"MIXED SAMPLES, MIXED DISTRIBUTIONS AND THE MONOTONE LIKELIHOOD RATIO PROPERTY"

by JOHN DOUGLAS RICKARD

A Thesis presented to the Australian National University

for the Degree of Master of Science

STATEMENT OF AUTHORSHIP

Except where otherwise indicated by references the contents of this thesis are entirely my own work.

J.D. Rickard

CONTENTS

CHAPTER ONE

SYNOPSIS AND STATEMENT OF ORIGINAL CONTENT

The main body of this thesis, in particular Chapters Three and Four, is devoted to the problems of estimation and classification encountered when dealing with mixtures of uniform distributions or with mixtures of samples from uniform populations. Apparently, such questions have not been considered explicitly in the literature, possibly because approaches using the well known methods of moments and maximum likelihood are conceptually quite straightforward. However, there are some technical difficulties involved, not to mention considerable labour, and so the full exposition of these methods should be of use to the practical experimenter.

The formal definitions of mixtures, as well as the question of identifiability, are introduced in §2.1. The next section of Chapter Two outlines sufficient interesting and varied applications of mixtures to convince the reader of their importance in modern statistical theory. Finally, §2.3 traces the historical development of estimation methods suitable for use in mixtures of normal distributions. The various approaches are compared in a critical light and brief mention is made of some important unresolved difficulties.

Then in Chapter Three, which deals with mixtures of two one-parameter uniform distributions, the moment equations are solved and the necessary conditions for meaningful estimates of the parameters are displayed. So that an experimenter may readily examine his errors, §3.3 is devoted to the derivation of the asymptotic distribution of the moment estimators. Moreover, it is also shown that, by increasing sample sizes, one can always expect meaningful roots with probability tending to one. In §3.2 and §3.4 the method of maximum likelihood is discussed under a mixed samples and mixed distributions model respectively, and in both cases simple procedures are possible. Unfortunately, the asymptotic errors cannot be easily computed and so no immediate comparison with the moment estimators is possible. An illustrative example completes the chapter.

Chapter Four follows similar lines, except that mixtures of two-parameter uniform distributions are discussed. This of course provides considerable additional complication, but here again the moment estimators and their asymptotic distributions are obtained, and conditions for meaningful solutions are displayed and discussed. In §4.2 the maximum likelihood method under a mixed samples model is shown to suffer from the weakness that the parameters are not all estimable.

The fifth and last chapter may be regarded as something of an addendum to the rest of the thesis. It begins with an introduction to the property of monotone likelihood ratio (M.L.R.) and an exposition of the rôle played by this property in the analysis of mixed samples by the maximum likelihood method. In §5.2 we derive a general family of distributions possessing an M.L.R. and similar to one originally suggested by Karlin (1957 (a)). The family is of sufficient generality to warrant the discussion of methods for estimating the parameter and determining confidence and tolerance intervals for the parameter. This is carried out in §5.3 and §5.4. In §5.5 original use is made of the general family developed in §5.2 to offer a partial answer to a question implicit in the paper of Heathcote (1972), viz the investigation of necessary conditions on the variance-covariance matrix for a multivariate normal distribution to possess an M.L.R. in some statistic.

Acknowledgement. The author wishes to express his sincere gratitude to Dr S. John, with whom he had many helpful discussions throughout the preparation of this thesis, and to Professor C.R. Heathcote who posed the problem considered in §5.5.

CHAPTER TWO

INTRODUCTION TO THE THEORY OF MIXED DISTRIBUTIONS AND MIXED SAMPLES

2.1 DEFINITIONS AND PRELIMINARIES

Let R_1^m be a measurable subset of Euclidean m-space and let $\mathcal{F} = \{F(x;\theta),\ \theta \in R_1^m\}$ where $F(x;\theta)$ is a cumulative distribution function in the variable x for each $\theta \in R_1^m$, and is also measurable on the product space of x and θ. Then for any nondegenerate m-dimensional c.d.f. G whose induced Lebesgue-Stieltjes measure μ_G assigns measure one to R_1^m, the c.d.f.

$$H(x) = \int_{R_1^m} F(x;\theta)\,dG(\theta) \tag{1.1}$$

is called a *G-mixture of* $\mathcal{F}$ or, more briefly, a *mixture*. Note that a c.d.f. G is degenerate if it concentrates all its mass at a single point of R^m. Also, $\mathcal{F}$ is assumed tacitly to contain at least two elements.

