IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY

Department of Mathematics

MULTIPIVOTAL MODELS WITH APPLICATIONS TO A SHAPE FITTING PROBLEM

by

Paul James Scott

Thesis submitted for the degree of
Doctor of Philosophy in the University of London

June, 1983

ABSTRACT

A class of probability models called Multipivotal Models is examined.

Pivotal Models, which are an extension of the Structured Models developed by Fraser, are considered first. The group structure of the transformations defined in Structured Models is found to be too strong and can be weakened to an algebraic structure called a Loop while still keeping most of the properties of the Structured Model. This includes preserving the existence of a "pivotal distribution" of the transformations from which pivotal probability statements about the transformations can be obtained. A pivotal probability measure is introduced which offers an explanation for some of the more unusual properties of pivotal probability. One possible interpretation of the pivotal distribution is also given.

Multipivotal models are introduced by combining pivotal models in a symmetrical step by step method, thus ensuring that the order of combination of the pivotal models is not important.

These multipivotal models provide a general method of fitting a shape from a class of geometrical shapes, such as ellipses or cones, to a set of data that lies approximately on a shape from the class; this specific application is called the shape fitting problem.

Multipivotal models lead to tests between various hypotheses of shape associated with Megalithic stone rings. Other examples are in manufacturing industry in supplying tests of the shape of engineered components within acceptable tolerances. Examples are presented from these fields of application.

ACKNOWLEDGEMENTS

I am indebted to Dr R Coleman for his guidance throughout the preparation of this work.

Thanks are due to Rank Taylor Hobson Limited for its financial support throughout the three years and to the Science and Engineering Research Council who jointly financed the first two years, without which this research at Imperial College would not have been possible.

I would like to thank my fellow students for useful discussions on statistics. In particular, I thank my friend Charles Taylor, for much good advice and for invaluable hours of discussion on various aspects of statistics.

Finally, I thank Ms Jane Waterfield and Ms Meena Maher for typing this manuscript efficiently and being so patient with me.

TABLE OF CONTENTS

Appendices. Page No.

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Increasingly many disciplines are producing data structures that take the form of a set of points lying approximately on the perimeter of an unknown geometrical shape which is believed to come from a particular class of geometrical shapes. Questions associated with this type of data structure are collectively known as Shape Fitting Problems.

The "true" shapes of Megqlithic stone rings, such as Avebury or Stonehenge, have been the subject of various hypotheses about their being circular or elliptical or whether their shape is a series of arcs of a circle. The analysis of shape fitting problems can suggest tests of the alternative hypotheses.

The need of manufacturing industry to test the quality of its products is increasing. As the shape of most engineered components is specified in a blueprint, the accuracy of the fit to the blueprint shape can confirm that the components have the specified shape to within acceptable tolerances.

The particular case when the unknown geometrical shape is believed to come from the class of all circles was shown in Scott's DIC thesis, Scott (1981), that this could be modelled by a conditional structural model (Fraser 1968 part II). This makes use of geometrical transformations which seem to be sensible for such geometrical problems.

With this particular case in mind, it was decided to extend the results of Fraser to encompass a wider class of geometrical shapes. The theory for this is described in Chapter two.

Section 2.1 begins with the definition of a class of probability models called Structured Models (Fraser 1971). The distinction between Structured Models and Classical Probability Models is made.

In section 2.2 a Pivotal Model is defined. It is a Structural Model that satisfies three conditions: a Partition condition, an Additivity condition and a Pivotal condition. This particular form of definition is that used by Dawid and Stone (1982).

Furthermore it is shown how upper and lower probabilities (Dempster 1966) can be constructed on the set of transformations of a structured model, when it satisfies the partition condition. If a Structured Model also satisfies the additivity condition, the upper and lower probabilities on any particular subset of transformations will be equal to each other. This was shown by Plante (1979). Because a Pivotal Model satisfies both the Partition and Additivity conditions, we can alternatively define a Pivotal Distribution on the set of transformations.

In section 2.3 some properties of Pivotal Models are explored. The idea of Equivalent Pivotal Models is introduced. It is shown that any two Equivalent Pivotal Models will generate the same Pivotal Distribution on their common parameter space. It is then demonstrated how to construct a mathematically simple Equivalent Pivotal Model for any given Pivotal Model. It is then assumed for the remainder of this thesis that any given pivotal model is already in its mathematically simple equivalent form.

The sample space is decomposed into what are called reference variables and transformation variables. (These are analogous to the sufficient statistics and ancillary statistics of classical statistical inference). By conditioning on the reference variables we obtain a Reduced Pivotal Model which contains all the relevant information about the parameters in the model. A binary operation, defined on the parameter space, is induced by this particular reduction of the model and this is shown to have the algebraic property called a Loop. Loops do have a connection with geometrical transformations (see Bruck (1971) for instance), which is one of the reasons for this particular avenue of exploration.

The Pivotal Distribution is then calculated using loop-invariant differentials and the resulting distribution is written in a form that is similar to the form of the distribution obtained for Structural Models (Fraser 1968).

Finally in this section the relationships between Pivotal and Structural Models and between the Fiducial and Pivotal Distributions are discussed.

Section 2.4 is concerned with developing the idea of Pivotal Measure. The concept arose in answer to the following question:

"If the Pivotal Distribution is a probability distribution then what is its event space, σ-algebra and probability measure?"

The resulting Pivotal Measurable sets are shown to have an expected posterior probability interpretation.

Pivotal Measurable Functions are then introduced as a way of extending Pivotal Probability to other parameter spaces and the resulting measurable sets are also shown to have an expected posterior probability interpretation.

Finally there is a discussion of the relationship between some of the ideas of Wilkinson (1977) and those introduced in this section.

In section 2.5 the Pivotal Model is generalised to include the composition of transformations, such that for a particular transformation if the others were known then the model would be a Pivotal Model. This generalisation is called a Multipivotal Model.

The Conditional Pivotal Distributions of each transformation in the composition, given that we know the other transformations of the composition, are calculated and combined using a symmetrical step by step procedure to obtain a joint distribution called the Multipivotal Distribution. Finally the interpretation and some problems of using the Multipivotal Distribution are discussed.

In Chapter three some alternative approaches for specific shape fitting problems that have been suggested are described and compared with the methods developed in Chapter two through the analysis of some real data from the engineering industry and archaeology.

We begin, in section 3.1, by examining some of the alternative methods for fitting a circle through a set of data that are believed to be approximately on a circle.

A multipivotal model is constructed for the problem. From this a multipivotal distribution is obtained. This can be used to construct intervals of the parameters that have various pivotal probabilities attached to them. These can be used to make various inferences about the parameters.

The first illustrative example is Brogar Megalithic Stone Ring (Figure 3.1.1). One hypothesis about this ring is that it is a circle with a diameter of 125

Megalithic Yards. Assuming this to be true, one can obtain estimates of the length of a megalithic yard. A comparison of various methods of fit is given together with their estimates of the value of the megalithic yard.

The second example is Avebury Megalithic Stone Ring (Figure 3.1.4). It was suggested by Thom et al (1976) that the ring consists of a series of arcs of a circle with specific centres and radii. This is investigated.

In section 3.2 we examine the ellipse fitting problem. Some of the alternative methods of fitting an ellipse through a set of data that are believed to lie approximately on an ellipse are given.

Two multipivotal models are constructed together with their multipivotal distributions. The distributions for the two models have different interpretations, so the choice of model depends on the final interpretation that is required.

The illustrative example used for this comes from the engineering industry: a mechanical component that is believed to have an elliptic cross-section with major and minor axes of 12.70 mm and 6.35 mm.

In section 3.3 the rectangle, limacon, cuboid and cone fitting problems are presented. A multipivotal model is constructed for each, together with its multipivotal distribution. These particular shape fitting problems illustrate some of the typical problems of shape fitting. The limacon fitting problem is an example where the data and the class of geometrical shapes are expressed in polar co-ordinates. The cuboid and cone fitting problems are examples of three dimensional shape fitting problems.

Finally, in chapter four, conclusions and suggestions for further work are discussed.

In Appendix one some theorems and definitions are given.

In Appendix two a summary of the theory of loop and loop invariant differentials is presented.

Appendix three contains an algorithm used in the circle and ellipse fitting problem and listings of the computer programs used in chapter three.

2 PIVOTAL AND MULTIPIVOTAL MODELS

We start this chapter by defining a class of probability models called Structured Models (Fraser 1971). We then restrict the Structured Models to a category called Pivotal Models (Dawid & Stone, 1982) which we examine in some detail. Later in this section we extend the pivotal models to a class called Multipivotal Models.

2.1 Structured Models

Structured Models are a theoretically useful class of probability models. They are also called Functional Models (Dawid & Stone, 1982). In Structured Models we observe a realisation of a random variable from a fixed known distribution which has undergone an unknown transformation taken from a known set of transformations. This set is such that each realisation and transformation produces a unique observation. We denote this model by

$$x = \theta o e \qquad (2.1.1)$$

where $e \varepsilon \Omega$ is a random variable from a known distribution $f(\cdot)$ and Ω is the sample space of e's, we observe $x \varepsilon \mathfrak{X}$ the "observed" sample space and $\{\phi_\theta : \theta \varepsilon \Theta\}$ the set of transformations: $\Omega \to \mathfrak{X}$ is given. A more extensive formal definition of a Structured Model may be found in appendix A1.1 and is due to Plante (1979).

Example 2.1.1 $N(\theta,1)$ model.

Let $f(\cdot)$ be the standard normal distribution $N(0,1)$ on $\Omega=\mathbb{R}$; and consider the family of location transformations

$$\{\phi_\theta : \theta\varepsilon\Theta = \mathbb{R}\} : \Omega = \mathbb{R} \to \mathfrak{X} = \mathbb{R}$$

such that if $e\varepsilon\Omega$ and $\theta\varepsilon\Theta$ then

$$\theta o e = \theta + e\varepsilon\mathfrak{X} \tag{2.1.2}$$

We now have a Structured Model with

$$x = \theta o e \;, \quad e \sim N(0,1) \tag{2.1.3}$$

where $x\varepsilon\mathfrak{X}$, $\theta\varepsilon\Theta$ and $e\varepsilon\Omega$ □

Example 2.1.2 $\mathrm{Bin}(3,\ 1-\theta)$ model.

Let $\mathfrak{X}$ be the set $\{0,1,2,3\}$, let $f(\cdot)$ be the rectangular distribution on $\Omega = (0,1)$ and consider the family of transformations $\{\Phi_\theta : \theta\varepsilon\Theta = (0,1)\}$
defined as follows. If $\theta\varepsilon\Theta$ and $e\varepsilon\Omega$ then

$$\theta o e = \begin{cases} 0 & \text{if } 0 < e \leqq t_1 \\ 1 & \text{if } t_1 < e \leqq t_2 \\ 2 & \text{if } t_2 < e \leqq t_3 \\ 3 & \text{if } t_3 < e < 1 \end{cases} \tag{2.1.4}$$

where $t_1 = \theta^3$, $t_2 = t_1 + 3\theta^2(1-\theta)$, $t_3 = t_2 + 3\theta(1-\theta)^2$.
We now have a Structured Model with

$$x = \theta \circ e \qquad e \sim \text{rect}(0,1) \tag{2.1.5}$$

where $x \in \mathfrak{X}$, $\theta \in \Theta$ and $e \in \Omega$ □

From the definition of a Structured Model, it is easily seen that the interpretation is different from that of a classical probability model. In the classical model the random variable is observed directly but its distribution is known only to come from a set of possible distributions, whereas in the Structured Model the random variable has a fixed distribution but we only observe the random variable after it has been transformed by a transformation taken from a set of transformations.

Of course we can always convert a Structured Model into a classical model fairly easily ensuring that there is a one to one correspondence between the transformation used in the Structured Model and the distribution used in the classical model. In the classical theory, x of example 2.1.1 would have a normal distribution with mean θ and variance 1 and x of example 2.1.2 would have a binomial $(3,1-\theta)$ distribution.

This conversion to a classical model may involve a loss of the information contained in the internal structure of the Structured Model.

The classical model ignores this internal structure. As a consequence it is possible to construct different Structured Models which correspond to the same classical model.

Example 2.1.3

Consider the Structured Model

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1, & 0 \\ \rho, & (1-\rho^2)^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \qquad (2.1.6)$$

with e_1, e_2 iid N(0,1)

where $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \varepsilon \mathbb{R}^2 = \mathfrak{X}$, $\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \varepsilon \mathbb{R}^2 = \Omega$

and $\left\{ \begin{pmatrix} 1, & 0 \\ \rho, & (1-\rho^2)^{\frac{1}{2}} \end{pmatrix} : \rho \varepsilon (-1,1) = \Theta \right\}$ a set of

transformations $\Omega \to \mathfrak{X}$

This model corresponds to the classical model

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right] \qquad (2.1.7)$$

Now x_1 and x_2 enter equation (2.1.7) symmetrically but equation (2.1.6) is not symmetric in x_1 and x_2, so if we exchange x_1 for x_2 in equation (2.1.6) we will obtain a Structured Model that is different from the original Structured Model but corresponds to the same classical model. □

2.2 Pivotal Models

We have defined the Structured Model and shown how it differs from the classical model. In this section we begin by defining upper and lower probabilities (Dempster 1966) on the set of transformations used in the Structured Model.

It is assumed throughout this section that we have no additional information about the model apart from observing an $x \varepsilon \mathfrak{X}$. If this is not the case then other methods of analysis might be more appropriate, e.g. if we had a prior distribution for θ then a Baysian analysis would be more appropriate.

Given an observation x and a Structured Model, the set of possible antecedents for $x \varepsilon \mathfrak{X}$ is

$$\Omega_x = \{e: e \varepsilon \Omega, \exists \theta \varepsilon \Theta \text{ s.t. } x = \theta o e\} \qquad (2.2.1)$$

that is to say, the subset of the sample space of e that can be transformed into the observed x. The set Ω_x might include the whole of the sample space Ω , but usually this is not the case so we restrict our attention to the set Ω_x. Since we know that the random variable must take its value on the set Ω_x we can limit the probability distribution on Ω to a conditional probability distribution on Ω_x.

To calculate the conditional probability distribution by conventional formulae we require either

a) that the probability given to the set Ω_x by $f(\cdot)$ be non-zero, or

b) that the sets $\{\Omega_x : x \in \mathfrak{X}\}$ form a partition of Ω where we partition Ω as follows.

Consider the set Ω_x of possible antecedents for $x \in \mathfrak{X}$ and the set of feasible observations from an $e \in \Omega$.

$$\mathfrak{X}_e = \{x:\ x \in \mathfrak{X}, \exists \theta \in \Theta \text{ s.t. } x = \theta o e\} \qquad (2.2.2)$$

then the sets Ω_x partition the space Ω.

$$\text{Iff } \forall e \in \Omega, \forall y,\ z \in \mathfrak{X}_e,\ \Omega_y = \Omega_z; \qquad (2.2.3)$$

that is to say iff each feasible observation $x \in \mathfrak{X}$, from any fixed element $e \in \Omega$, produces the same set of possible antecedents, namely Ω_x.

<u>Example 2.2.1</u> N(θ,1) model with two observations.

We consider the structured model

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \theta o \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad \text{with } e_1, e_2 \text{ iid } N(0,1)$$

where $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 = \mathfrak{X}$, $\underline{e} = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \in \mathbb{R}^2 = \Omega$ and $\theta \in \Theta = \mathbb{R}$

indexing a family of location transformations $\{\phi_\theta : \theta\varepsilon\mathbb{R}\} : \Omega\to\mathfrak{X}$ i.e. $\mathbb{R}^2\to\mathbb{R}^2$ defined to be

$$\theta o \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} \theta + e_1 \\ \theta + e_2 \end{pmatrix}$$

Now consider a fixed element $\underline{e}\varepsilon\Omega$. The set of feasible observations from $\underline{e}$, $\mathfrak{X}_{\underline{e}}$ takes the form

$$\mathfrak{X}_{\underline{e}} = \left\{ \begin{pmatrix} e_1 + \theta \\ e_2 + \theta \end{pmatrix} : \theta\varepsilon\mathbb{R} \right\}$$

Consider a general element from $\mathfrak{X}_{\underline{e}}$, $\underline{x} = \begin{pmatrix} e_1 + \theta_o \\ e_2 + \theta_o \end{pmatrix}$ $\theta_o\varepsilon\mathbb{R}$. This element will generate a set of possible antecedents $\Omega_{\underline{x}}$ which will take the form

$$\Omega_{\underline{x}} = \left\{ \begin{pmatrix} e_1 + \theta_o - \theta \\ e_2 + \theta_o - \theta \end{pmatrix} : \theta o,\ \theta\varepsilon\mathbb{R} \right\} \equiv \left\{ \begin{pmatrix} e_1 - \theta^* \\ e_2 - \theta^* \end{pmatrix} : \theta^*\varepsilon\mathbb{R} \right\}$$

One can easily see that the generated set of possible antecedents $\Omega_{\underline{x}}$ is independent of the element chosen from $\mathfrak{X}_{\underline{e}}$. Hence from (2.2.3) the sets $\Omega_{\underline{x}}$: $\underline{x}\varepsilon\mathfrak{X}$ partition the space Ω. These positions are a beam of straight lines parallel to the line $z_1 = z_2, (z_1, z_2)\varepsilon\mathbb{R}^2$. □

We will consider only those structured models where the conditional probability distribution on Ω_x, has been obtained by a partition. This is one of the restrictions on a structured model needed to obtain a Pivotal Model. The details of how to calculate the conditional probability on Ω_x for a Pivotal Model will be dealt with more fully later.

Having obtained the conditional distribution on Ω_x, we can proceed to obtain the upper and lower probabilities on Θ, the indexing set of the transformations.

Consider a set $C \subseteq \Theta$ (a measurable subset of Θ), that is to say, C is a subset of the possible indexes. Given that we have observed $x \in \mathfrak{X}$, we define the set of antecedents of x that are generated by the set $\{\phi_\theta\}$, $\theta \in C \subseteq \Theta$ as

$$\Omega_x(C) = \{e\colon e \in \Omega \text{ and for which } \theta \in C \text{ s.t. } x = \theta o e\} \tag{2.2.4}$$

Then $\Omega_x(C) \subseteq \Omega_x$ and one can see that the previous definition of Ω_x, the set of all antecedents of $x \in \mathfrak{X}$, coincides with the definition of $\Omega_x(\Theta)$.

We now define the upper probability of $C \subseteq \Theta$ given that we have observed $x \in \mathfrak{X}$, $P^+_x(C)$ say, to be the probability given to the set $\Omega_x(C)$ by the conditional probability distribution on Ω_x. We define the lower probability of $C \subseteq \Theta$ given that we have observed $x \in \mathfrak{X}$, $P^-_x(C)$ say to be

$$P^-_x(C) = 1 - P^+_x(\overline{C})$$

where $\overline{C}$ is the complement of C in Θ, that is to say, the set of indexes in Θ that are not in C.

<u>Example 2.2.2</u> (Example 2.1.2 continued) Bin(3,1-θ) model.

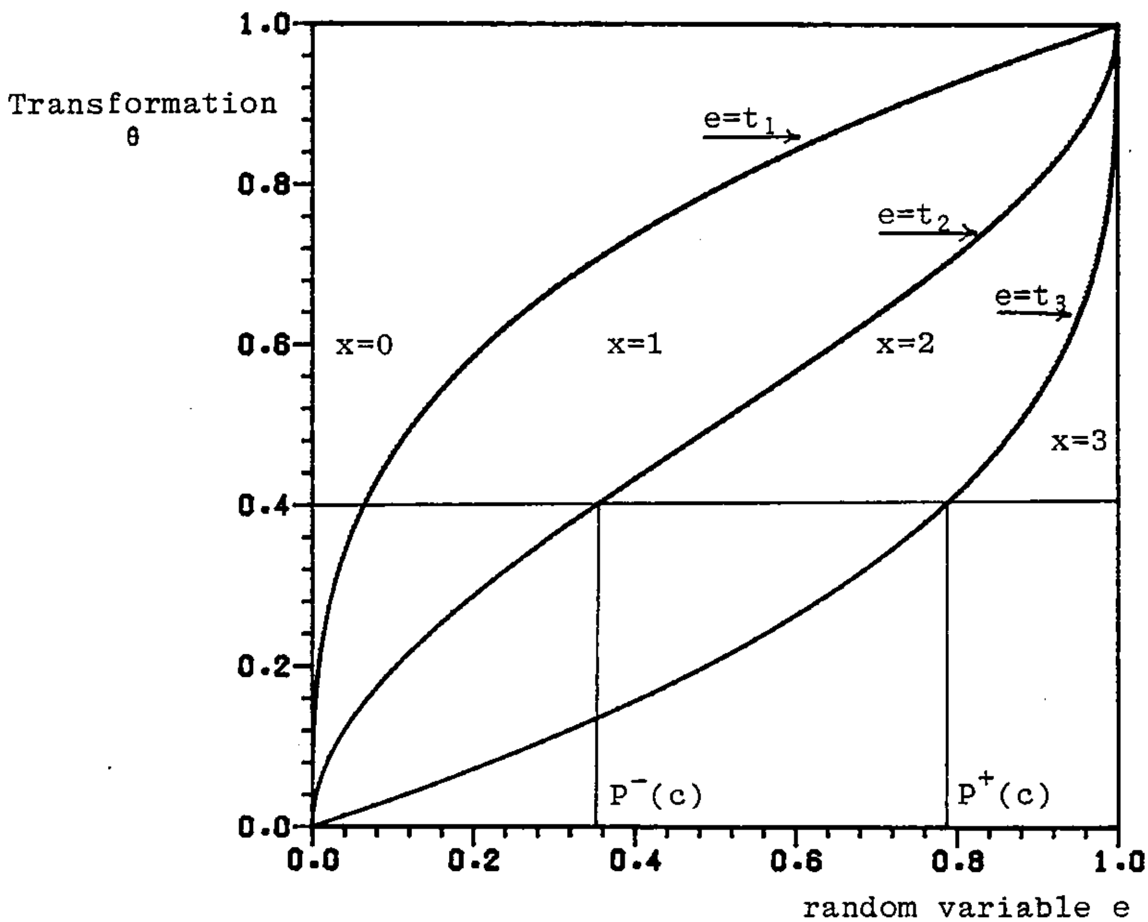
Consider the case when we have observed x and we are

Figure 2.1.1 Upper and Lower Probability Diagram for Example 2.2.2, Bin(3,1-θ)

The subset illustrated is $C = \{\theta : 0 \leqslant \theta \leqslant .4\} \subseteq \Theta$ when we observe x=2.

$P_2^+(C) = 0.784 = 1 - (1-0.4)^3$

$P_2^-(C) = 0.352 = (0.4)^3 + 3(0.4)^2(1-0.4)$



required to find the upper and lower probabilities for the set of indexes

$$C = \{\theta: 0 < \theta < .4\} \quad C \subset \Theta$$

The set of antecedents of each $x \varepsilon \mathfrak{X}$, Ω_x are all equal to the original space Ω , hence we have a trivial partition of Ω, namely the whole space Ω itself. As a result the conditional probability distribution will be the original probability distribution defined on Ω, namely

$$e \sim \text{rect}(0,1) \quad e \varepsilon \Omega = (0,1)$$

In the particular case in which we observe $x = 2$, the set of antecedents generated by the set $\{\phi_\theta\}$, $\theta \varepsilon C$ is easily shown to be

$$\Omega_2(C) = \{e: 0 < e < 0.784\}$$

where $0.784 = 1 - (1 - 0.4)^3$.

This is illustrated in figure 2.1.1. This set has probability 0.784, so the upper probability of C is just

$$P^+_2(C) = 0.784.$$

The complement of the set C is

$$\overline{C} = \{\theta: 0.4 < \theta < 1\}$$

The set of antecedents of $x = 2$ generated by $\{\phi_\theta\}$ $\theta\varepsilon\overline{C}$ can be shown to be

$$\Omega_2(\overline{C}) = \{e: 0.352 < e < 1\}$$

where $0.352 = (0.4)^3 + 3(0.4)^2(1 - 0.4)$.

This set has probability 0.648, so the lower probability of C is

$$P^-_2(C) = 1 - 0.648 = 0.352. \qquad \square$$

Example 2.2.3 (Example 2.1.1 continued) $N(\theta,1)$ model.

Consider the case when we have observed x and we are required to find the upper and lower probabilities for the set of indexes

$$C = \{\theta : 1 < \theta < 4.92\}$$

The set of antecedents of each $x\varepsilon\mathfrak{X}$, Ω_x, are all equal to the original space Ω. As a result the conditional probability distribution will be the original probability distribution defined on Ω, namely $e \sim N(0,1)$ $e\varepsilon\Omega = \mathbb{R}$.

In the particular case in which we observe $x = 2.96$, the set of antecedents generated by the set $\{\phi_\theta\}$ $\theta\varepsilon C$ can be easily shown to be

$$\Omega_{2.96}(C) = \{e: -1.96 < e < 1.96\}$$

This set has a probability of 0.9500, rounded to 4 dp so the upper probability of C will be

$$P^{+}_{2.96}(C) = 0.9500.$$

The complement of the set C is

$$\overline{C} = \{\theta: \theta \leq 1 \text{ or } \theta \geq 4.96\}$$

The set of antecedents of x = 2.96 generated by $\{\phi_\theta\}$ $\theta \varepsilon \overline{C}$ can be shown to be

$$\Omega_{2.96}(\overline{C}) = \{e: e \leq -1.96 \text{ or } e \geq 1.96\}$$

This set has a probability of 0.0500, rounded to 4 dp so the lower probability of C is

$$P^{-}_{2.96}(C) = 0.9500 \qquad \square$$

A more thorough discussion of upper and lower probabilities, together with some of their properties, is given in Dempster (1967). Their use in a decision theory framework can be found in Beran (1971), and although his upper and lower probabilities were obtained by a different mechanism than the one adopted here, his results still remain valid.

We now investigate the question: "When are the upper and lower probabilities equal to each other for all subsets of Θ, the indexing set of the transformations defined in the structured model?"

This is an important question because when they are we can define a probability distribution on Θ, the indexing set of the transformations.

The question of equal upper and lower probabilities was answered by Plante (1979) using the idea of additivity.

A structured model is called additive at the point $x \varepsilon \mathfrak{X}$ if and only if

$$\Omega_x(\{\theta_1\}) \cap \Omega_x(\{\theta_2\}) = \phi \qquad (2.2.5)$$

the empty set, whenever the elements θ_1 and θ_2 of Θ are different, that is to say the point $x \varepsilon \mathfrak{X}$ is additive if and only if for each random variable $e \varepsilon \Omega_x$ there is only one transformation ϕ_θ, $\theta \varepsilon \Theta$ that transforms it to x. A structured model is called additive if and only if it is additive at every point $x \varepsilon \mathfrak{X}$.

Plante in the above paper was then able to prove that the upper and lower probabilities are equal to each other for any (measurable) subset of Θ if and only if the structured model is additive.

<u>Example 2.2.4</u> (Example 2.2.2 continued) Bin(3, $1-\theta$) model.

Again consider the case in which we observe $x = 2$. The set of antecedents of $x = 2$ generated by ϕ_θ $\theta = 0.4$ can be shown to be

$$\Omega_2(\{0.4\}) = \{e\colon 0.352 < e < 0.784\}$$

and by ϕ_θ: $\theta = 0.5$ can be shown to be

$$\Omega_2(\{0.5\}) = \{e\colon 0.5 < e < 0.875\}$$

Hence the intersection of these two sets is

$$\Omega_2(\{0.4\}) \cap \Omega_2(\{0.5\}) = \{e\colon 0.5 < e < 0.784\}$$

which is non empty. This implies that the structured model is not additive, since it is not additive at the point $x = 2$. Hence the upper and lower probabilities need not be the same on a given subset of Θ. This fact was borne out in example 2.2.2 where an example was given of a subset of Θ that had different upper and lower probabilities. □

<u>Example 2.2.5</u> (Example 2.1.1 continued) $N(\theta,1)$ model.

Consider the case in which we have observed $x \varepsilon \mathfrak{X}$ the set of antecedents of x generated by an element ϕ_θ $\theta \varepsilon \Theta$ consists of the single point

$$\Omega_x(\theta) = \{e\colon e = x - \theta\}$$

To have a non empty intersection of two of the above sets for a fixed $x \varepsilon \mathfrak{X}$, we require

$$x - \theta_1 = x - \theta_2 \quad \theta_1, \theta_2 \varepsilon \Theta \quad \text{or } \theta_1 = \theta_2$$

This implies from condition (2.2.5) that the structured model is additive at the point x and as this is true for a general $x \in \mathfrak{X}$ the structured model is additive. Hence any (measurable) subset of Θ will have equal upper and lower probabilities which proved to be the case for the subset used in example 6. □

The additive and partition properties are restrictions on a structured model needed to obtain a Pivotal Model. The final restriction is the Pivotal condition (Dawid and Stone 1982).

A structured model satisfies the pivotal condition if and only if

$$x \in \mathfrak{X} \text{ and } \forall\, \theta \in \Theta\ \exists!\, e \in \Omega \text{ s.t. } x = \theta o e \qquad (2.2.6)$$

that is to say, for every observation $x \in \mathfrak{X}$ and transformation ϕ_θ: $\theta \in \Theta$, it is possible to solve $x = \theta o e$ for a unique $e \in \Omega$.

Notice that the pivotal condition implies that there exists a function T: $\mathfrak{X} \times \Theta \to \Omega$ such that

$$\text{if } x = \theta o e \text{ then } T(x,\theta) = e \qquad (2.2.7)$$

where $x \in \mathfrak{X}$, $\theta \in \Theta$ and $e \in \Omega$.

T is called the pivotal function. Since e has a known distribution on Ω, the pivotal function T corresponds to the classical definition of a pivot that is a function of the data and parameters that has a known distribution. Unlike classical models, where a model might have many maximal pivots a particular structured model that satisfies the pivotal condition will have only one maximal pivotal function.

<u>Example 2.2.6</u> (Example 2.1.1 continued) $N(\theta,1)$ model.

Given a random variable $e\varepsilon\Omega$ and a transformation ϕ_θ $\theta\varepsilon\Theta$, the observation generated is

$$x = \theta o e = \theta + e$$

By rearranging this we obtain $x - \theta = e$. The difference between the two real numbers is unique hence this example admits a pivotal function

$$T(x,\theta) = x - \theta$$

□

In summary: a Pivotal Model is a structured model with the following restrictions

a) The transformations ϕ_θ $\theta\varepsilon\Theta$ partition Ω the set of random variables.

b) The structured model is additive.

c) The structured model satisfies the pivotal condition.

Example 2.1.1, the N(θ,1) model, is an illustration of a pivotal model since it satisfies all three conditions (see examples 2.2.3, 2.2.5 and 2.2.6).

In the next section are developed some theoretical aspects of pivotal models.

