1/2, \sigma_{23} - 1/2)'$. We can write $\phi = A\beta$ where A is given by

$$\begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1/2 & 0 & -1/2 & 0 & -1/2 \\ 0 & 0 & 0 & 0 & 1/2 & -1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & -1/2 \end{bmatrix}$$

As before we can write $\mathbf{p} = \mathbf{1/6} + X^*\phi$ where $X^* = XA^{-1}$. We have seen previously an example of the construction of equivalent difference models by making use of the equivalence relations.

4.5 Example calculations

To illustrate the calculations involved, and to compare the approximate and exact models consider the data used in the example of chapter 3. For convenience the observed frequencies of the 6 permutations are given in the first row of Table 4.5. In order to apply the least squares procedure to fit the approximate model we use

$$X_1 = \begin{bmatrix} 0.1410 & 0.0000 & 0.0459 & -0.0919 \\ 0.1410 & -0.1410 & -0.0919 & 0.0459 \\ 0.0000 & 0.1410 & 0.0459 & 0.0459 \\ -0.1410 & 0.1410 & -0.0919 & 0.0459 \\ 0.0000 & -0.1410 & 0.0459 & 0.0459 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu_1 \\ \mu_{12} \\ V_{12} \\ V_{13} \end{bmatrix}$$

As before we have

$$Z_1 = \begin{pmatrix} \frac{135}{800} - \frac{1}{6} \\ \vdots \\ \frac{126}{800} - \frac{1}{6} \end{pmatrix} \qquad \hat{P}^{-1} = \begin{bmatrix} \frac{800}{135} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \frac{800}{126} \end{bmatrix} \qquad \hat{p}_6 = \frac{150}{800}$$

Table 4.5: Rankings of three cars

Order	123	132	213	231	312	321	Total
Observed Frequency	135	98	152	139	126	150	800
Expected Frequency (A)	141.0	100.9	158.1	135.2	123.8	141.0	800
Expected Frequency (E)	138.4	101.8	159.8	134.1	123.5	142.4	800

and we can calculate $V_{11}^{-1} = 800(\hat{P}^{-1} + \mathbf{1}\mathbf{1}'/\hat{p}_6)$ as

$$\begin{bmatrix} 9007 & 4267 & 4267 & 4267 & 4267 \\ & 10797 & 4267 & 4267 & 4267 \\ & & 8477 & 4267 & 4267 \\ & & & 8871 & 4267 \\ & & & & 9346 \end{bmatrix}$$

We calculate the initial estimate of β as

$$\hat{\beta} = (X_1'V_{11}^{-1}X_1)^{-1}X_1'V_{11}^{-1}Z_1 \doteq \begin{pmatrix} -0.067 \\ 0.115 \\ 0.186 \\ -0.032 \end{pmatrix}$$

By calculating $P = \mathbf{1}/6 + X\beta$ and then recalculating V_{11}^{-1} and iterating we find that in this case the values given above are correct to 4 decimal places. The expected frequencies obtained from this model are almost exactly the same as the observed frequencies, indicating a model which fits very well indeed. The correlation matrix calculated for these 4 estimates is:

$$\begin{bmatrix} 0.0036 & 0.5472 & 0.0808 & 0.0585 \\ & 0.0035 & 0.0046 & 0.0703 \\ & & 0.0104 & 0.5192 \\ & & & 0.0116 \end{bmatrix}$$

An analysis of the full model is given in Table 4.6.

It is reasonable to deduce from Table 4.6 that the parameters μ_1 and V_{13} are non significant. The estimates of the remaining parameters are calculated by altering

Table 4.6: Analysis of Full Model

Parameter	Estimate	Z^2
μ_1	-0.067	1.23
μ_2	0.115	3.74
V_{12}	0.186	3.31
V_{13}	-0.032	0.09

the design matrix X_1 in the appropriate way, that is by removing the first and last columns. The estimates obtained are $\widehat{\mu_2} = 0.152$ and $\widehat{V_{12}} = 0.208$. The expected frequencies of this model are shown in Table 4.5 (labelled A). The corresponding χ^2 value is 1.30 on 3 degrees of freedom. The correlation matrix obtained for the two estimates is

$$\begin{bmatrix} 0.0025 & -0.0590 \\ & 0.0076 \end{bmatrix}$$

The corresponding Z^2 statistics are 9.31 and 5.72 respectively, indicating no further reduction in the model is appropriate.

The interpretation of this model is reasonably straightforward. The positive value of μ_2 indicates again the preference for the current make of car. The positive value of V_{12} indicates an association in the rankings of cars 1 and 2.

Using numerical maximisation of the likelihood function as explained previously, we can calculate the estimates of the exact normal model. For the 4 parameter model the estimates obtained are

$$\widehat{\mu_1} = -0.061 \quad \widehat{\mu_2} = 0.109 \quad \widehat{V_{12}} = 0.191 \quad \widehat{V_{13}} = -0.033$$

We note the good agreement between these estimates and those obtained from the approximation. The estimates obtained by setting $\mu_1 = 0$ and $V_{13} = 0$ are $\widehat{\mu_2} = 0.143$ and $\widehat{V_{12}} = 0.207$. The expected frequencies corresponding to this model are given in Table 4.5 (labelled E). The χ^2 value is 1.25 on 3 d.f. Again we note the very close agreement with the approximate model.

For purposes of comparison with the logistic models of chapter 5, we can find the standardised bivariate model using relationships given earlier in this chapter as:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0.000 \\ 0.101 \end{pmatrix}, \begin{pmatrix} 1.000 & 0.604 \\ & 1.000 \end{pmatrix}\right]$$

Chapter 5

Alternative Distributions

5.1 Introduction

The normal model presented in chapter 4 is clearly a very flexible model for the analysis of permutations. It has the limitation of being a second order model, but this is not too great a restriction in the practical application of these types of models. The very general range of permissible values for the correlation matrix certainly is a feature of this model. When alternative distributions to the normal are sought, they are usually found to be quite restrictive in this area. The multinomial logit model based on the logistic distribution, mentioned in the review of transportation models, is an example of this.

The reason for searching for alternative distributions to base our model upon is also quite clear. The evaluation of normal probabilities is a time consuming task. For problems with large amounts of data and particularly cases involving covariate studies, the calculating effort can be prohibitive. The basis for this chapter of work is to investigate alternative distributions which provide explicit expressions for the permutation probabilities. The cost associated with this benefit is invariably the problem of restrictive correlation structures mentioned previously.

As is historically the case, we are particularly interested in distributions which are 'like' the normal distribution. The logistic is one such distribution that has been used extensively for this purpose, for example the probit and logit methods of biological assay. We saw in chapter one that the Bradley-Terry model arises from an assumption of logistically distributed differences. It was also shown that we could equivalently assume independent extreme value distributions for the original variables and arrive at the same expressions for the choice probabilities. It is also a consequence of the material presented in chapter two that the exponential model is linked to the extreme value distribution by a monotonic transformation and so is a model of some interest with regards to generalisation.

We have as a result three immediate choices for extension from the case of independent random variables to a model based on dependent variables. We will firstly consider a generalised logistic distribution due to Satherwaite and Hutchinson[34]. The second model is based on the generalised extreme value distribution introduced by McFadden[27] and we will conclude by looking at models derived from a multivariate exponential distribution due to Gumbel[15].

5.2 Bivariate logistic distributions

5.2.1 The three parameter model

Gumbel[16] introduced a bivariate logistic distribution which had a distribution function of the form

$$F(y_1, y_2) = \frac{1}{1 + e^{-y_1} + e^{-y_2}}$$

A summary of the details of this distribution are given in section 5 of chapter 42 of Johnson and Kotz[21]. In particular, we note that the correlation between Y_1 and Y_2 is fixed at $1/2$. From the previous chapter we remember that a correlation of $1/2$ between the difference variables is equivalent to independence of the original variables X_1, X_2, X_3. We can in fact show that if X_1, X_2, X_3 are independent extreme value variables with distribution function given by

$$F(x) = \exp(-e^{-x}) \qquad -\infty < x < \infty$$

then the joint distribution function of

$$Y_1 = X_1 - X_3 \qquad Y_2 = X_2 - X_3$$

is the one given by Gumbel's formulation.

In their paper of 1978, Satherwaite and Hutchinson[34] presented a generalisation of Gumbels distribution by raising his expression to an arbitrary power to give

$$F(y_1, y_2) = \frac{1}{(1 + e^{-y_1} + e^{-y_2})^{\nu}} \qquad \nu > 0$$

The distribution is not unlike the bivariate normal distribution. It is symmetric about $y_1 = y_2$ but not about $y_1 = -y_2$. The density function is

$$f(y_1, y_2) = \frac{\nu(\nu+1)e^{-y_1}e^{-y_2}}{(1 + e^{-y_1} + e^{-y_2})^{\nu+2}}$$

The marginal distributions are given by

$$F(y_1) = \frac{1}{(1 + e^{-y_1})^{\nu}}$$

with corresponding density

$$f(y_1) = \frac{\nu e^{-y_1}}{(1 + e^{-y_1})^{\nu+1}}$$

The moments of the distribution are given by

$$\mathrm{E}(Y_i) = \psi(\nu) + \mathrm{C}$$

$$\mathrm{Var}(Y_i) = \frac{\pi^2}{6} + \varsigma(2, \nu)$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = \frac{\varsigma(2,\nu)}{\varsigma(2,\nu) + \frac{\pi^2}{6}}$$

where

$$\psi(z) = \Gamma'(z)/\Gamma(z)$$

is the digamma function and

$$\varsigma(s, a) = \sum_{m=0}^{\infty} (m + a)^{-s}$$

is the generalised Riemann zeta function used here with $s = 2$; C is Eulers constant (=0.577...). We note that the correlation ρ tends to 1 as ν tends to 0, is 1/2 when ν =1 and tends to 0 as ν tends to ∞.

From our previous discussion of the bivariate normal distribution we know that a bivariate model on the differences

$$Y_1 = X_1 - X_3 \qquad Y_2 = X_2 - X_3$$

with equal variances and a correlation parameter ρ is a model of association in the rankings of elements 1 and 2. We can calculate probabilities for this three parameter model in the following way. Assume that (Y_1, Y_2) have a bivariate Satherwaite distribution with distribution function

$$\mathrm{F}(y_1, y_2) = \frac{1}{(1 + e^{-(y_1-v_1)} + e^{-(y_2-v_2)})^{\nu}}$$

We can for example calculate

$$\begin{aligned}
p_{123} &= \mathrm{P}(Y_1 > Y_2,\ Y_2 > 0) \\
&= \int_0^{\infty} \left[\frac{\partial \mathrm{F}}{\partial y_2}\right]_{y_1=y_2}^{y_1=\infty} dy_2 \\
&= \int_0^{\infty} \left[\frac{\nu e^{-(y_2-v_2)}}{(1 + e^{-(y_1-v_1)} + e^{-(y_2-v_2)})^{\nu+1}}\right]_{y_1=y_2}^{y_1=\infty} dy_2 \\
&= \int_0^{\infty} \left[\frac{\nu e^{-(y_2-v_2)}}{(1 + e^{-(y_2-v_2)})^{\nu+1}} - \frac{\nu e^{-(y_2-v_2)}}{(1 + e^{-y_2}(e^{v_1} + e^{v_2}))^{\nu+1}}\right] dy_2 \\
&= \left[\frac{1}{(1 + e^{-(y_2-v_2)})^{\nu}}\right]_{y_2=0}^{y_2=\infty} - \frac{e^{v_2}}{e^{v_1} + e^{v_2}} \left[\frac{1}{(1 + e^{-y_2}(e^{v_1} + e^{v_2}))^{\nu}}\right]_{y_2=0}^{y_2=\infty} \\
&= 1 - \frac{1}{(1 + e^{v_2})^{\nu}} - \frac{e^{v_2}}{e^{v_1} + e^{v_2}} \left(1 - \frac{1}{(1 + e^{v_1} + e^{v_2})^{\nu}}\right)
\end{aligned}$$

In a similar way we can derive the other probabilities so that we have the following expressions:

$$p_{132} = \frac{1}{(1 + e^{v_2})^{\nu}} - \frac{1}{(1 + e^{v_1} + e^{v_2})^{\nu}}$$

$$
\begin{aligned}
p_{213} &= 1 - \frac{1}{(1+e^{v_1})^{\nu}} - \frac{e^{v_1}}{e^{v_1}+e^{v_2}}\left(1 - \frac{1}{(1+e^{v_1}+e^{v_2})^{\nu}}\right) \\
p_{231} &= \frac{1}{(1+e^{v_1})^{\nu}} - \frac{1}{(1+e^{v_1}+e^{v_2})^{\nu}} \\
p_{312} &= \frac{e^{v_1}}{(e^{v_1}+e^{v_2})(1+e^{v_1}+e^{v_2})^{\nu}} \\
p_{321} &= \frac{e^{v_2}}{(e^{v_1}+e^{v_2})(1+e^{v_1}+e^{v_2})^{\nu}}
\end{aligned}
$$

We observe that the model of association between objects 1 and 2 is symmetric in these elements and so for example having calculated p_{123} we can obtain the expression for p_{213} by interchanging v_1 and v_2 in p_{123}.

The benefits associated with a model which gives explicit expressions for the probabilities are reasonably clear. Numerical maximisation of the likelihood function consumes less time, not only because of the time saved in calculating the individual probabilities, but also because it is possible to calculate analytic derivatives of the likelihood function. This means that rather than using search techniques to maximise the likelihood function it is possible to write down the analytic form of the likelihood equations and make use of numerical techniques for solving non-linear systems of equations. These methods are much faster in general than the simplex type of searches used up until this point.

As an example consider the likelihood function

$$\mathcal{L} = \prod p_{ijk}^{f_{ijk}}$$

and the corresponding loglikelihood function

$$\mathrm{L} = \log \mathcal{L} = \sum f_{ijk} \log p_{ijk}$$

For simplicity reparameterise so that

$$u_1 = e^{v_1} \qquad u_2 = e^{v_2}$$

We have then for example

$$p_{123} = 1 - \frac{1}{(1+u_2)^\nu} - \frac{u_2}{u_1+u_2}\left(1 - \frac{1}{(1+u_1+u_2)^\nu}\right)$$

To find the derivative of the log likelihood function we require the derivatives of this expression with respect to u_1, u_2 and ν. We can show that these are:

$$\begin{aligned}
\frac{\partial p_{123}}{\partial u_1} &= \frac{u_2}{u_1+u_2}\left[\frac{1}{u_1+u_2}\left(1 - \frac{1}{(1+u_1+u_2)^\nu}\right) - \frac{\nu}{(1+u_1+u_2)^{\nu+1}}\right] \\
\frac{\partial p_{123}}{\partial u_2} &= \frac{\nu}{(1+u_2)^{1+\nu}} - \frac{u_1}{(u_1+u_2)^2}\left(1 - \frac{1}{(1+u_1+u_2)^\nu}\right) \\
&\quad - \frac{u_2\nu}{(u_1+u_2)(1+u_1+u_2)^{\nu+1}} \\
\frac{\partial p_{123}}{\partial \nu} &= \frac{\log(1+u_2)}{(1+u_2)^\nu} - \frac{u_2\log(1+u_1+u_2)}{(u_1+u_2)(1+u_1+u_2)^\nu}
\end{aligned}$$

We can make use of the symmetry in the argument between objects 1 and 2. Since p_{123} and p_{231} are symmetric in u_1 and u_2 we note that

$$\begin{aligned}
\frac{\partial p_{213}}{\partial u_1} &= \left.\frac{\partial p_{123}}{\partial u_2}\right|_{u_1=u_2, u_2=u_1} \\
\frac{\partial p_{213}}{\partial u_2} &= \left.\frac{\partial p_{123}}{\partial u_1}\right|_{u_1=u_2, u_2=u_1} \\
\frac{\partial p_{213}}{\partial \nu} &= \left.\frac{\partial p_{123}}{\partial \nu}\right|_{u_1=u_2, u_2=u_1}
\end{aligned}$$

In addition, to simplify calculation, we use the relationship

$$p_{123} + p_{213} + p_{132} = 1 - \frac{1}{(1+u_1)^\nu}$$

which gives

$$\begin{aligned}
\frac{\partial p_{132}}{\partial u_1} &= \frac{\nu}{(1+u_1)^{1+\nu}} - \frac{\partial p_{123}}{\partial u_1} - \frac{\partial p_{213}}{\partial u_1} \\
\frac{\partial p_{132}}{\partial u_2} &= -\frac{\partial p_{123}}{\partial u_2} - \frac{\partial p_{213}}{\partial u_2} \\
\frac{\partial p_{132}}{\partial \nu} &= \frac{\log(1+u_1)}{(1+u_1)^{\nu}} - \frac{\partial p_{123}}{\partial \nu} - \frac{\partial p_{213}}{\partial \nu}
\end{aligned}$$