This is the definition of a mixture of distributions as given by Teicher (1961) who also rigorously defined the property of identifiability of mixtures. Let $\mathcal{G}$ denote a class of c.d.f.'s [G] of the type considered above, $\mathcal{H}$ the induced class of mixtures [H] and $\mathcal{J}$ the class of degenerate distributions in R^m. Then $\mathcal{H}$ will be called *identifiable in* $\mathcal{G}$ (with respect to $\mathcal{F}$) if (1.1) effects a one-to-one correspondence between $\mathcal{H} \cup \mathcal{F}$ and $\mathcal{G} \cup \mathcal{J}$; or equivalently, if the relationship

$$H = \int F(x;\theta)\,dG(\theta) = \int F(x;\theta)\,dG^*(\theta)$$

implies $G = G^*$ for all $G^* \in \mathcal{G} \cup \mathcal{J}$. If $\mathcal{H}$ is identifiable in the class of all $G \setminus \mathcal{J}$, it will simply be called *identifiable*.

It is often advisable to establish identifiability before considering questions of estimation and hypothesis testing. Otherwise, if a mixture is not uniquely determined, one might not be able to uniquely "identify" the mixture from a set of estimates of the parameters nor make a choice between possibly indistinguishable hypotheses.

Teicher (1960, 1961, 1963) gives some general conditions for identifiability as well as a discussion of the question for several specific classes of mixtures. He shows, among other things, that identifiability holds for finite mixtures of normal, gamma and one-parameter uniform distributions. Using Laplace transforms, Feller (1943) has shown that mixtures of Poisson distributions are identifiable. Boes (1966) has given necessary and sufficient conditions for the identifiability of certain finite mixtures.

We shall be mainly concerned in the sequel with finite mixtures. These can be regarded as a subclass of the following type of mixture. Let each $G \in \mathcal{G}$ be a step function with steps at say $\theta_1, \theta_2, \ldots$ (with $\Theta_1 = \{\theta_1, \theta_2, \ldots\}$ possibly depending on G), so that equation (1.1) reduces to

$$H(x) = H_G(x) = \sum_{i=1}^{\infty} \alpha_i F(x; \theta_i) \qquad (1.2)$$

where α_i is the mass assigned by G to $F(.; \theta_i)$. Following the terminology of Teicher (1960), such a mixture will be designated a *countable* mixture. If $\Theta_1 = \Theta_1(G)$ contains, for each $G \in \mathcal{G}$, only a finite number of elements, the result of equation (1.2) will be called a *finite* mixture. (Note that this is not meant to imply that the set R_1^m of possible values of θ is finite or countable.) In addition, we shall call the individual distributions $F(.;\theta_i)$ being mixed to produce a particular H_G the *components* of the mixture H_G. The term *compound distribution* is usually used synonomously with the term mixture of distributions.

In the third and fourth chapters of this thesis we shall be concerned with the problem of estimation in a mixture of two uniform distributions. When these uniform distributions are of the two-parameter type, the question of identifiability does not seem to have been resolved in the literature. However, a little non-rigorous inspection suggests that uniqueness of the representation will only be in doubt if the parameters satisfy certain strict functional relationships. It is likely that these will not generally be satisfied and so there will be a unique representation of the parameters in most situations. In any case, the experimenter will often be interested only in an estimate of the overall distribution and will not be concerned as to whether he has obtained unique estimates for the components of the mixture.