2.3 Properties of Pivotal Models

In the last section we defined pivotal models by placing restrictions on structured models. In this section we begin to explore some of their properties. We start by introducing the idea of equivalent pivotal models. Given a structured model

$$x = \theta o e \tag{2.3.1}$$

where $x \varepsilon \mathfrak{X}$, $e \varepsilon \Omega$ and θ indexes a set of transformations $\phi_\theta : \Omega \to \mathfrak{X}$, $\theta \varepsilon \Theta$

Consider another observable sample space $\tilde{\mathfrak{X}}$ such that there exists a one to one mapping α: $\tilde{\mathfrak{X}} \to \mathfrak{X}$. If α is a fixed and known mapping then this induces a structured model of the form

$$\tilde{x} = \theta \tilde{o} e \tag{2.3.2}$$

where $\tilde{x} \varepsilon \tilde{\mathfrak{X}}$, $e \varepsilon \Omega$ and θ indexes a set of transformations $\tilde{\phi}_\theta = \alpha^{-1} o \phi_\theta : \Omega \to \tilde{\mathfrak{X}}$, $\theta \varepsilon \Theta$

We then say that the structured model given in (2.3.2) is an equivalent structured model to the one given in (2.3.1). This leads to the definition of an equivalent structured model as follows:

Given a structured model, a second structured model is said to be an equivalent structured model to the first if there exists a one to one transformation of its observable sample space that converts it into the first structured model.

One can easily show that the above relationship is an equivalence relation on the set of structured models. The following theorem indicates why we have pursued the definition of equivalent structured models; it is proved in the appendices A1.2.

Theorem 2.3.1

a) Given a pivotal model, then any equivalent structured model will also be a pivotal model.

b) Given a (measurable) subset $C \subseteq \Theta$ and an observation $x \in \mathfrak{X}$ then the probability induced by x on C will be the same as that induced by the corresponding point $\tilde{x} \in \tilde{\mathfrak{X}}$ in the equivalent pivotal model on C.

This is an important theorem because it means that if we have a pivotal model we can take any equivalent pivotal model and still obtain the same probability statements about the (measurable) subsets of Θ. Hence if an equivalent pivotal model simplifies a particular pivotal model then we are allowed to consider it instead.

Corollary 2.3.1

Any pivotal model has an equivalent pivotal model with the following properties:

a) The observable sample space $\tilde{\mathfrak{X}}$ and the random variable sample space Ω are the same, that is $\tilde{\mathfrak{X}} = \Omega$.

b) One of the transformations $\tilde{\phi}_i$, $i\varepsilon\Theta$ say, is the identity transformation $\Omega \to \Omega$.

Proof

This can be achieved by taking the one to one mapping $\alpha : \tilde{\mathfrak{X}} = \Omega \to \mathfrak{X}$ to be a fixed transformation from the set $\{\phi_\theta , \theta\varepsilon\Theta\}$, $\alpha = \phi_i$ $i\varepsilon\Theta$ say. Then $\tilde{\phi}_\theta = \phi_i{}^{-1} o \phi_\theta : \Omega \to \Omega$ with $\theta\varepsilon\Theta$. When $\theta = i\varepsilon\Theta$, $\tilde{\phi}_i$ will be the identity transformation. Note from the definition of a pivotal model that all transformations $\phi_\theta, \theta\varepsilon\Theta$ will be one to one mappings of Ω onto $\mathfrak{X}$.

Example 2.3.1

Consider the pivtal model

$$x = \theta o e \qquad e \sim N(0,1)$$

with $x, e \varepsilon \Omega = \mathbb{R}$ where $\theta\varepsilon\Theta = \mathbb{R}$ indexes a set of transformations $\phi_\theta : \mathbb{R} \to \mathbb{R}$, $\theta\varepsilon\Theta$ defined to be

$$\theta o e = (\theta + e)^3 \qquad \theta\varepsilon\Theta \ , \ e\varepsilon\Omega$$

Note that the observed sample space and the random variable sample space are the same but that there is no identity element amongst the set of transformations $\phi_\theta : \theta\varepsilon\Theta$.

By taking the one to one mapping $\alpha = \phi_0 : \Omega\to\Omega$ we can obtain an equivalent pivotal model that takes the form

$$x = \theta \tilde{o} e \qquad e \sim N(0,1)$$

with x, $e\varepsilon\Omega = \mathbb{R}$ where $\theta\varepsilon\Theta = \mathbb{R}$ indexes a set of transformations $\tilde{\phi}_\theta = \phi_0^{-1} o \phi_\theta : \Omega\to\Omega$, $\theta\varepsilon\Theta$ defined to be

$$\theta \tilde{o} e = ((\theta+e)^3 - 0)^{1/3} = \theta+e \qquad \theta\varepsilon\Theta \ , \ e\varepsilon\Omega$$

This equivalent pivotal model now has the desired properties namely the observable sample space and the random variable sample space are the same and the transformation $\tilde{\phi}_0 : 0\varepsilon\Theta = \mathbb{R}$ is the identity element.

Note that the equivalent pivotal model is the one described in example 2.1.1 namely the $N(\theta,1)$ model. □

It will be assumed throughout the rest of this thesis, unless otherwise stated, that for a given pivotal model the above simplification has already been carried out.

The simplification enables us to rewrite the restrictions (a) - (c) given in section 2.2, on structured models to obtain pivotal models, in an equivalent form that makes it easier to explore the properties of pivotal models.

It can easily be shown that the following are equivalent to the above named restrictions and a brief outline is given afterwards.

Consider a structured model

$$x = \theta o e \quad \text{with } x,\ e \varepsilon \Omega \text{ , a sample space.}$$

e has a probability distribution f(·) say and $\theta \varepsilon \Theta$ indexes a set of transformations

$$\phi_\theta : \Omega \to \Omega \ , \ \theta \varepsilon \Theta \quad \text{with the following properties:}$$

P1. If h, $g \varepsilon \Theta$, $e \varepsilon \Omega$ then hoe = goe $\Rightarrow$ h = g

P2. $\forall$ h, $g \varepsilon \Theta$, $\forall\, e \varepsilon \Omega$, let z = go(hoe) then
$\exists\, h^* \varepsilon \Theta$ s.t. z = h*oe

P3. $\exists\, i \varepsilon \Theta$ s.t. $\forall\, e \varepsilon \Omega$ ioe = e

P4. $\forall\, h_1$, $h_2 \varepsilon \Theta$ $\forall\, e \varepsilon \Omega$ $\exists\, g \varepsilon \Theta$ s.t. $go(h_1 oe) = h_2 oe$

P5. $\forall\, z \varepsilon \Omega$, $\forall\, g \varepsilon \Theta$ $\exists!\, e \varepsilon \Omega$ s.t. z = goe

Note since we have $\mathfrak{X} = \Omega$ we can simplify the partition condition (2.2.3) into an equivalent definition as follows:

$$\forall e \varepsilon \Omega \quad \forall y \varepsilon\ \mathfrak{X}_e \quad \mathfrak{X}_y = \mathfrak{X}_e$$

P1 is the additivity condition in a slightly different form.

P2 is obtained by applying the simplified partition condition as follows:

Given $e \varepsilon \Omega$ we have $hoe \; \varepsilon \; \mathfrak{X}_e$. so by the simplified partition condition we have

$$\mathfrak{X}_{hoe} = \mathfrak{X}_e$$

but $z = go(hoe) \; \varepsilon \; \mathfrak{X}_{hoe}$ so $z \; \varepsilon \; \mathfrak{X}_e$ hence by the definition of $\mathfrak{X}_e$ (2.2.2) there exists $h^* \varepsilon \Theta$ s.t. $z = h^*oe$ hence P2.

P3 specifies the existence of an identity transformation amongst the set $\phi_\theta \;\; \theta \varepsilon \Theta$

P4 is obtained by applying the simplified pivotal condition twice. Given $e \varepsilon \Omega$ we have $h_k oe \; \varepsilon \; \mathfrak{X}_e$, $k = 1,2$; so by the simplified pivotal condition we have $\mathfrak{X}_{h_1 oe} = \mathfrak{X}_e = \mathfrak{X}_{h_2 oe}$; but by P3 we have $h_2 oe \; \varepsilon \; \mathfrak{X}_{h_2 oe}$ and so $h_2 oe \; \varepsilon \; \mathfrak{X}_{h_1 oe}$. Hence by the definition of $\mathfrak{X}e$ (2.2.2) we have that there exists $g \varepsilon \Theta$ s.t. $go(h_1 oe) = h_2 oe$. Hence P4.

P5 is the pivotal condition in a slightly different form.

Having defined the above properties we now consider obtaining the partitions Ω_x of Ω and the probability distribution of e conditional on $e \varepsilon \Omega_x$.

Consider the set

$$\Theta_x = \{\theta o x : \theta \varepsilon \Theta\} \quad x \varepsilon \Omega \qquad (2.3.3)$$

that is to say, the set of points that it is possible to transform x to, by the set of transformations $\{\phi_\theta : \theta \varepsilon \Theta\}$

<u>Lemma 2.3.2</u>

The sets $\{\Theta_x\}$ $x\varepsilon\Omega$ partition Ω, the same partition as $\{\Omega_x\}$ $x\varepsilon\Omega$

This lemma is proved in appendix A1.3.

To obtain the probability distribution of e conditional on $e\varepsilon\Theta_x$ (= Ω_x) a particular partition, we first have to transform the random variables into two variables; the first called the reference variable and denoted by D(e), describes the partition that a particular $e\varepsilon\Omega$ is in; the second, called the transformation variable and denoted by [e], describes the position it takes on that partition. From the joint distribution of the two variables D(e) and [e] we can obtain the probability distribution of the transformation variable [e] conditional on the reference variable D(e) and this conditional distribution will be the required probability distribution used to obtain probability statements concerning θ. The details of this argument are as follows:

Let the set $Q \subseteq \Omega$ be a cross-section to Ω/Θ, that is to say, the set Q has one point from each partition $\{\Theta_x : x \in \Omega\}$.

We define a reference variable $D(\cdot) : \Omega \to \Omega$ as follows:

$$\text{Let } D(e) = \Theta_e \cap Q \quad e \in \Omega \tag{2.3.4}$$

that is to say $D(e)$ $e\theta\Omega$ is the set of points that are in both the same partition as $e \in \Omega$ and the set Q; but from the definition of Q, the set $D(e)$ $e \in \Omega$ will consist only of one point from the sample space, so $D(e)$ $e \in \Omega$ is well defined.

<u>Lemma 2.3.3</u>

Let $D(e)$ $e \in \Omega$ be a reference variable as defined in (2.3.4) then $e \in \Omega$ $\theta \in \Theta$ $D(e) = D(\theta o e) \in \Theta e$.

<u>Proof</u>:

From the definition (2.3.3) of Θe, the point $\theta o e \in \Theta e$. Hence from Lemma 2.3.2 we have $\Theta_{\theta o e} = \Theta e$, that is to say, the point $\theta o e \in \Omega$ generates the same partition as the point $e \in \Omega$. Hence from the definition (2.3.4) of $D(e)$, we have $D(\theta o e) = D(e)$. Also from the definition of $D(e)$ we have $D(e) \in \Theta e$. Hence the Lemma is established. //

This tells us that all points in the same partition have the same reference variable, and, because this reference variable is a member of that partition, different partitions have different reference variables. We can therefore use the reference variable to label the partitions. We define the transformation variable $[\cdot] : \Omega \to \Theta$ as follows. Let $[e] \varepsilon \Theta$, $e \varepsilon \Omega$, be such that

$$[e] \circ D(e) = e \qquad (2.3.5)$$

for a given reference variable; that is to say, let $\phi_{[e]}$, $e \varepsilon \Omega$, $[e] \varepsilon \Theta$, be the unique element of the set $\{\phi_\theta : \theta \varepsilon \Theta\}$ that transforms D(e) back to e. Note existence is guaranteed by the fact that $e \varepsilon \Theta_{D(e)}$, and uniqueness by P1.

We can now define a binary operation $*: \Theta \times \Theta \to \Theta$ conditional on the reference point D(e) for each partition as follows:

If θ_1, $\theta_2 \varepsilon \Theta$ then by P2 there exists $\theta_{12} \varepsilon \Theta$ such that $\theta_1 \circ \{\theta_2 \circ D(e)\} = \theta_{12} \circ D(e)$. We use this fact to define $\theta_1 * \theta_2 = \theta_{12}$ where $*$ is conditional on D(e). The reason for this definition is given in the following Lemma.

<u>Lemma 2.3.4</u>

Let $[e] \varepsilon \Theta$ be a transformation variable, then $\forall\, \theta \varepsilon \Theta$ we have $\theta * [e] = [\theta \circ e]$ with $*$ conditional on D(e).

<u>Proof</u>:

$$\{\theta * [e]\} \circ D(e) = \theta \circ \{[e] \circ D(e)\} \quad \text{by definition}$$

$$= \theta \circ e = [\theta \circ e] \circ D(\theta \circ e) \quad \text{by (2.3.5)}$$

$= [\theta o e] \text{ o } D(e)$ by Lemma 2.3.3.

Hence $\theta * [e] = [\theta o e]$ by P1 with * conditional on D(e).

Hence Lemma. //

We can now show that it is possible to decompose a pivotal model $x = \theta o e$ where $x, e \varepsilon \Omega, \theta \varepsilon \Theta$ as follows. From Lemma 2.3.3 we have $D(x) = D(e)$ and from Lemma 2.3.4 $[x] = \theta * [e]$ with * conditional on D(x). Since we observe $x \varepsilon \Omega$ and hence know the reference point D(x), the binary operation, $*: \Theta \times \Theta \to \Theta$ defined above, is well defined. This particular decomposition will be called the Reduced Pivotal Model.

Example 2.3.2 (example 2.2.1 continued)

$N(\theta,1)$ model with two observations.

It is easily shown that this example is a pivotal model of the required form. It was demonstrated in example 2.2.1 that the partitions form a beam of straight lines parallel to the line $z_1 = z_2$ $(z_1,z_2) \varepsilon \Omega = \mathbb{R}^2$.
We can take Q to be the set

$$Q = \{(z_1,z_2) : z_1 = 0, (z_1,z_2) \varepsilon \mathbb{R}^2\}$$

that is to say Q is the line $z_1 = 0$ in $\mathbb{R}^2$. It is easily verified that the set Q is a cross-section of the partitions.

Given an observation $\underline{x} = (x_1,x_2) \varepsilon \Omega = \mathbb{R}^2$, the partition that $\underline{x}$ belongs to is the set

$$\Theta_{\underline{x}} = \{(x_1 - \theta, x_2 - \theta) : \theta \varepsilon \Theta = \mathbb{R}\}$$

hence the reference variable is the point

$$\begin{aligned} D(x_1,x_2) &= \{(x_1 - \theta, x_2 - \theta) : \theta \varepsilon \mathbb{R}\} \cap \\ &\quad \{(z_1,z_2) : z_1 = 0, (z_1,z_2) \varepsilon \mathbb{R}^2\} \\ &= \{(0, x_2 - x_1)\} \end{aligned}$$

From this the transformation variable is easily shown to be $[(x_1,x_2)] = x_1 \varepsilon \Theta = \mathbb{R}$ since the transformation ϕ_{x_1}, $x_1 \varepsilon \Theta$ transforms the point $(0, x_2 - x_1) \varepsilon \Omega$ back to the point (x_1,x_2).

We can now examine the binary operation $*: \Theta \times \Theta \rightarrow \Theta$ conditional on the point $(0, x_2 - x_1)$. Consider $\theta_1, \theta_2 \varepsilon \Theta = \mathbb{R}$ then from the definition of the binary operation we have

$$\begin{aligned} (\theta_1 * \theta_2) o (0, x_2 - x_1) &= \theta_1 o \{\theta_2 \, o (0, x_2 - x_1)\} \\ &= \theta_1 o (\theta_2, \theta_2 + x_2 - x_1) \\ &= (\theta_1 + \theta_2, \theta_1 + \theta_2 + x_2 - x_1) \\ &= (\theta_1 + \theta_2) o (0, x_2 - x_1) \end{aligned}$$

Thus we have $\theta_1 * \theta_2 = \theta_1 + \theta_2$ with $*$ conditional on $D(\underline{x})$. From this we can write the pivotal model in a reduced pivotal model form as follows

$$(0, x_2 - x_1) = D(\underline{x}) = D(\underline{e}) = (0, e_2 - e_1) \quad \text{and}$$

$$x_1 = [(x_1, x_2)] = \theta*[(e_1, e_2)] = \theta + e_1$$

with * conditional on $D(\underline{x}) = (0, x_2 - x_1)$. □

Before we calculate the joint probability distribution of the reference and transformation variables, we look at some of the properties of the binary operation $*: \Theta \times \Theta \to \Theta$ conditional on the point D(x), defined above.

<u>Theorem 2.3.2</u>

The binary operation $*: \Theta \times \Theta \to \Theta$ conditional on the point D(x) has the algebraic structure called "a loop". That is to say, it has the following properties:

<u>L1</u> $\forall \theta_1, \theta_2 \varepsilon \Theta$ then $\theta_1 * \theta_2 \varepsilon \Theta$

i.e. Θ is closed under the operation *.

<u>L2</u> $\exists i \varepsilon \Theta$ s.t. $\forall \theta \varepsilon \Theta$ $i*\theta = \theta = \theta*i$

i.e. Θ has an identity element, namely i.

<u>L3</u> $\forall \theta_1, \theta_2 \Theta$ $\exists!$ $h_1, h_2 \varepsilon \Theta$ s.t.

$\theta_1 * h_1 = \theta_2$, $h_2 * \theta_1 = \theta_2$

This theorem will be proved in appendix A1.4.

Two further binary operations $\Theta \times \Theta \to \Theta$ associated with a loop are left division \ and right division / defined as follows:

If $k = h*g$ $(k, h, g \varepsilon \Theta$ with $*$ conditional on $D(x))$ then $h \backslash k = g$ and $k/g = h$.

Note these binary operations are both well defined being the solutions of the equations given in L3.

The three binary operations $*, \backslash, /$ defined on Θ together form an algebra that enables us to manipulate the transformation variables with the transformations ϕ_θ, $\theta \varepsilon \Theta$. Further properties of loops are given in appendix 2 and Bruck (1971).

We now return to the question of obtaining the joint probability distribution of the reference and transformation variables. This will be achieved by using "loop-invariant" differentials.

The methods assume that the spaces Ω and Θ are locally compact topological spaces and the transformations $\theta_1 = g*h$, $\theta_2 = g\backslash h$, $\theta_3 = g/h$, $x = ho(goe)$, where $*$, $\backslash$, and $/$ are conditional on the point $D(x)$, are continuously differentiable with respect to $g, h \varepsilon \Theta$ and $e \varepsilon \Theta$, with $[e]$ also a continuous transformation. The derivatives are the appropriate Radon-Nikodym derivatives relative to the given measures on Ω and Θ; these measures are also assumed to be totally σ - finite on Ω and Θ .

Consider an element V at a point $z \varepsilon \Omega$. A transformation ϕ_θ $\theta \varepsilon \Theta$ applied to the point $z \varepsilon \Omega$ will change the "volume" of V by the positive Jacobian factor

$$J_{\Omega}(\theta:z), = \left|\frac{\partial(\theta o z)}{\partial(z)}\right| \quad \text{say.} \tag{2.3.6}$$

i.e. $d(\theta o z) = J_{\Omega}(\theta:z)dz$

The existence of the Jacobian, with the above property is assured by Theorem D §39 in Halmos (1950).
This Jacobian can be used to produce a loop-invariant differential dm(z) on Ω, where

$$dm(z) = \frac{dz}{J_{\Omega}([z]:D(z))} \tag{2.3.7}$$

Loop-invariant differentials on a space are a measure of "volume" that remain constant under any transformation of the space, from a set of possible transformations that are in a 1-1 correspondence with a loop structure.
It is shown in appendix A2.3 that the differential defined in (2.3.7) is indeed a loop-invariant differential on Ω, that is to say

$$\forall z \varepsilon \Omega \quad \forall \theta \varepsilon \Theta \qquad dm(\theta o z) = dm(z) \tag{2.3.8}$$

The transformations $\{\phi_{\theta} : \theta \varepsilon \Theta\}$ on the reference and transformations variables do not affect the reference variable by definition so any differential based only on the reference variable will automatically be a loop-invariant differential. The transformations affect only the transformation variables via the binary operation $*: \Theta \times \Theta \to \Theta$.

Consider an element V at a point $[z]\varepsilon\Theta$. A transformation ϕ_θ $\theta\varepsilon\Theta$ applied to the reference variable [z] will change the "volume" of V by the positive Jacobian factor

$$J_*(\theta:\ [z]\,|D(z)) = \left|\frac{\partial(\theta*[z])}{\partial([z])}\right| \quad \text{say} \qquad (2.3.9)$$

where * is conditional on D(z).

Again the existence of the Jacobian is assured by Theorem D §39 in Halmos (1950). This Jacobian can also be used to produce a loop-invariant differential $d\mu(\theta)$ on Θ as follows:

$$d\mu(\theta) = \frac{d\theta}{J_*(\theta:\ i\,|D(z))} \qquad (2.3.10)$$

where i is the identity element of the loop.

This differential is shown in appendix A2.3 to indeed be a loop-invariant differential on Θ i.e.

$$\forall g,h\varepsilon\Theta \qquad d\mu(g*h) = d\mu(h) \qquad (2.3.11)$$

where * is conditional on D(z).

Now consider the two invariant differentials as they apply to the element V and to images of V under the transformations from $\{\phi_\theta:\ \theta\varepsilon\Theta\}$ at the reference point D(z).

Let $\delta(D(z))$ be the ratio

$$dm(z) = \frac{dz}{J_\Omega([z]|D(z))} = \delta(D(z)) \frac{d[z]}{J_*([z],i|D(z))}$$

$$= \delta(D(z)) \, d\mu([z])$$

Since both differentials are invariant under the transformation from $\{\phi_\theta : \theta\varepsilon\Theta\}$ the equality holds throughout the orbit; thus we obtain

$$dz = \delta(D(z)) \frac{J_\Omega([z]|D(z))}{J_*([z],i|D(z))} \, d[z] \qquad (2.3.12)$$

The joint distribution of D(e), = D say and [e] $e\varepsilon\Omega$ can now be derived.

Let the random variable $e\varepsilon\Omega$ have a generalised probability density function f(.) say.

Note we might have to use the appropriate Radon-Nikodym derivative to obtain this.

Then from (2.3.12) we obtain

$$f(e)de = f([e]oD)\,\delta(D) \frac{J_\Omega([e]|D)}{J_*([e],i|D)} \, d[e]dD \qquad (2.3.13)$$

The right hand side of (2.3.13) is the joint probability density function of the reference and transformation variables D(e) and [e]. From this the conditional probability density function of [e] given D = D(e), g([e]|D) say, is easily calculated to be

$$g([e]|D)d[e] = k(D)\ f([e]oD)\ \frac{J_{\Omega}([e]|D)}{J_{*}([e],i|D)}\ d[e]$$

(2.3.14)

$$\text{where } \frac{1}{k(D)} = \int_{[e]\varepsilon\Theta} f([e]oD)\ \frac{J_{\Omega}([e]|D)}{J_{*}([e],i|D)}\ d[e]$$

the constant of proportionality.

The reduced pivotal model can now be written as

$$\begin{aligned} D(x) &= D(e) = D \\ [x] &= \theta * [e],\ *\ \text{conditional on } D \\ [e] &\sim g([e]|D) \end{aligned} \qquad (2.3.15)$$

with $x,e\varepsilon\Omega$, $\theta\varepsilon\Theta$

Example 2.3.3 (example 2.3.2 continued)

$N(\theta,1)$ model with two observations.

In example 2.3.2 it was shown that, given the particular reference variable defined there, the binary operation $*: \Theta \times \Theta \to \Theta$ is of the form $\theta_1 * \theta_2 = \theta_1 + \theta_2$ $\theta_1, \theta_2 \varepsilon \Theta$ =

From the definitions of the binary operations \ and / it can easily be shown that

$$\theta_1 \backslash \theta_2 = \theta_2 - \theta_1 \quad \text{and} \quad \theta_1 / \theta_2 = \theta_1 - \theta_2$$

Hence the Jacobian factors given in (2.3.6) and (2.3.9) are calculated to be

$$J_{\Omega}(\theta:\underline{z}) = \left|\frac{\partial(\theta o \underline{z})}{\partial(\underline{z})}\right| = \left|\frac{\partial(\theta+z_1, \theta+z_2)}{\partial(z_1, z_2)}\right| = 1$$

and $J_*(\theta:[z]|D(z)) = \left|\frac{\partial(\theta*[z])}{\partial([z])}\right| = \left|\frac{\partial(\theta+[z])}{\partial([z])}\right| = 1$

This enables us to calculate the conditional probability density function of $[\underline{e}]$ given $D(\underline{e}) = \begin{Bmatrix} d_1 \\ d_2 \end{Bmatrix} = D$

$$g([\underline{e}]|D) \propto \exp(-\tfrac{1}{2}\{([\underline{e}]+d_1)^2 + ([\underline{e}]+d_2)^2\})$$

$$\propto \exp(\frac{-([e]-\mu)^2}{2\sigma^2})$$

where $\mu = \frac{e_1 - e_2}{2}$ and $\sigma^2 = \tfrac{1}{2}$, since $\begin{Bmatrix} d_1 \\ d_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ e_2 - e_1 \end{Bmatrix}$

Hence the reduced pivotal model can be written as

$$D(\underline{x}) = \begin{Bmatrix} 0 \\ x_2 - x_1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ e_2 - e_1 \end{Bmatrix} = D(\underline{e}) = D$$

$$x_1 = \theta + e_1$$

$$[\underline{e}] = e_1 \qquad g([\underline{e}]|D) = N(\frac{e_1 - e_2}{2}, \tfrac{1}{2})$$ □

We can now proceed to obtain a confidence kernel on the indexing set. From this confidence kernel we can obtain probability statements about θ.

Given the reduced pivotal model

$$D(x) = D(e) = D$$

$$[x] = \theta*[e], \text{ * conditional on } D$$

$[e] \sim g([e]|D)$ with $x,e\varepsilon\Omega$, $\theta\varepsilon\Theta$

The value of [x] in the reduced pivotal model is known, so every value of the unknown [e] corresponds to a possible value for θ.

$$\theta = [x]/[e] \qquad [e] = \theta\backslash[x] \tag{2.3.16}$$

Thus a probability statement about [e] will also be a probability statement about θ. Hence the conditional probability density function of [e] given $g([e]|D)$ will induce a probability density function of θ given [x] and D called the Pivotal Distribution of θ given [x] and D.

Before we can calculate the pivotal distribution of θ given [x] and D, we need some further results about loop-invariant differentials.

Consider an element V at a point $\theta\varepsilon\Theta$. If we multiply θ by $h\varepsilon\Theta$ on the right hand side by the binary operation *. This will change the "volume" of V by the positive Jacobian factor

$$J_{*R}(\theta: h|D) = \left|\frac{\partial(h*\theta)}{\partial(h)}\right| \quad \text{say} \tag{2.3.17}$$

where * is conditional on D.

Again the existence of the Jacobian is assured by theorem D §39 in Halmos (1950).

The Jacobian can be used to produce a right loop-invariant differential $d\nu(\theta)$ on Θ as follows:

$$d\nu(\theta) = \frac{d\theta}{J_{*R}(\theta: i|D)} \qquad (2.3.18)$$

where i is the identity element of the loop.

This differential is shown in appendix A2.3 to be a right loop-invariant differential on Θ, i.e.

$$\forall g,h \in \Theta \qquad d\nu(h*g) = d\nu(h) \qquad (2.3.19)$$

where * is conditional on D.

Also we define $\Delta(\theta)$ to be the ratio of the right and left invariant differentials, i.e.

$$d\mu(\theta) = \Delta(\theta)\, d\nu(\theta) \qquad (2.3.20)$$

Hence by definition

$$\Delta(\theta) = \frac{J_{*R}(\theta, i|D)}{J_{*}(\theta, i|D)} \qquad (2.3.21)$$

The result that is required to calculate the pivotal distribution of θ given [x] and D is

$$d\mu(\theta\backslash h) = \Delta(\theta\backslash h)\, d\mu(\theta) \qquad (2.3.22)$$

This result is proved in appendix A2.4.

Hence the pivotal distribution of θ given [x] and D can be calculated to be:

$$g([e]|D)d[e] = K(D)f([e]oD)J_\Omega([e]|D)d\mu([e])$$

$$= k(D)f(\{\theta\backslash[x]\}oD)J_\Omega(\theta\backslash[x]|D)d\mu(\theta\backslash[x]) \text{ by (2.3.16)}$$

$$= k(D)f(\theta^{-1}ox)J_\Omega(\theta\backslash[x]|D)\Delta(\theta\backslash[x])d\mu(\theta) \text{ by (2.3.22)}$$

$$= k(D)f(\theta^{-1}ox)J_\Omega(\theta\backslash[x]|D)\frac{J_{*R}(\theta\backslash[x],i|D)d\theta}{J_*(\theta\backslash[x],i|D)J_*(\theta,i|D)} \quad (2.3.23)$$

by (2.3.21) and (2.3.10)

That is to say the pivotal distribution of θ given [x] and D is $p(\theta|D,[x])d\theta$

$$k(D)f(\theta^{-1}ox)J_\Omega(\theta\backslash[x]|D)\frac{J_{*R}(\theta\backslash[x],i|D)d\theta}{J_*(\theta\backslash[x],i|D)J_*(\theta,i|D)} \quad (2.3.24)$$

The question now remains about what happens if we had chosen a different reference variable. Do we still obtain the same pivotal distribution of θ given [x]' and D'(x), where D'(x) and [x]' are the new reference variable and its associated transformation variable? The answer is yes and this is proved in appendix A1.5.

We have now proved that the pivotal distribution of $\theta\varepsilon\Theta$ is unique given a value of $x\varepsilon\mathfrak{X}$ and we can thus simplify the expression of the confidence kernel by taking the reference point D to be x itself thus obtaining

$$p(\theta|x)d\theta =$$
$$k(x)f(\theta^{-1}ox)J_{\Omega}(\theta\backslash i\,|x)\frac{J_{*R}(\theta\backslash i,\ i\,|x)}{J_{*}(\theta\backslash i, i\,|x)}\ \frac{d\theta}{J_{*}(\theta, i\,|x)} \tag{2.3.25}$$

Example 2.3.4 (example 2.3.3 continued)

$N(\theta,1)$ model with two observations.