The remaining derivatives follow in a similar manner. We make use of the symmetry of p_{132} and p_{231} to write

$$\begin{aligned}
\frac{\partial p_{231}}{\partial u_1} &= \left.\frac{\partial p_{132}}{\partial u_2}\right|_{u_1=u_2,u_2=u_1} \\
\frac{\partial p_{231}}{\partial u_2} &= \left.\frac{\partial p_{132}}{\partial u_1}\right|_{u_1=u_2,u_2=u_1} \\
\frac{\partial p_{231}}{\partial \nu} &= \left.\frac{\partial p_{132}}{\partial \nu}\right|_{u_1=u_2,u_2=u_1}
\end{aligned}$$

Using the given expression for p_{321} we have

$$\begin{aligned}
\frac{\partial p_{321}}{\partial u_1} &= -\frac{u_2}{(u_1+u_2)(1+u_1+u_2)^{\nu}}\left[\frac{1}{u_1+u_2} + \frac{\nu}{(1+u_1+u_2)}\right] \\
\frac{\partial p_{321}}{\partial u_2} &= \frac{1}{(u_1+u_2)(1+u_1+u_2)^{\nu}}\left[\frac{u_1}{u_1+u_2} - \frac{\nu u_2}{(1+u_1+u_2)}\right] \\
\frac{\partial p_{321}}{\partial \nu} &= \frac{-u_2\log(1+u_1+u_2)}{(u_1+u_2)(1+u_1+u_2)^{\nu}}
\end{aligned}$$

from which the derivatives of p_{312} follow by symmetry as

$$\begin{aligned}
\frac{\partial p_{312}}{\partial u_1} &= \left.\frac{\partial p_{321}}{\partial u_2}\right|_{u_1=u_2,u_2=u_1} \\
\frac{\partial p_{312}}{\partial u_2} &= \left.\frac{\partial p_{321}}{\partial u_1}\right|_{u_1=u_2,u_2=u_1} \\
\frac{\partial p_{312}}{\partial \nu} &= \left.\frac{\partial p_{321}}{\partial \nu}\right|_{u_1=u_2,u_2=u_1}
\end{aligned}$$

Using these derivatives we can construct the likelihood equations as

$$\begin{aligned}
\sum \frac{f_{ijk}}{p_{ijk}} \frac{\partial p_{ijk}}{\partial u_1} &= 0 \\
\sum \frac{f_{ijk}}{p_{ijk}} \frac{\partial p_{ijk}}{\partial u_2} &= 0 \\
\sum \frac{f_{ijk}}{p_{ijk}} \frac{\partial p_{ijk}}{\partial \nu} &= 0
\end{aligned}$$

There are a number of routines available for solving this system of non-linear equations, such as the IMSL routine ZSPOW. We can obtain reasonable starting values for these routines from the Taylor series approximation and the least squares fitting procedure. These results will be presented when we consider the 4 parameter model shortly.

The easiest method for interpreting the parameters of the model is to standardise to unit variances. To achieve this we note

$$\mathrm{E}(Y_i) = \psi(\nu) + \mathrm{C} + v_i \qquad i = 1, 2$$

$$\mathrm{Var}(Y_i) = \frac{\pi^2}{6} + \varsigma(2, \nu) \qquad i = 1, 2$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = \frac{\varsigma(2, \nu)}{\varsigma(2, \nu) + \frac{\pi^2}{6}}$$

The standardised model is achieved by the transformation $Z_i = Y_i/\sigma_2$ where σ_2 is the standard deviation of Y_2. The standardised model values are given by

$$\mu_i = \frac{\psi(\nu) + \mathrm{C} + v_i}{\sqrt{\frac{\pi^2}{6} + \varsigma(2, \nu)}} \qquad i = 1, 2$$

$$\sigma_i^2 = 1 \qquad i = 1, 2$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = \frac{\varsigma(2, \nu)}{\varsigma(2, \nu) + \frac{\pi^2}{6}}$$

$\psi(z)$, the digamma function is a standard IMSL function and a convenient method for calculating $\varsigma(2,\nu)$, the generalised zeta function is given in an appendix to the paper of Satherwaite and Hutchinson[34]. An example of the use of this model is given at the end of the next section.

5.2.2 The four parameter model

By introducing a variance parameter to extend this model to 4 parameters we arrive at a model which is quite similar to the bivariate normal specification. The introduction of this additional parameter unfortunately means that we are no longer able to write explicit expressions for each of the probabilities, but the numerical work involved is much less than that required for the normal model.

Suppose that the joint distribution of Y_1 and Y_2 has distribution function

$$F(y_1,y_2) = \frac{1}{(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_2-v_2)})^{\nu}}$$

so that

$$\begin{aligned} \mathrm{Var}(Y_1) &= \sigma_1^2\left(\frac{\pi^2}{6}+\varsigma(2,\nu)\right) & \mathrm{Var}(Y_2) &= \frac{\pi^2}{6}+\varsigma(2,\nu) \\ E(Y_1) &= \sigma_1(\psi(\nu)+C)+v_1 & E(Y_2) &= \psi(\nu)+C+v_2 \end{aligned}$$

$$\rho = \mathrm{Corr}(Y_1,Y_2) = \frac{\varsigma(2,\nu)}{\varsigma(2,\nu)+\frac{\pi^2}{6}}$$

Calculation of the permutation probabilities associated with this model proceeds in the following way.

$$\begin{aligned} p_{123} &= \int_0^\infty \left[\frac{\partial F}{\partial y_1}\right]_{y_2=0}^{y_2=y_1} dy_1 \\ &= \int_0^\infty \left(\frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}} - \frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}e^{v_2})^{1+\nu}}\right) dy_1 \end{aligned}$$

$$= \int_0^\infty \frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}}dy_1 - \left[\frac{1}{(1+e^{-(y_1-v_1)/\sigma_1}+e^{v_2})^\nu}\right]_{y_1=0}^{y_1=\infty}$$

$$= \int_0^\infty \frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}}dy_1 + \frac{1}{(1+e^{v_1/\sigma_1}+e^{v_2})^\nu} - \frac{1}{(1+e^{v_2})^\nu}$$

The first term in this expression requires numerical evaluation, the details of which we shall consider shortly. Continuing with the expressions for the permutation probabilities we have:

$$\begin{aligned}
p_{132} &= \int_0^\infty \left[\frac{\partial F}{\partial y_1}\right]_{y_2=-\infty}^{y_2=0} dy_1 \\
&= \frac{1}{(1+e^{v_2})^\nu} - \frac{1}{(1+e^{v_1/\sigma_1}+e^{v_2})^\nu} \\
p_{213} &= \int_0^\infty \left[\frac{\partial F}{\partial y_1}\right]_{y_2=y_1}^{y_2=\infty} dy_1 \\
&= 1 - \frac{1}{(1+e^{v_1/\sigma_1})^\nu} - \int_0^\infty \frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}}dy_1
\end{aligned}$$

We note that the last term in this expression is the same as the one to be numerically evaluated in p_{123}.

$$\begin{aligned}
p_{231} &= \int_{-\infty}^0 \left[\frac{\partial F}{\partial y_1}\right]_{y_2=0}^{y_2=\infty} dy_1 \\
&= \frac{1}{(1+e^{v_1/\sigma_1})^\nu} - \frac{1}{(1+e^{v_1/\sigma_1}+e^{v_2})^\nu} \\
p_{312} &= \int_{-\infty}^0 \left[\frac{\partial F}{\partial y_1}\right]_{y_2=-\infty}^{y_2=y_1} dy_1 \\
&= \int_{-\infty}^0 \frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}}dy_1 \\
p_{321} &= \int_{-\infty}^0 \left[\frac{\partial F}{\partial y_1}\right]_{y_2=y_1}^{y_2=\infty} dy_1 \\
&= \frac{1}{(1+e^{v_1/\sigma_1}+e^{v_2})^\nu} - \int_{-\infty}^0 \frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}}dy_1
\end{aligned}$$

Once again the last term in this expression is the same integral as the one in the previous expression. As a result of these calculations we need only evaluate two integrals numerically i.e. the integrals over the intervals $(-\infty, 0)$ and $(0, \infty)$ of the integrand:

$$\frac{\nu e^{-(y_1-v_1)/\sigma_1}}{\sigma_1(1+e^{-(y_1-v_1)/\sigma_1}+e^{-(y_1-v_2)})^{1+\nu}}\,dy_1$$

A reasonably simple method of approximating these integrals exist. We firstly transform the integrand using the substitution $z = (y_1 - v_1)/\sigma_1$. Consider for example the second integral. After the substitution this integral becomes

$$\int_{-v_1/\sigma_1}^{\infty} \frac{\nu e^{-z}}{(1+e^{-z}+e^{-(\sigma_1 z+v_1-v_2)})^{\nu+1}}\,dz$$

We can bound the integrand by the function

$$h(z) = \frac{\nu e^{-z}}{(1+e^{-z})^{1+\nu}}$$

For positive z this function is a reasonably good approximation for $\sigma_1 > 1$ For negative z it is best when σ_1 is small. These affects are amplified when ν is not equal to 1.

We can evaluate the integral numerically to an accuracy of less than a given value ϵ (ignoring roundoff effects) by choosing a value k such that

$$\int_{k}^{\infty} \frac{\nu e^{-z}}{(1+e^{-z}+e^{-(\sigma_1 z+v_1-v_2)})^{\nu+1}}\,dz < \epsilon$$

Using the function bound we have

$$\int_{k}^{\infty} \frac{\nu e^{-z}}{(1+e^{-z})^{1+\nu}}\,dz < \epsilon$$

We can show that this leads to the equation

$$1 - \frac{1}{(1+e^{-k})^{\nu}} < \epsilon$$

Solving this equation gives

$$k > -\ln\left(\frac{1}{(1-\epsilon)^{1/\nu}} - 1\right)$$

For $\epsilon = 0.00001$ the values of k involved in this approximation range from approximately 10 for $\nu = 0.2$ to 14 for $\nu = 10$. The resultant definite integral can easily be evaluated using any of the standard techniques.

For the other integral i.e.

$$\int_{-\infty}^{-v_1/\sigma_1} \frac{\nu e^{-z}}{(1 + e^{-z} + e^{-(\sigma_1 z + v_1 - v_2)})^{\nu+1}}\, dz$$

we find k such that

$$\int_{-\infty}^{k} \frac{\nu e^{-z}}{(1 + e^{-z})^{1+\nu}}\, dz < \epsilon$$

Once again using the bounding function we require

$$\frac{1}{(1 + e^{-k})^{\nu}} < \epsilon$$

Solving this equation gives

$$k < -\ln\left(\frac{1}{\epsilon^{1/\nu}} - 1\right)$$

For $\epsilon = 0.00001$ the values of k involved in this approximation range from approximately -55 for $\nu = 0.2$ to -1 for $\nu = 10$.

Using the standardisation procedure outlined previously we obtain the following standard values:

$$E(Y_1) = \frac{\sigma_1(\psi(\nu) + C) + v_1}{\sqrt{\frac{\pi^2}{6} + \varsigma(2,\nu)}} \qquad E(Y_2) = \frac{\psi(\nu) + C + v_2}{\sqrt{\frac{\pi^2}{6} + \varsigma(2,\nu)}}$$

$$\mathrm{Var}(Y_1) = \sigma_1^2 \qquad \mathrm{Var}(Y_2) = 1$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = \frac{\varsigma(2,\nu)}{\varsigma(2,\nu) + \frac{\pi^2}{6}}$$

We calculate the series approximation for this model about the values which produce the uniform probability model. In the original parameterisation these values were $v_1 = v_2 = 0$ and $\nu = \sigma_1 = 1$. We will continue to use $u_1 = e^{v_1}$ and $u_2 = e^{v_2}$ We can find the Taylor series expansion in the parameters $\gamma = (u_1, u_2, \nu, \sigma_1)'$ as

$$\mathbf{p} = \frac{1}{6}\mathbf{1} + X(\gamma - \mathbf{1})$$

where the matrix X is given by

$$\begin{bmatrix} 1/9 & 1/36 & (\log 2 - \log(3)/3)/2 & \log(3)/4 - 1/6 \\ 1/9 & -5/36 & -\log(2)/2 + \log(3)/3 & 0 \\ 1/36 & 1/9 & (\log 2 - \log(3)/3)/2 & -\log(3)/4 + 1/6 \\ -5/36 & 1/9 & -\log(2)/2 + \log(3)/3 & 0 \\ 1/36 & -5/36 & -\log(3)/6 & -\log(1.5)/4 - 1/12 \\ -5/36 & 1/36 & -\log(3)/6 & \log(1.5)/4 + 1/12 \end{bmatrix}$$

The first three columns of this matrix are obtained by evaluating expressions for terms such as $\frac{\partial p_{123}}{\partial u_1}$ at $\gamma = \mathbf{1}$. We found these expressions when considering the 3 parameter model. To calculate the last column we need to evaluate expressions such as $\frac{\partial p_{123}}{\partial \sigma_1}$. We see that this requires us to evaluate

$$\frac{\partial}{\partial \sigma_1} \int_0^\infty \frac{\nu e^{-(y_1 - v_1)/\sigma_1}}{\sigma_1(1 + e^{-(y_1-v_1)/\sigma_1} + e^{-(y_1 - v_2)})^{1+\nu}} dy_1$$

Differentiating and setting $\gamma = \mathbf{1}$ we have

$$\int_0^\infty \frac{e^{-y}(y - 1 - 2e^{-y})}{(1 + 2e^{-y})^3} dy$$

We can evaluate this integral by showing that

$$\int_0^\infty \frac{ye^{-y}}{(1+2e^{-y})^3} dy \quad = \quad \int_0^\infty \frac{ye^{2y}}{(2+e^y)^3} dy \quad = \quad \int_1^\infty \frac{u \log u}{(u+2)^3} du$$

Using integration by parts we can show that this integral is

$$\int_1^\infty \frac{u+1}{u(u+2)^2}\,du$$

which we evaluate as $\log(3)/4 + 1/6$. For the other parts of the integral we obtain

$$\int_0^\infty \frac{e^{-y}}{(1+2e^{-y})^3}\,dy = \frac{8}{36} \qquad \int_0^\infty \frac{2e^{-2y}}{(1+2e^{-y})^3}\,dy = \frac{1}{9}$$

The other terms are calculated in a similar manner.

The interpretation of the parameters is not quite as straightforward as was the case in the normal model. The main reason for this is that the association parameters ν and σ_1 also appear as part of the mean structure. In general it is the case that ν represents association between objects 1 and 2 and σ_1 represents association between objects 2 and 3. The easiest way of interpreting the model is to consider the standardised model.

$$E(Y_1) = \frac{\sigma_1(\psi(\nu)+C)+v_1}{\sqrt{\frac{\pi^2}{6}+\varsigma(2,\nu)}} \qquad E(Y_2) = \frac{\psi(\nu)+C+v_2}{\sqrt{\frac{\pi^2}{6}+\varsigma(2,\nu)}}$$

$$\mathrm{Var}(Y_1) = \sigma_1^2 \qquad \mathrm{Var}(Y_2) = 1$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = \frac{\varsigma(2,\nu)}{\varsigma(2,\nu)+\frac{\pi^2}{6}}$$

We recall that a model of association betwee objects 1 and 2 is obtained by specifying

$$\mathrm{Var}(Y_1) = 1 \quad \to \quad \sigma_1 = 1 \quad \mathrm{cov}(Y_1, Y_2) > \frac{1}{2} \quad \to \quad \nu < 1$$

As a preliminary to the example in Chapter 6, consider a model with association between objects 1 and 3. In our previous discussion we saw that this model is given by

$$\mathrm{Var}(Y_1) = \delta \quad \mathrm{cov}(Y_1, Y_2) = \frac{\delta}{2}$$

It is clear that we can achieve this parameterisation by setting

$$\sigma_1^2 = \delta \quad \frac{\varsigma(2,\nu)}{\varsigma(2,\nu)+\frac{\pi^2}{6}} = \frac{\sigma_1}{2}$$

To estimate the parameter σ_1 we would need to solve the equation for ν using any one of the standard numerical techniques. We will consider an example involving this parameterisation in Chapter 6.

5.2.3 Example calculations

As before to illustrate the calculations involved, consider the data used in the example of Chapter 3. For convenience the observed frequencies of the 6 permutations are given in the first row of Table 5.1. In order to apply the least squares

Table 5.1: Rankings of three cars

Order	123	132	213	231	312	321	Total
Observed Frequency	135	98	152	139	126	150	800
Expected Frequency	143.6	108.0	157.1	138.0	116.6	136.7	800

procedure to fit the approximate model we calculate X_1, V_{11}, Z_1 as demonstrated in previous examples and calculate an estimate of $\beta = \gamma - 1$ where $\gamma = (u_1, u_2, \nu, \sigma_1)$ as

$$\hat{\beta} = (X_1' V_{11}^{-1} X_1)^{-1} X_1' V_{11}^{-1} Z_1 = \begin{pmatrix} 0.347 \\ 0.554 \\ -0.305 \\ -0.014 \end{pmatrix}$$

The corresponding estimates of the original parameters are

$$v_1 = 0.30 \quad v_2 = 0.44 \quad \nu = 0.70 \quad \sigma_1 = 0.99$$

The expected frequencies obtained from this model are almost exactly the same as the observed frequencies, indicating a model which fits very well indeed. The correlation matrix calculated for these 4 estimates is:

$$\begin{bmatrix} 0.0397 & 0.9429 & -0.9408 & 0.0958 \\ & 0.0368 & -0.9383 & -0.0108 \\ & & 0.0198 & -0.0645 \\ & & & 0.0026 \end{bmatrix}$$

An analysis of the full model is given in Table 5.2.