Consider a sample of size N which is assumed to consist of N observations on a random variable X with the c.d.f.

$$H(x) = \sum_{i=1}^{k} \alpha_i F(x,\theta_i) \tag{1.3}$$

In many circumstances it may be just as reasonable to assume that the sample is actually a mixture of k separate samples of sizes $N_1, N_2, \ldots, N_k$ on the k populations represented by their distribution functions as $F(x,\theta_1), F(x,\theta_2), \ldots, F(x,\theta_k)$. Indeed, such a model has recently received attention in the work of John (1970(a),(b),(c)), who has considered questions of classification of the observations and estimation of the parameters in the cases where the underlying distribution $F(x,\theta)$ is normal, gamma, binomial, Poisson, negative binomial and hypergeometric. As yet, the uniform distribution does not seem to have received any examination under either model, and so in later chapters we shall attempt to develop the appropriate methods and illustrate the ideas using various subclasses of this important distribution.

Before going on to discuss the practical uses of mixed distributions and mixed samples, let us first note the relationship between these two models. The sample from a compound distribution can be regarded as a random sub-sample from a sample of size infinity in the other model, so that the numbers of individuals originating from either population are random variables, whereas they are to be regarded as parameters in the mixed samples model. Hence, one would expect that estimates obtained by the use of methods appropriate to each model will be reasonably numerically close. As the maximum likelihood method presents some difficulties in the case of a compound distribution model, estimates obtained from the mixed samples model should be useful first approximations in the initiation of iterative processes.

2.2 APPLICATIONS OF MIXTURES

There has recently been increased interest shown in the area of mixed distributions. This is mainly due to the increasing number of interesting problems concerning mixtures which are being encountered in certain applications. We shall now turn our attention to a few examples of the latter.

In such areas of biology as conservation, ecology and taxonomy, it is often wished to study certain characteristics of a particular species. For this purpose one often needs to take a sample of individuals from their natural population and make numerical measurements of these characteristics. Because age is frequently difficult to ascertain in samples from wild populations and because a quantitative characteristic may vary markedly with age, the biologist is therefore dealing with a mixture of distributions, the mixing being done over a parameter depending on the unobservable variate "age". When sampling is carried out on natural populations of fish, for example, such problems have been considered by, among others, Cassie (1954), Buchanan-Wollaston and Hodgeson (1929), Oka (1954), Tanaka (1962) and Bhattacharya (1967).

A number of examples of mixtures which arise in the chemical industry have been given in the book by Medgyessy (1961). The applications given by this author include the use of finite mixtures of normal distributions in the investigation of absorption spectra and of electrophoretic separation of proteins.

Applications of mixtures to general industrial situations have been given by Hald (1952), who suggests that a mixture of normal distributions might be due to the use of different machines, changing attendants (e.g. day and night shift-workers), the properties of the raw material or to such time factors as wear.

Mixtures of distributions are also encountered in the area of life-testing and in acceptance-testing. Using a mixture of Weibull distributions, Kao (1959) discusses a life-testing problem involving electron tubes which are subject to a sudden or delayed failure. In the life-testing of radio equipment a similar problem is discussed by Mendenhall and

Hader (1958) and by Cox (1959). The problem of constructing acceptance-testing procedures when the true proportion of defectives varies from batch to batch is considered by Barnard (1954) and Vagholkar (1959).

Some further interesting examples of mixtures are given by Anscombe (1961) and Blischke (1962). The former author gives an application to a marketing problem, and the latter to a description of an ICBM weapons system. Also, for the use of mixtures in the description of accident proneness and other models of "contagion", the reader is referred to the review article by Blischke (1963).

It is hoped that the above examples will serve to convince the reader of the wide practical uses and general importance of mixtures. Before taking up the problem of estimation in mixtures of distributions, we conclude this section by noting that, where a mixed distributions model has been used exclusively in the past to analyse the above examples, the mixed samples model discussed in §2.1 may have been both more appropriate and more useful. In particular, the maximum likelihood method is often comparatively easy to apply under a mixed samples model (see John (1970(a))).

2.3 ESTIMATION IN MIXTURES OF DISTRIBUTIONS

The difficulty of constructing useful estimators for the parameters of mixtures of distributions has plagued research in mixture theory for many years. Many estimation techniques yield only an intractable set of equations. Though these can often be solved using computer techniques, the estimation problem is certainly not yet near a fully satisfactory degree of solution, especially for small sample sizes.