In example 2.3.3 it was shown that the reduced pivotal model takes the form

$$D(\underline{x}) = \begin{Bmatrix} 0 \\ x_2-x_1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ e_2-e_1 \end{Bmatrix} = D(\underline{e}) = D$$

$$x_1 = \theta + e_1$$

$$e_1 \sim N\left(\frac{e_1-e_2}{2}\ ,\tfrac{1}{2}\right)$$

The Jacobian factor given in (2.3.17) can be calculated to be

$$J_{*R}(\theta:\ h|D) = \left|\frac{\partial(h*\theta)}{\partial(h)}\right| = \left|\frac{\partial(h+\theta)}{\partial(h)}\right| = 1$$

Using this in 2.3.25 we obtain the pivotal distribution of θ given $\underline{x}$ to be

$$p(\theta|\underline{x}) \propto \exp\{-\tfrac{1}{2}((\theta+x_1)^2+(\theta+x_2)^2)\}$$

$$\propto \exp\{-\tfrac{1}{2}\frac{(\theta-\overline{x})^2}{\sigma^2}\}$$

where $\overline{x} = \dfrac{x_1+x_2}{2}$ $\quad \sigma^2 = \tfrac{1}{2}$, i.e.

$\theta \sim N(\overline{x}, \frac{1}{2})$ □

The pivotal models developed here are an extension of the structural models developed by Fraser (1968). In the structural models the binary operation $*: \Theta \times \Theta \to \Theta$ used in this section has the algebraic property of a group, which is a loop with the additional property of associativity, i.e.

$$\forall a,b,c \in \Theta \quad a*(b*c) = (a*b)*c$$

In all the examples of pivotal models so far presented, the loop induced is also a group. The following example results in a loop that is not a group.

Example 2.3.5

Let Θ and Ω be the set $\{0,1,2,3,4\}$. Let the distribution P defined on Ω be $\{P_\omega\}$ where $P_\omega = P(\omega \varepsilon \Omega)$ ($\omega = 0,1,\ldots,4$); and consider the family of transformations $\{\phi_\theta: \Omega \to \Omega \ , \ \theta \varepsilon \Theta\}$ defined as follows:

	ϕ_θ	Ω: 0	1	2	3	4
Θ	0	0	1	2	3	4
	1	1	4	3	2	0
	2	2	0	1	4	3
	3	3	2	4	0	1
	4	4	3	0	1	2

e.g. ϕ_1 takes $2 \to 3$

It can easily be shown that the model

$$x = \theta o e \quad e \sim P$$

where $x, e \varepsilon \Omega \ \theta \varepsilon \Theta$ is a pivotal model and satisfies properties P1 to P5. Consider the case where we have observed x = 2 and we wish to calculate the pivotal distribution of θ given x = 2.

The partition Ω_x is the trivial partition namely the whole space Ω itself. Without loss of generality we take the reference point D to be x itself.

This induces the loop *

*	0	1	2	3	4
0	0	1	2	3	4
1	1	0	3	4	2
2	2	3	4	1	0
3	3	4	0	2	1
4	4	2	1	0	3

In the appendices A2.2 it is shown that this particular loop is not a group.

The Jacobian factors given in (2.3.6), (2.3.9) and (2.3.17) are calculated to be

$$J_\Omega(\theta : D) = J_*(\theta, h : D) = J_{*R}(\theta : hD) = 1$$

Thus from (2.3.25) we obtain the pivotal distribution of θ given x = 2 to be

θ	0	1	2	3	4
$P(\theta\|x=2)$	p_2	p_3	p_0	p_1	p_4

□

Structural models are obtained if we replace property P2 by the stronger property P2' defined as follows.

<u>P2'</u> $\forall h,g\varepsilon\Theta$ $\exists h^{**}\varepsilon\Theta$ s.t. $\forall e\varepsilon\Omega$ let $z = go(hoe)$ then $z = h^{**}oe$

That is to say the element $h^{**}\varepsilon\Theta$ with the above defined property depends only on $h,g\varepsilon\Omega$ and is independent of the choice of $e\varepsilon\Omega$, unlike the element $h^*\varepsilon\Theta$ defined in property P2 which also could depend on the choice of $e\varepsilon\Omega$. Hence P2'$\Rightarrow$P2.

It is shown in Bruck (1971) that if a loop has the additional property of associativity then it is a group. Theorem 2.3.2 has already shown us that the induced binary operation $*: \Theta \times \Theta \to \Theta$ has the properties of a loop. Hence to prove that this binary operation is a group, when we replace property P2 by property P2', we need only show that the induced binary operation also has the property of associativity.

Consider any three elements h_1, h_2, $h_3\varepsilon\Theta$ and $e\varepsilon\Omega$

Now $h_1o(h_2o(h_3oe))$

$= h_1o(h_{23}oe)$ $\qquad$ h_{23} exists by P2'

$= h_{1,23}oe$ $\qquad$ $h_{1,23}$ exists by P2'

$= \{h_1*(h_2*h_3)\}oe$ $\qquad$ where * conditional on e but $h_1 o(h_2 o(h_3 oe))$

$= h_1 o(h_2 oz)$ where $z = h_3 oe$

$= h_{12} oz$ h_{12} exists by P2'

$= h_{12} o(h_3 oe) = h_{12,3} oe$ $h_{12,3}$ exists by P2'

$= \{(h_1 \tilde{*} h_2)*h_4\}oe$ where $*$ conditional on e

$\tilde{*}$ conditional on z

but $h_1 \tilde{*} h_2 = h_{12} = h_1 * h_2$

since the construction of h_{12} is independent of any element of Ω by P2'

hence $\{h_1*(h_2*h_3)\}\, oe = \{(h_1*h_2)*h_3\}oe$

thus $h_1*(h_2*h_3) = (h_1*h_2)*h_3$ by P1.

This proves that the binary operation * does indeed have the property of associativity.

The next example was chosen to illustrate the fact that the pivotal distribution has the same form as the fiducial distribution for a wide variety of cases.
It is also an example of where the induced binary operation is a continuous loop.

Example 2.3.6

Let $F(x|\theta)$ be a family of univariate cumulative distribution functions indexed by $\theta \varepsilon \Theta = (a,b)$. Suppose $F(x|\theta)$ is strictly increasing in x for each θ , strictly decreasing in θ for each x, and $F(x|a) = 0$, $F(x|b) = 1 \quad x\varepsilon\Omega$

We can write this model as a pivotal model as follows: let $y = F(x|\theta_o)$, where $\theta_o\varepsilon\Theta$ is fixed and known.

Define $\quad \theta oe = F(F^{-1}(e|\theta)|\theta o) \qquad (2.3.26)$

then it can be shown that

$$y = \theta oe \qquad e \sim \text{uniform}(0,1)$$

where $y, e\varepsilon\Omega = (0,1)$

is equivalent to the above model and is a pivotal model.

The partition induced on Ω is the trivial partition, namely the whole space Ω itself.

Thus every point $z\varepsilon\Omega$ can be transformed into any other point in Ω by a transformation given in (2.3.26). From this fact we have that

$$\frac{J_\Omega(\theta\backslash i|x)\ J_{*R}(\theta\backslash i, i|x)}{J_*(\theta, i|x)\ J_*(\theta\backslash i, i|x)}$$

$$= \frac{J_\Omega([e]|x) \qquad J_{*R}([e],[x]|x)}{J_*(\theta,[x]|x) \quad J_*([e],[x]|x)} \qquad \text{since } [x] = i \text{ and } \theta\backslash[x] = [e]$$

$$= \frac{\left|\frac{\partial([e]ox)}{\partial(x)}\right| \left|\frac{\partial([x]*[e])}{\partial([x])}\right|}{\left|\frac{\partial(\theta*[x])}{\partial([x])}\right| \left|\frac{\partial([e]*[x])}{\partial([x])}\right|} \qquad \text{by definition of the jacobians}$$

$$= \left|\frac{\partial([e]ox)}{\partial(\theta)}\right| \left|\frac{\partial([x])}{\partial(x)}\right| \qquad \text{after some rearranging.}$$

Now the second factor will be a constant since x is assumed to be fixed. Hence from (2.3.25) we find that the pivotal distribution of θ given x is

$$p(\theta|x) = k(x)f(\theta^{-1}ox)\left|\frac{\partial(\theta^{-1}ox)}{\partial(\theta)}\right| \text{since } [e]ox = \theta^{-1}ox$$

$$= k(x)\left|\frac{\partial(F(x|\theta))}{\partial(\theta)}\right|$$

which is of the same form as that obtained by the fiducial method.

□

2.4 Pivotal Measure and Pivotal Measurable Functions

In this section the idea of Pivotal Measure will be introduced. The concept arose in answer to the following question:

"If the pivotal distribution is a probability distribution then what is its event space, σ-algebra and probability measure?"

First we note that the transformation variable $[e]\varepsilon\Theta$ has a probability density function conditional on D, given by (2.3.14), and so will have an event space Θ, σ-algebra $\mathfrak{D}$ and probability measure $g_{[e]|D}(\cdot|\cdot)$, already defined. We also note that from (2.3.16) we have $[e] = \theta\backslash[x]$ with $\backslash$ conditional on D. From these two facts we can construct a σ-algebra, $\mathcal{P}$, on an event space $\Theta \times \Theta$ as follows:

For every $E^*\subseteq\Theta$ that is a member of the event space of [e] given D we define a set $\Theta^* \subseteq\Theta \times \Theta$ as follows:

$$(\theta,[x])\varepsilon\ \Theta^*\varepsilon\mathcal{P} \text{ iff } \theta\backslash[x]\varepsilon E^*\varepsilon\mathfrak{D} \qquad (2.4.1)$$

with $\backslash$ conditional on D, hence Θ^* is conditional on D.

It is easily seen that the sets $\{E^*: E^*\varepsilon\mathfrak{D}\}$ and $\{\Theta^*:\Theta^*\varepsilon\mathcal{P}\}$ are in a 1-1 correspondence with each other and hence it can be shown that the sets $\{\Theta^*\}$ do indeed form a σ-algebra on the event space $\Theta \times \Theta$.

We can construct a probability measure conditional on D, $P_M(\cdot|D)$ say, on the event space $\Theta \times \Theta$ as follows:

For every $\Theta^* \varepsilon \mathcal{P}$ let $E^* \varepsilon \mathcal{D}$ be the event in 1-1 correspondence with Θ^*, then let

$$P_M(\Theta^*|D) = \int_{[e]\varepsilon E^*} g([e]|D)d[e] \qquad (2.4.2)$$

Again since the sets $\{E^*: E^* \varepsilon \mathcal{D}\}$ and $\{\Theta^*: \Theta^* \varepsilon \mathcal{P}\}$ are in 1-1 correspondence, it can be shown that $P_M(\cdot|D)$ is a conditional probability measure defined on the event space $\Theta \times \Theta$.

The relationship of the event space $\Theta \times \Theta$, σ-algebra $\mathcal{P}$ and the conditional probability measure $P_M(\cdot|D)$ to pivotal models and the pivotal distribution will now be demonstrated.

Define the cross-section of an event $\Theta^* \varepsilon \mathcal{P}$ by a fixed $[x] \varepsilon \Theta$, $S(\Theta^*, [x], D) \subseteq \Theta$ say, as follows:

$$S(\Theta^*, [x], D) = \{\theta: (\theta, [x]) \varepsilon \Theta^*, \Theta^* \text{ conditional on } D\} \qquad (2.4.3)$$

This leads to the following lemma:

<u>Lemma 2.4.1</u>

$\forall \Theta^* \varepsilon \mathcal{P} \quad \forall [x] \varepsilon \Theta$

$$\int_{\theta \varepsilon S(\Theta^*, [x], D)} p(\theta | [x]oD) d\theta = P_M(\Theta^* | D)$$

that is to say the same probability is given to every cross-section of a fixed event $\Theta^* \varepsilon \mathcal{P}$ by the pivotal distribution and this is equal to the probability given to Θ^* by the probability measure $P_M(\cdot | D)$.

Proof:

Now $$\int_{\theta \varepsilon S(\Theta^*, [x], D)} p(\theta | [x]oD) d\theta =$$

$$= \int_{\theta \varepsilon S(\Theta^*, [x], D)} k(D) f(\theta^{-1} ox) J_\Omega(\theta \backslash [x] | D) \Delta(\theta \backslash [x]) d\mu(\theta)$$ by 2.3.24

Let $\theta = [x]/[e]$ or $[e] = \theta \backslash [x]$

Now $\theta \varepsilon S(\Theta^*, [x], D)$ iff $[e] \varepsilon E^* \varepsilon \mathcal{D}$ by 2.4.1 where E^* is the event in 1-1 correspondence with Θ^*.

Hence LHS = $$\int_{[e] \varepsilon E^*} k(D) f([e]oD) J_\Omega([e] | D) d\mu([e])$$ by 2.3.24

$$= \int_{[e] \varepsilon E^*} g([e] | D) d[e] = P_M(\Theta^* | D)$$ by 2.4.2

which proves the lemma. ||

This lemma supports our calling the sets $\Theta^* \varepsilon \mathcal{P}$ pivotal measurable sets.

If we consider each set $S(\Theta^*, [x], D)$ to be a particular cover of the parameter space Θ given that we have observed $x \varepsilon \Omega$, then we can calculate the posterior probability of the cover if we are given a conditional prior distribution of $\theta \varepsilon \Theta$ given $[e]$, $\Pi(\theta|[e])$ say. Denote the calculated posterior probability of $S(\Theta^*, [x], D)$ given a conditional prior distribution $\Pi(\theta|[e])$ by $\alpha\{S|\Pi\}$.

We can also calculate the generalised probability density of $[x]$ given D and $\Pi(\theta|[e])$, $\tilde{p}([x]|D, \Pi(\theta|[e]))$ say.

The following theorem, which is proved in appendix A1.6, gives one interpretation of the pivotal probability attached to each cover $S(\Theta^*, [x], D)$.

<u>Theorem 2.4.1</u>

For every conditional prior distribution $\Pi(\theta|[e])$,

$$E_{[x]|D}\{\alpha\{ S(\Theta^*, [x], D)|\Pi(\theta|[e])\}|D\} =$$

$$= \int_{[x]\varepsilon\Theta} \alpha\{ S(\Theta^*, [x], D)|\Pi(\theta|[e])\} \tilde{p}([x]|D)d[x]$$

$$= \int_{\theta\varepsilon S(\Theta^*, [x], D)} p(\theta|[x]oD)d\theta$$

that is to say the expected value of $\alpha\{S|\Pi\}$ taken over $[x] \varepsilon \Theta$ given D is equal to the pivotal probability of the set $S(\Theta^*, [x], D)$ given D and this is true no matter what conditional prior distribution $\Pi(\theta|[e])$ we might choose.

This property of the pivotal distribution, the specification of the expected posterior probability of the sets $S(\Theta^*, [x], D)$ rather than the actual posterior probability which is unknown, is a property Fisher gives to his fiducial probability (Fisher 1930). There are a great many other similarities to fiducial probability such as the dependence on pivots. The main difference is that Fisher uses classical probability models, whereas the pivotal model is a structured model. It was shown in section 2.1 that there is a difference between these two types of probability model. It is for this reason that the pivotal distribution is not called a fiducial distribution despite their having many properties in common.

<u>Example 2.4.1</u> (based on an example given in Buehler 1971)

Consider the pivotal model

$$x = \theta + e \qquad e \sim f(\cdot)$$

where $x, e \varepsilon \Omega = \mathbb{Z}$ $\theta \varepsilon \Theta = \mathbb{Z}$

$$\text{and } f(e) = \begin{cases} P \geq 0 & \text{if } e = 1 \\ 1-P \geq 0 & \text{if } e = -1 \\ 0 & \text{otherwise} \end{cases}$$

We note that this is already in a reduced pivotal model form.

It can be shown that there are four sets in the σ-algebra $\mathcal{P}$ namely $\{\phi, \Theta^*, \Theta_2^*, \Theta_1^* \cup \Theta_2^*\}$ where $\Theta_1^* = \{(z-1,z): z\varepsilon\Omega\}$, $\Theta_2^* = \{(z+1,z): z\varepsilon\Omega\}$ and they have the following pivotal probabilities:

$$P_M(\phi) = 0;\ P_M(\Theta_1^*) = P;\ P_M(\Theta_2^*) = 1-P;$$
$$P_M(\Theta_1^* \cup \Theta_2^*) = 1 \qquad (2.4.3)$$

We will now calculate the posterior probability of the sets $\{S(\Theta^*,x): x\varepsilon\mathbb{Z}\}$ given a conditional prior distribution of $\theta\varepsilon\Theta$ given $e\varepsilon\Omega$, $\Pi(\theta|e)$ say.

The joint distribution of e and θ is

$$\begin{cases} \Pi(\theta|e=1)P & e = 1,\ \theta\varepsilon\,\mathbb{Z} \\ \Pi(\theta|e=-1)(1-P) & e = -1,\ \theta\varepsilon\mathbb{Z} \end{cases}$$

and from this we can calculate the joint distribution of x and θ, $\tilde{P}(x,\theta)$ say.

$$\tilde{p}(x,\theta) = \begin{cases} \Pi(\theta|e=1)P & x = \theta+1 \\ \Pi(\theta|e=-1)(1-P) & x = \theta-1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.4.4)$$

The marginal distribution of x is

$$\tilde{P}(x) = \sum_{\theta \varepsilon \mathbb{Z}} \tilde{P}(x, \theta)$$

and the conditional distribution of θ given x satisfies

$$\tilde{P}(x,\theta) = \tilde{P}(\theta|x)\tilde{P}(x) \qquad (2.4.5)$$

The posterior probability of the set $S(\Theta_1^*,x)$ $x \varepsilon \mathbb{Z}$ will then be $\alpha\{S(\Theta_1^*,x)|\Pi(\theta|e)\} = \tilde{P}(\theta = x-1|x)$.

We can now verify theorem 2.4.1 for this particular example by calculating the expected posterior probability of the sets $\{S(\Theta_1^*,x): x\varepsilon\mathbb{Z}\}$

$$E_x(\alpha\{ (\Theta_1^*,x)|\Pi(\theta|e)\}) =$$

$$= \sum_{x\varepsilon \mathbb{Z}} \tilde{P}(\theta = x-1|x)\tilde{P}(x)$$

$$= \sum_{x\varepsilon \mathbb{Z}} \tilde{P}(\theta = x-1,x) \qquad \text{by } 2.4.5 \qquad (2.4.6)$$

$$= P \sum_{x\varepsilon \mathbb{Z}} \Pi(x-1|e=1) \qquad \text{by } 2.4.4$$

$$= P \quad = P_M(\Theta_1^*) \qquad \text{by } 2.4.3$$

Hence the theorem is verified for this particular example.

The following betting scheme was proposed (Buehler 1971) to defeat any "pivotal type" of argument with respect to this example. (Buehler used the value $P = \frac{1}{2}$, though this is not necessary).

"If $x \geq 1$ bet with even money $\theta = x-1$

If $x < 1$ refuse to bet."

If we have a conditional prior distribution $\Pi(\theta|e)$ this betting scheme is equivalent to reconditioning the joint distribution of x and θ to

$$\tilde{P}_*(x,\theta) = \begin{cases} k^{-1}\Pi(\theta|e=1) & 1 \leq x = \theta+1 \\ k^{-1}\Pi(\theta|e=-1) & 1 \leq x = \theta-1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.4.7)$$

where $k = \sum_{x=1}^{\infty} \{\Pi(\theta = x-1|e=1)+\Pi(\theta = x+1|e=-1)\}$

This is well defined as long as $k > 0$ otherwise no bet will take place.

The expected posterior probability of the sets $\{S(\theta|\theta_1{}^*,x)\ x\varepsilon\mathbb{Z}\}$ will be by 2.4.6

$$= \sum_{x\varepsilon\mathbb{Z}} \tilde{P}_*(x,\ \theta = x-1)$$

$$= k^{-1} \sum_{x=1}^{\infty} \Pi(\theta = x-1|e=1) \qquad \text{by (2.4.7)}$$

$$= \frac{\sum_{x=1}^{\infty} \Pi(\theta = x-1|e = 1}{\sum_{x=1}^{\infty} \{\Pi(\theta-x-1|e=1)+\Pi(\theta = x+1|e = -1)\}}$$

and this can take any value between 0 and 1. Thus it is possible to construct conditional prior distributions of θ given e where the "pivotal" gambler would win every bet, such as

$$\Pi(\theta|e=1) = \begin{cases} 1 & \theta = -4 \\ 0 & \text{otherwise} \end{cases}$$

$$\Pi(\theta|e=-1) = \begin{cases} 1 & \theta = 3133 \\ 0 & \text{otherwise} \end{cases}$$

To be fair the "pivotal" gambler would not accept the bet because he is being asked to put a pivotal probability on a non pivotal measurable set, namely

$$\{\{\theta = x+1, x\} : x \geq 1\} \notin \mathcal{P}$$ □

<u>Example 2.4.2 (example 2.1.1 continued) N(θ,1) model</u>

As a reminder this model takes the form

$$x = \theta + e \qquad e \sim N(0,1)$$

with $x, e \in \Omega = \mathbb{R} \quad \theta \in \Theta = \mathbb{R}$

The measure induced by the random variable e on the space Ω is the ordinary Lebesgue measure on $\mathbb{R}$. Thus from 2.4.1 given a Lebesgue measurable set E^* on $\mathbb{R}$, (-1.96, 1.96) say, we can construct a pivotal measurable set Θ^*

$(\theta,x)\varepsilon\Theta^*\varepsilon\mathcal{P}$ iff $x-\theta\varepsilon E^* = (-1.96,1.96)$

The cross-sections $S(\Theta^*,x)$ will take the form

$$S(\Theta^*,x) = (x-1.96,x+1.96) \quad x\varepsilon\Omega$$

The pivotal measure assigned to the pivotal measurable set Θ^* is

$$P_M(\Theta^*) = \int_{e\varepsilon E^*} \phi(e)de \qquad \text{by 2.4.1}$$

$$= \int_{-1.96}^{1.96} \phi(e)de \doteq .9500 \text{ to 4 d.p.}$$

where $\phi(.)$ is the standard normal density function. Thus by Theorem 2.4.1 the expected posterior probability of the sets $S(\Theta^*,x)$ when x is allowed to vary over Ω will be 95%. This is true whatever conditional prior distribution of $\theta\varepsilon\Theta$ given $e\varepsilon\Omega$ we might choose.

Suppose that we are told beforehand that the parameter $\theta\varepsilon\Theta$ takes values only on the interval $(0,3)\subset\Theta$. This in essence places a restriction on the set of conditional prior distributions of $\theta\varepsilon\Theta$ given $e\varepsilon\Omega$, namely to those distributions whose supports are contained in the interval (0,3). The expected posterior probability of the sets $S(\Theta^*,x)$ as x varies over Ω will still be 95% because Theorem 2.4.1 states that the expected posterior probability is independent of which conditional prior distribution one takes even if we restrict them to having supports contained in the interval (0,3). □

We now extend the idea of pivotal probabilities to parameters other than θ. This will be achieved using pivotal measurable functions.

Given a function $H: \Theta \times \Theta \to T$, where T is a topological space, we say that H is a pivotal measurable function if and only if for every open set t in T $H^{-1}(t) \subseteq \Theta \times \Theta$ is a pivotal measurable set.

Note this is just an application of the usual definition of a measurable function to the pivotal measure.

If H is a pivotal measurable function, for every open set $t \subseteq T$ we can assign a probability namely

$$P_M(H^{-1}(t)|D), = P_T(t|D) \text{ say} \qquad (2.4.8)$$

Hence if we can find a pivotal measurable function $H: \Theta \times \Theta \to T$ we can extend the probability $P_M(\cdot|D)$ to the topological space T to obtain $P_T(\cdot|D)$.

Returning to the question of extending the idea of pivotal probabilities, we would like to obtain parameters other than θ that also have the expected posterior probability interpretation.

Consider functions $H: \Theta \times \Theta \to T$ that take the form

$$H: (\theta,[x]) \to (\hat{\theta}(\theta,[x]),[x]) \in T \qquad (2.4.9)$$

where $\hat{\theta}$ is an arbitrary function of $(\theta,[x]) \in \Theta \times \Theta$, $\hat{\theta}: \Theta \times \Theta \to \hat{\Theta}$ a new parameter space, and where T is a topological space that is a direct product of the spaces $\hat{\Theta}$ and Θ .

For every open set $t \subseteq T$ we define its cross-section by a fixed $[x] \in \Theta$, $S_H(t,[x],D)$ say, as follows:

$$S_H(t,[x],D) = \{\hat{\theta}: (\hat{\theta},[x]) \in t\} \qquad (2.4.10)$$

This leads to the following lemma.

Lemma 2.4.2

If $H: (\theta,[x]) \to (\hat{\theta}(\theta,[x],[x]) \in T$

is a pivotal measurable function then for every open set $t \subseteq T$

$$\theta \in S(H^{-1}(t),[x],D) \text{ iff } \hat{\theta}(\theta,[x]) \in S_H(t,[x],D)$$

Proof:

Since H is a pivotal measurable function, for every open set $t \subseteq T$, $H^{-1}(t)$ is a pivotal measurable set, thus the cross-sections are well defined. By construction of H, the element $(\theta,[x]) \in H^{-1}(t)$ maps onto the element $(\hat{\theta}(\theta,[x]),[x])$ which is easily seen must be a member of the open set t. Now both these elements are in the appropriate cross-sections, hence

$$\theta\varepsilon\ S(H^{-1}(t),[x],D) \Rightarrow (\hat{\theta},[x])\varepsilon\ \ S_H(t,[x],D)$$

Furthermore by the construction of H, if the element $(\hat{\theta}(\theta,[x]),[x])\varepsilon t$ then (as can easily be seen) the element $(\theta,[x])$ must be a member of the pivotal measurable set $H^{-1}(t)\varepsilon\mathcal{P}$. Thus

$$\hat{\theta}(\theta,[x])\varepsilon\ \ S_H(t,[x],D)\Rightarrow\theta\varepsilon\ S(H^{-1}(t),[x],D)$$

which proves the lemma. ||

For a fixed $[x]\varepsilon\Theta$ it is easily seen that the set of cross-sections of the open sets of T form a σ-algebra on $\hat{\Theta}$. To each of these cross-sections we assign a pivotal probability conditional on [x] and D namely the same pivotal probability that was given to the corresponding cross-section of a pivotal measurable set mentioned in lemma 2.4.2. Hence we can construct a pivotal density conditional on [x] and D on the parameter space $\hat{\Theta}$ by taking the generalised probability density induced by the pivotal probabilities on the cross-sections of the open sets of T. We denote this conditional pivotal density on $\hat{\Theta}$ by $p_H(\hat{\theta}|[x],D)$.

<u>Lemma 2.4.3</u>

$\forall$ open sets $t\subseteq T$, $\forall[x]\varepsilon\Theta$

$$\int_{\theta \varepsilon S_H(t[x]D)} p_H(\hat{\theta}|[x],D)d\hat{\theta} = P_T(t|D)$$

<u>Proof</u>:

$$P_T(t|D) = P_M(H^{-1}(t)|D) \qquad \text{by 2.4.8}$$

$$= \int_{\theta \varepsilon S(H^{-1}(t),[x],D)} p(\theta|[x]D)d\theta \qquad \text{by lemma 2.4.1}$$

$$= \int_{\hat{\theta} \varepsilon S_H(t,[x],D)} p(\hat{\theta}|[x]D)d\hat{\theta} \qquad \text{by lemma 2.4.2 and by construction}$$

Hence the lemma. ||

Having now obtained the conditional pivotal density on the parameter space Θ, we wish to show that it has an expected posterior probability interpretation.

Again if we are given a conditional prior distribution of $\theta\varepsilon\Theta$ given [e], $\Pi(\theta|[e])$ say, then the joint distribution of θ and [x] given D, $\tilde{p}([x],\theta|D)$, is given by A1.6.3. From this joint distribution we can calculate (as in A1.6.4) the marginal distribution of [x] given D $\tilde{p}([x]|D)$, and the conditional distribution of θ given [x] and D, $\tilde{p}(\theta|[x],D)$ (in A1.6.5).

In a similar construction to that for obtaining the conditional pivotal density $p(\hat{\theta}|[x],D)$ from the conditional pivotal density $p(\theta|[x],D)$ we can construct from

$\tilde{p}(\theta|[x],D)$ the conditional distribution of $\hat{\theta}$ given [x] and D, $\tilde{p}(\hat{\theta}|[x],D)$ say. This conditional distribution will be defined on the σ-algebra constructed from the cross-sections $\{S_H(t,[x],D): t\varepsilon T\}$ for fixed $[x]\varepsilon\Theta$.

Thus given the conditional prior distribution $\Pi(\theta|[e])$ and considering each set $S_H(t,[x],D)$ as a particular cover of the parameter space $\hat{\Theta}$ given that we have observed $x\varepsilon\Omega$ we can calculate the posterior probability of the cover. Denote the calculated posterior probability of $S_H(t,[x],D)$ given a conditional prior distribution $\Pi(\theta|[e])$ by $\alpha_H\{S_H(t,[x],D)|\Pi(\theta|[e])\}$. The following theorem shows that the sets $S_H(t,[x],D)$ do indeed have an expected posterior probability interpretation.

Theorem 2.4.2

For every conditional prior distribution $\Pi(\theta|[e])$

$$E_{[x]|D}\{\alpha_H\{S_H(t,[x],D)|\Pi(\theta|[e])\}$$

$$= \int_{[e]\varepsilon\Theta} \alpha_H\{S_H(t,[x],D)|\Pi(\theta|[e])\}\ \tilde{p}([x]|D)d[x]$$

$$= \int_{\hat{\theta}\varepsilon S_H(t,[x],D)} p_H(\hat{\theta}|[x],D)d\hat{\theta}$$

Proof:

$$\int_{\hat{\theta}\varepsilon S_H(t,[x],D)} p_H(\hat{\theta}|[x],D)d\hat{\theta} = P_T(t|D) \qquad \text{by lemma 2.4.3}$$

$$= P_M(H^{-1}(t)|D) \qquad \text{by 2.4.8}$$

$$= E_{[x]|D}\{\alpha\{S(H^{-1}(t),[x],D)|\Pi(\theta|[e])\}\} \qquad \text{by A1.6.9}$$

$$= \int_{[x]\varepsilon\Theta} \tilde{p}([x]|D) \int_{\theta\varepsilon S(H^{-1}(t),[x],D)} \tilde{p}(\theta|[x],D)d\theta d[x] \qquad \text{by A1.6.6}$$

$$= \int_{[x]\varepsilon\Theta} \tilde{p}([x]|D) \int_{\hat{\theta}\varepsilon S_H(t,[x],D)} \tilde{p}_H(\hat{\theta}|[x],D)d\hat{\theta}d[x] \qquad \text{by construction and lemma 2.4.2}$$

$$= E_{[x]|D}\{\alpha_H\{S_H(t,[x],D)|\Pi(\theta|[e])\}\}$$

Hence theorem. ||

We have thus found a sufficient condition to extend the expected posterior probability to parameters other than $\theta\varepsilon\Theta$. Since we are only using $[x]\varepsilon\Theta$ as a label of the cross-sections of the open sets of T, without loss of generality we can take functions of the form:

$$H: (\theta,[x]) \to (\hat{\theta}(\theta,[x]), h([x])) \varepsilon T \qquad (2.4.11)$$

where $\hat{\theta}$ is an arbitrary function of $[x],\theta\varepsilon\Theta$ and h is a 1-1 function of [x].