Table 5.2: Analysis of Full Model

Parameter	Estimate	Z^2
$u_1 - 1$	0.347	3.02
$u_2 - 1$	0.554	8.33
$\nu - 1$	0.186	4.70
$\sigma_1 - 1$	0.032	0.07

It is reasonable to deduce from Table 5.2 that the parameter σ_1 is not significantly different from 1. For the sake of comparison we can also set

$$v_1 = -c - \psi(\nu)$$

this has the effect of producing a standardised mean μ_1 of zero, in line with previous models.

Using the numerical maximisation routine, we find that the estimates obtained are $\widehat{u_2} = 1.543$ and $\hat{\nu} = 0.851$. The expected frequencies of the model obtained using these parameters are shown in Table 5.1. This model gives a χ^2 value of 3.67 on 3 degrees of freedom.

We can find the standardised bivariate model using relationships given earlier in this chapter as:

$$E(Y_1) = 0 \qquad E(Y_2) = 0.082 \qquad \mathrm{Var}(Y_1) = 1 \qquad \mathrm{Var}(Y_2) = 1$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = 0.560$$

We note the close agreement between this standardised form and the one given for the normal distribuiton which was

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0.000 \\ 0.101 \end{pmatrix}, \begin{pmatrix} 1.000 & 0.604 \\ & 1.000 \end{pmatrix}\right]$$

The interpretation of this model is very similar to that of the normal model. The positive value of μ_2 indicates again the preference for the current make of car. The value of $\nu < 1$ as we saw previously indicates a positive association in the rankings of cars 1 and 2.

5.3 Extreme value distributions

5.3.1 The three parameter model

In an attempt to generalise the multinomial logit model, McFadden[27] introduced a family of distributions which he called the Generalised Extreme Value distributions (GEV). These are multivariate extreme value distributions which have marginal distribution functions of the form

$$F(x_i) = \exp -(e^{-x_i}) \qquad -\infty < x_i < \infty$$

We have already noted that the difference distribution of extreme value variables is the logistic and so it is possible to express these models in the logistic form.

Consider firstly the three parameter form of the GEV distribution of McFadden. The distribution function is given as

$$F(x_1, x_2, x_3) = \exp - \left(e^{-(x_3 - v_3)} + \left(e^{-(x_1 - v_1)/(1-p)} + e^{-(x_2 - v_2)/(1-p)}\right)^{1-p}\right)$$

where $0 < p < 1$ and the range of x_i is $(-\infty, \infty)$, $i = 1, 2, 3$. An equivalent formulation and a discussion of this distribution can also be found on page 254 of Johnson and Kotz[21] . It is known that in this form, $X_1 - X_2$ has a logistic distribution, from which it can be deduced that

$$\mathrm{Corr}(X_1, X_2) = 1 - (1-p)^2$$

We note that this model only incorporates positively dependent variables. This is quite a restriction in it's application.

The other moments of the distribution follow from the properties of the marginal distribution

$$F(x_i) = \exp -(e^{-(x_i - v_i)})$$

For $i = 1, 2, 3$ we therefore have

$$E(X_i) = v_i + C \qquad \mathrm{Var}(X_i) = \frac{\pi^2}{6}$$

where C is Eulers constant. We can show that the distribution of

$$Y_1 = X_1 - X_3 \qquad Y_2 = X_2 - X_3$$

has the following form:

$$F(y_1, y_2) = \frac{1}{1 + \left(e^{-(y_1 - \theta_1)/(1-p)} + e^{-(y_2 - \theta_2)/(1-p)}\right)^{1-p}}$$

where

$$\theta_1 = v_1 - v_3 \qquad \theta_2 = v_2 - v_3$$

The density function has the form

$$f(y_1, y_2) = \frac{e^{-y_1'} e^{-y_2'} d^{-p} \left[pd^{-1} + (2-p)d^{-p}\right]}{(1+d^{1-p})^3(1-p)}$$

where

$$y_i' = \frac{y_i - \theta i}{1-p} \qquad i = 1, 2$$

and

$$d = e^{-y_1'} + e^{-y_2'}$$

From the properties of the function x^p, $x > 0$ we can deduce that the density function is only positive for $0 \leq p < 1$ in the following way.

We compare the graphs of d^{-1} and d^{-p}. For $p < 0$ we see that for $d < 1$ we have $d^{-p} < 1$ and d^{-1} can be made arbitrarily large. We can therefore choose a value of d such that

$$pd^{-1} + (2-p)d^{-p} < 0$$

and hence the density is negative.

For $p > 1$ we note that the term $d^{-p}/(1-p)$ is negative. The term in the square brackets is positive for $1 < p \leq 2$ in which case the density is negative. For $p > 2$ and $d > 1$ we have

$$d^{-p} < d^{-1} \quad \rightarrow \quad pd^{-1} + (2-p)d^{-p} > 2d^{-p} > 0$$

and therefore the density is negative.

We can deduce the moments of this distribution from those of the extreme value distribution

$$E(Y_i) = \theta_i \qquad \mathrm{Var}(Y_i) = \frac{\pi^2}{3} \qquad i = 1,2$$

$$\mathrm{Corr}(Y_1, Y_2) = 1 - \frac{(1-p)^2}{2}$$

Since $0 \leq p < 1$ the possible correlation range for the difference distribution is (1/2,1). This means that unlike the normal distribution, there is no extension of the correlation structure from the use of the difference distribution.

We can find the permutation probabilities for this model in the following way.

$$\begin{aligned}
p_{123} &= \Pr(Y_1 > Y_2, Y_2 > 0) \\
&= \int_0^\infty \left[\frac{\partial F}{\partial y_2}\right]_{y_1=y_2}^{y_1=\infty} dy_2 \\
&= \int_0^\infty e^{-\frac{(y_2-\theta_2)}{1-p}} \left(\frac{e^{\frac{p(y_2-\theta_2)}{1-p}}}{(1+e^{-(y_2-\theta_2)})^2} - \frac{\left(ae^{-\frac{y_2}{1-p}}\right)^{-p}}{(1+(ae^{-\frac{y_2}{1-p}})^{1-p})^2} \right) dy_2 \\
&= \left[\frac{1}{1+e^{-(y_2-\theta_2)}}\right]_0^\infty - \left[\frac{e^{\frac{\theta_2}{1-p}}}{a(1+(ae^{\frac{-y_2}{1-p}})^{1-p})}\right]_0^\infty \\
&= \frac{e^{\theta_2}}{1+e^{\theta_2}} - \frac{e^{\frac{\theta_2}{1-p}}a^{-p}}{1+a^{1-p}}
\end{aligned}$$

where

$$a = e^{\frac{\theta_1}{1-p}} + e^{\frac{\theta_2}{1-p}}$$

In a similar way we can calculate the following probabilities:

$$\begin{aligned}
p_{132} &= \frac{1}{1+e^{\theta_2}} - \frac{1}{1+a^{1-p}} \\
p_{213} &= \frac{e^{\theta_1}}{1+e^{\theta_1}} - \frac{e^{\frac{\theta_1}{1-p}}a^{-p}}{1+a^{1-p}}
\end{aligned}$$

$$
\begin{aligned}
p_{213} &= \frac{1}{1+e^{\theta_1}} - \frac{1}{1+a^{1-p}} \\
p_{312} &= \frac{e^{\frac{\theta_1}{1-p}}}{a(1+a^{1-p})} \\
p_{321} &= \frac{e^{\frac{\theta_2}{1-p}}}{a(1+a^{1-p})}
\end{aligned}
$$

We note that we can use the fact the the distribution is symmetrical in y_1 and y_2 so that for example we can write down p_{213} from the expression given for p_{123} by interchanging θ_1 and θ_2.

The Taylor series approximation can be easily calculated following arguments presented previously. We write

$$
\mathbf{p} = \frac{1}{6}\mathbf{1} + X\gamma
$$

where $\gamma = (\theta_1, \theta_2, p)'$ and the matrix X is

$$
\begin{bmatrix}
\frac{4}{36} & \frac{1}{36} & \frac{\ln 2}{9} \\
\frac{4}{36} & \frac{-5}{36} & \frac{-2\ln 2}{9} \\
\frac{1}{36} & \frac{4}{36} & \frac{\ln 2}{9} \\
\frac{-5}{36} & \frac{4}{36} & \frac{-2\ln 2}{9} \\
\frac{1}{36} & \frac{-5}{36} & \frac{\ln 2}{9} \\
\frac{-5}{36} & \frac{1}{36} & \frac{\ln 2}{9}
\end{bmatrix}
$$

We note the similarity between this matrix and the one obtained from the normal distribution. The interpretation of the parameters of the model is also clear. θ_1 and θ_2 represent the preference for objects 1 and 2 relative to a zero preference for object 3. p is a measurement of association in the rankings of objects 1 and 2.

5.3.2 The four parameter model

To achieve a four parameter model in the $r = 3$ case we can generalise the distribution function in an obvious way to

$$F(x_1, x_2, x_3) = \exp - \left\{ \left(\left(e^{-x'_{12}} + e^{-x'_2}\right)^{1-p_{12}} + \left(e^{-x'_{13}} + e^{-x'_3}\right)^{1-p_{13}} \right) \right\}$$

$$x'_{1i} = \frac{x_1 - v_1}{1 - p_{1i}} \quad i = 2,3 \qquad x'_2 = \frac{x_2 - v_2}{1 - p_{12}} \quad x'_3 = \frac{x_3}{1 - p_{13}}$$

where we set $v_3 = 0$ to satisfy the constraint on the means.

Extending the distribution in this way has one major disadvantage. The 3 parameter model of the previous section can not be generated from this distribution by setting $p_{13} = 0$. We note that the marginal distribution obtained by setting $p_{13} = 0$ is not symmetric in y_1 and y_2, we observe this lack of symmetry in the probabilities and Taylor series expansion to follow. Despite these shortcomings the four parameter distribution is of some use in practice, since it does give explicit expressions for the probabilities and is a reasonably straightforward model to interpret.

The joint distribution of X_1 and X_i, $i = 1, 2$ is given by

$$F(x_1, x_i) = \exp \left\{ - \left(\left(e^{-(x_1 - v_1)/(1-p_{1i})} + e^{-(x_i - v_i)/(1-p_{1i})}\right)^{1-p_{1i}} + e^{-(x_1 - v_1)} \right) \right\}$$

We note that the marginal distributions are extreme value, in the case of X_1 the distribution function is given by

$$F(X_1) = \exp\left(-e^{-(X_1 - v_1 - \ln 2)}\right)$$

and so we have

$$\mathrm{E}(X_1) = v_1 + \ln 2 \quad \mathrm{E}(X_2) = v_2 \quad \mathrm{E}(X_3) = 0$$

$$\mathrm{Var}(X_i) = \frac{\pi^2}{6} \quad i = 1,2,3$$

As before we can find the joint distribution of the differences

$$Y_1 = X_1 - X_3 \quad Y_2 = X_2 - X_3$$

as

$$F(y_1, y_2) = \frac{(1 + e^{-(y_1-v_1)/(1-p_{13})})^{-p_{13}}}{(e^{-(y_1-v_1)/(1-p_{12})} + e^{-(y_2-v_2)/(1-p_{12})})^{1-p_{12}} + (1 + e^{-(y_1-v_1)/(1-p_{13})})^{1-p_{13}}}$$

where

$$0 < p_{12} < 1, \qquad 0 < p_{13} < 1$$

Looking at the marginal distributions we see firstly that for Y_2 we have the standard logistic distribution

$$F(y_2) = \frac{1}{1 + e^{-(y_2-v_2)}}$$

from which we note that

$$\mathrm{E}(Y_2) = v_2 \quad \mathrm{Var}(Y_2) = \frac{\pi^2}{3}$$

The marginal distribution of Y_1 takes the form

$$F(y_1) = \frac{\left(1 + e^{-(y_1-v_1)/(1-p_{13})}\right)^{-p_{13}}}{e^{-(y_1-v_1)} + (1 + e^{-(y_1-v_1)/(1-p_{13})})^{1-p_{13}}}$$

This distribution has some interesting features. Put

$$p = p_{13} \qquad Z = Y_1 - v_1$$

so that we have the standard form

$$F_p(z) = \frac{\left(1 + e^{-z/(1-p)}\right)^{-p}}{e^{-z} + (1 + e^{-z/(1-p)})^{1-p}}$$

It is convenient to write $F_p(z)$ in the equivalent form

$$F_p(z) = \frac{1}{1 + e^{-z/(1-p)} + e^{-z}\left(1 + e^{-z/(1-p)}\right)^p}$$

Consider firstly the limiting distribution as $p \to 1$. For $z < 0$ we see that the term

$$\left(1 + e^{-z/(1-p)}\right)$$

increases without bound as $p \to 1$ so that

$$\lim_{p \to 1} F_p(z) = 0 \qquad z < 0$$

For $z = 0$ we have

$$F_p(0) = \frac{1}{2 + 2^p}$$

which has a limit of $1/4$ as $p \to 1$ so that

$$\lim_{p \to 1} F_p(z) = \frac{1}{4} \qquad z = 0$$

For $z > 0$ we see that the term

$$\left(1 + e^{-z/(1-p)}\right)$$

tends to 1 as $p \to 1$ so that

$$\lim_{p \to 1} F_p(z) = \frac{1}{1 + e^{-z}} \qquad z > 0$$

The limit of this term as $z \to 0$ from above is $1/2$, we therefore have an interesting example of a limiting distribution which is discontinuous from both the left and right at 0.

The second point of interest involves the correlation of the original pair of variables X_1, X_3. Since we have

$$Y_1 = X_1 - X_3$$

we can write

$$\mathrm{Var}(Y_1) = \frac{\pi^2}{3}(1 - \mathrm{corr}(X_1, X_3))$$

Putting

$$Z_1 = Y_1 - v_1$$

so that

$$\mathrm{E}(Z_1) = \ln 2$$

we have

$$\mathrm{corr}(X_1, X_3) = 1 - \frac{3}{\pi^2}(\mathrm{E}(Z^2) - (\ln 2)^2)$$

where the distribution function of the variable Z is given as before by

$$F_p(z) = \frac{\left(1 + e^{-z/(1-p)}\right)^{-p}}{e^{-z} + \left(1 + e^{-z/(1-p)}\right)^{1-p}}$$

I have been unable to evaluate either $\mathrm{E}(Z^2)$, or equivalently $\mathrm{corr}(X_1, X_3)$, analytically. We can however calculate $\mathrm{E}(Z^2)$ numerically in the following way using the following known result obtained by integration by parts:

$$\mathrm{E}(Z^2) = -2\int_{-\infty}^{0} zF_p(z)\, dz + 2\int_{0}^{\infty} z(1 - F_p(z))\, dz$$

We can express the first of these integrals as

$$-2(1-p)^2 \int_{-\infty}^{0} \frac{z\left(1 + e^{-z}\right)^{-p}}{\left(1 + e^{-z}\right)^{1-p} + e^{-(1-p)z}}\, dz$$

To approximate this integral with error $< \epsilon$ we need to choose k s.t.

$$-2(1-p)^2 \int_{-\infty}^{k} \frac{z\left(1+e^{-z}\right)^{-p}}{\left(1+e^{-z}\right)^{1-p}+e^{-(1-p)z}}\, dz < \epsilon$$

As an approximation for $z < 0$ we can show that

$$\frac{z\left(1+e^{-z}\right)^{-p}}{\left(1+e^{-z}\right)^{1-p}+e^{-(1-p)z}} < \frac{ze^{pz}}{2e^{-(1-p)z}} = ze^{pz}$$

so that we find k s.t.

$$-2(1-p)^2 \int_{-\infty}^{k} ze^{z}\, dz < \epsilon$$

i.e.

$$(k-1)e^{k} > -\frac{\epsilon}{(1-p)^2}$$

Given a value of ϵ we can easily solve this equation for k and hence evaluate the integral by any of the standard techniques.

For $z > 0$ we require k s.t.

$$\int_{k}^{\infty} 2z(1-F_p(z))\, dz < \epsilon$$

we can show that

$$\begin{aligned} 1-F_p(z) &= \frac{e^{-z}+\left(1+e^{-z/(1-p)}\right)^{-p}e^{-z/(1-p)}}{e^{-z}+\left(1+e^{-z/(1-p)}\right)^{1-p}} \\ &< e^{-z}+e^{-z/(1-p)} \end{aligned}$$

so that we require k s.t.

$$\int_{k}^{\infty} e^{-z}+e^{-z/(1-p)}\, dz < \epsilon$$

This gives the following equation for k

$$2\left[(k+1)e^{-k}+\frac{e^{-k/(1-p)}}{1-p}\left(k+\frac{1}{1-p}\right)\right] < \epsilon$$

Once again we could solve this equation by standard means. We note that the first term in the bracket, i.e. $(k+1)e^{-k}$, is the dominant term and for $0 \leq p \leq 0.99$ using $k = 15$ gives an error (ignoring roundoff) less than 10^{-6}.