For the purpose of illustrating the above general comments we shall now concern ourselves with outlining and discussing some of the many attempts which have been made to estimate the parameters in mixtures of normal distributions. The amount of effort which has been expended on mixtures of normals is indicative of the complications generally encountered in the mixture problem. For the details of estimation in mixtures of non-normal distributions, particularly discrete distributions, the reader is referred to Blischke (1963).

Here we shall be interested only in finite mixtures; usually, in fact, with mixtures of two normal distributions. The density function in the latter case is given by

$$p(x) = (2\pi)^{-\frac{1}{2}}\left\{ \frac{\alpha}{\sigma_1} \exp\left[-\tfrac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] + \frac{(1-\alpha)}{\sigma_2}\exp\left[-\tfrac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right\} \quad (3.1)$$

where $0 < \alpha < 1$. Moment estimators for the five parameters α, μ_1, μ_2, σ_1 and σ_2 were constructed by Pearson (1894). This was the first serious attempt at dealing with a mixture problem. Pearson was motivated by the thought that it would be useful if frequency curves (based on large numbers of observations) which deviated from the normal frequency curves could be dissected into two or more normal components. Of course, as Pearson pointed out, the method of moments is not suitable if the number of components is greater than two. This is because the use of higher and higher moments will always result in both increased errors and increased analytical difficulties. However, the case of two normal components is of special interest in the context of the theory of evolution as considered by Pearson; for "a family probably breaks up first into two species, rather than three or more, owing to the pressure at a given time of some particular form of natural selection."

Pearson was well aware that, in a practical situation, his method suffered from the possibility of multiple solutions. These arise firstly because one must solve a nonic which may not yield a unique root and, secondly, because other models such as the limit of a binomial distribution may also provide a reasonable fit to the data. Pearson suggested, with considerable reservations, a theoretical criterion for the resolution of this problem whereby one chooses that solution which gives to the compound curve a sixth moment nearest in value to that of the frequency curve.

In the days before computers, the arithmetical difficulties posed by the calculation of moments upto the sixth order, coupled with the difficulty of extracting the roots of a nonic were considered most unfortunate. Indeed, Charlier (1906) comments: "The solution of an equation of the ninth degree, where almost all powers to the ninth, of the unknown quantity are existing is, however, a very laborious operation. Mr Pearson has indeed possessed the energy to perform this heroic task in some instances.... But I fear that he will have very few successors, if the dissection of the frequency curve into two components is not very urgent." Charlier recommended a modified method of elimination which leads to two equations to be solved simultaneously for two unknowns. These equations involve one polynomial of the third degree in one unknown and the quotient of two other similar polynomials. The simplification is only marginal, however, and Charlier remarks: "there is enough labour to discourage an inquirer from operating a mathematical dissection of a given frequency curve."

Possibly the simplest method in the absence of a computer program is that given by Strömgren (1934) using the "half-invariants" (certain functions of the moments) computed by Burrau (1934). The method is not straightforward, as it is based on a fairly lengthy process involving graphical and tabular aids, special procedures for refining the accuracy of the solutions, as well as a considerable amount of trial and error. Graphical aids for some special cases have been given by Preston (1953), using a procedure based on the relationship between skewness and kurtosis, and also by Sittig (1948) and Weichselberger (1951).

The moment equations for a mixture of three normals (under certain simplifying assumptions which result in only four unknown parameters) has been given by Pollard (1934). For a mixture of two normals with equal variances, Rao (1952) has displayed an iterative solution to the moment equations using Newton's method. Even in this case, the calculations are still quite laborious.

Charlier and Wicksell (1924) succeeded in considerably simplifying Pearson's results for a mixture of two normal distributions, and Cohen (1967) presents a further simplification whereby one follows an iterative procedure which circumvents a direct solution of Pearson's nonic and requires only the solution of a cubic for a unique negative root. Cohen also discusses a method using the first four moment equations followed by either a conditional maximum likelihood or a conditional minimum chi-square method of estimation. This eliminates use of the fifth moment with its large variance and is suggested as a compromise between reliable estimating procedures and tractability of the estimating equations.