If H is a pivotal measurable function then the cross-sections of the open sets of T with fixed h([x]) will have an expected posterior probability interpretation.

It is still an open question whether this is still true if h is an arbitrary function of [x], with H being a pivotal measurable function.

Note H being a pivotal measurable function is a necessary condition otherwise we are unable to transfer the pivotal probability onto the open sets of T.

<u>Example 2.4.3 (example 2.1.1 continued) N(θ,1) model</u>

Consider taking the function H: $\Theta \times \Theta \to T$ to be of the form H: $(\theta,x) \to (\hat{\theta} = \theta^3, x) \in T = \mathbb{R} \times \mathbb{R}$ and define the open sets of T to be of the form:

$$t \subseteq T \text{ is an open set iff } \exists S^* \subseteq \mathbb{R} \text{ a Lebesgue measurable set s.t. } (\hat{\theta},x) \in t \text{ iff } \exists s \in S^* \text{ s.t. } \hat{\theta} = (x-s)^3 \quad (2.4.12)$$

It can easily be shown that this does define a topology on T.

We will now show that H is a pivotal measurable function. Let t be an open set of T and let $S^* \in \mathbb{R}$ a Lebesgue measurable set that satisfies the conditions given in 2.4.12 for this particular open set t. From 2.4.12 the set $H^{-1}(t)$ will take the form

$(\theta, x) \in H^{-1}(t)$ iff $\exists s \in S^*$ s.t. $\theta = x - s$

This is precisely the condition given in example 2.4.2 for the pivotal measurable sets. Hence H is a pivotal measurable function.

The pivotal density for $\theta \in \Theta$ conditional on x is

$$p(\theta|x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-(\theta-x)^2}{2}\right\}$$

From this the pivotal density for $\hat{\theta} = \theta^3$ conditional on x is easily calculated to be

$$p_H(\hat{\theta}|x) = \frac{1}{\sqrt{2\pi}} \frac{1}{3\hat{\theta}^{2/3}} \exp\left\{\frac{-(\hat{\theta}^{1/3}-x)^2}{2}\right\}$$

and this will have an expected posterior probability interpretation. □

Wilkinson (1977) extended fiducial probability to what he called "pivotally equivalent functions". From arguments he gave in his paper, it can be seen that these "pivotally equivalent functions" are isomorphic to the functions $\hat{\theta}(\theta,[x])$ given in (2.4.9) when H is a pivotal measurable function. In the same paper Wilkinson also mentioned the Noncoherence Principle:

"The inferential implications of observational data alone (with sampling distribution known in parametric form) are noncoherent, in that they cannot be represented by a single inferential probability distribution on the parameter space."

We have shown above that we can extend pivotal probability by the idea of pivotal measurable functions. It is possible to have a function that is not pivotal measurable with respect to a given model. We then need another model for which it is a pivotal measurable function. We can then represent our inferential implications with this new model. Hence the Noncoherence Principle is a consequence of the pivotal measure. Examples of non pivotal measurable functions (non pivotally equivalent functions) are given in Wilkinson (1977, p129).

2.5 Multipivotal Models

In this section we define the multipivotal model which is a generalisation of the pivotal model discussed in the previous sections.

The multipivotal model takes the form

$$x = \theta_1 o \theta_2 o \ldots o \theta_m o e \tag{2.5.1}$$

where e is a random variable from a fixed known distribution f(.) and $x, e \varepsilon \Omega$ a sample space.

Each θ_i, i = 1, ..., m is an index element of a set of transformations $\{\phi_{\theta_i} : \Omega \rightarrow \Omega;\ \theta_i \varepsilon \Theta_i\}$ where Θ_i is the indexing set.

For convenience define

$$y_i = \begin{cases} \theta_{i-1}^{-1} o \ldots o \theta_1^{-1} o x & i = 2, \ldots, m \\ x & i = 1 \end{cases} \tag{2.5.2}$$

and
$$d_i = \begin{cases} \theta_{i+1} o \ldots o \theta_m o e & i = 1, \ldots, m-1 \\ e & i = m \end{cases} \tag{2.5.3}$$

Three restrictions are placed on the set of transformations in a multipivotal model namely:

I For a fixed j = 1,...,m. If the values $\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_m$ become known then the reduced model

$$y_j = \theta_j o d_j \qquad (2.5.4)$$

is a pivotal model.

<u>II</u> For a fixed j = 1,...,m-1 and k = j+1,...,m.
If the values $\theta_1,\ldots, \theta_{j-1}, \theta_{k+1},\ldots, \theta_m$ become known, then the reduced model

$$y_j = \theta_j o\ldots o\theta_k o d_k \qquad (2.5.5)$$

is not a pivotal model.

<u>III</u> For fixed $e\varepsilon\Omega$

$$\theta_1 o\ldots o\theta_m oe = \Psi_1 o\ldots o\Psi_m oe$$

where $\theta_i, \Psi_i \ \ \varepsilon\Theta_i \quad i = 1,\ldots,m$

$$\Rightarrow \forall\ i = 1,\ldots,m \quad \theta_i = \Psi_i \qquad (2.5.6)$$

<u>Example 2.5.1</u>

Consider the model

$$(x,y) = \theta oro(e,d) \qquad (e,d) \sim f(\cdot) \qquad (2.5.7)$$

with $(x,y), (e,d)\varepsilon\Omega = \mathbb{R}^2$.

Let $(V,W)\varepsilon\Omega$. We define the transformations $\phi_\theta:\Omega\to\Omega$ and $\phi_r:\Omega\to\Omega$ as follows:

$$\theta o(V,W) = (V\cos\theta - W\sin\theta,\ V\sin\theta + W\cos\theta)$$

$\{\phi_\theta:\theta\varepsilon\Theta_\theta \equiv 0 \leqq \theta < 2\Pi\}$ being the set of rotations about the origin, and

$$ro(V,W) = (rV,W)$$

$\{\phi_r:\ r\varepsilon\Theta_r = \mathbb{R}^+\}$ being the set of enlargements in the X direction about the Y-axis.

Note both of these sets of transformations individually form a group on Ω. As a result each will form a reduced pivotal model of the form (2.5.4) if the value of the index of the other transformation becomes known. The last two requirements, namely the combination of the two sets of transformations not forming a pivotal model and the uniqueness of the combination of the transformations are both easily demonstrated. Hence the model (2.5.7) is a multipivotal model. □

We first derive the pivotal distributions of θ_i conditional on $\theta_1,\ldots,\ \theta_{i-1},\ \theta_{i+1},\ldots,\ \theta_m;\ i = 1,\ldots,m$ being known. From (2.5.4) we can rewrite the multipivotal model (2.5.1) as

$$y_i = \theta_i o d_i$$

We first though have to find the distribution of d_i; call this distribution $f_i(\cdot)$.

Let $J_{i,\Omega}(\theta_i|D_i)$ be the Jacobian $J_\Omega(\theta|D)$, defined by (2.3.6) applied to the set of transformations $\{\phi_{\theta_i}: \theta_i \varepsilon \Theta_i\}$ $i = 1,\ldots,m$. (Similarly we define $J_{i,*}(\theta_i, h_i|D_i)$, $J_{i,*R}(\theta_i: h_i|D_i)$, $\mu_i(\theta_i)$, $\nu_i(\theta_i)$ and $\Delta_i(\theta_i)$ $i = 1,\ldots,m$).

From (2.5.1) and (2.5.2) and using standard transformation of variables we obtain

$$f_i(d_i) = f(\theta_m^{-1} o \ldots o \; {}_{i+1}^{-1} o d_i)$$

$$\prod_{j=i+1}^{m} J_{j,\Omega}(\theta_j, id_j: \; \theta_{j-1}^{-1} o \ldots o \theta_{i+1}^{-1} o d_i) \tag{2.5.8}$$

where ϕ_{id_j} is the identity transform of the jth set of transformations.

Substituting this into equation (2.3.25) we obtain the pivotal distribution of θ_i conditional on x and $\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_m$ as

$$p(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_m, x) = \tag{2.5.9}$$

$$= k_i f(\theta_m^{-1} o \ldots o \theta_1^{-1} o x) \frac{J_{i,\Omega}(\theta_i \backslash id_i | y_i) J_{i,*R}(\theta_i \backslash id_i, id_i | y_i)}{J_{i,*}(\theta_i \backslash id_i, id_i | y_i) J_{i,*}(\theta_i, id_i | y_i)}$$

$$\prod_{j=i+1}^{m} J_{j,\Omega}(\theta_j \backslash id_j : y_j)$$

where k_i is the constant of proportionality which is dependent on $\theta_i, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_m$ and x.

Having obtained all the conditional pivotal distributions, we obtain the joint pivotal distribution (which we call the Multipivotal Distribution) as the following product of ratios:

$$\frac{p(\theta_1, \ldots, \theta_m | x)}{p(h_1, \ldots, h_m | x)} =$$

$$\prod_{i=1}^{m} \left\{ \frac{p(\theta_i | \theta_1, \ldots, \theta_{i-1}, h_{i+1}, \ldots, h_m, x)}{p(h_i | \theta_1, \ldots, \theta_{i-1}, h_{i+1}, \ldots, h_m, x)} \right\} \qquad 2.5.10)$$

where $\theta_i, h_i \varepsilon \Theta_i$ $i = 1, \ldots, m$ and $h_1, \ldots, h_m$ are known fixed index elements.

Note the conditional pivotal distributions on the right hand side of (2.5.10) have to satisfy certain conditions for a consistent probability structure to be specified. These conditions are stated in a theorem due to Hammersley and Clifford discussed with a simple proof in Besag (1974). It can be shown that the conditional pivotal distributions, for the multipivotal model, always satisfy these conditions. We thus have a consistent probability structure and equation (2.5.10) is well defined.

If we take $\phi_{h_1},\ldots,\phi_{h_m}$ to be the identity transformations in the respective set of transformations $\{\phi_{\theta i}: \theta_i \varepsilon \Theta_i\}$ $i = 1,\ldots,m$ and assume that none of the probabilities in the denominator of equation (2.5.6) are zero, then from equations (2.5.9) and (2.5.10) we can write the multipivotal distribution of $\theta_1,\ldots,\theta_m$ conditional on knowing $x\varepsilon\Omega$ in the form:

$$\frac{p(\theta_1,\ldots,\theta_m|x)}{p(id_1,\ldots,id_m|x)} = \frac{f(\theta_m^{-1}o\ldots o\theta_1^{-1}ox)}{f(x)}$$

$$\prod_{i=1}^{m} \frac{J_{i,\Omega}(\theta_i\backslash id_i|y_i)J_{i,*R}(\theta_i\backslash id_i,id_i|y_i)}{J_{i,*}(\theta_i\backslash id_i,id_i|y_i)J_{i,*}(\theta_i,id_i|y_i)} \qquad (2.5.11)$$

Thus up to a constant of proportionality we have the multipivotal distribution for the parameters $\theta_1,\ldots,\theta_m$ given an observed $x\varepsilon\Omega$.

Example 2.5.2 (example 2.5.1 continued)

In example 2.5.1 it was stated that both sets of transformations $\{\phi_\theta: \theta\varepsilon\Theta_\theta \equiv [0,2\Pi]\}$ and $\{\phi_r: r\varepsilon\Theta_r = \mathbb{R}^+\}$ form groups on the sample space Ω. The binary operations $\underset{\theta}{*}$ and $\underset{r}{*}$ induced by the respective groups (see section 2.3) will both be independent of the choice of the reference variable (as a consequence of property P2' of section 2.3). The reduced binary operations will therefore take the form:

Given $\theta,\Psi\varepsilon\Theta_\theta;\theta \quad \Psi = \theta+\Psi \bmod 2\Pi$

Hence $\theta \backslash_{\theta} \Psi = \Psi - \theta \bmod 2\Pi$ and $\theta /_{\theta} \Psi = \theta - \Psi \bmod 2\Pi$,
and given $r, R \varepsilon \Theta_r$ $\; r \cdot R = rR$.
Hence $r \backslash_{r} R = r^{-1}R$ and $r /_{r} R = rR^{-1}$

From this we can calculate the relevant Jacobians:

$$J_{\theta,\Omega}(\theta|(V,W)) = \left|\frac{\partial(V\cos\theta - W\sin\theta, V\sin\theta + W\cos\theta)}{\partial(V,W)}\right| = 1$$

$$J_{\theta,*}(\theta;\Psi) = \left|\frac{\partial(\theta+\Psi)}{\partial(\Psi)}\right| = 1$$

$$J_{\theta,*R}(\theta;\Psi) = \left|\frac{\partial(\Psi+\theta)}{\partial(\Psi)}\right| = 1$$

$$J_{r,\Omega}(r|(V,W)) = \left|\frac{\partial(rV,W)}{\partial(V,W)}\right| = r$$

$$J_{r,*}(r,R) = \left|\frac{\partial(rR)}{\partial(R)}\right| = r$$

$$J_{r,*R}(r,R) = \left|\frac{\partial(Rr)}{\partial(R)}\right| = r$$

Substituting these into equation (2.5.11) we obtain the multipivotal distribution of $\theta\varepsilon\Theta_\theta$ and $r\varepsilon\Theta_r$ given an observation (x,y) as

$$\frac{p(\theta,r|(x,y))}{p(0,1|(x,y))} = \frac{f(r^{-1} o \theta^{-1} o (x,y))}{f((x,y))}\frac{1}{r^2} \qquad \square$$

Historically the method of combining the conditional pivotal models given in equation (2.5.10) is related to "step by step procedures" of Fisher (1956 (1973 edition)). But unlike most step by step procedures it does not depend on the order that the parameters are entered into the process, there being only one resulting multipivotal distribution.

The interpretation of the multipivotal distribution is a little more complicated than for a pivotal distribution. From the construction of the multipivotal distribution from the set of conditional pivotal distributions defined in (2.5.9), we see that it will have a conditional expected posterior probability interpretation for each of the parameters θ_i given x and $\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_n$ $i = 1,\ldots,n$, being the interpretation for the individual conditional pivotal distributions.

Is there more than just this conditional expected posterior probability interpretation? For instance, is there a joint expected posterior probability interpretation? Unfortunately from the above combination of conditional pivotal distributions we have to answer "no", as the following example demonstrates.

Example 2.5.3

Consider the model

$$\begin{pmatrix} x & , & y \\ 0 & , & z \end{pmatrix} = \begin{pmatrix} R_1 , & 0 \\ 0 & ,R_2 \end{pmatrix} \begin{pmatrix} 1 & , & Q \\ 0 & , & 1 \end{pmatrix} \begin{pmatrix} e & , & d \\ 0 & , & c \end{pmatrix}$$

$$\begin{pmatrix} e & , & d \\ 0 & , & c \end{pmatrix} \sim f(\cdot) \tag{2.5.12}$$

where $\begin{pmatrix} x, & y \\ 0, & z \end{pmatrix}$, $\begin{pmatrix} e, & d \\ 0, & c \end{pmatrix}$ $\varepsilon\Omega \equiv \begin{pmatrix} \mathbb{R}^+, & \mathbb{R} \\ 0, & \mathbb{R}^+ \end{pmatrix}$

$R_1, R_2 \varepsilon \mathbb{R}^+ \quad Q \varepsilon R.$

The following set of transformations form groups:

$$\theta_1 = \begin{pmatrix} R_1, & 0 \\ 0, & R_2 \end{pmatrix} \varepsilon\Theta_1 = \begin{pmatrix} \mathbb{R}^+, & 0 \\ 0, & \mathbb{R}^+ \end{pmatrix} ; \theta_2 = \begin{pmatrix} 1, & Q \\ 0, & 1 \end{pmatrix} \varepsilon\Theta_2 = \begin{pmatrix} 1, & \mathbb{R} \\ 0, & 1 \end{pmatrix}$$

$$\text{and } \theta_{12} = \begin{pmatrix} R, & R_1 Q \\ 0, & R_2 \end{pmatrix} = \begin{pmatrix} R_1, & 0 \\ 0, & R_2 \end{pmatrix} \begin{pmatrix} 1, & Q \\ 0, & 1 \end{pmatrix} \varepsilon\Theta_{12} = \begin{pmatrix} \mathbb{R}^+, & \mathbb{R} \\ 0, & \mathbb{R}^+ \end{pmatrix}$$

If we relax condition II (2.5.5) of the definition of the multipivotal model then we have two routes for obtaining a distribution of the parameters R_1, R_2 and Q. The first is to consider the model as a pivotal model using the group based on the space Θ_{12}. The distribution will then have a joint expected posterior probability interpretation. The second way is to obtain the conditional pivotal distributions of θ_1 given θ_2, and θ_2 given θ_1. Then using equation (2.5.11) we combine them to obtain a distribution of R_1, R_2 and Q that will have a conditional expected posterior probability interpretation. We will show that these two approaches give different distributions. Hence in general the distributions that have conditional expected posterior probability and joint expected posterior probability interpretations will be different.

We consider first the model to be a pivotal model.

Now $\begin{pmatrix} R_1, R_1Q \\ 0, R_2 \end{pmatrix} \begin{pmatrix} U, V \\ 0, W \end{pmatrix} = \begin{pmatrix} R_1U, R_1QW+R_1V \\ 0, R_2W \end{pmatrix} \varepsilon\Omega$

where $\begin{pmatrix} R_1, R_1Q \\ 0, R_2 \end{pmatrix} = \theta_{12}\varepsilon\Theta_{12}$ and $\begin{pmatrix} U, V \\ 0, W \end{pmatrix} = D\varepsilon\Omega$

Hence $J_\Omega(\theta_{12}|D) = \left|\frac{\partial(R_1U, R_1QW+R_1V, R_2W)}{\partial(U,V,W)}\right| = R_1{}^2R_2$

Now for $\begin{pmatrix} R_1, R_1Q \\ 0, R_2 \end{pmatrix} \varepsilon\Theta_{12}$ and $h_{12} = \begin{pmatrix} r_1, r_1q \\ 0, r_2 \end{pmatrix} \varepsilon\Theta_{12}$

we have $\begin{pmatrix} R_1, R_1Q \\ 0, R_2 \end{pmatrix} \begin{pmatrix} r_1, r_1q \\ 0, r_2 \end{pmatrix} = \begin{pmatrix} R_1r_1, R_1r_1q+R_1Qr_2 \\ 0, R_2r_2 \end{pmatrix}$

i.e. $(R_1,R_2,Q)(r_1,r_2,Q) = (R_1r_1, R_2r_2, q+\frac{r_2}{r_1}Q)$

Hence $J_*(\theta_{12}|h_{12}) = \left|\frac{\partial(R_1r_1, R_2r_2, q+\frac{r_2}{r_1}Q)}{\partial(r_1,r_2,q)}\right| = R_1R_2$

and $J_{*R}(\theta_{12}|h_{12}) = \left|\frac{\partial(r_1R_1, r_2R_2, Q+\frac{R_2}{R_1}q)}{\partial(r_1,r_2,q)}\right| = R_2{}^2$

Hence substituting the above into equation (2.3.25) we obtain the pivotal distribution

$$p(R_1,R_2,Q \mid x,y,z) \propto f\left\{\begin{pmatrix} R_1, R_1Q \\ 0, R_2 \end{pmatrix}^{-1} \begin{pmatrix} x, y \\ 0, z \end{pmatrix}\right\} \frac{1}{R_1{}^2R_2{}^3}$$

(2.5.13)

Secondly we consider the model as a combination of two conditional pivotal models.

Now $\begin{pmatrix} R_1, & 0 \\ 0 & ,R_2 \end{pmatrix}\begin{pmatrix} U & , & V \\ 0 & , & W \end{pmatrix} = \begin{pmatrix} R_1U, R_1V \\ 0 \quad ,R_2W \end{pmatrix} \quad \varepsilon\Omega$

where $\begin{pmatrix} R_1, & 0 \\ 0 & ,R_2 \end{pmatrix} = \theta_1\varepsilon\Theta_1$ and $\begin{pmatrix} U & , & V \\ 0 & , & W \end{pmatrix} = D\varepsilon\Omega$

Hence $J_\Omega(\theta_1|D) = \left|\dfrac{\partial(R_1U,R_1V,R_2W)}{\partial(U,V,W)}\right| = R_1{}^2R_2$

Now consider $\begin{pmatrix} R_1, & 0 \\ 0 & ,R_2 \end{pmatrix}$, $\begin{pmatrix} r_1, & 0 \\ 0 & ,r_2 \end{pmatrix} = h_1\varepsilon\Theta_1$

$$\begin{pmatrix} R_1, & 0 \\ 0 & ,R_2 \end{pmatrix}\begin{pmatrix} r_1, & 0 \\ 0 & ,r_2 \end{pmatrix} = \begin{pmatrix} R_1r_1, & 0 \\ 0 & ,R_2r_2 \end{pmatrix}$$

i.e. $(R_1,R_2)o(r_1,r_2) = (R_1r_1,R_2r_2)$

Hence $J_*(\theta_1|h_1) = \left|\dfrac{\partial(R_1r_1,R_2r_2)}{\partial(r_1,r_2)}\right| = R_1R_2$

and $J_{*R}(\theta_1|h_1) = \left|\dfrac{\partial(r_1R_1,r_2R_2)}{\partial(r_1,r_2)}\right| = R_1R_2$

Now $\begin{pmatrix} 1 & , & Q \\ 0 & , & 1 \end{pmatrix}\begin{pmatrix} U & , & V \\ 0 & , & W \end{pmatrix} = \begin{pmatrix} U & , & V+QW \\ 0 & , & W \end{pmatrix} \varepsilon\Omega$

where $\begin{pmatrix} 1 & , & Q \\ 0 & , & 1 \end{pmatrix} = \theta_2\varepsilon\Theta_2$ $\quad \begin{pmatrix} U & , & V \\ 0 & , & W \end{pmatrix} = D\varepsilon\Omega$

Hence $J_\Omega(\theta_2|D) = \left|\dfrac{\partial(U,V+QW,W)}{\partial(U,V,W)}\right| = 1$

Now consider $\begin{pmatrix}1 & , Q\\ 0 & , 1\end{pmatrix}$, $\begin{pmatrix}1 & , q\\ 0 & , 1\end{pmatrix} = h_2 \varepsilon \Theta_2$

$$\begin{pmatrix}1 & , Q\\ 0 & , 1\end{pmatrix}\begin{pmatrix}1 & , q\\ 0 & , 1\end{pmatrix} = \begin{pmatrix}1 & , Q+q\\ 0 & , 1\end{pmatrix}$$

i.e. $(Q)o(q) = (Q+q)$

Hence $J_*(\theta_2|h\ ) = \left|\frac{\partial(Q+q)}{\partial(q)}\right| = 1$

and $J_{*R}(\theta_2|h_2) = \left|\frac{\partial(q+Q)}{\partial(q)}\right| = 1$

Substituting the above into equation (2.5.11) we obtain the following distribution:

$$\tilde{p}(R_1,R_2,Q\,|x,y,z) \propto f\left\{\begin{pmatrix}R & ,R\ Q\\ 0 & ,\ R\end{pmatrix}^{-1}\begin{pmatrix}x & , & y\\ 0 & , & z\end{pmatrix}\right\}\frac{1}{R_1^{\,3}R_2^{\,2}} \tag{2.5.14}$$

$$\propto \frac{R_2}{R_1}\, p(R_1,R_2,Q\,|x,y,z)$$

so $\tilde{p}(R_1,R_2,Q\,|x,y,z) \neq p(R_1,R_2,Q\,|x,y,z)$.

□

By imbedding the model into a pivotal model, it is possible sometimes to obtain a distribution that has a joint expected posterior probability interpretation. The following example illustrates this procedure.

Example 2.5.4

Consider the model

$$x = \theta_1 o \theta_2 o e \qquad e \sim f(\cdot) \tag{2.5.15}$$

where $x, e \in \Omega = \{0,1,2,3,4,5\}$, $\theta_1 \in \{i,\alpha\}$, $\theta_2 \in \{i,\beta\}$ and the transformations ϕ_θ are given by the following table:

	Ω 0	1	2	3	4	5
ϕ_i	0	1	2	3	4	5
ϕ_α	1	0	3	2	5	4
ϕ_β	5	2	1	4	3	0

This model is a multipivotal model since each set of transformations $\{\phi_{\theta_1}\}$ and $\{\phi_{\theta_2}\}$ form groups of order 2, but the joint model is not a pivotal model.

By considering the set of four transformations $\{\phi_{\theta_{12}} : \theta_{12} \varepsilon \{i,\alpha,\beta,\alpha\beta\}\}$ obtained by taking every combination of the two sets of transformations, we see that by adding just two transformations ϕ_γ and ϕ_δ to this set we obtain a pivotal model (2.5.16) into which the model of (2.5.15) is imbedded, where

$$x = \theta_{12} o e \qquad e \sim f(\cdot) \tag{2.5.16}$$

where $x, e \varepsilon \Omega = \{0,1,2,3,4,5\}$ and θ_{12} indexes a set of transformations $\{\phi_{\theta_{12}} : \theta_{12} \varepsilon (i,\alpha,\beta,\alpha\beta,\gamma,\gamma)\}$ defined to be

		Ω					
		0	1	2	3	4	5
	i	0	1	2	3	4	5
	α	1	0	3	2	5	4
θ_{12}	β	5	2	1	4	3	0
	$\alpha\beta$	2	5	4	1	0	3
	γ	4	3	0	5	1	2
	δ	3	4	5	0	2	1

From this pivotal model we can obtain a pivotal distribution that has a joint expected posterior probability interpretation of the parameter θ_{12} and hence of the parameters θ_1 and θ_2. This is because we have shown in theorem 2.4.1 that the expected posterior probability interpretation is true no matter what conditional prior distribution $\Pi(\theta_{12}|e)$ we might choose. In this particular case we know that the support of this unknown conditional prior distribution $\Pi(\theta_{12}|e)$ lies in the set $\{i,\alpha,\beta,\alpha\beta\}$.

Thus by imbedding the multipivotal model into a pivotal model we have obtained a distribution of the parameters θ_1 and θ_2 that has a joint expected posterior probability interpretation. □

Unfortunately there can be various ways of embedding some multipivotal models in a pivotal model which leads to different distributions on the parameters. An example of this will be given in section 3.2 the ellipse fitting problem.

Note in the case of finite and countably finite multipivotal models whatever the imbeddings there will be a unique parameter distribution, that derived from combining the conditional pivotal distribution using (2.5.10). This is because all the Jacobian products

$$\frac{J_{i,\Omega}(\theta_i \ id_i|y_i)J_{i,*R}(\theta_i \ id_i,id_i|y_i)}{J_{i,*}(\theta_i \ id_i,id_i|y_i)J_{i,*}(\theta_i,id_i|y_i)} \quad i = 1,\ldots,m$$

take the value one.

In summary, with multipivotal models we have shown how to obtain the multipivotal distribution which will have a conditional expected posterior probability interpretation and noted that in general this distribution will not have a joint expected posterior probability interpretation. We have also shown that sometimes it is possible to obtain a distribution of the parameters in the model that has a joint expected posterior probability interpretation, but with a general multipivotal model this is not always possible.

3 APPLICATIONS OF MULTIPIVOTAL MODELS TO SHAPE FITTING

In this chapter we look at the problem of fitting a shape from a class of geometrical shapes, such as circles or cones, to a set of data that lies approximately on a shape from the class. This specific application is called the Shape Fitting Problem.

We will illustrate how the methods developed in the previous chapter can be applied to the shape fitting problem for various classes of geometrical shapes and will compare the results obtained by these methods with other techniques. The examples given come from the engineering industry and Avebury and Brogar Megalithic stone rings.

3.1 The Circle Fitting Problem

In this section we look at fitting a circle through a set of data that lies approximately on a circle or arc of a circle.

When the members of the class of geometrical shapes, defined in the shape fitting problem, are all two dimensional we will call the problem a two dimensional shape fitting problem. For such problems we will assume that the set of data points is in the form of a profile.

A profile is defined as a finite collection of data points lying in a fixed plane. Each data point is given rectangular co-ordinates (x,y).

This definition of a profile was given in Scott (1981) and will be used throughout this thesis and is the one used in the engineering industry, even though this is not the usual meaning.

Let the data points for the circle fitting problem be the profile (x_i,y_i) $i = 1,\ldots,n$. The radius and co-ordinates of the centre of the "true" circle are taken to be unknown, so the problem is reduced to one of obtaining estimates or "confidence" intervals of these particular parameters.

The circle fitting problem was discussed in some detail in Scott (1981) where two methods were proposed. The first was a curve fitting technique and was called "Modified Least Squares". It consisted of minimising with respect to a, b and c the function

$$\sum_{i=1}^{n} d_i^2 \qquad (3.1.1)$$

where $d_i = \dfrac{x_i^2 + y_i^2}{2} - ax_i - by_i - c \qquad i = 1,\ldots,n.$

The values of a, b and c that achieve this minimisation $\hat{a}$, $\hat{b}$ and $\hat{c}$ say provide estimates of the co-ordinates of the centre of the circle $(\hat{a},\hat{b})$ and an estimate of the radius $\hat{r} = (2\hat{c} + \hat{a}^2 + \hat{b}^2)^{\frac{1}{2}}$.

Note Angel and Barber (1977) minimised the function

$$\sum_{i=1}^{n} (d_i - \overline{d})^2$$

with respect to a and b, where

$$\overline{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$$

and solved $\overline{d} = 0$ for c to obtain estimates of a, b and c; this can be shown to be equivalent to the modified least squares procedure.

The second method consisted of writing the circle fitting problem in terms of a conditional structural model, as defined in Fraser (1968, part II), as follows:

$$(\underline{x},\underline{y}) = [a,b,r]o(\underline{e},\underline{d}) \qquad (3.1.2)$$
$$(\underline{e},\underline{d}) \sim f(\cdot|\underline{v})$$

where $(\underline{x},\underline{y}), (\underline{e},\underline{d}) \in \Omega = \mathbb{R}^n \times \mathbb{R}^n$. $[a,b,r] \in \Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ index a set of transformations $\phi_\theta : \Omega \to \Omega \;\; \theta \in \Theta$ defined to be $[a,b,r]o(\underline{e},\underline{d}) = (a\underline{1} + r\underline{e},\, b\underline{1} + r\underline{d})$. This set of transformation form a group under composition and the errors have a distribution function $f(\cdot|\underline{v})$ indexed by the parameters $\underline{v}$.