A graph of the relationship between the calculated correlation and p is shown. We can calculate

$$\lim_{p \to 1} \operatorname{corr}(X_1, X_3)$$

by comparing $F_p(z)$ with the standard logistic distribution

$$F(z) = \frac{1}{1+e^{-z}}$$

For $z < 0$ we can show that

$$F_p(z) < e^{pz/(1-p)}$$

so that the first term in the integral for $\mathrm{E}(Z^2)$ is less than

$$\int_{-\infty}^{0} -2ze^{pz/(1-p)}dz = \frac{2(1-p)^2}{p^2}$$

which tends to 0 as $p \to 1$.

For $z > 0$ we can write

$$\begin{aligned}
2\int_0^\infty z(1-F_p(z))\,dz &= 2\int_0^\infty z(1-F_p(z)+F(z)-F(z))\,dz \\
&= 2\int_0^\infty z(1-F(z))\,dz + 2\int_0^\infty z(F(z)-F_p(z))\,dz
\end{aligned}$$

The first term in this expression is known from the standard logistic distribution to be $\pi^2/6$. The second term is positive since

$$F_z - F_p(z) = \frac{e^{-z}}{2} + \frac{e^{-z/(1-p)}}{2} > 0$$

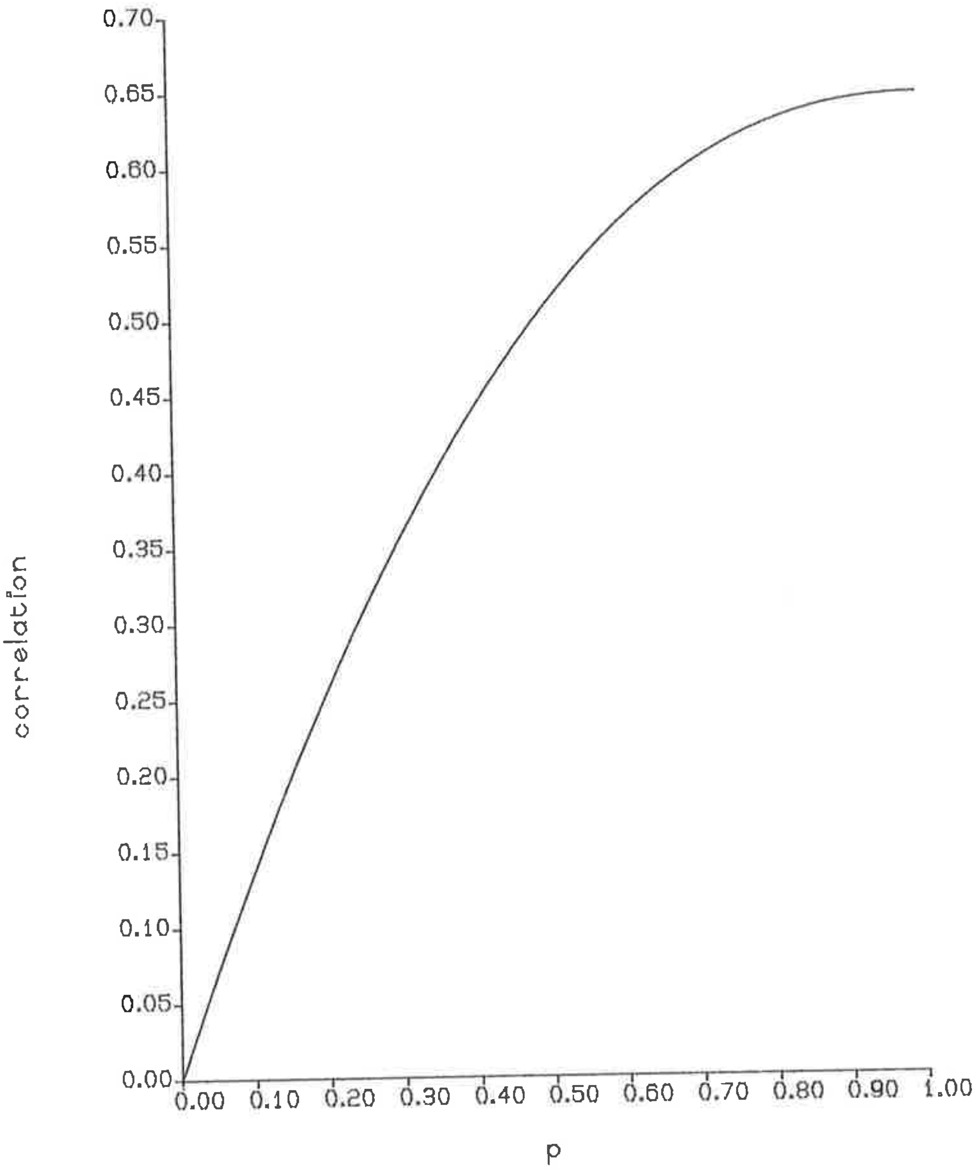


Fig 5.1 Correlation vs. p for the extreme value distribution.

In addition we can also show that

$$F_z - F_p(z) < e^{-z/(1-p)}$$

so that

$$2\int_0^\infty z(F(z) - F_p(z))\,dz < 2\int_0^\infty ze^{-z/(1-p)}\,dz = 2(1-p)^2$$

which tends to 0 as $p \to 1$.

We can therefore conclude that

$$\lim_{p\to 1} \mathrm{E}(Z^2) = \frac{\pi^2}{6}$$

Using previously stated relations we deduce that

$$\lim_{p\to 1} \mathrm{corr}(X_1, X_3) = \frac{1}{2} + \frac{(\ln 2)^2}{\pi^2/3} = 0.64604$$

The range of correlation is quite limited in comparison with the normal distribution, but for practical purposes the model is of some use since we find that the correlations involved are often reasonably small.

We can calculate the permutation probabilities corresponding to this model by standard integrals to be

$$\begin{aligned}
p_{123} &= \frac{e^{v_2}}{e^{v_2} + e^{v_3}} - \frac{e^{v_2/(1-p_{12})} a_{12}^{-p_{12}}}{a_{12}^{1-p_{12}} + a_{13}^{1-p_{13}}} \\
p_{132} &= \frac{e^{v_3}}{e^{v_2} + e^{v_3}} - \frac{e^{v_3/(1-p_{13})} a_{13}^{-p_{13}}}{a_{12}^{1-p_{12}} + a_{13}^{1-p_{13}}} \\
p_{213} &= \frac{e^{v_1} + e^{v_1/(1-p_{13})} a_{13}^{-p_{13}}}{e^{v_1} + a_{13}^{1-p_{13}}} - \frac{e^{v_1/(1-p_{12})} a_{12}^{-p_{12}} + e^{v_1/(1-p_{13})} a_{13}^{-p_{13}}}{a_{12}^{1-p_{12}} + a_{13}^{1-p_{13}}} \\
p_{231} &= \frac{e^{v_3/(1-p_{13})} a_{13}^{-p_{13}}}{e^{v_1} + a_{13}^{1-p_{13}}} - \frac{e^{v_3/(1-p_{13})} a_{13}^{-p_{13}}}{a_{12}^{1-p_{12}} + a_{13}^{1-p_{13}}} \\
p_{312} &= \frac{e^{v_1} + e^{v_1/(1-p_{12})} a_{12}^{-p_{12}}}{e^{v_1} + a_{12}^{1-p_{12}}} - \frac{e^{v_1/(1-p_{12})} a_{12}^{-p_{12}} + e^{v_1/(1-p_{13})} a_{13}^{-p_{13}}}{a_{12}^{1-p_{12}} + a_{13}^{1-p_{13}}}
\end{aligned}$$

$$p_{321} = \frac{e^{v_2/(1-p_{12})} a_{12}^{-p_{12}}}{e^{v_1} + a_{12}^{1-p_{12}}} - \frac{e^{v_2/(1-p_{12})} a_{12}^{-p_{12}}}{a_{12}^{1-p_{12}} + a_{13}^{1-p_{13}}}$$

where

$$a_{12} = e^{v_1/(1-p_{12})} + e^{v_2/(1-p_{12})} \qquad a_{13} = e^{v_1/(1-p_{13})} + e^{v_3/(1-p_{13})}$$

It is useful to find the Taylor series approximation for these probabilities. From expressions given previously we note that the standardized model is

$$\mathrm{E}(Y_1) = \frac{\sqrt{3}(v_1 + \ln(2))}{\pi} \qquad \mathrm{E}(Y_2) = \sqrt{3}V_2$$

$$\mathrm{Var}(Y_1) = 1 - \mathrm{corr}(X_1, X_3) \qquad \mathrm{Var}(Y_2) = 1$$

$$\mathrm{corr}(Y_1, Y_2) = \frac{\mathrm{corr}(X_1, X_2) - \mathrm{corr}(X_1, X_3) + 1}{2}$$

The values of the parameters $\gamma = (v_1, v_2, p_{12}, p_{13})$ which correspond to the independence or uniform probability model are

$$\gamma_0 = (-\ln(2), 0, 0, 0)$$

We can show that the Taylor series expansion of the permutation probabilities of the form

$$\mathbf{p} = \frac{1}{6}\mathbf{1} + X(\gamma - \gamma_0)$$

~~the matrix X~~ is produces a matrix X ~~given~~ as follows:

$$\begin{bmatrix} \frac{1}{9} & \frac{1}{36} & -(2\ln(2)-1.5\ln(3))/9 & (\ln(2)-1.5\ln(3))/9 \\ \frac{1}{9} & \frac{-5}{36} & (\ln(2)-1.5\ln(3))/9 & -(2\ln(2)-1.5\ln(3))/9 \\ \frac{1}{36} & \frac{1}{9} & \ln(2)/9 & -5\ln(2)/36+\ln(3)/8 \\ \frac{-5}{36} & \frac{1}{9} & (\ln(2)-1.5\ln(3)/9 & \ln(2)/36+\ln(3)/24 \\ \frac{1}{36} & \frac{-5}{36} & -5\ln(2)/36+\ln(3)/8 & \ln(2)/9 \\ \frac{-5}{36} & \frac{1}{36} & \ln(2)/36+\ln(3)/24 & (\ln(2)-1.5\ln(3))/9 \end{bmatrix}$$

We have the familiar interpretation of p_{12} and p_{13} as parameters of association in the rankings of the pairs (1,2) and (1,3).

5.3.3 Example calculations

As before to illustrate the calculations involved, consider the data used in the example of Chapter 3. For convenience the observed frequencies of the 6 permutations are given in the first row of Table 5.3. In order to apply the least squares

Table 5.3: Rankings of three cars

Order	123	132	213	231	312	321	Total
Observed Frequency	135	98	152	139	126	150	800
Expected Frequency (A)	136.2	104.4	159.4	133.7	120.0	146.3	800
Expected Frequency (B)	140.9	101.8	157.3	133.9	121.0	145.1	800

procedure to fit the approximate four parameter model we calculate X_1, V_{11}, Z_1 as demonstrated in previous examples and calculate an estimate of $\beta = \gamma - \gamma_0$ where

$\gamma - \gamma_0 = (v_1 + \ln(2), v_2, p_{12}, p_{13})$ as

$$\hat{\beta} = (X_1'V_{11}^{-1}X_1)^{-1}X_1'V_{11}^{-1}Z_1 = \begin{pmatrix} -0.066 \\ 0.124 \\ 0.172 \\ 0.013 \end{pmatrix}$$

The correlation matrix calculated for these 4 estimates is:

$$\begin{bmatrix} 0.0046 & 0.5601 & -0.0259 & 0.1372 \\ & 0.0047 & -0.1477 & 0.2075 \\ & & 0.0067 & 0.3441 \\ & & & 0.0075 \end{bmatrix}$$

An analysis of the full model is given in Table 5.4. It is reasonable to deduce from

Table 5.4: Analysis of Full Model

Parameter	Estimate	Z^2
$v_1 + \ln(2)$	-0.066	0.95
v_2	0.124	3.3
p_{12}	0.172	4.38
p_{13}	0.013	0.02

Table 5.4 that we can omit the parameters $v_1 + \ln(2)$ and p_{13}. Using the numerical maximisation routine, we find that the estimates obtained for the remaining parameters are $\widehat{v_2} = 0.147$ and $\widehat{p_{12}} = 0.163$. The expected frequencies of the model obtained using these parameters are shown in Table 5.3 (labelled A). This model gives a χ^2 value of 1.35 on 3 degrees of freedom.

We can find the standardised bivariate model using relationships given earlier in this chapter as:

$$E(Y_1) = 0 \qquad E(Y_2) = 0.081 \qquad \mathrm{Var}(Y_1) = 1 \qquad \mathrm{Var}(Y_2) = 1$$

$$\rho = \text{Corr}(Y_1, Y_2) = 0.606$$

Again we note the close agreement between this standardised form and the ones given previously for the normal and Satherwaite distributions.

If we fit a similar model using instead the expressions from the 3 parameter model we obtain the following estimates: $\widehat{v_2} = 0.161$ and $\widehat{p_{12}} = 0.116$. The expected frequencies of the model obtained using these parameters are shown in Table 5.3 (labelled B). This model gives a χ^2 value of 1.13 on 3 degrees of freedom. We note the agreement between the two models, it is usually the case that there is little difference. The standardised bivariate model using relationships given earlier in this chapter for the three parameter model is:

$$E(Y_1) = 0 \qquad E(Y_2) = 0.089 \qquad \text{Var}(Y_1) = 1 \qquad \text{Var}(Y_2) = 1$$

$$\rho = \text{Corr}(Y_1, Y_2) = 0.610$$

5.3.4 The $r = 4$ extreme value model

The generalisation of the extreme value distribution to $r = 4$ is clear. We can construct an appropriate distribution function by multiplying together functions of the form

$$\exp - \left(e^{-\frac{(X_i - v_i)}{1-p_{ij}}} + e^{-\frac{(X_j - v_j)}{1-p_{ij}}} \right)^{1-p_{ij}}$$

for as many of the pairs (1,2), (1,3), (1,4), (2,3), (2,4) as we wish. Unfortunately the permutation probabilities cannot generally be expressed in closed form. We can however make good use of the exponential form of the distribution function to achieve a significant simplification in the numerical work involved.

As an example consider a model which has a single association parameter in it. We will discuss the calculation of the permutation probabilities and illustrate the numerical simplifications that are possible. The use of the model will be illustrated in the $r = 4$ example of chapter 6. Consider the following distribution function.

$$F(x_1, x_2, x_3, x_4) = \exp - \left(\left(e^{-\frac{(x_1 - v_1)}{1-p_{12}}} + e^{-\frac{(x_2 - v_2)}{1-p_{12}}} \right)^{1-p_{12}} + e^{-(x_3 - v_3)} + e^{-(x_4 - v_4)} \right)$$

For this model we have

$$\mathrm{E}(X_i) = v_i + C \quad \mathrm{Var}(X_i) = \frac{\pi^2}{6} \quad \mathrm{corr}(X_1, X_2) = 1 - (1 - p_{12})^2$$

This function is symmetric within the subgroups of indices (1,2) and (3,4). This means that having calculated for example p_{1234} we can use this symmetry to calculate p_{1243}, p_{2134} and p_{2143}. Divide the 24 permutations accordingly into 4 groups of 6.

Group 1	Group 2	Group 3	Group 4
1234	1243	2143	2134
1324	1423	2413	2314
1342	1432	2431	2341
3412	4312	4321	3421
3142	4132	4231	3241
3124	4123	4213	3214

It is clear that when we have an algorithm for calculating the probabilities from within any one of these groups we can calculate the other probabilities by symmetry. We can show that to evaluate the integrals in the first group for example we need only evaluate 1 integral numerically. We can evaluate these probabilities as

$$\begin{aligned} p_{1234} &= \int_{-\infty}^{\infty} \int_{-\infty}^{x_3} \int_{x_3}^{\infty} \int_{x_2}^{\infty} f(x_1, x_2, x_3, x_4)\, dx_1 dx_2 dx_4 dx_3 \\ &= \frac{e^{v_3}}{e^{v_3} + e^{v_4}} \left[\frac{e^{v_2}}{e^{v_2} + e^{v_3} + e^{v_4}} - \frac{e^{v_2/(1-p_{12})} \left(e^{v_1/(1-p_{12})} + e^{v_2/(1-p_{12})} \right)^{-p_{12}}}{\left(e^{v_1/(1-p_{12})} + e^{v_2/(1-p_{12})} \right)^{1-p_{12}} + e^{v_3} + e^{v_4}} \right] \end{aligned}$$

We notice that some of the probabilities can be evaluated exactly. The essential feature of the calculation which makes this possible is that when we write the limits of integration for the variables x_1 and x_2 the first occuring pair of limits have to be elements of the set $(-\infty, \infty, x_1, x_2)$. In the case just presented the first occuring set of limits for this pair of variables was (x_2, ∞) and so we were able to evaluate the integral. As another example we have

$$\begin{aligned}
p_{3412} &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\int_{-\infty}^{x_1} f(x_1,x_2,x_3,x_4)\,dx_2dx_1dx_4dx_3 \\
&= \frac{e^{v_3}e^{v_4}e^{v_1/(1-p_{12})}}{b_{12}(a_{12}+e^{v_4})(a_{12}+e^{v_3}+e^{v_4})}
\end{aligned}$$

where

$$b_{12} = e^{\frac{v_1}{1-p_{12}}} + e^{\frac{v_2}{1-p_{12}}} \qquad a_{12} = \left(e^{\frac{v_1}{1-p_{12}}} + e^{\frac{v_2}{1-p_{12}}}\right)^{1-p_{12}}$$

$$\begin{aligned}
p_{1324} &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_3}\int_{x_4}^{x_3}\int_{x_3}^{\infty} f(x_1,x_2,x_3,x_4)\,dx_1dx_2dx_4dx_3 \\
&= e^{v_3}\left(\frac{e^{v_2}}{(e^{v_2}+e^{v_4})(e^{v_2}+e^{v_3}+e^{v_4})} - \frac{1}{(a_{12}+e^{v_3}+e^{v_4})}\right) \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{x_3} e^{-(x_3-v_3)}e^{-(x_4-v_4)}\exp-(g(x_3,x_4))\,dx_4dx_3
\end{aligned}$$

where

$$g(x_3,x_4) = e^{-(x_3-v_3)} + e^{-(x_4-v_4)} + \left(e^{-(x_3-v_3)/(1-p_{12})} + e^{-(x_4-v_2)/(1-p_{12})}\right)^{1-p_{12}}$$

We can transform this integral by using the substitution

$$Y_1 = X_4 - X_3 \quad Y_2 = X_3$$

to give

$$e^{v_3+v_4}\int_{-\infty}^{0} e^{-y_1}\int_{-\infty}^{\infty} e^{-2y_2}\exp-(e^{-y_2}a)\,dy_2dy_1$$

where

$$a = e^{v_3} + e^{-(y_1 - v_4)} + \left(e^{\frac{v_1}{1-p_{12}}} + e^{-\left(\frac{y_1 - v_2}{1-p_{12}}\right)}\right)^{1-p_{12}}$$

We can show that this gives the following integral which we denote by

$$\mathcal{R}(v_1, v_2, v_3, v_4) = e^{v_3+v_4} \int_{-\infty}^{0} \frac{e^{-y_1}}{a^2}\, dy_1$$

We can evaluate this integral numerically in the following way. Find k s.t.

$$e^{v_3+v_4} \int_{-\infty}^{k} \frac{e^{-y_1}}{a^2}\, dy_1 < \epsilon$$

We can show that

$$\frac{e^{-y_1}}{a^2} < \frac{e^{y_1}}{4}$$

so that we require k s.t.

$$\frac{e^{v_3+v_4}}{4} \int_{-\infty}^{k} e^{y_1}\, dy_1 = \frac{e^{v_3+v_4} e^{k}}{4} < \epsilon$$

$$\rightarrow k < \ln\left(\frac{4\epsilon}{e^{v_3} + e^{v_4}}\right)$$

In a similar way we can find the probabilities of the remaining permutations in group 1.

$$\begin{aligned}
p_{1342} &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\int_{x_3}^{\infty} f(x_1, x_2, x_3, x_4)\, dx_1 dx_2 dx_3 dx_4 \\
&= \frac{e^{v_3} e^{v_4}}{(e^{v_2} + e^{v_4})(e^{v_2} + e^{v_3} + e^{v_4})} - \mathcal{R}(v_1, v_2, v_3, v_4) \\
p_{3142} &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_3}\int_{-\infty}^{x_4}\int_{x_4}^{x_3} f(x_1, x_2, x_3, x_4)\, dx_1 dx_2 dx_3 dx_4 \\
&= \mathcal{R}(v_1, v_2, v_3, v_4) - \frac{e^{v_3} e^{v_4}}{(e^{v_3} + e^{v_4} + a_{12})(e^{v_4} + a_{12})} \\
p_{3124} &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_3}\int_{x_4}^{x_2}\int_{x_2}^{x_3} f(x_1, x_2, x_3, x_4)\, dx_1 dx_2 dx_3 dx_4 \\
&= \frac{e^{v_3}}{a_{12} + e^{v_3} + e^{v_4}}\left(1 - \frac{e^{v_2/(1-p_{12})} b_{12}^{-p_{12}}}{a_{12} + e^{v_4}}\right) - \mathcal{R}(v_1, v_2, v_3, v_4)
\end{aligned}$$

The probabilities in the other groups can be calculated by symmetrical procedures as explained previously.

The standardised trivariate model is

$$\mu_i = \frac{\sqrt{3}v_i}{\pi} \quad i = 1,3 \qquad \mathrm{Var}(Y_i) = 1 \quad i = 1,3$$

$$\mathrm{corr}(Y_1, Y_2) = 1 - \frac{(1-p_{12})^2}{2} \qquad \mathrm{corr}(Y_1, Y_3) = \mathrm{corr}(Y_2, Y_3) = \frac{1}{2}$$

5.4 Exponential models

5.4.1 The $r = 3$ model

The third source of exponential type models is the multivariate exponential distribution. We have seen previously that the common logistic model can be developed from exponentially distributed variables. Once again the main criteria for the selection of a particular distribution is that it produces closed form expressions for the permutation probabilities. As an example of such a model we consider the system of distributions due to Morgenstern, discussed by Gumbel[15].

The Morgenstern system of bivariate distributions is given in general form by the distribution function :

$$F(x,y) = F(x)G(y)\left(1 + \alpha(1 - F(x))(1 - G(y))\right) \qquad -1 \leq \alpha \leq 1$$

where the marginal distributions of X and Y are $F(x)$ and $G(y)$ respectively. The correlation between X and Y is related to the parameter α by

$$\mathrm{corr}(X,Y) = \frac{\alpha}{4}$$

X and Y are independent if, and only if, $\alpha = 0$. Since $|\alpha| \leq 1$ the correlation is restricted to the range $(-1/4, 1/4)$.

For our particular purpose consider the following distribution as a model for the case $r = 3$.

$$F(u_1, u_2, u_3) = \prod_{i=1}^{3} F(u_i)\left[1 + \theta_{12}(1 - F(u_1))(1 - F(u_2)) + \theta_{13}(1 - F(u_1))(1 - F(u_3))\right]$$

where

$$F(u_i) = 1 - e^{-\alpha_i u_i} \qquad u_i \geq 0$$

The marginal distributions are of standard exponential form, with mean $1/\alpha_i$. The corresponding density function is

$$f(u_1, u_2, u_3) = \prod_{i=1}^{3} f(u_i)\left[1 + \theta_{12}B_{12} + \theta_{13}B_{13}\right]$$

where

$$B_{1j} = (1 - 2F(u_1))(1 - 2F(u_j)) \quad j = 2, 3$$

$$f(u_i) = \alpha_i e^{-\alpha_i u_i} \qquad u_i \geq 0$$

The constraints on the parameters θ_{12} and θ_{13} can be written down as

$$-1 \leq \theta_{12} + \theta_{13} \leq 1 \qquad -1 \leq \theta_{12} - \theta_{13} \leq 1$$

The marginal distributions of (U_1, U_i), $i = 2, 3$ are members of the Morgenstern family since they take the form

$$F(u_1, u_i) = F(u_1)F(u_i)\left(1 + \alpha(1 - F(u_1))(1 - F(u_i))\right) \qquad -1 \leq \alpha \leq 1$$

The marginal properties of these variables are summarised as

$$\mathrm{E}(U_i) = \frac{1}{\alpha_i} \quad \mathrm{Var}(U_i) = \frac{1}{\alpha_i^2}$$

$$\mathrm{corr}(U_1, U_2) = \frac{\theta_{12}}{4} \quad \mathrm{corr}(U_1, U_3) = \frac{\theta_{13}}{4}$$

To rewrite this model in the logistic form, consider firstly the transformation

$$X_i = -\ln(U_i) \qquad i = 1, 2, 3$$

This transformation is based upon the results mentioned in Chapter 2, we recall that the distribution of X_i will be extreme value. We can write down the density function of $\mathbf{X}$ as

$$f(\mathbf{x}) = \prod_{i=1}^{3} h(x_i)\,[1 + \theta_{12}B_{12} + \theta_{13}B_{13}]$$

where

$$h(x_i) = e^{-(x_i - \ln \alpha_i)} e^{-e^{-(x_i - \ln \alpha_i)}} \qquad -\infty < x_i < \infty$$

and

$$B_{1j} = (2e^{-e^{-(x_1 - \ln \alpha_1)}} - 1)(2e^{-e^{-(x_j - \alpha_j)}} - 1) \qquad j = 2, 3$$

The marginal distributions are $h(x_i)$ which are extreme value distributions with parameters

$$\mathrm{E}(X_i) = \ln(\alpha_i) + C \quad \mathrm{Var}(X_i) = \frac{\pi^2}{6} \qquad i = 1, 2, 3$$

where C is Euler's constant.

We can, as before, find the equivalent bivariate specification by considering the distribution of the variables:

$$Y_1 = X_1 - X_3$$

$$Y_2 = X_2 - X_3$$

From the above expectations and variances of the variables X_i we can deduce that

$$\mathrm{E}(Y_1) = \ln(\alpha_1/\alpha_3) \quad \mathrm{Var}(Y_1) = \frac{\pi^2}{3} - 2\theta_{13}(\ln 2)^2$$

$$\mathrm{E}(Y_2) = \ln(\alpha_2/\alpha_3) \quad \mathrm{Var}(Y_2) = \frac{\pi^2}{3}$$

We can derive the following standardised density and distribution functions for the variables

$$Z_i = Y_i - \mathrm{E}(Y_i)$$

$$F(z_1, z_2) = \left[\frac{1+\theta_{12}+\theta_{13}}{1+e^{-z_1}+e^{-z_2}} + \frac{\theta_{12}}{1+2e^{-z_1}+2e^{-z_2}} - \frac{\theta_{12}}{1+2e^{-z_1}+e^{-z_2}} - \frac{\theta_{12}}{1+e^{-z_1}+2e^{-z_2}} \right.$$
$$\left. + \frac{2\theta_{13}}{2+2e^{-z_1}+e^{-z_2}} - \frac{\theta_{13}}{1+2e^{-z_2}+e^{-z_2}} - \frac{2\theta_{13}}{2+e^{-z_1}+e^{-z_2}}\right]$$

$$f(z_1, z_2) = 2e^{-z_1}e^{-z_2}\left[\frac{1+\theta_{12}+\theta_{13}}{(1+e^{-z_1}+e^{-z_2})^3} + \frac{4\theta_{12}}{(1+2e^{-z_1}+2e^{-z_2})^3} - \frac{2\theta_{12}}{(1+2e^{-z_1}+e^{-z_2})^3} \right.$$
$$\left. - \frac{2\theta_{12}}{(1+e^{-z_1}+2e^{-z_2})^3} + \frac{4\theta_{13}}{(2+2e^{-z_1}+e^{-z_2})^3} - \frac{2\theta_{13}}{(1+2e^{-z_2}+e^{-z_2})^3} - \frac{2\theta_{13}}{(2+e^{-z_1}+e^{-z_2})^3}\right]$$

We see that the marginal distribution of Z_1 is given by

$$F(z_1) = \frac{1+2\theta_{13}}{1+e^{-z_1}} - \frac{\theta_{13}}{1+2e^{-z_1}} - \frac{2\theta_{13}}{2+e^{-z_1}}$$

For the marginal distribution of Z_2 we have the standard logistic distribution.

By considering the distribution of $Y_1 - Y_2$ we can show that the correlation of Y_1 and Y_2 is

$$\rho = \frac{\pi^2/6 + \theta_{12}(\ln 2)^2 - \theta_{13}(\ln 2)^2}{\sqrt{\frac{\pi^2}{3}(\frac{\pi^2}{3} - 2\theta_{13}(\ln 2)^2)}}$$

The standardised form of the model obtained by scaling the variables so that $\mathrm{Var}(Y_2) = 1$ gives

$$\mathrm{E}(Y_i) = \frac{\ln(\alpha_i/\alpha_3)}{\sqrt{\frac{\pi^2}{3}}} \qquad \mathrm{Var}(Y_1) = 1 - \frac{6\theta_{13}(\ln 2)^2}{\pi^2}$$

To find the region in which the density function is non-negative it is convenient to consider the behaviour of f on the line

$$\theta_{13} = k\theta_{12} \qquad -\infty < k < \infty$$

To simplify the notation put

$$x_1 = e^{-z_1} \qquad x_2 = e^{-z_2}$$

and use the following abbreviation:

$$(a\ b\ c) = (a + bx_1 + cx_2)$$

We can write

$$f(x_1, x_2) = 2x_1x_2g(x_1, x_2) \qquad x_1 \geq 0, x_2 \geq 0$$

where

$$g(x_1, x_2) = \frac{1}{(1\ 1\ 1)^3} + \theta_{12}h(x_1, x_2, k)$$

$$h(x_1, x_2, k) = \frac{1}{(1\ 1\ 1)^3} + h_1(x_1, x_2) + \frac{k}{(1\ 1\ 1)^3} + kh_2(x_1, x_2)$$

$$h_1(x_1, x_2) = \frac{4}{(1\ 2\ 2)^3} - \frac{2}{(1\ 2\ 1)^3} - \frac{2}{(1\ 1\ 2)^3}$$

$$h_2(x_1, x_2) = \frac{4}{(2\ 2\ 1)^3} - \frac{2}{(1\ 2\ 1)^3} - \frac{2}{(2\ 1\ 1)^3}$$

We are interested in the values of θ_{12}, θ_{13} for which g and hence f is positive for all $x_1, x_2 \geq 0$. Let A be the region in which h is positive and A' be the region in which h is negative. It is easy to show that both of these regions are non empty.

When $h = 0$ we see that

$$g(x_1, x_2) = \frac{1}{(1 + x_1 + x_2)^3} > 0$$

For g to be positive it is clear that we require

$$\theta_{12} \geq \max_A \frac{-1}{(1 + x_1 + x_2)^3 h(x_1, x_2, k)} = \frac{-1}{\max_A (1 + x_1 + x_2)^3 h(x_1, x_2, k)}$$

and

$$\theta_{12} \leq \min_{A'} \frac{-1}{(1 + x_1 + x_2)^3 h(x_1, x_2, k)} = \frac{-1}{\min_{A'} (1 + x_1 + x_2)^3 h(x_1, x_2, k)}$$

The problem then is essentially one of finding the maximum and minimum values of the function

$$r(x_1, x_2, k) = (1\,1\,1)^3 h(x_1, x_2, k) = 1 + k + (1\,1\,1)^3 h_1(x_1, x_2) + k(1\,1\,1)^3 h_2(x_1, x_2)$$

Consider the values of r at its boundaries:

Boundary (x_1, x_2)	Value of r
(0,0)	$1 - 3k/4$
(0,∞)	k-3/4
(∞,0)	-3(1+k)/4
(∞, ∞)	(17-22k)/54

It is easy to see that the maximum value of r at the boundary varies with k in the following way:

$$k \leq 1 \quad \text{max value of r } = 1 - 3k/4$$

$$k > 1 \quad \text{max value of r } = k - 3/4$$

Similarly the minimum value of r at the boundary is

$$k \leq 0 \quad \text{min value of r } = k - 3/4$$

$$k > 0 \quad \text{min value of r } = -3(1+k)/4$$

To investigate the possibility of absolute extremal values existing within the region of interest we can use the following two easily proven identities:

Result 1

$h_1(x_1, x_2) < 0$ within the interior of the region $(0, \infty) \times (0, \infty)$.

The function is equal to zero only in the limiting case $x_2 \to \infty$. An equivalent result holds for $h_2(x_1, x_2)$ with the 0 value occurring at $x_1 = 0, x_2 = 0$.