From this model a conditional structured distribution of [a,b,r] given $\underline{v}$ $(\underline{x},\underline{y})$ was obtained,

$$p\{[a,b,r]|\underline{v},(\underline{x},\underline{y})\} =$$

$$k\{\underline{v},(\underline{x},\underline{y})\}f\{[a,b,r]^{-1}o(\underline{x},\underline{y})|\underline{v}\}r^{-(2n+1)} \qquad (3.1.3)$$

where $k\{\underline{v},(\underline{x},\underline{y})\}$ is the constant of proportionality.

Also a marginal likelihood function for $\underline{v}$.

$$L\{(\underline{x},\underline{y})|\underline{v}\} = 1/k\{\underline{v},(\underline{x},\underline{y})\} \qquad (3.1.4)$$

An estimate of the parameters $\underline{v}$,$\hat{\underline{v}}$ say was obtained by maximising the marginal likelihood function with respect to the $\underline{v}$'s.

The values of $\hat{\underline{v}}$ were then substituted into the conditional structured distribution (3.1.3), from which it is possible to make various inferences about the parameters a, b and r. For instance it was suggested that one possible way to obtain estimates of the parameters a, b and r was to use those values that maximise the conditional structured distribution.

Finally one possible distribution was suggested for $f(\cdot|\underline{v})$ which was based on the assumption that each point (e_i,d_i) $i = 1,\ldots,n$ was independent and identically distributed radially from the origin as a Normal distribution with mean 1 and variance v.

Several authors have suggested using maximum likelihood. This requires a probability model for the distribution of the data points. For example, Mardia and Holmes (1980) use a construction based on maximising entropy to obtain the probability density function of a single data point $(x_i, y_i) = \underline{x}_i$ of the form

$$p(\underline{x}_i | \underline{a}, \Sigma, k) =$$

$$C(k)|\Sigma|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}k\{(\underline{x}_i-\underline{a})^T \Sigma^{-1}(\underline{x}_i-\underline{a})-1\}^2] \qquad (3.1.5)$$

where C(k) is the constant of proportionality; $\underline{a}^T = (a,b) \varepsilon \mathbb{R}^2$; Σ is a 2x2 symmetric nonsingular matrix, and k is a positive concentration parameter.

The density has a mode on the quadratic form

$$(\underline{x}-\underline{a})^T \; \Sigma^{-1}(\underline{x}-\underline{a})-1 \qquad (3.1.6)$$

For the circle fitting problem $\Sigma = r\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Estimates of these parameters are obtained by taking those values $\hat{a}$, $\hat{b}$, $\hat{r}$ and $\hat{k}$ say, that maximise the likelihood function, assuming the data points are independent, given by (3.1.5). Then $(\hat{a},\hat{b})$ estimates the centre of the "true" circle and $\hat{r}$ its radius; $\hat{k}$ gives an indication of how well a fit the model was.

The following is suggested as an alternative to the above methods for the circle fitting problem. It is based on the work of Chapter two.

We start by defining a "datum profile".

A datum profile is a profile whose data points (s_i,t_i) $i = 1,\ldots,n$ are assumed to lie on the perimeter of a known geometrical shape.

In pivotal models (and multipivotal models) we will assume that the unknown random variables, mentioned in (2.1.1), lie approximately on a datum profile with a known error distribution about the datum profile. The error distribution giving an indication of how good a fit the random variables are to the datum profile.

For the circle fitting problem we will assume that the datum profile is a unit circle with its centre at the origin. In the multipivotal model it is convenient to assume that each point in the error distribution is independent identically distributed radially, about the datum profile, as a truncated N(1,1) distribution, restricted to positive values. The truncation is because radial directions have to be positive.

Transforming the distribution to cartesian co-ordinates we obtain

$$f(\underline{e},\underline{d}) \propto \exp\{-\tfrac{1}{2}\sum_{i=1}^{n}[(d_i^2+e_i^2)^{\frac{1}{2}}-1]^2\}\prod_{i=1}^{n}\frac{1}{(d_i^2+e_i^2)^{\frac{1}{2}}} \tag{3.1.7}$$

The full multipivotal model we shall consider for the circle fitting problem is as follows:

$$(\underline{x},\underline{y}) = [a,b,r]_1 o [\sigma]_2 o (\underline{e},\underline{d}) \qquad (3.1.8)$$

$$(\underline{e},\underline{d}) \sim f(.) \text{ given by } (3.1.7)$$

where $(\underline{x},\underline{y}),(\underline{e},\underline{d}) \varepsilon \Omega = \mathbb{R}^n x \mathbb{R}^n$ and both $[a,b,r]_1 \varepsilon \Theta_1$ and $[\sigma]_2 \varepsilon \Theta_2$ index the transformations $\{\phi_{\theta_1} : \theta_1 \varepsilon \Theta_1 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ and $\{\phi_{\theta_2} : \theta_2 \varepsilon \Theta_2 = \mathbb{R}^+\}$ respectively, defined on the ith element of $(\underline{s},\underline{t}) \varepsilon \Omega$ to be

$$[a,b,r]_1 o (s_i,t_i) = (a+rs_i,\ b+rt_i) \qquad (3.1.9)$$

$$[\sigma]_2 o (s_i,t_i) = \left(\frac{\{(u_i-1)\sigma+1\}s_i}{u_i} , \frac{\{(u_i-1)\sigma+1\}t_i}{u_i} \right) \qquad (3.1.10)$$

where $u_i = \sqrt{s_i^2+t_i^2}$ with $[a,b,r]_1 \varepsilon \Theta_1$, $[\sigma]_2 \varepsilon \Theta_2$ fixed.

Both sets of transformations form groups on their respective indexing spaces namely

$$[A,B,R]_1 * [a,b,r]_1 = [A+Ra,\ B+Rb,\ Rr]_1 \varepsilon \Theta_1 \qquad (3.1.11)$$

when $[A,B,R]_1 , [a,b,r]_1 \varepsilon \Theta_1$

and $\quad [\Sigma]_2 * [\sigma]_2 = [\Sigma\sigma]_2 \varepsilon \Theta_2 \qquad (3.1.12)$

when $[\Sigma]_2 , [\sigma]_2 \varepsilon \Theta_2$

The group action of $\{\phi_{\theta_2}:\theta_2\varepsilon\Theta_2\}$ on the sample space Ω has the effect of altering the dispersion of the unknown random variables about the datum profile in a radial direction and can be used to measure the goodness of fit of the multipivotal model to the data, resembling the concentration parameter k given in (3.1.5).

The group action of $\{\phi_{\theta_1}:\theta_1\varepsilon\Theta_1\}$ on the sample space Ω has the effect of enlarging a profile by a scaling factor of r and translating the origin to (a,b). Since we have assumed that the unknown random variables lie approximately on a unit circle with a centre at the origin, the transformed profile will be approximately on a circle with a radius r and a centre at the point (a,b).

Having defined our model we can now proceed to calculate the multipivotal distribution for the parameters using the results from Chapter two.

From (3.1.9), (3.1.10), (3.1.11) and (3.1.12) we can calculate the relevant Jacobians:

$$J_{1,\Omega}([a,b,r] \mid (\underline{s},\underline{t})) = r^{2n}$$

$$J_{1,*}([a,b,r], [0,0,1]) = r \qquad (3.1.13)$$

$$J_{1,*R}([a,b,r], [0,0,1]) = r$$

$$J_{2,\Omega}([\sigma]_2|(\underline{s},\underline{t})) = \sigma^n \prod_{i=1}^{n} \left[\frac{(u_i-1)\sigma+1}{u_i}\right]$$

$$J_{2,*}([\sigma]_2,[1]_2) = \sigma \qquad (3.1.14)$$

$$J_{2,*R}([\sigma]_2,[1]_2) = \sigma$$

where $[a,b,r]_1$, $[0,0,1]_1 \varepsilon \Theta_1$ $[\sigma]_2$, $[1]_2 \varepsilon \Theta_2$ $[0,0,1]_1$, $[1]_2$ being the identity elements of the respective groups and $(\underline{s},\underline{t}) \varepsilon \Omega$ with $u_i = \sqrt{s_i^2+t_i^2}$ $i = 1,\ldots,n$.

Substituting the above into equation (2.5.11), we obtain the multipivotal distribution for the parameters as follows:

$$p(a,b,r,\sigma|(\underline{x},\underline{y})) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left[\frac{r_i}{r} - 1\right]^2\right\} \frac{1}{(\sigma r)^{n+1}} \prod_{i=1}^{n} \frac{1}{r_i} \qquad (3.1.15)$$

where $r_i = \sqrt{\{(x_i-a)^2+(y_i-b)^2\}}$

We will use the multipivotal distribution to construct the shortest interval of the parameters a,b,r given σ that has approximately 95% probability attached to it by the distribution, knowing that from results given in section 2.5 the probability assigned to a particular interval of the parameters by the distribution has a conditional expected posterior probability interpretation. The approximation is due to the fact that the interval is calculated from an asymptotic expansion of the multipivotal distribution about its maximum with respect to the parameters. The details of the calculation of the maximum are as follows. For computational convenience we reparameterise the distribution to

$$\exp\left\{-\frac{1}{2\tau^2}\sum_{i=1}^{n}[r_i-r]^2\right\}\frac{1}{\tau^{n+1}}\prod_{i=1}^{n}\frac{1}{r_i} \quad . \qquad (3.1.16)$$

where $\tau = r\sigma$ and $r_i = \sqrt{[(x_i-a)^2+(y_i-n)^2]}$.

Taking minus the log of the reparameterised distribution we obtain an equivalent minimisation problem:

$$\min_{a,b,r,\tau}\left\{\frac{1}{2\tau^2}\sum_{i=1}^{n}[r_i-r]^2 + (n+1)\log\tau + \sum_{i=1}^{n}\log r_i\right\} \qquad (3.1.17)$$

The above equation has to be solved numerically; an iterative procedure was used, details can be found in Appendix 3.1. The starting values were taken to be the solutions of the modified least squares method (3.1.1), this being a non iterative procedure.

Denote the values of the parameters that maximise the multipivotal distribution by $\hat{a},\hat{b},\hat{r}$ and $\hat{\sigma}$. (Note $\hat{\sigma}$ is calculated by $\hat{\sigma} = \hat{\tau}/\hat{r}$ where $\hat{\tau}$ is the value that minimises (3.1.17) together with $\hat{a},\hat{b}$ and $\hat{r}$).

To calculate an approximate 95% pivotal probability interval of a,b,r given $\hat{\sigma}$, we use a multivariate normal approximation to the multipivotal distribution about its maximum. The multivariate normal distribution has its mean at the maximum and its covariance matrix equal to the inverse of the matrix of second derivatives of minus the log of the multipivotal distribution calculated at its maximum.

The interval we will use is the central 95% probability ellipsoid of the multivariate normal distribution. Experience has shown that this gives an adequate approximation to the desired interval. It is possible to improve on this interval by using an iterative procedure based on gaussian quadrature which uses hermitian polynomials (see Naylor & Smith, 1982), but this is computationally very heavy and so will not be used.

A listing of a computer program that calculates the maximum of the multipivotal distribution and the covariance matrix of the multivariate distribution is given in Appendix 3.2.

The first numerical example is the calculation for the Brogar Megalithic stone ring shown in Fig. 3.1.1. This ring was surveyed by Thom and Thom (1973), who gave the positions of each stone in polar co-ordinates, with radii measured to the nearest tenth of a foot and angles to the nearest tenth of a degree. In table 3.1.2 the positions of the stones in rectangular co-ordinates are given to the nearest hundredth of a foot. It was decided to give the position of the stones to an extra decimal place so that the conversion to rectangular co-ordinates does not have an effect on the final results.

Thom in the above paper suggested that the builders of the Brogar ring intended that its shape be a circle with a diameter of 125 MY (Megalithic Yards). Assuming this hypothesis of exactly 125 MY to be correct, the fitting of a circle provides an estimate of the value of the Megalithic Yard in feet.

Figure 3.1.1 Plan of Brogar Megalithic Stone Ring

Brogar is on the Orkney Islands. The plan indicates for certain stones identification numbers given to them by archaeologists.

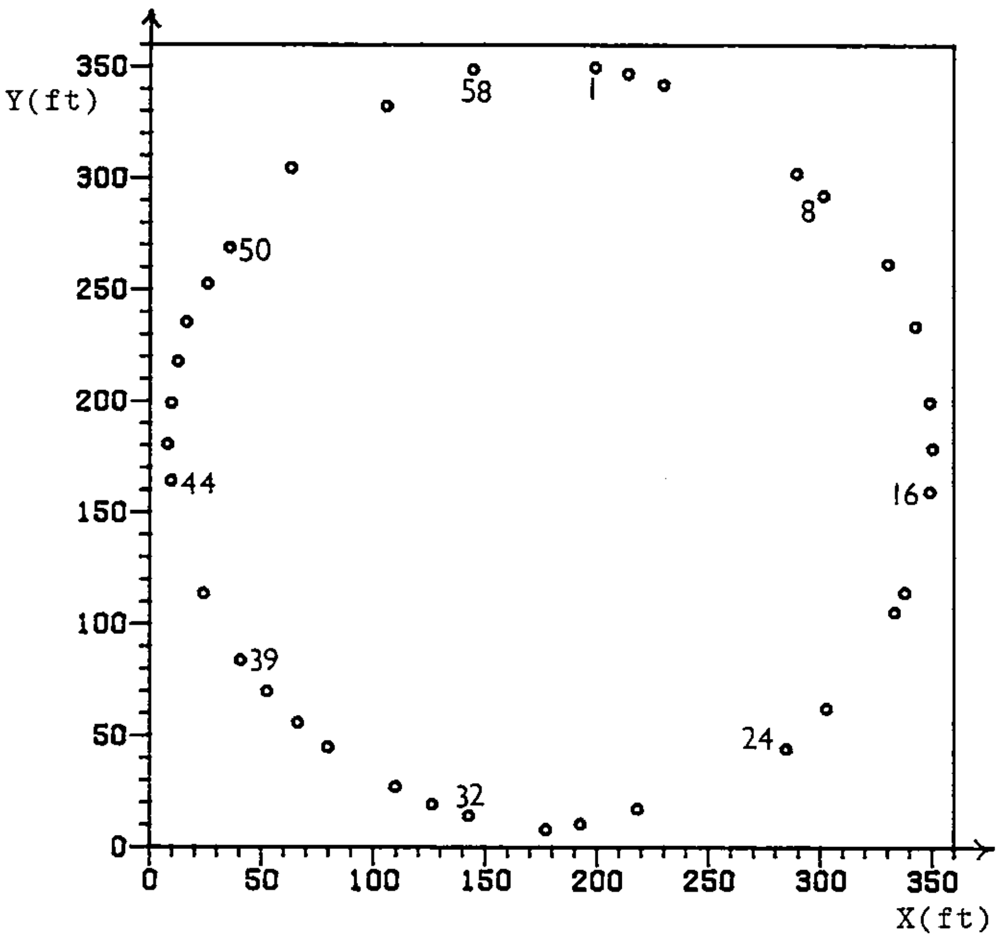


Table 3.1.2 Position of the stones in Brogar Ring

Stone	X	Y
1	199.06	349.93
2	213.99	347.08
3	229.84	342.01
7	289.31	302.26
8	301.36	292.38
10	330.28	261.59
12	342.64	233.79
14	349.07	199.56
15	350.40	178.81
16	349.07	159.54
19	337.78	114.32
20	333.29	105.57
22	302.70	62.34
24	284.58	44.20
28	218.16	17.32
29	192.44	10.56
30	177.00	8.03
32	142.63	14.16
33	126.25	19.39
34	109.72	27.19
36	79.66	44.89
37	66.23	55.84
38	52.37	69.83
39	40.56	83.81
41	23.88	113.73
44	9.43	164.18
45	7.90	180.60
46	9.67	199.11
47	12.50	217.75
48	16.32	235.40
49	25.64	252.64
50	35.55	268.87
53	63.23	304.79
56	105.65	332.44
58	144.41	348.89

Co-ordinates are all in feet.

We shall use two alternative methods for comparison with our pivotal procedure. These are firstly the one Thom developed in his original paper, which is an iterative scheme that is equivalent to a least squares fit of a circle in a radial direction. That is to say we find those values of a, b and r that minimise

$$\sum_{i=1}^{n} \{\sqrt{[(x_i-a)^2+(y_i-b)^2]}-r\}^2 \qquad (3.1.18)$$

where (x_i,y_i) $i = 1,\ldots,n$ are the positions of the stones. The second is Modified Least Squares (3.1.1).

The results are given in the following table.

Table 3.1.3 Brogar Ring Results

	Thom	MLS	Pivotal $\hat{\sigma} = 1.03E-2$
a	179.52	179.52	179.50
b	180.26	180.26	180.24
r	170.01	170.01	170.00
MY	2.7202	2.7202	2.7200

All entries are in feet.

The approximate 95% Pivotal Probability ellipsoid calculated from the multivariate approximation is:

$$(\underline{\theta}-\underline{\theta}_0)^T \Sigma^{-1}(\underline{\theta}-\underline{\theta}_0) < 7.81 \qquad (3.1.19)$$

where $\underline{\theta}^T = (a,b,r)$, $\underline{\theta}_0^T = (\hat{a},\hat{b},\hat{r})$

$$\text{and } \Sigma = \begin{pmatrix} 4.0739E-2 & & \\ 1.0988E-4, & 4.6989E-3 & \\ 1.6826E-3, & 5.8814E-3, & 2.2793E-3 \end{pmatrix}$$

the covariance matrix of the approximation.

As can easily be seen all three methods give similar results; this is because Brogar approximates very closely to a full circle. This is not true for our next example.

Avebury Megalithic stone ring (Fig. 3.1.4) is the largest and most complicated in Europe. Details of the survey of this ring can be found in Thom and Thom (1978). Table 3.1.5 gives the positions of the stones in rectangular co-ordinates to the nearest tenth of a foot. As can easily be seen Avebury is far from circular. Thoms' geometrical interpretation of this site consisted of a series of arcs of circles. When the centres of the arcs are linked they form right angle Pythagorean triangles, that is to say the lengths of the sides of the triangles take integer values when measured in Megalithic Yards. The radii of the arcs themselves are also assumed to be integer multiples of a MY. This extremely complicated design is explained in greater detail in Thom et al (1976).

The alternative comparisons we shall employ on the individual arcs are: the Modified Least Squares method (3.1.1) and one by Thom, who fitted the whole set of arcs simultaneously, again by an iterative method which is described in Thom and Thom (1978). The results are given in Table 3.1.6.

There is a large discrepancy between Thom's method and the others. We conclude that it is very unlikely that Thom's geometrical interpretation of Avebury is correct.

Figure 3.1.4 Plan of Avebury Megalithic Stone Ring

Avebury is in Wiltshire. The plan indicates for certain stones identification numbers given to them by archaeologists.

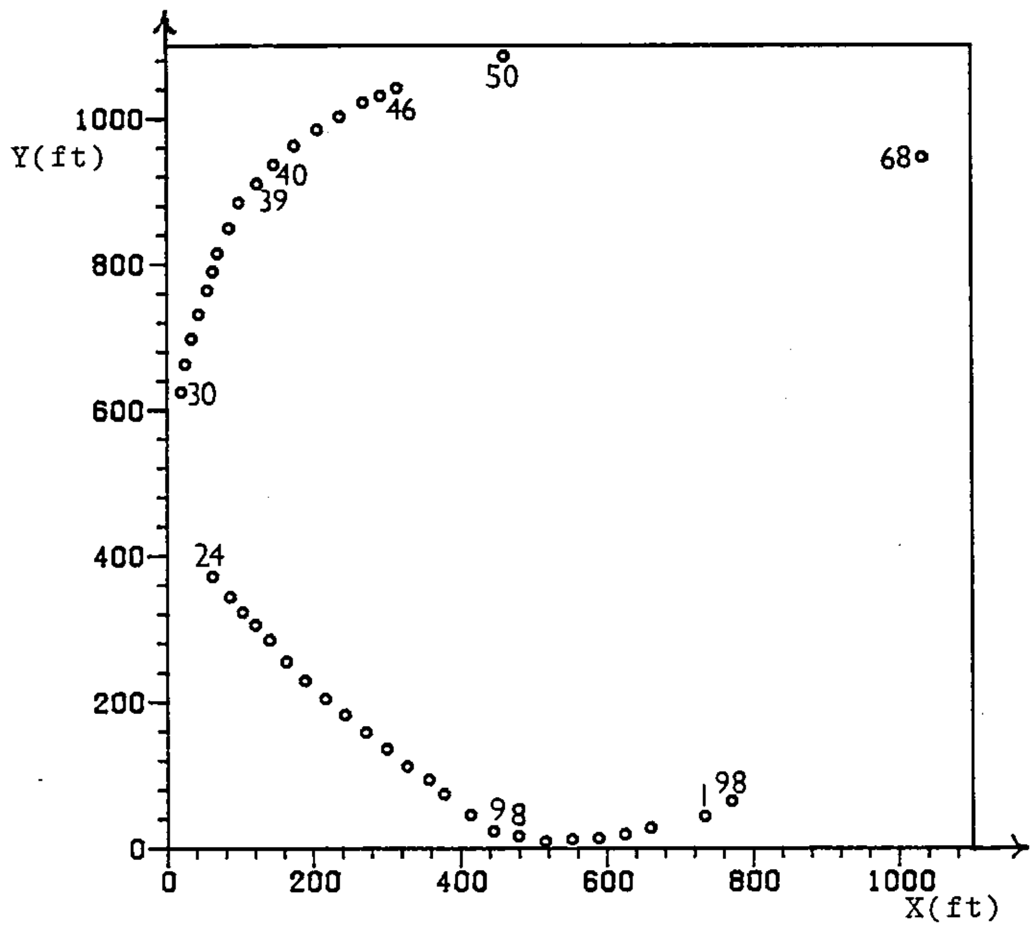


Table 3.1.5 Position of the stones in Avebury Ring

Stone	X	Y
1	733.7	44.0
3	659.7	28.0
4	624.2	19.3
5	588.4	13.9
6	441.6	12.3
7	515.1	9.5
8	478.0	16.6
9	445.3	23.4
10	413.8	46.2
11	377.9	74.1
12	357.1	94.1
13	327.7	112.4
14	300.6	136.2
15	272.0	158.8
16	243.5	183.0
17	216.3	205.0
18	188.9	229.8
19	163.5	255.5
20	140.0	285.0
21	120.6	305.7
22	103.1	323.1
23	85.9	344.0
24	61.8	371.3

continued overleaf

30	19.3	624.4
31	24.9	663.0
32	33.3	698.3
33	43.7	731.3
34	55.5	764.4
35	62.9	790.1
36	69.2	815.0
37	85.0	849.8
38	98.5	884.6
39	123.6	910.5
40	146.8	936.9
41	175.2	962.4
42	206.7	984.7
43	237.6	1002.9
44	270.3	1022.5
45	292.5	1031.2
46	315.8	1042.0
50	461.1	1085.4
68	1033.4	946.2
98	769.9	64.9

Co-ordinates are all in feet.

Table 3.1.6 Avebury Ring Results

	Thom	MLS	Pivotal Max.	Σ covariance matrix
Arc 1 Stones 1-8			$\hat{\sigma}$ = 2.86E-3	
a	520.9	538.6	537.2	3.522E0
b	720.8	585.0	596.9	1.137E1, 3.683E1
r	707.7	573.0	584.8	-4.176E-1, -1.99E0, 8.089E2
Arc 2 Stones 9-24			$\hat{\sigma}$ = 9.93E-4	
a	1612.8	1375.9	1386.9	9.969E2
b	1697.1	1446.8	1459.1	-1.075E-1, 2.503E0
r	2041.5	1697.6	1714.0	6.705E-2, 1.708E0, 2.593E3
Arc 3 Stones 30-39			$\hat{\sigma}$ = 3.55E-3	
a	723.2	667.1	690.9	1.638E-1
b	538.6	561.1	552.7	5.605E-2, 6.424E0
r	707.7	648.6	673.5	1.562E-1, -2.074E0, 1.089E3
Arc 4 Stones 40-46			$\hat{\sigma}$ = 1.13E-3	
a	586.6	503.2	504.6	4.773E-2
b	386.9	548.5	546.2	7.513E-2, 3.366E-1
r	707.7	527.4	530.1	2.520E-2, -1.479E-1, 2.589E2

All entries in the table are in feet.

3.2 The Ellipse Fitting Problem

In this section we look at fitting an ellipse through a set of data that lies approximately on an ellipse or part of an ellipse.

Let the data points for the problem be a profile (x_i, y_i) $i = 1,\ldots,n$. The lengths and orientation of the major and minor axes and the centre of the "true" ellipse are taken to be unknown. The problem is reduced to one of obtaining estimates or "confidence" intervals for these particular parameters.

Modified Least Squares is easily adapted to the ellipse fitting problem. It consists of minimising with respect to A, B, C, D and E the function

$$\sum_{i=1}^{n} d_i^2 \tag{3.2.1}$$

where $d_i = x_i^2 + Ax_iy_i + By_i^2 + Cx_i + Dy_i + E$.

Denote those values of the parameters that minimise (3.2.1) by $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$ and $\hat{E}$. From these it is not difficult to calculate the lengths and orientation of the major and minor axes and the co-ordinates of the centre of the fitted ellipse.

The probability model (3.1.5) for the distribution of profile points used by Mardia & Holmes was originally constructed for the ellipse fitting problem, with Σ being any non-singular positive definite matrix. The values of the parameters $\underline{a}$, Σ and k that maximise the resulting likelihood function can then be used to obtain estimates of the lengths and orientation of the major and minor axes as well as the co-ordinates of the centre of the "true" ellipse.

There are two possible multipivotal models that could be used for the ellipse fitting problem. The choice of which one to use should depend on the interpretation of the resulting multipivotal distribution that is required for the particular application being used. We will present both models here.

In both of the alternative models, the datum profile will be the same as that used for the circle fitting problem, namely a unit circle with its centre at the origin. The distribution of the unknown random variables (e_i, d_i) $i = 1, \ldots, n$ about this particular datum profile is also given by (3.1.7).

The first multipivotal model we shall consider for the ellipse fitting problem is as follows:

$$(\underline{x}, \underline{y}) = [M, \underline{\alpha}]_1 \circ [\sigma]_2 \circ (\underline{e}, \underline{d}) \qquad (3.2.2)$$

$$(\underline{e}, \underline{d}) \sim f(\cdot) \text{ given by } (3.1.7)$$

where $(\underline{x}, \underline{y}), (\underline{e}, \underline{d}) \in \Omega = \mathbb{R}^n \times \mathbb{R}^n$ and both

$[M,\underline{\alpha}]_1 \varepsilon \Theta_1$ and $[\sigma]_2 \varepsilon \Theta_2$ index the transformations $\{\phi_{\theta_1} : \theta_1 \varepsilon \Theta_1 = \Sigma \times \mathbb{R}^2$ and $\{\phi_{\theta_2} : \theta_2 \varepsilon \Theta_2 = \mathbb{R}^+\}$ respectively, where Σ is the space of non-singular positive definite 2x2 real matrices.

The first set of transformations $[M,\underline{a}]_1 \varepsilon \Theta_1$ is defined on the ith element of $(\underline{s},\underline{t}) \varepsilon \Omega$ to be

$$[M,\underline{\alpha}]_1 o(s_i,t_i) = (m_1 s_i + m_2 t_i + a, m_3 s_i + m_4 t_i + b) \quad (3.2.3)$$

where $M = \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix} \varepsilon \Sigma$ and $\underline{\alpha} = \begin{pmatrix} a \\ b \end{pmatrix} \varepsilon \mathbb{R}^2$ are fixed

and the second set of transformations $[\sigma]_2 \varepsilon \Theta_2$ is defined by (3.1.10).

The relevant Jacobians for the calculation of the multipivotal distribution are (3.1.14) and

$$J_{1,\Omega}([M,\underline{\alpha}]_1 | (\underline{s},\underline{t})) = (m_1 m_4 - m_2 m_3)^n$$

$$J_{1,*}([M,\underline{\alpha}]_1, [I,\underline{0}]_1) = (m_1 m_4 - m_2 m_3)^3 \quad (3.2.4)$$

$$J_{1,*R}([M,\underline{\alpha}]_1, [I,\underline{0}]_1) = (m_1 m_4 - m_2 m_3)^2$$

where $[M,\underline{\alpha}]_1$, $[I,\underline{0}]_1 \varepsilon \Theta_1$; $[I,\underline{0}]_1$ being the identity element of this particular set of transformations and $(\underline{s},\underline{t}) \varepsilon \Omega$.

Substituting the above into equation (2.5.11), we obtain the multipivotal distribution for the parameters as follows:

$$p([M,\underline{a}]_1,[\sigma]_2\ (\underline{x},\underline{y})) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2}(U_i-1)^2\right\} \frac{1}{(m_1m_4-m_2m_3)^{n+2}} \frac{1}{\sigma}^{n+1} \prod_{i=1}^{n} \frac{1}{U_i}$$

(3.2.5)

where $U_i = \sqrt{\{(x_i-a,y_i-b)\ M^TM\ (x_i-a,y_i-b)^T\}}$

The second multipivotal model we shall consider for the ellipse fitting problem is as follows:

$$(\underline{x},\underline{y}) = [a,b]_1 o[\theta]_2 o[r_1,r_2]_3 o[\sigma]_4 o(\underline{e},\underline{d}) \qquad (3.2.6)$$

$(\underline{e},\underline{d}) \sim f(\cdot)$ given by (3.1.7)

where $(\underline{x},\underline{y})$, $(\underline{e},\underline{d})\varepsilon\Omega = \mathbb{R}^n \times \mathbb{R}^n$ and $[a,b]_1$, $[\Psi]_2$, $[r_1,r_2]_3$ and $[\sigma]_4$ index the transformations $\{\phi_{\theta_1}:\theta_1\varepsilon\Theta_1 = \mathbb{R}^2\}$, $\{\phi_{\theta_2}:\theta_2\varepsilon\Theta_2 = [0,2\pi)\}$, $\{\phi_{\theta_3}:\theta_3\varepsilon\Theta_3 = \mathbb{R}^+ \times \mathbb{R}^+\}$ and $\{\phi_{\theta_4}:\theta_4\varepsilon\Theta_4 = \mathbb{R}^+\}$ respectively, which are defined on the ith element of $(\underline{s},\underline{t})\varepsilon\Omega$ to be:

$$[a,b]_1 o(s_i,t_i) = (s_i+a,t_i+b) \qquad (3.2.7)$$

$$[\Psi]_2 o(s_i,t_i) = (s_i\cos\Psi-t_i\sin\Psi,s_i\sin\Psi+t_i\cos\Psi)$$

(3.2.8)

$$[r_1,r_2]_3 o(s_i,t_i) = (r_1s_i,r_2t_i) \qquad (3.2.9)$$

and $[\sigma]_4 o(s_i,t_i)$ defined by (3.1.10).