Result 2

$$7/4 + (1\ 1\ 1)^3 h_i(x_1, x_2) \geq 0 \quad x_1, x_2 \geq 0, \quad i = 1, 2$$

We will find the values of k for which the maximum value of r occurs at the boundary. For $k > 1$ we require the solution of the equation

$$r(x_1, x_2, k) \leq k - 3/4 \quad \forall x_1, x_2 \geq 0$$

This gives

$$-7/4 - (1\ 1\ 1\ )^3 h_1(x_1, x_2) - k(1\ 1\ 1\ )^3 h_2(x_1, x_2) \geq 0 \quad \forall x_1, x_2 \geq 0$$

Using result 1 we deduce

$$k \geq \frac{-7/4 - (1\ 1\ 1\ )^3 h_1(x_1, x_2)}{(1\ 1\ 1\ )^3 h_2(x_1, x_2)} \quad \forall x_1, x_2 \geq 0$$

i.e.

$$k \geq \max_{x_1,x_2 \geq 0} \frac{-7/4 - (1\ 1\ 1\ )^3 h_1(x_1,x_2)}{(1\ 1\ 1\ )^3 h_2(x_1,x_2)}$$

Numerically we can show that the maximum value of this function is 1.1319. This maximum occurs at $x_1 = 1.096$, $x_2 = 0.927$. For $k \geq 1.1319$ the absolute maximum value of $r = (1\ 1\ 1)^3 h(x_1, x_2, k)$ is $k - 3/4$ so that in this range of k we have

$$\theta_{12} \geq -\frac{1}{k - 3/4} \qquad \rightarrow \qquad \theta_{13} - \frac{3}{4}\theta_{12} \geq -1$$

For $k \leq 1$ we find by a similar argument that the maximum value at the boundary $(1 - 3k/4)$ is an absolute maximum for $k \leq 0.8835 = 1/1.1319$. This symmetry is consistent with the symmetry of the original distribution. For $k \leq 0.8835$ we then have

$$\theta_{12} \geq -\frac{1}{1 - 3k/4} \qquad \rightarrow \qquad \theta_{12} - \frac{3}{4}\theta_{13} \geq -1$$

For the region defined by $0.8835 < k < 1.1319$ we need to calculate

$$m' = \max_{x_1,x_2 \geq 0} r(x_1, x_2, k)$$

After a little simplification the equations for the stationary points of r can be written as

$$\begin{aligned}
\frac{\partial r}{\partial x_1} - \frac{\partial r}{\partial x_2} &= \frac{2}{(1\,1\,2)^4} - \frac{2}{(1\,2\,1)^4} + k\left(\frac{4}{(2\,2\,1)^4} - \frac{2}{(1\,2\,1)^4}\right) = 0 \\
\frac{\partial r}{\partial x_2} &= \frac{4}{(1\,2\,2)^3} - \frac{2}{(1\,2\,1)^3} - \frac{2}{(1\,1\,2)^3} + k\left(\frac{4}{(2\,2\,1)^3} - \frac{2}{(1\,2\,1)^3} - \frac{2}{(2\,1\,1)^3}\right) \\
&\quad -(1\,1\,1)\left(\frac{8}{(1\,2\,2)^4} - \frac{6}{(1\,1\,2)^4} - \frac{2k}{(2\,1\,1)^4}\right) = 0
\end{aligned}$$

It doesnt seem possible to solve these equations analytically but we can do so numerically. We can find the maximum value m' and so we have

$$\theta_{12} > \frac{-1}{m'} \qquad 0.8835 < k < 1.1319$$

For the minimum values of r consider firstly the case $k \geq 0$. The values of k for which the absolute minimum occurs at the boundary are given by

$$r(x_1, x_2, k) \geq -3(1+k)/4 \quad \forall x_1, x_2 \geq 0$$

With simplification this becomes

$$7/4 + (1\ 1\ 1\ )^3 h_1(x_1, x_2) + k(7/4 + (1\ 1\ 1\ )^3 h_2(x_1, x_2)) \geq 0 \quad \forall x_1, x_2 \geq 0$$

Using result 2 we see that this is true for all k. We therefore have

$$\theta_{12} \leq \frac{-1}{-3(1+k)/4} \quad k \geq 0$$

so that

$$\theta_{12} + \theta_{13} \leq \frac{4}{3}$$

For $k < 0$ a similar argument leads to the region defined by

$$\theta_{13} - \frac{3}{4}\theta_{12} \leq -1 \quad k < 0$$

The region formed by these constraints is shown in Figure 5.2. We note that it is an extension of the original pair of constraints which were

$$-1 \leq \theta_{12} + \theta_{13} \leq 1 \qquad -1 \leq \theta_{12} - \theta_{13} \leq 1$$

The calculation of the permutation probabilities based on this model is straightforward. As we saw in Chapter 2, it is convenient when working with the exponential density to define, for example:

$$p_{123} = \Pr(U_1 < U_2 < U_3)$$

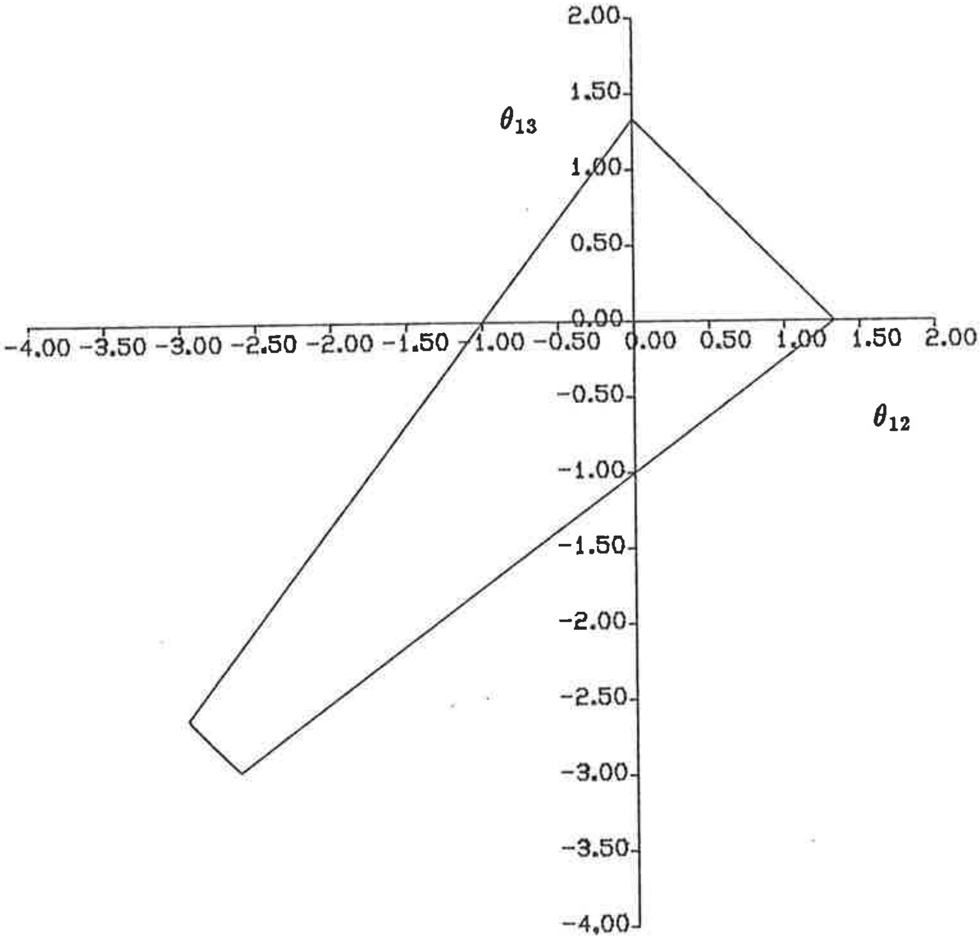


Fig 5.2 The constrained region of θ_{12}, θ_{13} within which the logistic density function is positive

When we make the transformation $X_i = -\ln U_i$ we have probabilities calculated in the usual way i.e.

$$p_{123} = \Pr(X_1 > X_2 > X_3)$$

We have

$$p_{123} = \int_0^\infty \int_{u_1}^\infty \int_{u_2}^\infty f(u_1, u_2, u_3)\, du_3 du_2 du_1$$

We can express $f(u_1, u_2, u_3)$ as a sum of terms which have the general form

$$e^{-(au_1+bu_2+cu_3)}$$

and using the result that

$$\int_0^\infty \int_{u_1}^\infty \int_{u_2}^\infty e^{-(au_1+bu_2+cu_3)}\, du_3 du_2 du_1 = \frac{1}{c(b+c)(a+b+c)} = h(a, b, c)$$

we can see that

$$p_{123} = \alpha_1\alpha_2\alpha_3\left(h(\alpha_1, \alpha_2, \alpha_3) + \theta_{12}g_1(\alpha_1, \alpha_2, \alpha_3) + \theta_{13}g_2(\alpha_1, \alpha_2, \alpha_3)\right)$$

where

$$g_1(\alpha_1, \alpha_2, \alpha_3) = h(\alpha_1, \alpha_2, \alpha_3) - 2h(2\alpha_1, \alpha_2, \alpha_3) - 2h(\alpha_1, 2\alpha_2, \alpha_3) + 4h(2\alpha_1, 2\alpha_2, \alpha_3)$$

$$g_2(\alpha_1, \alpha_2, \alpha_3) = h(\alpha_1, \alpha_2, \alpha_3) - 2h(2\alpha_1, \alpha_2, \alpha_3) - 2h(\alpha_1, \alpha_2, 2\alpha_3) + 4h(2\alpha_1, \alpha_2, 2\alpha_3)$$

It is useful to write this in the form

$$p_{123} = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2+\alpha_3)(\alpha_2+\alpha_3)} + \theta_{12}f_1(\alpha_1, \alpha_2, \alpha_3) + \theta_{13}f_2(\alpha_1, \alpha_2, \alpha_3)$$

where

$$f_i(\alpha_1, \alpha_2, \alpha_3) = \alpha_1\alpha_2\alpha_3 g_i(\alpha_1, \alpha_2, \alpha_3) \qquad i = 1, 2$$

The first term in this expression is of course the one obtained from the independent logistic model as discussed in Chapter 1. The association parameters θ_{12}, θ_{13} cause additive perturbations to the independence model, the magnitudes of which are dependent upon the mean values $1/\alpha_i$.

We can obtain expressions for the remaining probabilities by permuting elements in the expressions for h. We note that since the density $f(u_1, u_2, u_3)$ is not symmetrical in $\alpha_1, \alpha_2, \alpha_3$ we can not simply permute the elements of the functions g_1 and g_2. As an example we see that

$$p_{231} = \alpha_1\alpha_2\alpha_3 \left(h(\alpha_2, \alpha_3, \alpha_1) + \theta_{12}a_1(\alpha_1, \alpha_2, \alpha_3) + \theta_{13}a_2(\alpha_1, \alpha_2, \alpha_3)\right)$$

where

$$a_1(\alpha_1, \alpha_2, \alpha_3) = h(\alpha_2, \alpha_3, \alpha_1) - 2h(\alpha_2, \alpha_3, 2\alpha_1) - 2h(2\alpha_2, \alpha_3, \alpha_1) + 4h(2\alpha_2, \alpha_3, 2\alpha_1)$$

$$a_2(\alpha_2, \alpha_3, \alpha_1) = h(\alpha_2, \alpha_3, \alpha_1) - 2h(\alpha_2, \alpha_3, 2\alpha_1) - 2h(\alpha_2, 2\alpha_3, \alpha_1) + 4h(\alpha_2, 2\alpha_3, 2\alpha_1)$$

The simple form of the probability expressions makes calculation of derivatives and the Taylor series expansion quite straightforward. As in Chapter 2 it is most convenient to use the constraint

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

for identifiability. This means that α_i represents the probability of selection of object i. With this constraint we can find the Taylor series expansion of the probabilities in terms of the parameters $\gamma = (\alpha_1, \alpha_2, \theta_{12}, \theta_{13})$ about the values γ_0=(1/3,1/3,0,0). The expansion is

$$\mathbf{p} = \frac{1}{6}\mathbf{1} + X(\gamma - \gamma_0)$$

where the matrix X is given by

$$\begin{bmatrix} 3/4 & 1/2 & 1/60 & -2/60 \\ 1/4 & -1/2 & -2/60 & 1/60 \\ 1/2 & 3/4 & 1/60 & 1/60 \\ -1/2 & 1/4 & -2/60 & 1/60 \\ -1/4 & -3/4 & 1/60 & 1/60 \\ -3/4 & -1/4 & 1/60 & -2/60 \end{bmatrix}$$

5.4.2 Example calculations

As before to illustrate the calculations involved, consider the data used in the example of Chapter 3. For convenience the observed frequencies of the 6 permutations are given in the first row of Table 5.5. In order to apply the least squares

Table 5.5: Rankings of three cars

Order	123	132	213	231	312	321	Total
Observed Frequency	135	98	152	139	126	150	800
Expected Frequency (A)	141.5	101.8	156.7	133.9	120.8	145.3	800

procedure to fit the approximate four parameter model we calculate X_1, V_{11}, Z_1 as demonstrated in previous examples and calculate an estimate of $\beta = \gamma - \gamma_0$ where $\gamma - \gamma_0 = (\alpha_1 - 1/3, \alpha_2 - 1/3, \theta_{12}, \theta_{13})$ as

$$\hat{\beta} = (X_1' V_{11}^{-1} X_1)^{-1} X_1' V_{11}^{-1} Z_1 = \begin{pmatrix} -0.031 \\ 0.037 \\ 0.575 \\ 0.083 \end{pmatrix}$$

The correlation matrix calculated for these 4 estimates is:

$$\begin{bmatrix} 0.0002 & -0.4496 & -0.0800 & -0.1233 \\ & 0.0002 & -0.0533 & 0.1984 \\ & & 0.0799 & 0.4993 \\ & & & 0.0912 \end{bmatrix}$$

Table 5.6: Analysis of Full Model

Parameter	Estimate	Z^2
$\alpha_1 - 1/3$	-0.031	6.09
$\alpha_2 - 1/3$	0.037	8.85
θ_{12}	0.575	4.14
θ_{13}	0.083	0.08

An analysis of the full model is given in Table 5.6.

It is reasonable to deduce from Table 5.6 that we can omit the parameter θ_{13}. In addition we can achieve a model similar to those seen previously for this data by setting $\alpha_1 = \alpha_3$ Taking into account the constraint $\sum \alpha_i = 1$ the parameterisation required is then

$$\alpha_2 = \alpha \qquad \alpha_1 = \alpha_3 = \frac{(1-\alpha)}{2}$$

Using the numerical maximisation routine, we find that the estimates obtained are $\hat{\alpha} = 0.370$ and $\widehat{\theta_{12}} = 0.541$. The expected frequencies of the model obtained using these parameters are shown in Table 5.5 (labelled A). This model gives a χ^2 value of 1.15 on 3 degrees of freedom.

We can find the standardised bivariate model using relationships given earlier in this section as:

$$E(Y_1) = 0 \qquad E(Y_2) = 0.089 \qquad \mathrm{Var}(Y_1) = 1 \qquad \mathrm{Var}(Y_2) = 1$$

$$\rho = \mathrm{Corr}(Y_1, Y_2) = 0.579$$

Again we note the close agreement between this standardised form and the ones given previously.

It is clear that in this small example that there is little difference in the normal, Satherwaite logistic, extreme value and exponential models. Figures 5.3(a) and (b) illustrate the density functions corresponding to these 4 models, standardised for comparison. These pictures reinforce what we have seen in the example calculations. The distributions have similar shapes although we notice the obvious lack of symmetry about $y_1 = -y_2$ in the logistic case. In chapter 6 we shall consider another simple example which illustrates some of the problems with the restrictive ranges of correlation that some of these models suffer from.

5.4.3 Higher order models

Unlike any of the previous models, the generalisation of the exponential models to higher orders presents no problem in terms of the analytic calculation of permutation probabilities. We can easily generalise the multivariate exponential distribution to a form such as

$$F(u_1,\ldots,u_n) = \prod_{i=1}^{n} F(u_i)\left[1 + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n} \theta_{ij}(1 - F(u_i))(1 - F(u_j))\right]$$

It is also clear that we could produce models of higher order which allowed for groupings of 3 or more objects, rather than just interaction between pairs. Since the corresponding density function is of the form

$$f(u_1,\ldots,u_n) = \prod_{i=1}^{n} f(u_i)\left[1 + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n} \theta_{ij}(1 - 2F(u_i))(1 - 2F(u_j))\right]$$

the probability expressions are once again formed as the summation of simple exponential integrals of the form previously discussed. We note that the restric-

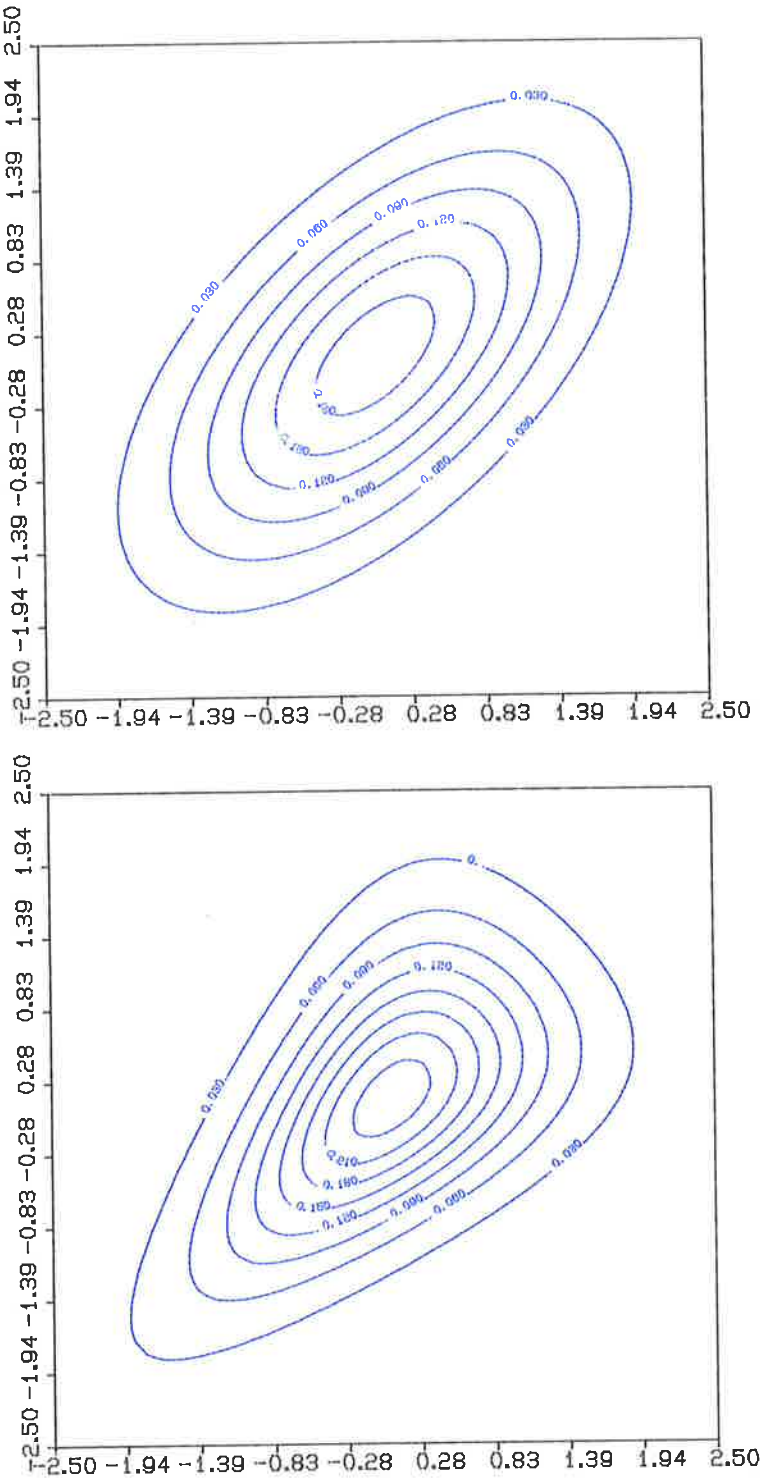
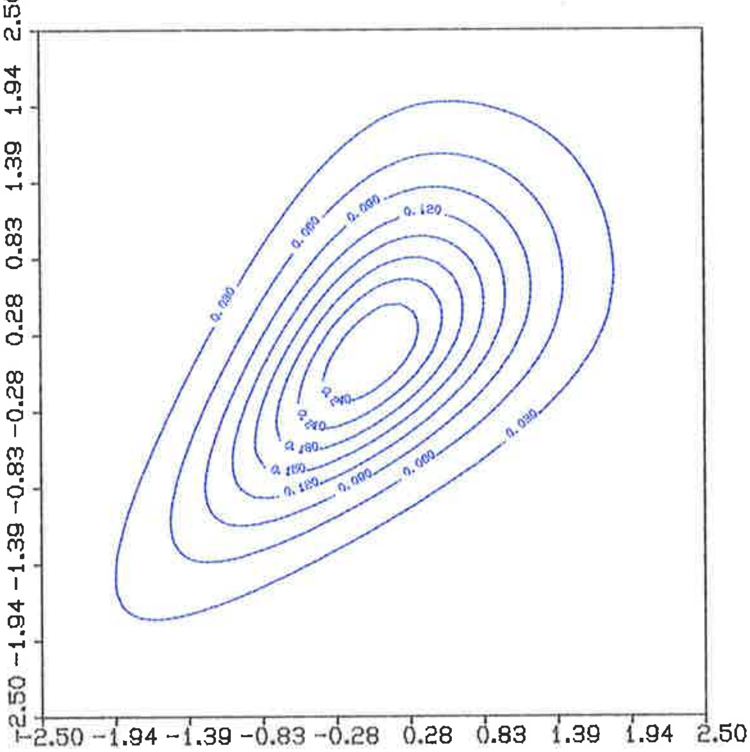


Fig 5.3(a) A comparison of the standardised distributions : the Normal (top) and the Satherwaite Logistic.



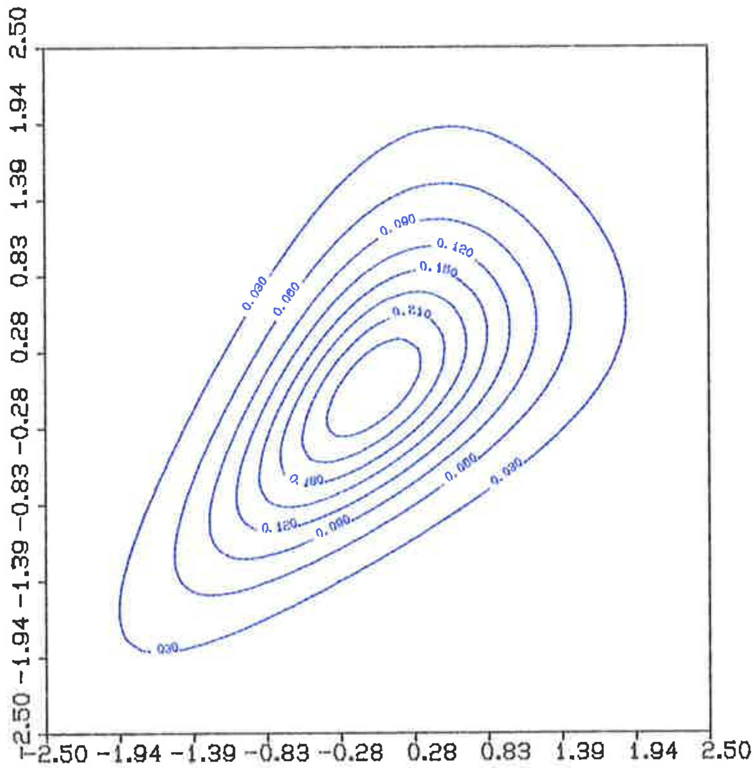


Fig 5.3(b) A comparison of the standardised distributions : logistic distributions derived from the Extreme value model (top) and the Exponential.

tions on the range of admissible values of the parameters may be quite stringent. These models do have a reasonable interpretation as approximations outside of the allowable correlation range, and can be quite useful for problems in higher dimensions.

As preliminary work for an example to be considered in the next chapter, consider the following distribution as a model for $r = 4$, with a single correlation parameter.

$$F(u_1, u_2, u_3, u_4) = \prod_{i=1}^{4} F(u_i)\left[1 + \theta_{12}(1 - F(u_1))(1 - F(u_2))\right]$$

Following a similar argument to that presented previously, we can show that the equivalent standardised model has parameters:

$$\mathrm{E}(Y_i) = \frac{\ln(\alpha_i/\alpha_4)}{\sqrt{\frac{\pi^2}{3}}} \qquad \mathrm{Var}(Y_i) = 1$$

$$\mathrm{corr}(Y_1, Y_2) = \frac{\pi^2/6 + \theta_{12}(\ln 2)^2}{\frac{\pi^2}{3}} = \frac{1}{2} + \frac{3\theta_{12}(\ln 2)^2}{\pi^2}$$

$$\mathrm{corr}(Y_1, Y_3) = \mathrm{corr}(Y_2, Y_3) = \frac{1}{2}$$

As an example of the form of the probability expressions we have

$$p_{1234} = \alpha_1\alpha_2\alpha_3\alpha_4\left[h(\alpha_1, \alpha_2, \alpha_3, \alpha_4) + \theta_{12}\left(h(\alpha_1, \alpha_2, \alpha_3, \alpha_4)\right.\right.$$

$$\left.\left. -2h(2\alpha_1, \alpha_2, \alpha_3, \alpha_4) - 2h(\alpha_1, 2\alpha_2, \alpha_3, \alpha_4) + 4h(2\alpha_1, 2\alpha_2, \alpha_3, \alpha_4)\right)\right]$$

where

$$h(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \frac{1}{\alpha_4(\alpha_4 + \alpha_3)(\alpha_4 + \alpha_3 + \alpha_2)(\alpha_4 + \alpha_3 + \alpha_2 + \alpha_1)}$$

Chapter 6

Further examples

6.1 Introduction

In this chapter we will consider two further applications of the order statistics models to the analysis of permutations. The two simple data sets are chosen to illustrate particular aspects of the models.

The first example is some data used by Plackett[32] relating to voting patterns in local government elections. The example is of interest mainly due to the effect of a negative correlation that appears in the various models, rather than as a definitive study on voting patterns in council elections. The data is used in this sense for illustrative purposes much as it was originally by Plackett[32].

The second example comes from the car clinic data, kindly provided by Data Sciences Pty. Ltd., Melbourne. From the 8 cars evaluated four have been selected to illustrate the application of the $r = 4$ models.

6.2 Council voting data

6.2.1 The Normal model

The 1973 Local Government Elections in England and Wales were originally studied by Brook and Upton[5] and Upton and Brook[40]. The data refers to parties with 3 candidates in the Tyne-Wear area. The candidates are labelled abc according to the order in which they appear on the ballot paper. A particular outcome in terms of the number of votes received is given by a permutation e.g. cba which indicates that c received most votes and a received least. These outcomes are totalled over all parties with three candidates and over all wards in the Tyne-Wear area, and are given in the first row of Table 6.1. We see that in 232 of the 948 occasions where a party fielded three candidates the resultant order at the poll was the same as the order on the ballot paper. Plackett fits a first order logistic model

Table 6.1: Orders of three candidates in elections

Order	123	132	213	231	312	321	Total
Obs Frequency	232	136	174	151	114	141	948
Exp Frequency (Logit)	201	143	204	149	124	127	948
Exp Frequency (Approx)	221.4	144.6	181.3	144.6	108.0	148.1	948
Exp Frequency (Norm)	226.6	136.2	181.2	146.6	111.3	146.1	948
Exp Frequency (Sath)	218.2	134.5	190.7	141.1	113.6	149.9	948
Exp Frequency (Extr)	223.5	133.8	186.6	138.6	116.7	148.8	948
Exp Frequency (Expo)	225.9	130.0	184.4	142.1	118.1	147.5	948

which gives expected frequencies as shown in Table 6.1. This model is basically an independent variable model. This model gives a value of 11.91 for the standard

χ^2 test, on 3 degrees of freedom. We note that the main discrepancies between the expected and observed frequencies are in the first and third permutations i.e. 123 and 213, indicating that an association parameter may be required to give a better summary of the data. In some circumstances we may interpret these sorts of discrepancies as being due to a phenomena known as the 'donkey vote' where some preferences are allocated to candidates simply according to the position that they occupy on the ballot paper. In general those candidates placed higher on the ballot paper have an increased chance of attracting a vote by this method. We might therefore expect that the mean values for the variables X_1, X_2 and X_3 would be decreasing. Since another aspect of the 'donkey vote' is to increase the chance of permutations which follow the 'natural order' of the paper, here 123, we might also expect to see evidence of association in the rankings such as positive association between places 1 and 2 or 2 and 3, or negative association between places 1 and 3.

We begin by fitting the approximate normal model using the least squares procedure as explained in chapter 4. We define X_1 as before and we have

$$Z_1 = \begin{pmatrix} \frac{232}{948} - \frac{1}{6} \\ \vdots \\ \frac{114}{948} - \frac{1}{6} \end{pmatrix} \qquad \hat{P}^{-1} = \begin{bmatrix} \frac{948}{232} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \frac{948}{114} \end{bmatrix} \qquad \hat{p}_6 = \frac{141}{948}$$

We can calculate $V_{11}^{-1} = 948(\hat{P}^{-1} + \mathbf{1}\mathbf{1}'/\hat{p}_6)$ as

$$\begin{bmatrix} 10248 & 6374 & 6374 & 6374 & 6374 \\ & 12982 & 6374 & 6374 & 6374 \\ & & 11539 & 6374 & 6374 \\ & & & 12325 & 6374 \\ & & & & 14257 \end{bmatrix}$$

We calculate the initial estimate of β as

$$\hat{\beta} = (X_1' V_{11}^{-1} X_1)^{-1} X_1' V_{11}^{-1} Z_1 = \begin{pmatrix} 0.273 \\ 0.277 \\ 0.018 \\ -0.298 \end{pmatrix}$$

By calculating $P = \mathbf{1}/6 + X\beta$ and then recalculating V_{11}^{-1} and iterating we find that in this case the values given above are correct to 3 decimal places. The χ^2 statistic obtained for this model is 2.23 on 1 degree of freedom indicating a model which fits reasonably well. The equivalent test on the exact 4 parameter normal model gives a very similar result. This test amounts basically to a test of 'normality' for the data. The correlation matrix calculated for these 4 estimates is:

$$\begin{bmatrix} 0.0031 & 0.5198 & 0.0816 & -0.0260 \\ & 0.0027 & 0.1511 & 0.0993 \\ & & 0.0084 & 0.4660 \\ & & & 0.0093 \end{bmatrix}$$

An analysis of the 4 parameters is given in Table 6.2

Table 6.2: Analysis of Full Model

Parameter	Estimate	Z^2
μ_1	0.273	24.29
μ_2	0.277	28.04
V_{12}	0.018	0.04
V_{13}	-0.298	9.59

It is clear from Table 6.2 that the parameter V_{12} is non significant. It is also apparent that we could regard μ_1 and μ_2 as being equal. Making these changes and putting $\mu = \mu_1 = \mu_2$, we can write the approximate model as $\mathbf{p} = \mathbf{1}/6 + X\beta$

where

$$X = \begin{bmatrix} a & -2b \\ 0 & b \\ a & b \\ 0 & b \\ -a & b \\ -a & -2b \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu \\ V_{13} \end{bmatrix}$$

with

$$a = \frac{1}{\sqrt{4\pi}} \qquad b = \frac{1}{4\pi\sqrt{3}}$$

We can extract the 5×2 array X_1 from this expression and fit the approximate model as previously. The estimates obtained are $\hat{\mu} = 0.274$ and $\widehat{V_{13}} = -0.307$. The expected frequencies of the model obtained are shown in Table 6.1 (labelled Approx). This model gives a χ^2 value of 2.27 on 3 degrees of freedom. The correlation matrix obtained for these estimates is

$$\begin{bmatrix} 0.0021 & -0.0166 \\ & 0.0072 \end{bmatrix}$$

The corresponding Z^2 statistics are 35.0 and 13.2 respectively, indicating no further reduction in the model is appropriate.

The interpretation of this model is reasonably straightforward. The positive value of μ relative to a zero value for μ_3 indicates preference for the first two positions on the ballot paper. As we suggested earlier, a negative value of V_{13} indicates a tendency of allocation of preference to the 'natural' ordering 123.

Using numerical maximisation of the likelihood function as explained previously, we can calculate the estimates of the exact normal model. For the 4 parameter model the estimates obtained are

$$\widehat{\mu_1} = 0.315 \quad \widehat{\mu_2} = 0.299 \quad \widehat{V_{12}} = -0.043 \quad \widehat{V_{13}} = -0.355$$

The estimates obtained by setting $\mu_1 = \mu_2 = \mu$ and $V_{12} = 0$ are $\widehat{\mu} = 0.303$ and $\widehat{V_{13}} = -0.325$. The expected frequencies corresponding to this model are given in Table 6.1 (labelled Norm). The χ^2 value is 0.80 on 3 d.f. We note the close agreement with the approximate model.

For purposes of comparison with the logistic models of chapter 5, we can find the standardised bivariate normal model using relationships given earlier in this chapter as:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0.214 \\ 0.214 \end{pmatrix}, \begin{pmatrix} 1.326 & 0.663 \\ & 1.000 \end{pmatrix}\right]$$

We recall that a model with the association parameter V_{13} is characterised in the standardised bivariate form by the model $\sigma_1^2 = 1 - V_{13}$ and $\sigma_{12} = \frac{1-V_{13}}{2}$.