The relevant Jacobians for the calculation of the multipivotal distribution are (3.1.14) and

$$J_{1,\Omega}([a,b]_1|(\underline{s},\underline{t})) = 1$$

$$J_{1,*}([a,b]_1,[0,0]_1) = 1 \qquad (3.2.10)$$

$$J_{1,*R}([a,b]_1,[0,0]_1) = 1$$

$$J_{2,\Omega}([\Psi]_2|(\underline{s},\underline{t})) = 1$$

$$J_{2,*}([\Psi]_2,[0]_2) = 1 \qquad (3.2.11)$$

$$J_{2,*R}([\Psi]_2,[0]_2) = 1$$

$$J_{3,\Omega}([r_1,r_2]_3|(\underline{s},\underline{t})) = (r_1r_2)^n$$

$$J_{3,*}([r_1,r_2]_3,[1,1]_3) = r_1r_2 \qquad (3.2.12)$$

$$J_{3,*R}([r_1,r_2]_3,[1,1]_3) = r_1r_2$$

where $[a,b]_1$, $[0,0]_1 \varepsilon \Theta_1$, $[\Psi]_2$, $[0]_2 \varepsilon \Theta_2$ and $[r_1,r_2]_3$, $[1,1]_3 \varepsilon \Theta_3$; $[0,0]_1$, $[0]_2$ and $[1,1]_3$ being the identity elements of the respective set of transformations and $(\underline{s},\underline{t}) \varepsilon \Omega$.

Substituting the above into equation (2.5.11), we obtain the multipivotal distribution for the parameters as follows:

$$p([a,b]_1,[\Psi]_2,[r_1,r_2]_3,[\sigma]_4 | (\underline{x},\underline{y})) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2}(\sqrt{[v_i^2+w_i^2]}-1)\right\}\frac{1}{(\sigma r_1 r_2)^{n+1}} \prod_{i=1}^{n} \frac{1}{\sqrt{[v_i^2+w_i^2]}}$$

(3.2.13)

where $(v_i,w_i) = [r_1,r_2]_3^{-1}o[\Psi]_2^{-1}o[a,b]_1^{-1}o(x_i,y_i)$.

As mentioned before the interpretation of the two above multipivotal distributions are different. From the first (3.2.2), it is possible to construct intervals of the parameters M and $\underline{\alpha}$ that have joint expected posterior probability interpretations given σ. From M and $\underline{\alpha}$ it is possible to calculate the parameters r_1, r_2, Ψ, a and b which denote the length and orientation of the major and minor axes and the co-ordinates of the centre of the "true" ellipse respectively. Thus we can construct intervals of these parameters that also have a joint expected posterior probability interpretation given σ.

The second multipivotal distribution (3.2.13) does not have the joint expected posterior probability interpretation given σ but a conditional expected posterior probability interpretation for each of the parameters indexing a particular set of transformations given the values of the parameters indexing the other sets of transformations. For example it is possible to construct intervals of the parameters r_1 and r_2 given Ψ, a, b and σ that have the above interpretation, since r_1 and r_2 index the third set of transformations defined by the model (3.2.6).

The example we will use to illustrate the ellipse fitting problem comes from the engineering industry. The blueprint of a certain engineered component specifies that its shape should be an ellipse with major and minor radii of 12.7 mm and 6.35 mm respectively with a tolerance of ± 2 μm. Table 3.2.1 gives the measured profile of a particular component measured in millimetres. We wish to find if this particular part conforms to the blueprint specification.

The multipivotal model we will use is the second one given in (3.2.6), this is because the only parameters we are interested in are the lengths of the major and minor axes. The other parameter values are mainly due to setting up errors when the component was originally measured.

Again we will use a multivariate normal approximation of the multipivotal distribution abouts its maximum with respect to the parameters a, b, r_1, r_2, Ψ and σ, to calculate an approximate 95% conditional pivotal probability interval of r_1 and r_2 given a, b, Ψ and σ. For computational convenience we reparameterise the distribution to:

Table 3.2.1 Profile Measurements in mm of an Engineered Components

X	Y
0.329387	-0.333805
0.562751	-0.270705
0.797104	-0.210446
1.032404	-0.153019
1.268597	-0.098341
1.505636	-0.046362
1.743488	0.002926
1.982115	0.049570
2.221479	0.093604
2.461551	0.135054
2.702301	0.173944
2.943700	0.210303
3.185721	0.244162
3.428342	0.275545
3.671542	0.304457
3.915303	0.330915
4.159610	0.354922
4.404448	0.376483
4.649794	0.395663
4.895650	0.412392
5.141997	0.426718
5.388830	0.438640
5.636141	0.448155
5.883926	0.455252
6.132184	0.459934
6.380912	0.462194
6.630114	0.462003
6.879787	0.459375
7.129941	0.454276
7.380583	0.446681
7.631721	0.436560
7.883365	0.423907
8.135527	0.408686
8.388226	0.390843
8.641482	0.370342
8.895314	0.347149
9.149739	0.321236
9.404794	0.292511
9.660505	0.260932
9.916909	0.226422

$$\exp\left\{-\frac{1}{2\sigma^2}(U_i-1)^2\right\}(q_1q_3-q_2^2)^{\frac{n+1}{2}}\frac{1}{\sigma}^{n+1}\prod_{i=1}^{n}\frac{1}{U_i}$$

(3.2.14)

where $U_i = \left[(x_i-a, y_i-b)\begin{pmatrix} q_1 & q_2 \\ q_2 & q_3 \end{pmatrix}\begin{pmatrix} x_i-a \\ y_i-b \end{pmatrix}\right]^{1/2}$

and $q_1 = \frac{\cos^2\psi}{r_1^2} + \frac{\sin^2\psi}{r_2^2}$

$q_2 = \cos\psi\sin\psi\left[\frac{1}{r_2^2} - \frac{1}{r_1^2}\right]$

$q_3 = \frac{\sin^2\psi}{r_1^2} + \frac{\cos^2\psi}{r_2^2}$

Taking minus the log of the reparameterised distribution we obtain an equivalent minimisation problem

$$\min_{a,b,q_1,q_2,q_3,\sigma}\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(U_i-1)+(n+1)\log\sigma - \frac{(n+1)}{2}\log(q_1q_3-q_2^2)\right\} \qquad (3.2.15)$$

The above equation has to be solved numerically. An iterative procedure was used, details can be found in Appendix 3.1. The starting values were calculated from the modified least squares estimates for the ellipse fitting problem.

Denote the values of the parameters that maximise the multipivotal distribution (3.2.13) by $\hat{a}$, $\hat{b}$, $\hat{r}_1$, $\hat{r}_2$, $\hat{\psi}$ and $\hat{\sigma}$. These can be calculated from the values of

a, b, q_1, q_2, q_3 and σ that minimise equation (3.2.15).

Table 3.2.2 summarises the results for this particular component giving the blueprint values, the modified least squares estimates and the values of the multi-pivotal distribution.

Table 3.2.2 Results for the component whose measurements are given in Table 3.2.1.

	Blueprint	MLS	Pivotal max.
r_1	12.700 mm	12.70046 mm	12.70057 mm
r_2	6.350 mm	6.35054 mm	6.35054 mm
a	-	6.48642 mm	6.48680 mm
b	-	-5.88812 mm	-5.88812 mm
Ψ	-	-0.00045°	0.00038°
σ	-	-	1.3665E-6

The approximate 95% conditional pivotal probability ellipsoid for r_1 and r_2 given $\hat{a}$, $\hat{b}$, $\hat{\Psi}$ and $\hat{\sigma}$, calculated from the multivariate approximation is

$$(\underline{\theta}-\underline{\theta}_0)^T \, \Sigma^{-1}(\underline{\theta}-\underline{\theta}_0) < 5.99$$

where $\underline{\theta}^T = (r_1,r_2)$, $\underline{\theta}_0{}^T = (\hat{r}_1,\hat{r}_2)$ (3.2.16)

$$\text{and } \Sigma = \begin{pmatrix} 9.048\text{E-}10, & \\ -1.183\text{E-}8, & 1.547\text{E-}7 \end{pmatrix}$$

is the covariance matrix of the multivariate normal approximation. This is the inverse of the matrix of second derivatives with respect to r_1 and r_2 of the log of the reciprocal of the multipivotal density evaluated at its maxi. A listing of a computer program that calculates this maximum and calculates the covariance matrix of the multivariate normal distribution is given in Appendix 3.3.

There is no evidence from the results in the talk that the component does not conform to the blueprint specification and so this component would not be rejected for being the wrong shape.

3.3 Other Shape Fitting Problems

In this section we briefly describe some other shape fitting problems. These include higher dimensional as well as two dimensional problems. For each particular class of geometrical shapes we will construct a multipivotal model to illustrate how the methods developed in chapter two are easily adapted to shape fitting.

Example 3.3.1 The rectangle fitting problem

In the rectangle fitting problem we look at fitting a rectangle through a set of data that lies approximately on a rectangle. Let the data points for the rectangle fitting problem be the profile (x_i, y_i) $i = 1,\ldots,n$. The lengths and orientation of the sides and the centre of the "true" rectangle, denoted by $2r$, $2r_2$, Ψ, a and b respectively, are taken to be unknown, so the problem is reduced to obtaining estimates or "confidence" intervals of these particular parameters.

The multipivotal model we will use for the rectangle fitting problem is similar to the second model for the ellipse fitting problem given by (3.2.6). The datum profile is assumed to be a square with an area of 4 units, its sides parallel to x-y axes and its centre at the origin.

We will assume that the unknown random variables

(e_i,d_i) $i = 1,\ldots,n$ defined in the model have an error distribution about the datum profile of the following form:

$$f(\underline{e},\underline{d}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (U_i-1)^2\right\} \qquad (3.3.1)$$

where $U_i = \max\{|e_i|,|d_i|\}$ $i = 1,\ldots,n$, that is to say they have a truncated N(1,1) distribution, restricted to positive values, about the nearest edge of the datum profile. The full multipivotal model we shall consider is as follows:

$$(\underline{x},\underline{y}) = [a,b]_1 o[\Psi]_2 o[r_1,r_2]_3 o[\sigma]_4 o(\underline{e},\underline{d}) \qquad (3.3.2)$$

$$(\underline{e},\underline{d}) \sim f(\cdot) \text{ given by } (3.3.1)$$

where $(\underline{x},\underline{y})$, $(\underline{e},\underline{d})\varepsilon\Omega = \mathbb{R}^n \times \mathbb{R}^n$ and $[a,b]_1$, $[\Psi]_2$, $[r_1,r_2]$, $[\sigma]_4$ index the transformations $\{\phi_{\theta_1}:\theta_1\varepsilon\Theta_1 = \mathbb{R}^2\}$, $\{\phi_{\theta_2}:\theta_2\varepsilon\Theta_2 = [0,2\Pi)\}$, $\{\phi_{\theta_3}:\theta_3\varepsilon\Theta_3 = \mathbb{R}^+ \times \mathbb{R}^+\}$ and $\{\phi_{\theta_4}:\theta_4\varepsilon\Theta_4 = \mathbb{R}^+\}$ respectively, which are defined on the ith element of $(\underline{s},\underline{t})\varepsilon\Omega$ to be:

$[a,b]_1 o(s_i,t_i)$ defined by (3.2.7)

$[\Psi]_2 o(s_i,t_i)$ defined by (3.2.8)

$[r_1,r_2]_3 o(s_i,t_i)$ defined by (3.2.9)

$$[\sigma]_4 o(s_i,t_i) = \left(\frac{\{(v_i-1)\sigma+1\}s_i}{v_i}\,,\ \frac{\{(v_i-1)\sigma+1\}t_i}{v_i}\right) \qquad (3.3.3)$$

where $v_i = \max\{|s_i|,|t_i|\}$

This last set of transformations has the effect of altering the dispersion of the unknown random variables about the datum profile.

Having defined our model we can now proceed to calculate the multipivotal distribution for the parameters using the results from chapter two.

The relevant Jacobians for the calculation are given by equations (3.2.10), (3.2.11), (3.2.12) and by

$$J_{4,\Omega}([\sigma]_4|(\underline{s},\underline{t})) = \sigma^n \prod_{i=1}^{n} \frac{(v_i-1)\sigma+1}{v_i}$$

$$J_{4,*}([\sigma]_4,[1]_4) = \sigma \qquad (3.3.4)$$

$$J_{4,*R}([\sigma]_4,[1]_4) = \sigma$$

where $[\sigma]_4$, $[1]_4 \varepsilon\ \Theta_4$, $[1]_4$ being the identity element of the set of transformations and $(\underline{s},\underline{t})\varepsilon\Omega$.

Substituting the above into equation (2.5.11) we obtain the multipivotal distribution for the parameters as follows:

$$p([a,b]_1,[\Psi]_2,[r_1,r_2]_3,[\sigma]_4|(\underline{x},\underline{y})) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(U_i-1)^2\right\}\frac{1}{(r_1r_2\sigma)^{n+1}}\prod_{i=1}^{n}\left\{\frac{(U_i-1)\sigma^{-1}+1}{U_i}\right\}$$

(3.3.5)

$$\text{where } U_i = \max\left[\left|\frac{(x_i-a)\cos\Psi+(y_i-b)\sin\Psi}{r_1}\right|, \left|\frac{(y_i-b)\cos\Psi-(x_i-a)\sin\Psi}{r_2}\right|\right]$$

The multipivotal distribution can then be used to obtain intervals of the parameters that have 95% pivotal probability attached to them, so that various inferences about the parameters can be made. □

The next example is slightly different from the previous examples in that the class of geometrical shapes is defined in polar co-ordinates.

Example 3.3.2 The Limacon Fitting Problem

In the limacon fitting problem we look at fitting a limacon through a set of data that lies approximately on a limacon.

A limacon is a figure whose equation in polar co-ordinates is

$$r = R + A\sin\Psi + B\cos\Psi \qquad (3.3.6)$$
$$R \in \mathbb{R}^+ ;\ A,B \in \mathbb{R}$$

The limacon is a useful figure because it is a linear approximation, in polar co-ordinates, to a circle of radius R when the centre, in cartesian co-ordinates, at (A,B) is close to the origin (relative to R) i.e. when A,B <<R.

Let the data points for this problem also be measured in polar co-ordinates (r_i, θ_i) $i = 1,\ldots,n$. The multi-pivotal model we will use for this particular problem takes the following form:

$$(\underline{r},\underline{\theta}) = [A,B,R]_1 \circ [\sigma]_2 \circ (\underline{e},\underline{\Psi}) \tag{3.3.7}$$

where $(\underline{r},\underline{\theta}), (\underline{e},\underline{\Psi}) \in \Omega = \mathbb{R}^{+n} \times [0,2\Pi)^n$.

The unknown random variables (e_i, Ψ_i) $i = 1,\ldots,n$ are assumed to lie approximately on a unit circle with its centre at the origin and an error distribution about this shape assumed to be

$$f(\underline{e},\underline{\Psi}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (e_i - 1)^2\right\} \tag{3.3.8}$$

that is to say each point in the error distribution is independent identically distributed radially about the unit circle as a truncated N(1,1) distribution restricted to positive values.

The transformations indexed by $[A,B,R]_1$ and $[\sigma]_2$ are defined on the ith element of $(\underline{s},\underline{w}) \in \Omega$ to be:

$$[A,B,R]_1 \circ (s_i, w_i) = (Rs_i + A\sin w_i + B\cos w_i, w_i) \tag{3.3.9}$$

$$[\sigma]_2 \circ (s_i, w_i) = (\{s_i - 1\}\sigma + 1, w_i) \tag{3.3.10}$$

where $[A,B,R]_1 \in \Theta_1 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ and $[\sigma]_2 \in \Theta_2 = \mathbb{R}^+$

The relevant Jacobians for the calculation of the multipivotal distribution are:

$$J_{1,\Omega}([A,B,R]_1 | (\underline{s},\underline{w})) = R^n$$

$$J_{1,*}([A,B,R]_1,[0,0,1]_1) = R^3 \tag{3.3.11}$$

$$J_{1,*R}([A,B,R]_1,[0,0,1]_1) = R$$

$$J_{2,\Omega}([\sigma]_2 | (\underline{s},\underline{w})) = \sigma^n$$

$$J_{2,*}([\sigma]_2,[1]_2) = \sigma \tag{3.3.12}$$

$$J_{2,*R}([\sigma]_2,[1]_2) = \sigma$$

where $[A,B,R]_1$, $[0,0,1]_1 \in \Theta_1$ and $[\sigma]_2$, $[1]_2 \in \Theta_2$; $[0,0,1]_1$ and $[1]_2$ being the identity elements of their respective set of transformations.

Substituting the above into equation (2.5.11) we obtain the multipivotal distribution for the parameters as follows:

$$p([A,B,R]_1,[\sigma] | (\underline{r},\underline{\theta})) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2 R^2} \sum_{i=0}^{n} (r_i - A\sin\theta_i - B\cos\theta_i - R)^2\right\} \frac{1}{(\sigma R)^{n+1}} \tag{3.3.13}$$

The multipivotal distribution can then be used to obtain intervals of the parameters with the usual interpretation.

Note the value of the parameters that maximise the multipivotal distribution coincide with the least squares solution for the parameters A, B and R. □

All the previous examples have been two dimensional shape fitting problems. There is no reason at all to restrict our attention to two dimensional problems, so the final two examples are both three dimensional shape fitting problems.

These both illustrate the extension of multipivotal models to model higher dimensional shape fitting problems.

Example 3.3.3 The Cuboid Fitting Problem

In the cuboid fitting problem we look at fitting a cuboid through a set of data that lies approximately on a cuboid. Let the data points for the cuboid fitting problem be the 3D profile (x_i, y_i, z_i) $i = 1, \ldots, n$.

A 3D profile is defined as a finite collection of data points lying in the Euclidian space $\mathbb{R}^3$. Each data point is given by its cartesian co-ordinates in the space. This definition is an obvious extension of a profile defined in section 3.1.

The lengths and orientation of the sides and the centre of the "true" cuboid, denoted by $2l_1$, $2l_2$, $2l_3$, α, β, γ, a, b and c respectively, are taken to be unknown, so the problem is reduced to that of obtaining estimaes or "confidence" intervals of these particular parameters.

The datum profile for the problem is assumed to be a cube with a volume of 8 units, its sides parallel to the axes and centre at the origin.

We will assume that the unknown random variables (u_i, v_i, w_i) $i = 1,\ldots,n$ defined in the multipivotal model have an error distribution about the datum profile of the following form:

$$f(\underline{u},\underline{v},\underline{w}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(e_i-1)^2\right\} \qquad (3.3.14)$$

where $e_i = \max\{|u_i|,|v_i|,|w_i|\}$ $i = 1,\ldots,n.$

That is to say they have a truncated N(1,1) distribution restricted to positive values, about the nearest face of the datum profile. The full multipivotal model we shall consider is as follows:

$$(\underline{x},\underline{y},\underline{z}) = [a,b,c]_1 \circ [\alpha]_2 \circ [\beta]_3 \circ [\gamma]_4 \circ [l_1, l_2, l_3]_5 \circ$$

$$[\sigma]_6 \circ (\underline{u},\underline{v},\underline{w}) \qquad (3.3.15)$$

$$(u,v,w) \sim f(.) \text{ given by } (3.3.14)$$

where $(\underline{x},\underline{y},\underline{z})$, $(\underline{u},\underline{v},\underline{w}) \in \Omega = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ and

$[a,b,c]_1, [\alpha]_2, [\beta]_3, [\gamma]_4, [l_1,l_2,l_3]_5, [\sigma]_6$ index the transformations $\{\phi_{\theta_1}: \theta_1 \varepsilon\ \Theta_1 = \mathbb{R}^3\}$, $\{\phi_{\theta_2}: \theta_2 \varepsilon\ \Theta_2 = [0,2\Pi)\}$, $\{\phi_{\theta_3}: \theta_3 \varepsilon\ \Theta_2 = [0,2\Pi)\}$, $\{\phi_{\theta_4}: \theta_4 \varepsilon\ \Theta_4 = (0,2\Pi)\}$, $\{\phi_{\theta_5}: \theta_5 \varepsilon\ \Theta_5 = \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+\}$ and $\{\phi_{\theta_6}: \theta_6 \varepsilon\ \Theta_6 = \mathbb{R}^+\}$ respectively, which are defined on the ith element of $(\underline{r},\underline{s},\underline{t}) \varepsilon \Omega$ to be:

$$[a,b,c]_1 o(r_i,s_i,t_i) = (r_i+a, s_i+b, t_i+c) \qquad (3.3.16)$$

$$[\alpha]_2 o(r_i,s_i,t_i) = (r_i, s_i\cos\alpha - t_i\sin\alpha, s_i\sin\alpha + t_i\cos\alpha) \qquad (3.3.17)$$

$$[\beta]_3 o(r_i,s_i,t_i) = (r_i\cos\beta - t_i\sin\beta, s_i, r_i\sin\beta + t_i\cos\beta) \qquad (3.3.18)$$

$$[\gamma]_4 o(r_i,s_i,t_i) = (r_i\cos\gamma - s_i\sin\gamma, r_i\sin\gamma + s_i\cos\gamma, t_i) \qquad (3.3.19)$$

$$[l_1,l_2,l_3]_5 o(r_i,s_i,t_i) = (l_1 r_i, l_2 s_i, l_3 t_i) \qquad (3.3.20)$$

$$[\sigma]_6 o(r_i,s_i,t_i) = (q_i r_i, q_i s_i, q_i t_i) \qquad (3.3.21)$$

where $q_i = \dfrac{(d_i-1)\sigma+1}{d_i}$ and $d_i = \max\{|r_i|, |s_i|, |t_i|\}$.

The action of the various set of transformations on the samples space Ω is as follows:

The first set has the effect of moving the origin to point (a,b,c). The next three rotate the space, about the x, y and z axes respectively. The fifth set of transformations enlarge the space by a factor of l_1 in the x direction and by l_2 and l_2 in the y and z directions respectively. The final set of transformations has the effect of altering the dispersion of the unknown random variables about the datum profile.

Having defined our model we can now proceed to calculate the relevant Jacobians to obtain the multipivotal distribution for the parameters.

$$J_{5,\Omega}([l_1,l_2,l_3]_5 | (\underline{r},\underline{s},\underline{t}) = l_1{}^n l_2{}^n l_3{}^n$$

$$J_{5,*}([l_1,l_2,l_3]_5[1,1,1]_5) = l_1 l_2 l_3 \qquad (3.3.22)$$

$$J_{5,*R}([l_1,l_2,l_3]_5,[1,1,1]_5) = l_1 l_2 l_3$$

$$J_{6,\Omega}([\sigma]_6 | (\underline{r},\underline{s},\underline{t})) = \sigma^{2^n} \sum_{l=1}^{n} q_i$$

$$J_{6,*}([\sigma]_6,[1]_6) = \sigma \qquad (3.3.23)$$

$$J_{6,*R}([\sigma]_6,[1]_6) = \sigma$$

where $[l_1,l_2,l_3]_5$, $[1,1,1]_5 \varepsilon\ \Theta_5$; $[\sigma]_6$, $[1]_6 \varepsilon\ \Theta_6$, $[1,1,1]_5$ and $[1]_6$ being the identity elements of the respective set of transformations and $(\underline{r},\underline{s},\underline{t}) \varepsilon \Omega$. All the other relevant Jacobians take the value one.

Substituting the above into equation (2.5.11), we obtain the multipivotal distribution for the parameters as follows:

$$p([a,b,c]_1,[\alpha]_2,[\beta]_3,[\gamma]_4,[l_1,l_2,l_3]_5,[\sigma]_6\ (\underline{x},\underline{y},\underline{z}))$$

$$\propto\ \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(h_i-1)^2\right\}\frac{1}{(\sigma l_1 l_2 l_3)^{n+1}}\prod_{i=1}^{n}\left\{\frac{(h_i-1)\sigma^{-1}+1}{h_i}\right\} \tag{3.3.24}$$

where $h_i = \max\{|p_i|,|q_i|,|r_i| : (p_i,q_i,r_i) = [l_1,l_2,l_3]_5^{-1} o [\gamma]_4^{-1} o [\beta]_3^{-1} o [\alpha]_2^{-1} o [a,b,c]_1^{-1} o (x_i,y_i,z_i)\}$

This multipivotal distribution can then be used to construct intervals of the parameters that have various measures of pivotal probability attached to them, so that inferences about the parameters can be made. □

Example 3.3.4 The Cone Fitting Problem

In the cone fitting problem we look at fitting a cone with circular cross-section through a set of data that lies approximately on a cone. Let the data points for the cone fitting problem be the 3D profile (x_i,y_i,z_i) $i = 1,\dots,n$.

The position of the peak, the orientation of the axis of symmetry and the ratio of the width to the height of the cone, denoted by a, b, c, α, β and l respectively, are taken to be unknown, so the problem is reduced to obtaining estimates or "confidence" intervals of these particular parameters.

The datum profile for the problem is assumed to be the cone given by the following equation:

$$x^2+y^2 = z^2$$

that is to say a cone that has its peak at the origin, the axis of symmetry is the z axis and the ratio of the width to the height of the cone of one.

The unknown random variables (u_i, v_i, w_i) $i = 1,\ldots,n$ are assumed to lie approximately on the datum profile, with an error distribution given by

$$f(\underline{u},\underline{v},\underline{w}) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (\sqrt{[u_i{}^2+v_i{}^2]}-|w_i|^2\right\} \qquad (3.2.25)$$

that is to say each point has a truncated $N(|w_i|,1)$ distribution restricted to positive values, in a radial direction about the z axis in the x-y plane through $z = w_i$. The full multipivotal model we shall consider is as follows:

$$(\underline{x},\underline{y},\underline{z}) = [a,b,c]_1 o[\alpha]_2 o[\beta]_3 o[1]_4 o[\sigma]_5 o(\underline{u},\underline{v},\underline{w}) \qquad (3.3.26)$$

$$(\underline{u},\underline{v},\underline{w}) \sim f(\cdot) \text{ given by } (3.3.25)$$

where $(\underline{x},\underline{y},\underline{z})$, $(\underline{u},\underline{v},\underline{w}) \varepsilon \Omega = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ and $[a,b,c]_1$, $[\alpha]_2$, $[\beta]_3$, $[1]_4$, $[\sigma]_5$ index the transformations $\{\phi_{\theta_1}:\theta_1 \varepsilon\ \Theta_1 = \mathbb{R}^3\}$, $\{\phi_{\theta_2}:\theta_2 \varepsilon\ \Theta_2 = [0,2\pi)\}$, $\{\phi_{\theta_3}:\theta_3 \varepsilon\ \Theta_3 = [0,2\pi]\}$, $\{\phi_{\theta_4}:\theta_4 \varepsilon\ \Theta_4 = \mathbb{R}^+\}$ and $\{\phi_{\theta_5}:\theta_5 \varepsilon\ \Theta_5 = \mathbb{R}^+\}$ respectively, which are defined on the ith element of $(\underline{r},\underline{s},\underline{t})\varepsilon\Omega$ to be

$[a,b,c]_1 o(r_i,s_i,t_i)$ defined by (3.3.16)

$[\alpha]_2 o(r_i,s_i,t_i)$ defined by (3.3.17)

$[\beta]_3 o(r_i,s_i,t_i)$ defined by (3.3.18)

$$[1]_4 o(r_i,s_i,t_i) = (1r_i, 1s_i, t_i) \qquad (3.3.27)$$

$$[\sigma]_5 o(r_i,s_i,t_i) = (q_i r_i, q_i s_i, t_i) \qquad (3.3.28)$$

where $q_i = \dfrac{(\sqrt{[r_i^2+s_i^2]}-|t_i|)\sigma+|t_i|}{\sqrt{[r_i^2+s_i^2]}}$

The relevant Jacobians for the calculation of the multipivotal distribution are:

$$J_{4,\Omega}([1]_4|(\underline{r},\underline{s},\underline{t})) = 1^{2n}$$

$$J_{4,*}([1]_4,[1]_4) = 1 \qquad (3.3.29)$$

$$J_{4,*R}([1]_4,[1]_4) = 1$$

$$J_{5,\Omega}([\sigma]_5|(\underline{r},\underline{s},\underline{t})) = \sigma^n \prod_{i=1}^{n} q_i$$

$$J_{5,*}([\sigma]_5,[1]_5) = \sigma \qquad (3.3.30)$$

$$J_{5,*R}([\sigma]_5,[1]_5) = \sigma$$

where $[1]_4$, $[1]_4 \in \Theta_4$ and $[\sigma]_5$, $[1]_5 \in \Theta_5$; $[1]_4$ and $[1]_5$ being the identity elements of the respective set of transformations and $(\underline{r},\underline{s},\underline{t}) \in \Omega$.

All the other relevant Jacobians take the value one.

Substituting the above into equation (2.5.11), we obtain the multipivotal distribution for the parameters as follows:

$$p([a,b,c]_1,[\alpha]_2,[\beta]_3,[1]_4,[\sigma]_5|(\underline{x},\underline{y},\underline{z})) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\sqrt{[e_i^2+d_i^2]}-|h_i|)^2\right\}\frac{1}{\sigma^{n+1}l^{2n+1}}\prod_{i=1}^{n}k_i \tag{3.3.31}$$

where $(e_i,d_i,h_i) = [1]_4^{-1}o[\beta]_3^{-1}o[\alpha]_2^{-1}o[a,b,c]_1^{-1}o(x_i,y_i,z_i)$

and $k_i = \dfrac{(\sqrt{[e_i^2+d_i^2]}-|h_i|)\sigma^{-1}+|h_i|}{\sqrt{[e_i^2+d_i^2]}}$

This multipivotal distribution can be used to construct intervals of the parameters that have various measures of pivotal probability attached to them. □

As all the examples show, multipivotal models adapt very easily to shape fitting problems, making use of the geometrical properties of each particular problem.

4 CONCLUSIONS AND FURTHER WORK

The main conclusion that one can drawn from the previous chapter is that multipivotal models provide a very versatile way to handle shape fitting problems, making use of the geometrical properties in each particular case. For those problems where alternative analyses are available the method compares favourably. In the circle and ellipse fitting problems the modified least squares method does provide a non-iterative alternative which has the advantage of speed of computation over the iterative multipivotal method. However it does have the disadvantage compared with the multipivotal method of not being able to take into account information about the distribution of the data points about the "true" shape.

One suggestion for further work is to use simulation to compare the robustness of alternative methods of fit for particular shape fitting problems against various extreme forms of data. For instance it is known that as the angle of arc of a circle is reduced the estimates of the radius and co-ordinates of the centre of the "true" circle become less reliable. A simulation study would indicate which of the various methods of fit of an arc of a circle is the most robust against this type of data.

Finally on the more theoretical side, it would be desirable to find some criteria for the choice of the "best" interval of the parameters that has a fixed pivotal probability. In chapter three the shortest interval criterion was chosen, but this has problems because under repeated trials the resulting interval might not remain the shortest. One possible solution might be to incorporate pivotal and multipivotal models in a decision theory framework.

APPENDICES

A1.1 Definition of Structured Model

A structured model is an ordered sequence

$$\mu = [(\Omega,\mathcal{A},P),(\mathfrak{X},\mathcal{B}),(\Theta,\mathcal{C}),\{\phi_\theta\}]$$

where $(\Omega,\mathcal{A},P)$ is a probability space, $\mathfrak{X}$ and Θ are a sample space and a parameter space each being associated with a σ-field $\mathcal{B}$ and $\mathcal{C}$ respectively, $\{\phi_\theta\colon \theta\varepsilon\Theta\}$ are measurable functions from Ω to $\mathfrak{X}$ and sets of the form:

$$\bigcup_\theta \phi_\theta^{-1}(\{x\}) \qquad \theta\varepsilon C\varepsilon\mathcal{C}\ ;\ x\varepsilon\mathfrak{X}$$

$$\bigcup_\theta \phi_\theta(\Omega) \qquad \theta\varepsilon C\varepsilon\mathcal{C} \qquad \text{and}$$

$$\{\theta : \theta\varepsilon\Theta \text{ and } \phi_\theta^{-1}(\{x\}) \neq \phi\}$$

are all measurable.

A1.2 Theorem 2.3.1

(a) Given a Pivotal Model, then any equivalent structured model to it will also be a Pivotal Model.

(b) Given a (measurable) subset $C\subsetneq\Theta$ and an observation $x\varepsilon\mathfrak{X}$ then the probability induced by $\tilde{x}\varepsilon\tilde{\mathfrak{X}}$ the corresponding point in the equivalent Pivotal Model.

Proof:

Before we prove this theorem we need a couple of results:

R1 Given a point $\tilde{x}\varepsilon\tilde{X}$ and a (measurable) subset $C\subseteq\Theta$ then $\Omega_{\tilde{x}}(C) = \Omega_{\alpha(\tilde{x})}(C)$.

R2 Given a point $e\varepsilon\Omega$ then there exists a 1-1 correspondence between X_e and $\tilde{X}_e$ such that:

if $x\varepsilon X_e \Rightarrow \alpha^{-1}(x)\varepsilon\tilde{X}_e$ and
if $\tilde{x}\varepsilon\tilde{X}_e \Rightarrow \alpha(\tilde{x})\varepsilon X_e$.

Both these results are easily demonstrated from the definition of $\tilde{\phi}_\theta:\theta\varepsilon\Theta = \alpha^{-1}\phi_\theta:\theta\varepsilon\Theta$

Proof of theorem: We have to prove the structured model generated by the 1-1 transformation α satisfies the three restrictions on a structured model for it to be a pivotal model.

a) Partition condition

Since the transformations $\{\phi_\theta:\theta\varepsilon\Theta\}$ partition Ω we have $\forall e\varepsilon\Omega$ $\forall x,y\varepsilon X_e \Rightarrow \Omega_y = \Omega_x$ A1.2.1

Now consider $\forall e\varepsilon\Omega$ $\forall \tilde{x},\tilde{y}\,\varepsilon\,\tilde{X}_e$. Consider an element $e_o\varepsilon\Omega_{\tilde{x}}$. From R1 and the fact $\Omega_x = \Omega_x(\Theta)$ we have $e_o\varepsilon\Omega_{\alpha(\tilde{x})}$. From R2 we have $\alpha(\tilde{x})$, $\alpha(\tilde{y})\varepsilon\ X_e$ hence from A1.2.1 we have $e_o\varepsilon\Omega_{\alpha(\tilde{y})}$.

Hence we get $\Omega_{\tilde{x}} \subseteq \Omega_{\tilde{y}}$. By symmetry we also get $\Omega_{\tilde{y}}\ \Omega_{\tilde{x}}$, hence $\Omega_{\tilde{x}} = \Omega_{\tilde{y}}$, hence the transformations $\{\tilde{\phi}_\theta : \theta \varepsilon \Theta\}$ partition the space Ω.

b) Additivity condition

Let $e \varepsilon \Omega_{\tilde{x}}(\{\theta_1\}) \cap \Omega_{\tilde{x}}(\{\theta_2\})$ for an $\tilde{x} \varepsilon \tilde{X}$ where $\tilde{x} \varepsilon \tilde{X}$ and $\theta_1 \neq \theta_2 \varepsilon \Theta$.

From R1 we obtain

$$e \varepsilon \Omega_{\alpha(\tilde{x})}(\{\theta_1\}) \text{ and } e \varepsilon \Omega_{\alpha(\tilde{x})}(\{\theta_2\}) \text{ hence}$$

$$e \varepsilon \Omega_{\alpha(\tilde{x})}(\{\theta_1\}) \cap \Omega_{\alpha(\tilde{x})}(\{\theta_2\})$$

but this is impossible since by additivity

$$\Omega_{\alpha(\tilde{x})}(\{\theta_1\})\ \Omega_{\alpha(\tilde{x})}(\{\theta_2\}) = \phi$$

hence the induced model is also additive.

c) Pivotal condition

Since the original model satisfies the pivotal condition we have $\forall\, \tilde{x} \varepsilon \tilde{X}$ and $\forall \theta \varepsilon \Theta$ $\exists ! e \varepsilon \Omega$ s.t. $\alpha(\tilde{x}) = \theta o e$. Rearranging we arrive at $\tilde{x} = \alpha^{-1}(\phi_\theta o e)$.

Now since α is a 1-1 mapping of $\tilde{X} \rightarrow X$ the induced structured model satisfies the pivotal condition.

Hence the induced structured model is a pivotal model because it too satisfies the three conditions (a), (b) and (c). Hence the first part of the theorem.

We need only consider equality of the upper probability for the two models because we have already proved that both models satisfy the additivity condition and hence that the upper and lower probabilities are equal to each other for every (measurable) subset of Θ for both models.

From R1 we have that the partitions induced by the transformations $\{\phi_\theta : \theta\varepsilon\Theta\}$ and $\{\tilde{\phi}_\theta : \theta\varepsilon\Theta\}$ are the same, hence the conditional probability induced will also be the same on each partition. R1 also shows that the set of antecedents produced by a (measurable) subset C Θ from observations $\tilde{x}\varepsilon\tilde{\mathfrak{X}}$ and $x = \alpha(\tilde{x})\varepsilon\mathfrak{X}$ are also the same, hence from the definition of upper probability the upper probabilities induced will also be the same. Hence theorem. ||

A1.3 <u>Lemma 2.3.2</u>

The sets $\{\Theta_x : x\varepsilon\Omega\}$ partition Ω, the same partition of Ω as the sets $\{\Omega_x : x\varepsilon\mathfrak{X}\}$.

<u>Proof</u>:

We first have to prove that if the sets Θ_x and $\Theta_{x'}$, $x,x'\varepsilon\Omega$ have a point in common, $x''\varepsilon\Omega$ say then $\Theta_x = \Theta_{x'}$.

The sets $\{\Theta_x\}$ $x \varepsilon \Omega$ will then partition Ω. We need only prove:

$$\text{if } x'' \varepsilon \Theta_x \text{ then } \Theta_{x''} = \Theta_x \tag{A1.3.1}$$

We start by proving the following:

$$\text{if } x'' \varepsilon \Theta_x \text{ then } \Theta_{x''} \subseteq \Theta_x \tag{A1.3.2}$$

Given $h \varepsilon \Theta$ $hox'' \varepsilon$ $ho\Theta ox$ by definition of x''. Now $ho\Theta_x = \Theta_x$ by P2 hence A1.3.2. To prove A1.3.1 we need only prove that $x \varepsilon \Theta_{x''}$ and then from A1.3.2 $\Theta_x \subseteq \Theta_{x''}$ hence $\Theta_{x''} = \Theta_x$.

Since $x'' \varepsilon \Theta_x$ we have, there exist $h \varepsilon \Theta$ such that $x'' = hox$. Now from P3 and P4 there exists $h^* \varepsilon \Theta$ s.t. $h^*o(hox) = x$, that is $x = h^*ox''$ hence $x \varepsilon \Theta_{x''}$.

So we have now proved that the sets $\{\Theta_x\}$ $x \varepsilon \Omega$ partition the space Ω, we only have to prove that this is the same partition as $\{\Omega_x\}$ $x \varepsilon \Omega$. We achieve this by proving that:

$$\forall\, x \varepsilon \Omega \quad \Omega_x = \Theta_x \tag{A1.3.3}$$

Consider a point $e \varepsilon \Omega_x$ by definition there exists $\theta \varepsilon \Theta$ s.t. $x = \theta oe$, hence $x \varepsilon \Theta_e$. From A1.3.1 we have $\Theta_x = \Theta_e$. Now from P3 $e \varepsilon \Theta_e = \Theta_x$ hence

$$\Omega x \subseteq \Theta x \tag{A1.3.4}$$

Consider a point $x' \varepsilon \Theta_x$ by definition there exist $h \varepsilon \Theta$

s.t. $x' = hox$, by P3 and P2 there exist a $\theta_1 \varepsilon \Theta$ s.t. $\theta_1 o(hox) = x$ that is $\theta_1 ox' = x$ so $x' \varepsilon \Omega_x$ and hence $\Theta_x \quad \Omega_x$ then from A1.3.4 we have proved A1.3.3 hence result. ||

A1.4 Theorem 2.3.2

The binary operation $*: \Theta \times \Theta \to \Theta$ conditional on the point D(x), defined in §2.3, has the algebraic structure called "a loop" that is to say it has the following properties:

L1 $\forall \theta_1, \theta_2 \varepsilon \Theta$ then $\theta_1 * \theta_2 \varepsilon \Theta$ that is to say Θ is closed under the operation *.

L2 $\exists i \varepsilon \Theta$ s.t. $\forall \theta \varepsilon \Theta \;\; i*\theta = \theta = \theta * i$ that is to say Θ has an identity element, namely i.

L3 $\forall \theta_1, \theta_2 \varepsilon \Theta \quad \exists! h_1, h_2 \varepsilon \Theta$ s.t.
$\theta_1 * h_1 = \theta_2, \quad h_2 * \theta_1 = \theta_2$.

Proof:

Before we can prove this theorem we need the following results:

$$\forall h_1, h_2 \varepsilon \Theta \quad \forall e \varepsilon \Omega \quad \exists! g_1, g_2 \varepsilon \Theta$$

$$\text{s.t.} \quad g_1 o(h_1 oe) = h_2 oe \qquad h_1 o(g_2 oe) = h_2 oe \tag{A1.4.1}$$

By P4 we already have $\exists g_1 \varepsilon \Theta$ s.t. $g_1 o(h_1 oe) = h_2 oe$ so we only have to prove uniqueness of this element. Assume $\exists g_1' \varepsilon \Theta$ s.t.

$$g_1' o(h_1 oe) = h_2 oe = g_1 o(h_1 oe)$$

Let $h_1 oe = e' \varepsilon \Omega$ then we have $g_1' oe' = g_1 oe'$ and by P1 we must have $g_1' = g_1$ hence uniqueness so we have proved the first part of the result.

To prove the existence of $g_2 \varepsilon \Theta$ consider the following. Let $z = h_2 oe$, by P5 $\exists ! d \varepsilon \Omega$ s.t. $h_1 od = z$. Define $\theta \varepsilon \Theta$ to be that element of Θ that satisfies $\theta o(h_1 od) = d$, θ exists by P3 and P4 and is unique by P1 hence

$$d = \theta o(h_1 od) = \theta oz = \theta o(h_2 oe) = g_2 oe \qquad \text{(A1.4.2)}$$

g_2 exists by P2. Now $g \varepsilon \Theta$ has the desired property since $h_1 o(g_2 oe) = h_1 od$ by A1.4.2
$= z = h_2 oe$ by definition.

To prove uniqueness consider $g_2' \varepsilon \Theta$ s.t.

$$h_1 o(g_2' oe) = h_2 oe = h_1 o(g_2 oe)$$

Since $d \varepsilon \Omega$ defined above is unique we have $g_2' oe = d = g_2 oe$, by P1 we have $g_2' = g_2$, hence result.

We can now proceed to prove the theorem.

Proof of theorem

L1 By the definition of the binary operation $*: \Theta \times \Theta \to \Theta$ given in §2.3 this is true hence L1.

L2 Let $i \in \Theta$ be the identity element mentioned in P3 then we have $g \in \Theta$

$(i*g)oD(x) = io(goD(x))$ by definition

$= ioz$

where $z = goD(x)$.

Now $ioz = z$ by P3

$= goD(x)$

hence $i*g = g$ by P1

$(g*i)oD(x) = go(ioD(x))$ by definition

$= goD(x)$ by P3

thus $g*i = g$ by P1

hence L2.

L3 By A1.4.1 $\exists! g_1 \in \Theta$ s.t. $g_1 o(h_1 oD(x)) = h_2 oD(x)$

hence $(g_1 * h_1)oD(x) = g_1 o(h_1 oD(x)) = h_2 oD(x)$.

Thus $g_1 * h = h_2$ by P1, hence the first part of L3.

By A1.4.1 $!g_2 \in \Theta$ s.t. $h_1 o(g_2 oD(x)) = h_2 oD(x)$

hence $(h_1 * g_2)oD(x) = h_1 o(g_2 oD(x)) = h_2 oD(x)$.

Thus $h_1 * g = h_2$ by P1, hence the second part of L3, hence L3, hence the theorem. ||

A1.5 Proof of the uniqueness of the pivotal distribution of θ given x

Since we have observed $x\varepsilon\Omega$ we can take the reference point in the partition Ω_x to be x itself. We will now prove that the pivotal distribution of $\theta\varepsilon\Theta$ given the reference variable $x\varepsilon\Omega$ and its associated transformation variable $i\varepsilon\Theta$, is the same as the pivotal distribution of $\theta\varepsilon\Theta$ given a general reference variable $D\varepsilon\Omega_x$ and its associated transformation variable $[x]\varepsilon\Theta$.

Before we proceed with the proof we do need a slight change in notation. Denote by $\underset{x}{*}: \Theta \times \Theta \rightarrow \Theta$ the loop obtained when we take x to be the reference variable and $\underset{D}{*}: \Theta \times \Theta \rightarrow \Theta$ the loop obtained when we take D to be the reference variable.

Since $D\varepsilon\Omega_x$ there exists uniquely $k\varepsilon\Theta$ such that

$$D = kox \tag{A1.5.1}$$

Let $a,b\varepsilon\Theta$ consider $a\underset{D}{*}b = c$, then from the definition of $\underset{D}{*}$ $\Theta \times \Theta \rightarrow \Theta$ we have

$$(a\underset{D}{*}b)oD = ao(boD) = ao(bo(kox))$$

$$= coD = co(kox)$$

hence $$a\underset{x}{*}(b\underset{x}{*}k) = C\underset{x}{*}k \tag{A1.5.2}$$

From this we obtain

$$a \underset{D}{*} b = \{a \underset{x}{*} (b \underset{x}{*} k)\} /_{x} k \qquad (A1.5.3)$$

where $/_{x}$ is right division associated with $\underset{x}{*}$.

If $a \underset{D}{*} b = c$ then $a \backslash_{D} c = b$, and from A1.5.2 we obtain

$$a \backslash_{D} c = \{a \backslash_{x} (c \underset{x}{*} k)\} /_{x} k \qquad (A1.5.4)$$

where $\backslash_{D}$ and $\backslash_{x}$ are the left divisions associated with $\underset{D}{*}$ and $\underset{x}{*}$ respectively.

Consider $(\theta \backslash_{D} [x])oD = z$ say.

Now $\theta oz = \{\theta \underset{D}{*} (\theta \backslash_{D} [x])\}oD$

$= [x]oD = x$ from lemma A2.1.1

Thus from lemma 2.3.4 we obtain $\theta \underset{x}{*} [z] = i$ or $[z] = \theta \backslash_{x} i$ hence we obtain

$$(\theta \backslash_{D} [x])oD = (\theta \backslash_{D} i)ox \qquad (A1.5.5)$$

Consider $J_{\Omega}(\theta \backslash_{D} [x] | D) = \left| \frac{\partial(\{\theta \backslash_{D} [x]\}oD)}{\partial(D)} \right|$

$$= \left| \frac{\partial(\{\theta \backslash_{x} i\}ox)}{\partial(x)} \right| \left| \frac{\partial(x)}{\partial(kox)} \right|$$ from A1.5.5 and A1.5.1.

Hence we obtain

$$J_{\Omega}(\theta_{D}\backslash[x]|D) = AJ_{\Omega}(\theta_{x}\backslash i|x) \qquad (A1.5.6)$$

where A is a constant independent of $\theta\varepsilon\Theta$.

Consider $d\mu(\cdot)$ the left loop invariant differential on Θ based on the loop $\underset{x}{*}$ $\Theta \times \Theta\to\Theta$.

$$\begin{aligned} \text{Now } d\mu(h\underset{D}{*}\theta) &= d\mu(\{h\underset{x}{*}(\theta\underset{x}{*}k)\}/_{x}k) && \text{by A1.5.3} \\ &= \frac{1}{\Delta(k)}\, d\mu(h\underset{x}{*}(\theta\underset{x}{*}k)) && \text{by A2.4.6} \\ &= \frac{1}{\Delta(k)}\, d\mu(\theta\underset{x}{*}k) && \text{by A2.3.4} \\ &= d\mu(\theta) && \text{by A2.4.4} \end{aligned}$$

Hence $d\mu(.)$ is also a left loop invariant differential on Θ based on the loop $\underset{D}{*}$: $\Theta \times \Theta\to\Theta$ and must therefore only differ by a multiplicative constant from any other left loop invariant differential on Θ based on the loop $\underset{D}{*}$: $\Theta \times \Theta\to\Theta$ that is to say

$$d\mu'(\theta) = Bd\mu(\theta) \qquad (A1.5.7)$$

where B is a constant independent of θ.

Consider $d\nu(\cdot)$ the right loop invariant differential on Θ based on the loop $\underset{x}{*}$ $\Theta \times \Theta\to\Theta$

$$\begin{aligned} \text{Now } d\nu(\theta\underset{D}{*}h) &= d\nu(\{\theta\underset{x}{*}(h\underset{x}{*}k)\}/_{x}k) && \text{by A1.5.3} \\ &= d\nu(\theta\underset{x}{*}(h\underset{x}{*}k)) && \text{by A2.4.9} \\ &= d\nu(\theta) && \text{by A2.3.6} \end{aligned}$$

Hence $d\nu(.)$ is also a right loop invariant differential on Θ based on the loop $\underset{D}{*}\ \Theta \times \Theta \to \Theta$ and also must only differ by a multiplicative constant from any other right loop invariant differential on Θ based on the loop $\underset{D}{*}\ \Theta \times \Theta \to \Theta$. From this fact and A1.5.7 we obtain

$$\Delta'(\theta) = C\Delta(\theta) \qquad \text{(A1.5.8)}$$

where $\Delta'(.)$ and $\Delta(.)$ are the ratios of the left and right loop invariant differentials on Θ based on $\underset{D}{*}$ and $\underset{x}{*}$ respectively and C is a constant independent of $\theta \varepsilon \Theta$.

From this we obtain

$$\Delta'(\theta \underset{D}{\backslash} [x]) = C\Delta(\theta \underset{D}{\backslash} [x])$$

From A1.5.4 $\quad = C\Delta(\{\theta \underset{x}{\backslash} \{[x] \underset{x}{*} k\}\} \underset{x}{/} k)$

From $[x] \underset{x}{*} k = i$ $\quad = C\Delta(\{\theta \underset{x}{\backslash} i\} \underset{x}{/} k)$

From A2.4.3 $\quad = C\dfrac{\Delta(\theta \underset{x}{\backslash} i)}{\Delta(k)} \qquad$ (A1.5.9)

Consider the pivotal distribution of θ given D and [x] $p(\theta|D,[x])$ from (2.3.23).

$$= k(D)f(\theta^{-1} o x)J_{\Omega}(\theta \underset{D}{\backslash} [x] | D)\Delta(\theta \underset{D}{\backslash} [x])d\mu(\theta)$$

From A1.5.6, A1.5.7, A1.5.9 and A1.5.4 we obtain

$$= k(D)f(\theta^{-1} o x)AJ_{\Omega}(\theta \underset{x}{\backslash} i | D)\frac{C}{\Delta(k)}\Delta(\theta \underset{x}{\backslash} i)Bd\mu(\theta)$$

$$= k(x)f(\theta^{-1} o x)J_{\Omega}(\theta \underset{x}{\backslash} i | D)\Delta(\theta \underset{x}{\backslash} i)d\mu(\theta)$$

where $k(x) = k(D)AB\frac{C}{\Delta(k)}$, the constant of proportionality,

$= p(\theta|x,i)$ by 2.3.23, the pivotal distribution of θ given x, hence result. ||

A1.6 <u>Theorem 2.4.1</u>

For every conditional prior distribution $\pi(\theta|[e])$

$$E_{[x]|D}\{\alpha\{S(\theta^*,[x],D)|\pi(\theta|[e])\}|D\}$$

$$= \int_{[x]\varepsilon\Theta} \alpha\{S(\theta^*,[x],D)|\pi(\theta|[e])\}\tilde{p}([x]|D,\pi(\theta|[e]))\, d[x]$$

$$= \int_{\theta\varepsilon S(\theta^*,[x],D)} p(\theta|[x],D)d\theta$$

<u>Proof</u>:

Let $x = \theta o e \quad e \sim f(e)$ (A1.6.1)

$e,x\varepsilon\Omega \quad \theta\varepsilon\Theta$ be a pivotal model, and let $\pi(\theta|[e])$ be the conditional prior distribution for $\theta\varepsilon\Theta$ given $[e]$.

By 2.3.15 we can rewrite A1.6.1 in terms of a reduced pivotal model:

$D(x) = D(e) = D$

$[x] = \theta*[e]$ * conditional on D

$[e] \sim g([e]|D) \quad \theta \sim \pi(\theta|[e])$ (A1.6.2)

with $x,e\varepsilon\Omega \quad \theta\varepsilon\Theta$

We first calculate the joint distribution of [x] and θ given D, $\tilde{p}([x],\theta|D)$ say, as follows:

The joint distribution of [e] and θ given D is

$$g([e]|D)\Pi(\theta|[e])d\theta d[e] =$$

$$= \Pi(\theta|[e])k(D)f([e]oD)J_{\Omega}([e]|D)d\mu([e])d\theta$$

by 2.3.14

Letting $[e] = \theta\backslash[x]$ we obtain

$$\tilde{p}([x],\theta|D)d[x]d\theta = \quad (A1.6.3)$$

$$= \Pi(\theta|\theta\backslash[x])k(D)f(\theta^{-1}ox)J_{\Omega}(\theta\backslash[x]|D)d\mu([x])d\theta$$

by A2.4.5

The marginal distribution of [x] given D thus becomes

$$\tilde{p}([x]|D) = \int_{\theta\varepsilon\Theta} \tilde{p}([x],\theta|D)d\theta \quad (A1.6.4)$$

and the conditional distribution of θ given [x] and D is

$$\tilde{p}(\theta|[x],D) = \frac{\tilde{p}([x],\theta|D)}{\tilde{p}([x]|D)} \quad (A1.6.5)$$

The posterior probability of a particular set $S(\theta^*,[x],D) = S$ say given a conditional prior distribution $\Pi(\theta|[e])$ is

$$\alpha\{S(\theta^*,[x],D)|\Pi(\theta|[e])\} =$$

$$= \int_{\theta\varepsilon S(\theta^*,[x],D)} \tilde{p}(\theta|[x],D)d\theta \quad (A1.6.6)$$

And so the expected posterior probability is

$$E_{[x]|D}\{\alpha\{S(\theta^*,[x],D)|\Pi(\theta|[e])\}\} =$$

$$= \int_{[x]\varepsilon\Theta} \int_{\theta\varepsilon S} \tilde{p}([x],\theta|D)d\theta d[x]$$ by A1.6.6 and A1.6.5

(A1.6.7)

$$= \int_{[x]\varepsilon\Theta} \int_{\theta\varepsilon S} \Pi(\theta|\theta\backslash[x])k(D)f(\theta^{-1}ox)J_{\Omega}(\theta\backslash[x]|D)d\theta d\mu[x]$$

by A1.6.3

Let $\theta = [x]/[e]$.

Now $d\theta = J_*(\theta,i|D)d\mu(\theta) = J_*\frac{([x]/[e],i|D)}{\Delta([e])}d\mu([e])$

by A2.4.8

Also $\theta\varepsilon S(\theta^*,[x],D)$ iff $[e]\varepsilon E^*\varepsilon\,\mathfrak{D}$ by 2.4.1

where E^* is the event from $\mathfrak{D}$ in 1-1 correspondence by 2.4.1, thus A1.6.7 becomes

$$\int_{[x]\varepsilon\Theta} \int_{[e]\varepsilon E^*} \Pi([x]/[e]|[e])k(D)f([e]oD)J_{\Omega}([e]|D)$$

$$J_*\frac{([x]/[e],i|D)}{\Delta([e])}d\mu[e]d\mu[x]$$

Changing the order of integration we obtain

$$= \int_{[e]\varepsilon E^*} \int_{[x]\varepsilon\Theta} k(D)\Pi([x]/[e]|[e])f([e]oD)J_\Omega([e]|D) \frac{J_*([x]/[e],i|D)}{\Delta([e])} d\mu[x]d\mu[e] \quad (A1.6.8)$$

Now let $[x] = \theta*[e]$

$$\frac{J_*([e]/[x],i|D)}{\Delta([e])} d\mu[x] = \frac{J_*(\theta,i|D)}{\Delta([e])} d\mu(\theta*[e])$$

$$= J_*(\theta,i|D)d\mu(\theta) \qquad \text{by A2.4.4}$$

$$= d\theta$$

Also for fixed [e],[x] taking values on Θ implies θ taking values on Θ. Thus A1.6.8 becomes

$$\int_{[e]\varepsilon E^*} k(D)f([e]oD)J_\Omega([e]|D) \int_{\theta\varepsilon\Theta}\Pi(\theta|[e])d\theta d\mu[e]$$

$$= \int_{[e]\varepsilon E^*} k(D)f([e]oD)J_\Omega([e]|D)d\mu[e]$$

$$= P_M(\Theta^*|D) \qquad \text{by 2.4.2} \quad (A1.6.9)$$

$$= \int_{\theta\varepsilon S(\Theta^*,[x],D)} p(\theta|[x]oD)d\theta \qquad \text{by lemma 2.4.1}$$

Hence result. ||

A2.1 Definition and some properties of loops

A binary operation * on a non-empty set Θ is defined to be a single-valued mapping from some subset of $\Theta \times \Theta$ into Θ and we denote this mapping by $\theta_1 * \theta_2$ where $\theta_1, \theta_2 \varepsilon \Theta$.

A groupoid $\Theta,*$ is a system consisting of a non-empty set Θ and a binary operation * on Θ such that $\theta_1 * \theta_2$ is defined in Θ for all $\theta_1, \theta_2 \varepsilon \Theta$.

We define a loop $\Theta,*$ as a groupoid with the following properties:

L1 $\forall \theta_1, \theta_2 \varepsilon \Theta$ then $\theta_1 * \theta_2 \Theta$ that is to say Θ is closed under the operation *.

L2 $\exists i \varepsilon \Theta$ s.t. $\forall \theta \varepsilon \Theta$ $i * \theta = \theta = \theta * i$ that is to say Θ has an identity element namely i.

L3 $\forall \theta_1, \theta_2 \varepsilon \Theta$ $\exists ! h_1, h_2 \varepsilon \Theta$ s.t. $\theta_1 * h_1 = \theta_2$, $h_2 * \theta_1 = \theta_2$

An example of a loop is given in A2.2.

Two further binary operations defined on Θ which are associated with a loop are the following:

The operation of left division ($\backslash$)

$$a \backslash b = c \quad \text{iff} \quad a * c = b \qquad (A2.1.1)$$

and the operation of right division (/)

$$a/b = c \quad \text{iff} \quad c*b = a \qquad (A2.1.2)$$

These binary operations are both well defined being the solutions of the equations given in L3. An example of the operations / and \ is also given in A2.2.

The three binary operations *,\,/ defined on Θ together form an algebra. The following lemma gives some of their properties.

Lemma A2.1.1

$\forall h,g \varepsilon \Theta$

i) $(h/g)*g = h$

ii) $h*(h\backslash g) = g$

Proof:

i) Let $k = h/g$ then by definition we have $h = k*g$, hence $(h/g)*g = k*g = h$, hence part (i).

ii) Let $k = h\backslash g$ then by definition we have $g = h*k$, hence $h*(h\backslash g) = h*k = g$, hence part (ii).

Hence lemma. ||

A2.2 Example of a loop and its associated binary operations \ and /

Consider the set $\Theta = \{0,1,2,3,4\}$ and the binary operation * defined to be

*	0	1	2	3	4
0	0	1	2	3	4
1	1	0	3	4	2
2	2	3	4	1	0
3	3	4	0	2	1
4	4	2	1	0	3

This can easily be shown to be a loop with the element $0 \varepsilon \Theta$ being the identity element of the loop.

Note the binary operation * defined on Θ does not form a group since it does not have the property of associativity for consider the following:

$$(3*2)*4 = 0*4 = 4$$

but $$3*(2*4) = 3*0 = 3$$

hence the above loop is not a group.

The binary operations of left and right division can easily be shown to be

/	0	1	2	3	4
0	0	1	3	4	2
1	1	0	4	2	3
2	2	4	0	3	1
3	3	2	1	0	4
4	4	3	2	1	0

\	0	1	2	3	4
0	0	1	2	3	4
1	1	0	4	2	3
2	4	3	0	1	2
3	2	4	3	0	1
4	3	2	1	4	0

A2.3 Loop-invariant differentials

Consider the differential $dm(z) = \dfrac{dz}{J_\Omega([z]):D(z)}$ defined on Ω where $J_\Omega(\theta:z) = \left|\dfrac{\partial(\theta o z)}{\partial(z)}\right|$ as defined in 2.3.6.

Consider $J_\Omega(h*g: D) = \left|\dfrac{\partial([h*g]oD)}{\partial(D)}\right|$

$$= \left|\frac{\partial(ho(goD))}{\partial(goD)}\right| \left|\frac{\partial(goD)}{\partial(D)}\right|$$

$$= J_\Omega(h: goD)J_\Omega(g: D) \qquad \text{(A2.3.1)}$$

We can now show that the differential dm(z) is a loop-invariant differential that is to say

$$dm(\theta oz) = dm(z) \quad \forall\, \theta\varepsilon\Theta \quad \forall\, z\varepsilon\Omega$$

Now $\quad dm(\theta oz) = \dfrac{d(\theta oz)}{J_\Omega([\theta oz]: D(\theta oz))}$

$$= \frac{d(\theta oz)}{J_\Omega(\theta*[z]: D(z))}$$

By A2.3.1 and definition of $J_\Omega(.|.)$

$$= \frac{J_\Omega(\theta:z)dz}{J_\Omega(\theta:[z]oD(z))J_\Omega([z]:D(z))}$$

$$= \frac{dz}{J_\Omega([z]:D(z))} = dm(z) \qquad \text{(A2.3.2)}$$

Hence dm(z) is a loop-invariant differential on Ω..