6.2.2 The Logistic models

We recall from our discussion of the Satherwaite bivariate logistic model, that we obtain a model of association between objects 1 and 3 using the following parameterisation:

$$\sigma_1^2 = \delta \quad \frac{\varsigma(2,\nu)}{\varsigma(2,\nu) + \pi^2/6} = \frac{\sqrt{\delta}}{2}$$

In addition we can obtain the equal first and second means model by setting

$$v_1 = v \quad v_2 = v + (\sigma_1 - 1)(\psi(\nu) + C)$$

The estimates of the two parameters obtained by doing this are:

$$\widehat{v} = 0.590 \qquad \widehat{\sigma_1^2} = 1.240$$

The expected frequencies corresponding to this model are given in Table 6.1 (labelled Sath). The χ^2 value is 3.56 on 3 d.f. The standardised bivariate model (written using the variance matrix notation of the normal distribution) is:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \left[\begin{pmatrix} 0.155 \\ 0.155 \end{pmatrix}, \begin{pmatrix} 1.240 & 0.620 \\ & 1.000 \end{pmatrix} \right]$$

The negative correlation causes no problem to the bivariate logistic model. The range of correlation available in this model parallels that available in the normal model. This model is very useful as an alternative to the normal distribution in the $r = 3$ case. It gives similar results and is much easier to calculate. In examples where covariates are involved this could be quite important. It is limited to the case $r = 3$ as generalisation of the distribution to higher dimensions does not seem to provide a useful model. Interpretation of the parameters in the model is not as easy as it is for the normal model.

In the case of the extreme value model (using here the 4 parameter version) we obtain a model of association between objects 1 and 3 using the association parameter p_{13} and setting the parameter $p_{12} = 0$. To achieve an equal means model we notice from the standardised form that we need to set

$$v_1 = \mu - \ln(2) \qquad v_2 = \mu$$

The estimates of the parameters obtained by doing this are:

$$\hat{\mu} = 0.318 \qquad \widehat{p_{13}} = -.209$$

The expected frequencies corresponding to this model are given in Table 6.1 (labelled Extr). The χ^2 value is 2.78 on 3 d.f. We notice a problem here. The distribution from which this model was derived has a condition that $0 < p_{13} < 1$

which corresponds to a model which admits only positive correlation. In this sense then we have an invalid model since $p_{13} < 0$. The model is however valid in the sense that the probabilities are positive and sum to 1. We obviously cannot interpret it as an order statistic model but it is of some limited use as a model of association.

We have then the obvious problem of a restricted range of correlation with this model. If we are a little liberal in our interpretation it does have the advantage of giving closed form expressions for the 4 parameter r=3 model, it is perhaps of some use as an initial guide in the selection of an appropriate model. We will see the application to the $r = 4$ case in the next section.

For the exponential model we once again obtain a model of association between objects 1 and 3 using the association parameter θ_{13} and setting the parameter $\theta_{12} = 0$. To achieve an equal means model we notice from the standardised form that we need to set $\alpha_1 = \alpha_2 = \mu$. The estimates of the parameters obtained by doing this are:

$$\widehat{\mu} = 0.368 \qquad \widehat{\theta_{13}} = -.712$$

The expected frequencies corresponding to this model are given in Table 6.1 (labelled Expo). The χ^2 value is 2.02 on 3 d.f. The standardised bivariate model (written using the variance matrix notation of the normal distribution) is:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \left[\begin{pmatrix} 0.184 \\ 0.184 \end{pmatrix}, \begin{pmatrix} 1.208 & 0.604 \\ & 1.000 \end{pmatrix} \right]$$

We notice once again that this is very similar to the two previous forms. In this case the negative correlation causes no problem although we are starting to get close to the lower boundary of -1 as the smallest allowable value for θ_{13}. The range of correlation available in this model is restricted compared to what it is in the normal and Satherwaite models. It does have an advantage over the extreme value model in that it allows both positive and negative correlation, although the range available is not that wide in either direction. In many practical situations however we do not see very large correlations. The model again has the useful property of closed form probability expressions. In the higher dimensions it is very versatile to work with as we have mentioned previously. It is a useful alternative to the normal distribution except in those circumstances where a wider range of correlation is required.

6.3 Car clinic example

6.3.1 The Normal model

The following example was kindly provided through the co-operation of Data Sciences Pty. Ltd., Melbourne. As part of a car clinic, 800 people were asked to give their preference for a set of motor vehicles. The rankings of 4 such vehicles are given in Table 6.3.

The only information that was supplied with the data about these 4 particular cars is that car 4 is an experimental model and that cars 1 and 2 can be regarded as having similar characteristics. Using the iterative least squares procedure explained previously, the approximate normal model was fitted. The maximum number of parameters in the normal model is 8. The parameter estimates from

Table 6.3: Rankings of 4 cars

Order	Observed frequency	Expected Frequencies			
		Approximate model	Normal model	Extreme model	Exponential model
1234	37	35.8	35.6	32.8	33.9
1243	36	35.8	35.6	32.8	33.9
1324	41	40.5	40.6	40.9	39.6
1342	44	46.0	47.0	45.7	45.4
1423	41	40.5	40.6	40.9	39.6
1432	49	46.0	47.0	45.7	45.4
2134	31	31.4	31.7	31.9	32.4
2143	26	31.4	31.7	31.9	32.4
2314	40	32.6	32.7	35.8	34.6
2341	33	33.6	33.4	32.8	32.4
2413	33	32.6	32.7	35.8	34.6
2431	35	33.6	33.4	32.8	32.4
3124	38	34.1	33.8	34.0	34.0
3142	38	36.1	34.8	35.4	36.3
3214	25	30.6	30.8	31.3	31.6
3241	22	28.2	28.2	27.1	28.1
3412	33	27.8	27.2	28.4	27.8
3421	25	23.4	24.3	23.9	23.8
4123	39	34.1	33.8	34.0	34.0
4132	23	36.1	34.8	35.4	36.3
4213	26	30.6	30.8	31.3	31.6
4231	34	28.2	28.2	27.1	28.1
4312	30	27.8	27.2	28.4	27.8
4321	21	23.4	24.3	23.9	23.8

both the approximate and exact normal models are given in Table 6.4, along with the Z^2 statistics formed from the approximate model calculations for which the following is the correlation matrix:

$$\begin{bmatrix} 0.0032 & 0.452 & 0.516 & -0.023 & -0.044 & -0.099 & 0.008 & 0.005 \\ & 0.0032 & 0.493 & 0.041 & 0.004 & -0.048 & 0.004 & -0.005 \\ & & 0.0032 & 0.010 & 0.060 & -0.041 & 0.010 & 0.009 \\ & & & 0.0130 & 0.602 & 0.595 & 0.593 & 0.595 \\ & & & & 0.0083 & 0.497 & 0.494 & 0.316 \\ & & & & & 0.0084 & 0.286 & 0.491 \\ & & & & & & 0.0084 & 0.493 \\ & & & & & & & 0.0083 \end{bmatrix}$$

Table 6.4: Analysis of Full Model

Parameter	Approximation estimate	Normal model estimate	Z^2
μ_1	0.194	0.200	11.71
μ_2	0.009	0.020	0.03
μ_3	0.024	0.025	0.18
V_{12}	-0.230	-0.220	4.08
V_{13}	0.010	0.016	0.01
V_{14}	-0.052	-0.028	0.33
V_{23}	-0.035	-0.025	0.15
V_{24}	-0.004	0.007	0.002

The χ^2 goodness of fit statistics on 15 d.f. are 14.1 and 14.6 respectively.

It is clear from Table 6.4 that a much simpler model can be obtained by setting all of the parameters but μ_1 and V_{12} equal to zero. For this reduced model, the least squares estimates are 0.179 and -0.203, compared with the normal estimates which are 0.186 and -0.206. For the approximate model the correlation matrix for these two estimates is

$$\begin{bmatrix} 0.0021 & -0.003 \\ & 0.0048 \end{bmatrix}$$

The frequencies for this 2 parameter normal model are given in Table 6.3 (labelled Normal) and for the 2 parameter approximate model (labelled Approximate). The χ^2 goodness of fit statistics on 21 d.f. are 15.04 and 14.79 respectively indicating a model which fits reasonably well.

The standardised trivariate model is given as:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0.132 \\ 0.000 \\ 0.000 \end{pmatrix}, \begin{pmatrix} 1.000 & 0.397 & 0.500 \\ & 1.000 & 0.500 \\ & & 1.000 \end{pmatrix}\right]$$

The interpretation of the model is reasonably straightforward. A value of $\mu_1 > 0$ relative to $\mu_2 = \mu_3 = \mu_4 = 0$ indicates a preference for the first car. The negative value of V_{12} indicates a negative association between the first two cars. This is surprising in the light of the opinion offered with the data which suggested that they had similar characteristics.

6.3.2 The logistic models

In this sub-section we will use the models derived in sections 5.3 and 5.4 which dealt with the case of a single association parameter.

In the case of the extreme value model we obtain a model of association between objects 1 and 2 using the association parameter p_{12}. We set $v_2 = v_3 = v_4 = 0$. The estimates of the parameters are:

$$\widehat{v_1} = 0.196 \qquad \widehat{p_{12}} = -0.141$$

The expected frequencies corresponding to this model are given in Table 6.3 (labelled Extreme). The χ^2 value is 15.0 on 21 d.f.

We once again notice that the negative value of p_{12} means that our model here is not a proper order statistics model. We also note that the model is valid in the sense that the probabilities are positive and sum to 1. The comments made in the previous section regarding this model are also appropriate here.

In the case of the exponential model we obtain a model of association between objects 1 and 2 using the association parameter θ_{12}. To achieve a model which has a single mean parameter $\mu = \alpha_1$ we notice from the standardised form that to have $\alpha_2 = \alpha_3 = \alpha_4$ we need to set

$$\alpha_2 = \alpha_3 = \alpha_4 = \frac{1-\mu}{3}$$

The estimates of the two parameters are obtained as:

$$\widehat{\mu} = 0.289 \qquad \widehat{\theta_{12}} = -.595$$

The expected frequencies corresponding to this model are given in Table 6.3. The χ^2 value is 15.9 on 21 d.f. The standardised bivariate model (written using the variance matrix notation of the normal distribution) is:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \left[\begin{pmatrix} 0.109 \\ 0.000 \\ 0.000 \end{pmatrix}, \begin{pmatrix} 1.000 & 0.413 & 0.500 \\ & 1.000 & 0.500 \\ & & 1.000 \end{pmatrix} \right]$$

We notice that this is very similar to the normal form. In this case the negative correlation causes no problem. The fact that this model has the useful property of closed form probability expressions makes it noticeably faster to calculate than the normal model even in what is a fairly simple situation.

6.4 Conclusion

These reasonably simple examples were chosen to illustrate some aspects of the order statistics model, in particular the interpretation of the parameters and the method of calculation. The parametric approach to the analysis of permutations deals quite simply with the dependence between the rankings, which is an advantage over a standard non-parametric analysis. In situations where we wish to compare two or more populations, or where covariates are involved, the model based approach has obvious advantages.

The limitations of this approach centre mainly on the numerical difficulties associated with the normal model. Although some useful logistic models have been developed, we saw that in common with the multinomial probit model discussed in Chapter 1, the obvious need is for a generalised logistic distribution which meets the dual criteria of having a suitably varied correlation structure, but being simple enough to calculate closed form expressions for the permutation probabilities.

Bibliography

[1] Block H. and Marshack, J. (1960). Random orderings and stochastic theories of response. In I. Olkin and others (Eds.), *Contributions to Probability and Statistics*. Stanford, Calif.: Stanford Univ. Press, 1960.

[2] Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete blocks. I. The method of paired comparisons. *Biometrika.*, **47**, 324-345.

[3] Bradley, R.A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics.*, **9**, 22-38.

[4] Bradley, R.A. (1976). Science, statistics and paired comparisons. *Biometrics.*, **32**, 213-232.

[5] Brook, D. and Upton, G.J.G. (1974). Biases in local government elections due to position on the ballot paper. *Appl.Statist.*, **23**, 414-419.

[6] Clark, C.E. (1961). The greatest of a finite set of random variables. *Oper.Res.*, **9**, 145-162.

[7] Daganzo, C.F. (1979). *Multinomial Probit; The theory and Its application to Demand Forecasting*. Academic Press, New York.

[8] Daley, D.J. (1974). Computation of bi- and tri-variate normal integrals. *Appl.Statist.*, **23**, 435-438.

[9] Dansie, B. R., (1983). A note on permutation probabilities. *J. R. Statist. Soc. B.*, **43**, 22-24.

[10] Dansie, B.R. (1985). Parameter estimability in the multinomial probit model, *Transpn. Res.-B.*, **19B**, 526-528.

[11] Dansie, B.R., (1986). Normal order statistics as permutation probability models. *Appl. Statist.*, **35**, 269-275.

[12] Davidson, R.R. (1969). On a relation between two representations of a model for paired comparisons. *Biometrics.*, **25**, 597-599.

[13] Davidson, R.R. and Farquhar , P.H. (1976). A bibliography on the method of paired comparisons. *Biometrics.*, **32**, 241-252.

[14] Godwin, H.J. (1949). Some low moments of order statistics. *Ann.Math.Satist.*, **20**, 279-85.

[15] Gumbel, E.J. (1960). Bivariate exponential distributions. *J.A.S.A.*, **55**, 698-707.

[16] Gumbel, E.J. (1961). Bivariate logistic distributions. *J.A.S.A.*, **56**, 335-349.

[17] Henery, R.J. (1981). Permutation probabilities as models for horse races. *J. R. Statist. Soc.*,B., **43**, 86-91.

[18] Henery, R.J. (1981). Place probabilities in normal order statisics models for horse races. *J.Appl. Prob.,* **18**, 839-852.

[19] Henery, R.J. (1983). Permutation probabilities for gamma random variables. *J.Appl. Prob.,* **20**, 822-834.

[20] Horowitz, J.L., Sparmann, J.M. and Daganzo, C.F. (1982). An investigation of the Clark approximation of the multinomial probit model. *Trans. Sci.,* **16**, 382-401.

[21] Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* New York:Wiley.

[22] Jones, H.L. (1948). Exact lower moments of order statistics in small samples from a normal population. *Ann.Math.Stat.,* **19**, 270-273.

[23] Luce, R.D. and Suppes, P. (1965). Preference, Utility and Subjective Probability. In R. Luce, R. Bush and E. Galanter (Eds.), *Handbook of Mathematical Psychology, III.* Wiley, New York.

[24] Luce, R.D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology.,* **15**, 215-233.

[25] McFadden, D. (1976). Conditional logit analysis of qualitative choice behaviour. In *Frontiers in econometrics.,* (Edited by P.Zarembka). Academic Press, New York.

[26] McFadden, D. (1976). Quantal choice analysis: a survey. *Annals of Economic and Social measurement.*, **5**, 363-390.

[27] McFadden, D. (1978). Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models.* (Edited by A. Karlqvist, L Lundquist, F Snickers and J. Weibull). North Holland, Amsterdam.

[28] Mosteller, F. (1951). Remarks on the method of paired comparisons I. *Psychometrika.*, **16**, 3-9.

[29] O'Neill, R. (1971). Function minimisation using a simplex procedure. *Appl.Statist.*, **20**, 338-345.

[30] Pendergrass, R.N. and Bradley, R.A. (1960). Ranking in triple comparisons. In I. Olkin and others (Eds.), *Contributions to probability and statistics.* Stanford, Calif.: Stanford Univ. Press, 1960. Pp.331-351.

[31] Pettitt, A.N. (1982). Inference for the linear model using a likelihood function based on ranks. *J.R. Statist. Soc.*,B, **44**, 234-243.

[32] Plackett, R.L. (1975). The analysis of permutations, *Appl. Statist.*, **24**, 193-202.

[33] Russell, N.S., Farrier, D.R. and Howel, J. (1985). Evaluation of multinormal probabilities using fourier series expansions, *Appl.Statist.*, **34**, 49-53.

[34] Satherwaite, S.P. and Hutchinson, T.P. (1978). A generalisation of Gumbel's bivariate logistic distribution. *Metrika.*, **25**, 163-170.

[35] Schervish, M.J. (1984). Multivariate normal probabilities with error bound. *Appl.Statist.*, **33**, 81-94.

[36] Schervish, M.J. (1985). Correction to 'Multivariate normal probabilities with error bound.' *Appl.Statist.*, **34**, 103-104.

[37] Schulman, R.S. (1979). Ordinal data : An alternative distribution. *Psychometrika.*, **44**, 3-20.

[38] Tallis, G.M. and Dansie, B.R. (1983). An alternative approach to the analysis of permutations. *Appl. Statist.*, **32**, 110-114.

[39] Thurstone, L.L. (1927). Psychophysical analysis. *Amer.J.Psychol.*, **38**, 368-389.

[40] Upton, G.J.G and Brook, D. (1974). The importance of positional voting bias in British elections. *Polit.Studies.*, **22**, 178-190.

[41] Yellott, J.I. (1977). The relationship between Luce's choice axiom, Thurstones theory of comparitive judgement, and the double exponential distribution. *Journal of Mathematical Psychology.*, **15**, 109-144.

Errata

Page 19 There are other constraints on the α's and β's, since the probabilities p satisfy $0 \leq p \leq 1$. The stated constraints are for the purpose of making the parameters identifiable, and not aliased.

Page 20 The same comments as above apply to the general model.

Page 58 For additional information, I am including the variance of the estimators on the diagonal of the correlation matrix. This is a non-standard way of writing this matrix.

Page 79, line 10 *its* not it's.

Page 83, line 8 *cannot*, all one word

Page 118 a *phenomenon* not phenomena