Consider the differential $d\mu(\theta) = \frac{d\theta}{J_*(\theta:\ i|D)}$ defined on Θ where $J_*(\theta:\ h|D) = \left|\frac{\partial(\theta*h)}{\partial(h)}\right|$ as defined in 2.3.9 where * is conditional on D.

Consider $J_*(h*g,i|D) = \left|\frac{\partial[(h*g)*i]}{\partial[i]}\right|$

$$= \left|\frac{\partial(h*(g*i))}{\partial(g*i)}\right| \left|\frac{\partial(g*i)}{\partial(i)}\right|$$

$$= J_*(h,g|D)J_*(g,i|D) \qquad \text{(A2.3.3)}$$

We can now show that the differential $d\mu(\theta)$ is a left loop invariant differential on Θ that is to say $d\mu(h*\theta) = d\mu(\theta)$ where * is conditional on D.

Now $\quad d\mu(h*\theta) = \frac{d(h*\theta)}{J_*(h*\theta,i|D)}$

By A2.3.3 and definition of $J_*(\ )$

$$= \frac{J_*(h,\theta|D)d\theta}{J_*(h,\theta|D)J_*(\theta,i|D)}$$

$$= \frac{d\theta}{J_*(\theta,i|D)} = d\mu(\theta) \qquad \text{(A2.3.4)}$$

hence $d\mu(\ )$ is a left loop-invariant differential on Θ.

Consider the differential $d\nu(\theta) = \dfrac{d\theta}{J_{*R}(\theta:i|D)}$ defined on Θ, where $J_{*R}(\theta:h|D) = \left|\dfrac{\partial(h*\theta)}{\partial(h)}\right|$ as defined in 2.3.17 where $*$ is conditional on D.

Consider $J_{*R}(g*h,i|D) = \left|\dfrac{\partial(i*\{g*h\}}{\partial(i)}\right|$

$$= \left|\frac{\partial(\{i*g\}*h)}{\partial(i*g)}\right| \left|\frac{\partial(i*g)}{\partial(i)}\right|$$

$$= J_{*R}(h,g|D)J(g,i|D) \qquad \text{(A2.3.5)}$$

We can now show that the differential $d\nu(\theta)$ is a right loop invariant differential on Θ that is to say $d\nu(\theta*h) = d\nu(\theta)$ where $*$ is conditional on D.

Consider $d\nu(\theta*h) = \dfrac{d(\theta*h)}{J_{*R}(\theta*h,i|D)}$

By A2.3.5 and definition of $J_{*R}(\ )$

$$= \frac{J_{*R}(h,\theta|D)d\theta}{J_{*R}(h,\theta|D)J_{*R}(\theta,i|D)}$$

$$= \frac{d\theta}{J_{*R}(\theta,i|D)} = d\nu(\theta) \qquad \text{(A2.3.6)}$$

hence $d\nu(\theta)$ is a right loop invariant differential on Θ.

A2.4 Some properties of loop invariant differentials

We first examine $\Delta(.)$ the ratio of the right and left loop invariant differentials as defined in (2.3.20). Consider $\Delta(h*g)$

$$= \frac{J_{*R}(h*g,i|D)}{J_*(h*g,i|D)} \qquad \text{by 2.3.21}$$

$$= \frac{J_{*R}(g,h|D)}{J_*(h,g|D)} \frac{J_{*R}(h,i|D)}{J_*(g,i|D)} \qquad \text{by A2.3.3 and A2.3.5}$$

Now it can be shown that

$$\frac{J_{*R}(g,h|D)}{J_*(h,g|D)} = \frac{J_{*R}(g,i|D)}{J_*(h,i|D)}$$

hence $\Delta(h*g) = \frac{J_{*R}(h,i|D)}{J_*(h,i|D)} \frac{J_{*R}(g,i|D)}{J_*(g,i|D)} = \Delta(h)\Delta(g)$ (A2.4.1)

Consider $\Delta(h*\{h\backslash g\}) = \Delta(g)$ by lemma A2.1.1

but $\Delta(h*\{h\backslash g\}) = \Delta(h)\Delta(h\backslash g)$ by A2.4.1

hence $\Delta(h\backslash g) = \frac{\Delta(g)}{\Delta(h)}$ if $\Delta(h) \neq 0$ (A2.4.2)

Consider $\Delta(\{g/h\}*h) = \Delta(g)$ by lemma A2.1.1

but $\Delta(\{g/h\}*h) = \Delta(g/h)\Delta(h)$ by A2.4.1

hence $\Delta(g/h) = \frac{\Delta(g)}{\Delta(h)}$ if $\Delta(h) \neq 0$ (A2.4.3)

For the rest of this section we examine some of the properties of $d\mu(.)$ and $d\nu(.)$ the left and right loop invariant differentials as defined in (2.3.10) and (2.3.18) respectively.

Consider $d\mu(\theta*h) = \Delta(\theta*h)d\nu(\theta*h)$ by 2.3.20

$= \Delta(h)\Delta(\theta)d\nu(\theta)$ by A2.4.1 and A2.3.6

thus we have $\underline{d\mu(\theta*h) = \Delta(h)d\mu(\theta)}$ by 2.3.20 (A2.4.4)

Consider $d\mu(h*\{h\backslash\theta\}) = d\mu(\theta)$ by lemma A2.1.1

but $d\mu(h*\{h\backslash\theta\}) = d\mu(h\backslash\theta)$ by A2.3.4

hence $\underline{d\mu(h\backslash\theta) = d\mu(\theta)}$ (A2.4.5)

Consider $d\mu(\{\theta/h\}*h) = d\mu(\theta)$ by lemma A2.1.1

but $d\mu(\{\theta/h\}*h) = \Delta(h)d\mu(\theta/h)$ by A2.4.4

hence $\underline{d\mu(\theta/h) = \frac{1}{\Delta(h)} d\mu(\theta) \text{ if } \Delta(h) \neq 0}$ (A2.4.6)

Consider $d\mu(\theta*\{\theta\backslash h\}) = d\mu(\theta\backslash h)$ by A2.3.4

but $d\mu(\theta*\{\theta\backslash h\}) = \Delta(\theta\backslash h)d\mu(\theta)$

hence $\underline{d\mu(\theta\backslash h) = \Delta(\theta\backslash h)d\mu(\theta)}$ (A2.4.7)

Consider $d\mu(\{h/\theta\}*\theta) = d\mu(\theta)$ by A2.3.4

but $d\mu(\{h/\theta\}*\theta) = \Delta(\theta)d\mu(h/\theta)$ by A2.4.4

hence $\underline{d\mu(h/\theta) = \frac{1}{\Delta(\theta)} d\mu(\theta) \text{ if } \Delta(\theta) \neq 0}$ (A2.4.8)

Consider $d\nu(\{\theta/h\}*h) = d\nu(\theta/h)$ by A2.3.6

but $= d\nu(\theta)$ by lemma A2.1.1

hence $\underline{d\nu(\theta/h) = d\nu(\theta)}$ (A2.4.9)

A3.1 An Algorithm to Solve a certain Minimisation Problem

The problem is the minimise with respect to $\tau > 0$ and $\underline{\theta} = (\theta_1,\ldots,\theta_p)$, functions of the following form:

$$s(\underline{\theta},\tau) = \frac{1}{2\tau^2}\sum_{i=1}^{m} f_i^2(\underline{\theta}) + g(\underline{\theta}) + \lambda\log\tau \qquad (A3.1.1)$$

where $g(\cdot)$ and $f_i(\cdot)$ $i = 1,\ldots,m$ are arbitrary functions such that $g(\cdot)$ is slowly varying compared to $\sum_{i=1}^{m} f_i^2(\cdot)$.

The algorithm used is a two stage iterative process. The first stage is to assume that the $\underline{\theta}$ are fixed. The problem then reduces to minimising (A3.1.1) with respect to τ. This is achieved by differentiating with respect to τ and setting this to zero; we thus obtain:

$$\frac{\lambda}{\tau} - \frac{1}{\tau^3}\sum_{i=1}^{m} f_i^2(\underline{\theta}) = 0$$

or rearranging:

$$\tau^2 = \frac{1}{\lambda}\sum_{i=1}^{m} f_i^2(\theta_i) \qquad (A3.1.2)$$

assuming $\tau,\tau^{-1} \neq 0$.

The value of τ given in (A3.1.2) minimises this particular reduced problem.

The second stage is to assume τ is fixed and that we have an estimate $\underline{\theta}_n$ say of the value of the parameters that minimise (A3.1.1).

We will obtain an improved estimate of the parameters, $\underline{\theta}_{n+} = \underline{\theta}_n + \underline{\delta\theta}_n$ say, as follows:

Taking linear expansions of $g(\cdot)$ and $f_i(\cdot)$ $i = 1,\ldots,n$ (A3.1.1) becomes:

$$\frac{1}{2\tau^2}(\underline{f}_n - X_n\underline{\delta\theta}_n)^2 + g(\underline{\theta}_n) + \underline{g}_n\underline{\delta\theta}_n + \lambda\log\tau \qquad \text{(A3.1.3)}$$

where $\underline{f}_n^T = (f_1(\underline{\theta}_n),\ldots,f_m(\underline{\theta}_n))$

$$\underline{g}_n^T = \left(\frac{\partial g(\cdot)}{\partial\theta_1},\ldots,\frac{\partial g(\cdot)}{\partial\theta_p}\right)\Bigg|_{\underline{\theta} = \underline{\theta}_n}$$

$$X_n = \left(\begin{matrix} \dfrac{-\partial f_1(\cdot)}{\partial\theta_1},\ldots, & \dfrac{-\partial f_1(\cdot)}{\partial\theta_p} \\ \vdots & \vdots \\ \dfrac{-\partial f_m(\cdot)}{\partial\theta_1},\ldots, & \dfrac{-\partial f_m(\cdot)}{\partial\theta_p} \end{matrix}\right)_{\underline{\theta} = \underline{\theta}_m}$$

We require to minimise (A3.1.3) with respect to $\underline{\delta\theta}_n$. This is achieved by differentiating with respect to $\underline{\delta\theta}_n$ and setting this to zero, we thus obtain:

$$\frac{1}{\tau^2}(X_n^T X_n\underline{\delta\theta}_n - X_n^T\underline{f}_n) + \underline{g}_n = 0$$

Rearranging we obtain:

$$\underline{\delta\theta}_n = (X_n^T X_n)^{-1} X_n^T \underline{f}_n - \tau^2 (X_n^T X_n)^{-1} g_n \qquad (A3.1.4)$$

and the improved estimate becomes

$$\underline{\theta}_{n+1} = \underline{\theta}_n + \underline{\delta\theta}_n \qquad (A3.1.5)$$

Note in the calculation of $(X_n^T X_n)^{-1}$ it is often convenient to decompose the X_n matrix to

$$X_n = Q_n R_n$$

where Q_n is a mxp matrix whose columns are orthogonal and R_n is a pxp upper triangular matrix. This can be achieved using Givens transformations, see Gentleman (1973).

The iterative procedure to minimise (A3.1.1) thus becomes:

Given a estimate, $\underline{\theta}_n$ say, of the parameters $\hat{\underline{\theta}}$, where $\hat{\underline{\theta}}$ and $\hat{\tau}$ are the values of the parameters that achieve the desired minimisation.

By fixing $\underline{\theta}_n$ calculate τ_n from (A3.1.2), then fixing τ_n calculate $\underline{\theta}_{n+1}$ by (A3.1.4) and (A3.1.5). Again fixing $\underline{\theta}_{n+1}$ we calculate τ_{n+1} from (A3.1.2) etc, etc. The procedure can be started by an initial estimate of $\hat{\underline{\theta}}$, $\underline{\theta}_o$ say.

Experience has shown that $\underline{\theta}_n, \tau_n$ converge to the required values $\hat{\underline{\theta}}, \hat{\tau}$.

APPENDIX 3.2

Circle fitting program using Pivotal Probability

Written for the BBC Microcomputer.

```
10@%=&20509
20DIM Z(1000)
30DIM R(3,3)
40DIM D(3)
50DIM X(3)
60REM INPUTS DATA FROM TAPE
70INPUT"FILE NAME",A$
80XX=OPENIN(A$)
90INPUT#XX,N
100INPUT#XX,V
110IF V*N>1000 THEN PRINT"DATA SET TOO LARGE":STOP
120FORI=1TON
130PRINT:PRINT"VECTOR ";I
140FORJ=0TO(V-1)
150INPUT#XX,ZZ
160PRINT ZZ;" ";
170Z(J*N+I)=ZZ
180NEXT:NEXT
190REM SELECTS DATA TO BE ANALYSED
200PRINT 'N;" VECTORS WITH ";V;" ELEMENTS"
210INPUT "INPUT CUT; N1 TO N2 "N1,N2
220IF N1<=0 OR N1>=N2 OR N2>1000 THEN210
230N3=N2-N1+2
240REM FITS LEAST SQUARES CIRCLE
250PROCCLEAR
260X(0)=1.0
270FOR I=N1 TO N2
280X(1)=Z(I):X(2)=Z(N+I)
290PRINT X(1);" ";X(2)
300X(3)=(X(1)*X(1)+X(2)*X(2))/2
310W=1
320FOR J=0 TO 2
330IF W=0 OR X(J)=0 THEN 440
340DR=D(J)
350D(J)=D(J)+W*X(J)*X(J)
360C=DR/D(J)
370S=W*X(J)/D(J)
380W=W*C
390FOR K=(J+1) TO 3
400XK=X(K)
410X(K)=X(K)-X(J)*R(J,K)
420R(J,K)=C*R(J,K)+S*XK
430NEXT
440NEXT
450D(3)=D(3)+W*X(3)*X(3)
460NEXT
470X(2)=R(2,3)
480X(1)=R(1,3)-R(1,2)*X(2)
490X(0)=R(0,3)-R(0,1)*X(1)-R(0,2)*X(2)
500C=X(0):A=X(1):B=X(2)
510RR=SQR(2*C+A*A+B*B)
520PROCPRINT(A$)
```

```
530REM SOLVES FOR PIVOTAL ESTIMATES
540PRINT "          A                 B          RADIUS"
550T2=0
560KK=0
570REPEAT
580KK=KK+1
590PROCCLEAR
600RL=RR:AL=A:BL=B
610X(0)=1:GA=0:GB=0
620FOR I=N1 TO N2
630XX=Z(I)-A:YY=Z(N+I)-B
640RI=SQR(XX*XX+YY*YY)
650X(1)=XX/RI:X(2)=YY/RI
660X(3)=RI-RR
670GA=GA+X(1)/RI:GB=GB+X(2)/RI
680W=1
690FOR J=0 TO 2
700IF W=0 OR X(J)=0 THEN 810
710DR=D(J)
720D(J)=D(J)+W*X(J)*X(J)
730C=DR/D(J)
740S=W*X(J)/D(J)
750W=W*C
760FOR K=(J+1) TO 3
770XK=X(K)
780X(K)=X(K)-X(J)*R(J,K)
790R(J,K)=C*R(J,K)+S*XK
800NEXT
810NEXT
820D(3)=D(3)+W*X(3)*X(3)
830NEXT
840T22=D(3)/N3
850T2=SQR(T22)
860GB=GB-R(1,2)*GA
870R(2,3)=R(2,3)+T22*GB/D(2)
880R(1,3)=R(1,3)+T22*GA/D(1)
890X(2)=R(2,3)
900X(1)=R(1,3)-R(1,2)*X(2)
910X(0)=R(0,3)-R(0,1)*X(1)-R(0,2)*X(2)
920RR=RR+X(0):A=A+X(1):B=B+X(2)
930PRINT A;" ";B;" ";RR
940ER=ABS(RL-RR)+ABS(AL-A)+ABS(BL-B)
950UNTIL (ER<1E-5 OR KK>15)
960PROCPRINT("PIVOTAL ESTIMATES")
970@%=&10509
980VDU2:PRINT"S.D. OF CO-ORDS ";T2
990T2=T2/RR:PRINT"CON-PARAMETER    ";T2
```

```
1000REM CALCULATES DECOMPOSED HESSIAN MATRIX AT ESTIMATES
1010PROCCLEAR
1020FOR I=N1 TO N2
1030X(0)=2*(Z(I)-A)/(RR*RR)
1040X(1)=2*(Z(N+I)-B)/(RR*RR)
1050X(2)=RR*(X(0)*X(0)+X(1)*X(1))/2
1060W=1
1070FOR J=0 TO 1
1080IF W=0 OR X(J)=0 THEN 1190
1090DR=D(J)
1100D(J)=D(J)+W*X(J)*X(J)
1110C=DR/D(J)
1120S=W*X(J)/D(J)
1130W=W*C
1140FOR K=(J+1) TO 2
1150XK=X(K)
1160X(K)=X(K)-X(J)*R(J,K)
1170R(J,K)=C*R(J,K)+S*XK
1180NEXT
1190NEXT
1200D(2)=D(2)+W*X(2)*X(2)
1210NEXT
1220REM INVERTS HESSIAN MATRIX
1230R(0,2)=R(0,1)*R(1,2)-R(0,2)
1240R(0,1)=-R(0,1):R(1,2)=-R(1,2)
1250T22=T2*T2
1260D(0)=D(0)/T22:D(1)=D(1)/T22:D(2)=D(2)/T22
1270R(2,2)=R(0,2)*R(0,2)/D(0)+R(1,2)*R(1,2)/D(1)+1/D(2)
1280R(1,2)=R(0,1)*R(0,2)/D(0)+R(1,2)/D(1)
1290R(1,1)=R(0,1)*R(0,1)/D(0)+1/D(1)
1300R(0,2)=R(0,2)/D(0)
1310R(0,1)=R(0,1)/D(0)
1320R(0,0)=1/D(0)
1330PRINT '"COVARIANCE MATRIX"'
1340PRINT "A  ";R(0,0)
1350PRINT "B  ";R(0,1);" ";R(1,1)
1360PRINT "R  ";R(0,2);" ";R(1,2);" ";R(2,2)
1370@%=&20509 :VDU3:GOTO200
1380DEF PROCCLEAR
1390FOR I=0 TO 3
1400FOR J=0 TO 3
1410R(J,I)=0
1420NEXT:D(I)=0:NEXT
1430ENDPROC
1440DEF PROCPRINT(B$)
1450VDU2
1460 PRINTTAB(15);B$
1470PRINT "NUMBERS ";N1;" TO ";N2
1480PRINT'"CO-ORDS OF CENTRE ";A;" , ";B
1490PRINT "RADIUS ";RR;''
1500VDU3
1510ENDPROC
```

APPENDIX 3.3

Ellipse fitting program using Pivotal Probability

Written for the BBC Microcomputer.

```
 10@%=&20509
 20DIM Z(1000)
 30DIM R(5,5)
 40DIM D(5)
 50DIM X(5)
 60INPUT"FILE NAME",A$
 70XX=OPENIN(A$)
 80INPUT#XX,N
 90INPUT#XX,V
100IF V*N>1000 THEN PRINT"DATA SET TOO LARGE":STOP
110FORI=1TON
120PRINT:PRINT"VECTOR ";I
130FORJ=0TO(V-1)
140INPUT#XX,ZZ
150PRINT ZZ;" ";
160Z(J*N+I)=ZZ
170NEXT:NEXT
180PRINT 'N;" VECTORS WITH ";V;" ELEMENTS"
190INPUT "INPUT CUT; N1 TO N2 "N1,N2
200IF N1<=0 OR N1>=N2 OR N2>1000 THEN190
210N3=N2-N1+2
220X(0)=1.0
230FOR I=N1 TO N2
240X(1)=Z(I):X(2)=Z(N+I)
250PRINT X(1);" ";X(2)
260X(3)=X(2)*X(2):X(4)=X(2)*X(1):X(5)=X(1)*X(1)
270W=1
280FOR J=0 TO 4
290IF W=0 OR X(J)=0 THEN 400
300DR=D(J)
310D(J)=D(J)+W*X(J)*X(J)
320C=DR/D(J)
330S=W*X(J)/D(J)
340W=W*C
350FOR K=(J+1) TO 5
360XK=X(K)
370X(K)=X(K)-X(J)*R(J,K)
380R(J,K)=C*R(J,K)+S*XK
390NEXT
400NEXT
410D(5)=D(5)+W*X(5)*X(5)
420NEXT
430X(4)=R(4,5)
440X(3)=R(3,5)-R(3,4)*X(4)
450X(2)=R(2,5)-R(2,3)*X(3)-R(2,4)*X(4)
460X(1)=R(1,5)-R(1,2)*X(2)-R(1,3)*X(3)-R(1,4)*X(4)
470X(0)=R(0,5)-R(0,1)*X(1)-R(0,2)*X(2)-R(0,3)*X(3)-R(0,4)*X(4)
```

```
480PROCANG(-X(4),1,X(3))
490AA=X(4)*CS*SN+X(3)*SN*SN-CS*CS
500BB=X(3)*CS*CS-X(4)*CS*SN-SN*SN
510CC=X(2)*SN+X(1)*CS
520DD=X(2)*CS-X(1)*SN
530EE=X(0)
540PRINT X(4);"    ";AA
550PRINT X(3);"    ";BB
560PRINT X(2);"    ";CC
570PRINT X(1);"    ";DD
580PRINT X(0);"    ";EE
590AR=-CC/(2*AA):BR=-DD/(2*BB)
600Q=AR*AR*AA+BR*BR*BB-EE
610R1=SQR(ABS(Q/AA)):R2=SQR(ABS(Q/BB))
620A=AR*CS-BR*SN:B=AR*SN+BR*CS
630PROCPRINT(A$)
640PROCCLEAR
650REM CALULATES PIVOTAL ESTIMATES
660KK=0
670Q1=(CS/R1)^2+(SN/R2)^2
680Q2=CS*SN*(1/(R2*R2)-1/(R1*R1))
690Q3=(SN/R1)^2+(CS/R2)^2
700REPEAT
710KK=KK+1
720PROCCLEAR
730AL=A:BL=B
740Q1L=Q1:Q2L=Q2:Q3L=Q3
750G0=0:G1=0:G2=0:G3=0:G4=0
760FOR I=N1 TO N2
770 XX=Z(I)-A:YY=Z(N+I)-B
780UI=SQR(Q1*XX*XX+2*Q2*XX*YY+Q3*YY*YY)
790X(0)=(Q1*XX+Q2*YY)/UI:G0=G0-X(0)/UI
800X(1)=(Q3*YY+Q2*XX)/UI:G1=G1-X(1)/UI
810X(2)=-XX*XX/(2*UI):G2=G2-X(2)/UI
820X(3)=-XX*YY/UI:G3=G3-X(3)/UI
830X(4)=-YY*YY/(2*UI):G4=G4-X(4)/UI
840X(5)=UI-1
850W=1
860FOR J=0 TO 4
870IF W=0 OR X(J)=0 THEN 980
880DR=D(J)
890D(J)=D(J)+W*X(J)*X(J)
900C=DR/D(J)
910S=W*X(J)/D(J)
920W=W*C
930FOR K=(J+1) TO 5
940XK=X(K)
950X(K)=X(K)-X(J)*R(J,K)
960R(J,K)=C*R(J,K)+S*XK
970NEXT
980NEXT
990D(5)=D(5)+W*X(5)*X(5)
1000NEXT
```

```
1010S2=D(5)/N3:S1=SQR(S2)
1020QQ=N3/(Q1*Q3-Q2*Q2)
1030G0=G0*S2:G1=G1*S2
1040G2=(G2-Q3*QQ/2)*S2
1050G3=(G3+QQ*Q2)*S2
1060G4=(G4-Q1*QQ/2)*S2
1070G1=G1-R(0,1)*G0
1080G2=G2-R(1,2)*G1-R(0,2)*G0
1090G3=G3-R(2,3)*G2-R(1,3)*G1-R(0,3)*G0
1100G4=G4-R(3,4)*G3-R(2,4)*G2-R(1,4)*G1-R(0,4)*G0
1110R(0,5)=R(0,5)-G0/D(0)
1120R(1,5)=R(1,5)-G1/D(1)
1130R(2,5)=R(2,5)-G2/D(2)
1140R(3,5)=R(3,5)-G3/D(3)
1150R(4,5)=R(4,5)-G4/D(4)
1160X(4)=R(4,5)
1170X(3)=R(3,5)-R(3,4)*X(4)
1180X(2)=R(2,5)-R(2,3)*X(3)-R(2,4)*X(4)
1190X(1)=R(1,5)-R(1,2)*X(2)-R(1,3)*X(3)-R(1,4)*X(4)
1200X(0)=R(0,5)-R(0,1)*X(1)-R(0,2)*X(2)-R(0,3)*X(3)-R(0,4)*X(4)
1210A=A+X(0):B=B+X(1)
1220Q1=Q1+X(2):Q2=Q2+X(3):Q3=Q3+X(4)
1230PRINT A;" ";B'Q1;" ";Q2;" ";Q3
1240ER=ABS(Q1-Q1L)+ABS(Q2-Q2L)+ABS(Q3-Q3L)
1250ER=ER+ABS(A-AL)+ABS(B-BL)
1260UNTIL (ER<2E-5 OR KK>15)
1270PROCANG(2*Q2,Q3,-Q1)
1280RR1=Q1+Q3:RR2=(Q1-Q3)/T
1290R1=SQR(ABS(2/(RR1+RR2)))
1300R2=SQR(ABS(2/(RR1-RR2)))
1310PROCPRINT("PIVOTAL ESTIMATES")
1320@%=&10509
1330VDU2:PRINT"CON-PARAMETER ";S1
1340REM CALCULATES HESSIAN MATRIX
1350PROCCLEAR
1360FOR I=N1 TO N2
1370XD=Z(I)-A:YD=Z(N+I)-B
1380XX=XD*CS+YD*SN:YY=YD*CS-XX*SN
1390UI=SQR((XX/R1)^2+(YY/R2)^2)
1400X(0)=-XX^2/(UI*R1^3)
1410X(1)=-YY^2/(UI*R2^3)
1420W=1
1430DR=D(0)
1440D(0)=D(0)+W*X(0)*X(0)
1450C=DR/D(0)
1460S=W*X(0)/D(0)
1470W=W*C
1480XK=X(1)
1490X(1)=X(1)-X(0)*R(0,1)
1500R(0,1)=C*R(0,1)+S*XK
1510D(1)=D(1)+W*X(1)*X(1)
1520NEXT
```

```
1530REM INVERTS HESSIAN MATRIX
1540D(0)=D(0)/S2:D(1)=D(1)/S2
1550R(0,0)=1/D(0)
1560R(1,1)=(R(0,1)^2)/D(0)+1/D(1)
1570R(0,1)=-R(0,1)/D(0)
1580PRINT'"COVARIANCE MATRIX"'
1590PRINT "R1  ";R(0,0)
1600PRINT "R2  ";R(0,1);" ";R(1,1)
1610VDU3:@%=&20509:GOTO180
1620STOP
1630DEFPROCCLEAR
1640FOR I= 0 TO 5
1650FOR J= 0 TO 5
1660R(J,I)=0
1670NEXT
1680D(I)=0
1690NEXT
1700ENDPROC
1710DEF PROCANG(A,B,C)
1720T=A/(B+C)
1730ANG=90*ATN(T)/PI
1740T=1/SQR(1+T*T)
1750CS=SQR((1+T)/2)
1760SN=SQR((1-T)/2)
1770ENDPROC
1780DEF PROCPRINT(B$)
1790VDU2
1800PRINTTAB(15);B$
1810PRINT "NUMBERS ";N1;" TO ";N2
1820PRINT'"RADII   ";R1;" ";R2
1830PRINT "CENTRE  ";A;" ";B
1840PRINT "ANGLE   ";ANG''
1850VDU3
1860ENDPROC
```

REFERENCES

ANGELL, I and BARBER, J (1977). An algorithm for fitting circles and ellipses to megalithic stone rings, *Science and Archael.* 20, 11-16.

BERAN, R J (1971). On distribution-free statistical inference with upper and lower probabilities, *Ann. Math. Stat.* 42, 157-168.

BESAG, J E (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *J R Statist. Soc. B* 36, 192-236.

BRUCK, R H (1971). *A Survey of Binary Systems*, Berlin: Springer-Verlag.

BUEHLER, R J (1971). Measuring information and uncertainty (with discussion). In *Foundations of Statistical Inference*, pp330-341, eds. Godamke, V P and Sprott, D A. Toronto: Holt, Rinehurt and Winston.

DAWID, A P and STONE, M (1982). The functional-model basis of fiducial inference (with discussion), *Ann. Statist.* 10, 1054-1074.

DEMPSTER, A P (1966). New methods for reasoning towards posterior distribution based on sample data, *Ann. Math. Stat.* 37, 355-374.

FISHER, R A (1930). Inverse probability, *Proc. Camb. Phil. Soc.* 26, 528-535.

FISHER, R A (1956). *Statistical Methods and Scientific Inference*, New York: Hafner (1973 ed.)

FRASER, D A S (1968). *The Structure of Inference*, New York: Wiley.

FRASER, D A S (1971). Events, information processing and the structural model (with discussion), In *Foundations of Statistical Inference*, pp 32-55, eds Godamke, V P and Sprott, D A. Toronto: Holt, Rinehurt and Winston.

GENTLEMAN, W M (1973). Least squares computations by Givens transformations without square roots. *J. Inst. Maths. Applics.* 12, 329-336.

HALMOS, P R (1950). *Measure Theory*. Princeton: D Van Nostrand.

MARDIA, K V and HOLMES, D (1980). A statistical analysis of megalithic data under elliptic pattern. *J R Statist. Soc. A.* 143, 293-302.

NAYLOR, J C and SMITH, A F M (1982). Applications of a method for the efficient computation of posterior distributions. *Appl. Statist.* 31, 214-225.

PLANTE, A (1979). On the validation of fiducial techniques. *Can. J Statist.* 7, 217-226.

SCOTT, P J (1981). Statistical analysis of circular profiles. MSc and DIC dissertation, Imperial College, University of London.

THOM, A and THOM, A S (1973). A megalithic lunar observatory in Orkney. The ring of Brogar and its cairns. *J Hist. Astron.* 4, 111-123.

THOM, A and THOM, A S (1978). *Megalithic remains in Britain and Brittany*. Oxford: Clarendon Press.

THOM, A, THOM A S and FOORD, T R (1976). Avebury (1): A new assessment of the geometry and metrology of the ring, *J Hist. Astron.* 7, 183-192.

WILKINSON, G N (1977). On resolving the controversy in statistical inference (with discussion), *J R Statist. Soc. B* 39, 119-171.