With the advent of electronic computers and the modern theory of estimation it has become possible to apply techniques involving iteration to the maximum likelihood equations, for example, instead of the moment equations. This has been done by Rao (1948) when dealing with the special case $\sigma_1 = \sigma_2$. It has also been carried out more recently by Hasselblad (1966) for a finite mixture of $k \geq 2$ normal components with arbitrary means and variances. Both a method of steepest descents and Newton's method are discussed, the latter converging very rapidly provided one is sufficiently close to the roots and the former providing quite a reliable though slow convergence even from less accurate initial approximations. The author relies on a technique for estimating the parameters of a singly truncated sample as given by Hald (1949) to provide the necessary initial approximations to the roots. In the discussion following Hasselblad's paper, Cohen points out among other things, that even for k=2 sample sizes of 1000 are desirable, while 400 or 500 would probably be the minimum possible if any degree of accuracy in the estimation is to be achieved. This can be readily seen from the formulae given by Hasselblad for the approximation of the variance-covariance matrix of the estimators.

Under a mixed samples model, John (1970(a)) discusses estimation in a mixture of two normal populations with equal variances. He shows that the maximum likelihood estimates of the parameters are considerably easier to obtain under this model than under a mixed distributions model. Unfortunately, their sampling distribution presents some difficulties which have not yet been resolved.

Among the more unusual approaches to estimation in the finite mixture problem are the graphical procedures described by Harding (1949) and Cassie (1954), using a plot of the empirical distribution function on probability paper. These methods involve estimation schemes based on the fact that the tails of a mixed distribution are sometimes essentially the tails of a single distribution. Estimation of the parameters of the components is carried out by extrapolation back from the respective tails of the mixture. Thus Cassie estimates each subpopulation in order from the left and then subtracts it from the cumulative distribution. Since accurate estimates of the tail probabilities are required, these methods are almost certainly quite inefficient, even in comparison with the moment procedure.

An extraordinary approach to the problem of identifying the components of a finite mixture of normal distribution functions is due to Medgyessy (1961). He uses a technique called by him the "variance reduction method". As noted in the review by Mallows (1962), the basic construction is entirely analytic, however, being completely devoid of any statistical considerations. The estimators resulting from this method have not been investigated for their statistical properties, and so it is not known whether they are at all efficient.

It should now be apparent that there are many problems encountered in estimating the parameters of mixtures which have yet to be resolved in a completely satisfactory manner. Problems of hypothesis testing and identifiability of mixtures are further areas which are in need of thorough investigation. In subsequent chapters more specialised questions will arise as we develop methods for estimation in mixtures of uniform distributions.

CHAPTER THREE

ESTIMATION IN A MIXTURE OF ONE-PARAMETER UNIFORM DISTRIBUTIONS

3.1 THE METHOD OF MOMENTS UNDER A MIXED SAMPLES MODEL

Consider a set of N observations $\{x_1, x_2, \ldots, x_N\}$, an unknown number N_1 of which are observations on the uniform distribution $U[0,\theta_1]$ on the interval $[0,\theta_1]$, and N_2 are observations on $U[0,\theta_2]$, where $N_1 + N_2 = N$. From this set of N observations we wish both to estimate the unknown parameters θ_1, θ_2, N_1, N_2 and to attempt a classification of the observations according to that uniform distribution from which they are taken. Let us assume that θ_1, θ_2 and all the observations are positive. Without loss of generality it can be further assumed that $\theta_1 < \theta_2$.

To apply the method of moments we require estimates $\tilde{N}_1$, $\tilde{N}_2$ of N_1 and N_2 which satisfy

$$\tilde{N}_1 + \tilde{N}_2 = N, \tag{1.1}$$

as well as the following three equations wherein $\tilde{\theta}_1$, $\tilde{\theta}_2$ denote estimates of θ_1 and θ_2:

$$\begin{aligned} A &= \sum_{r=1}^{N} x_r = \frac{1}{2}\tilde{N}_1\tilde{\theta}_1 + \frac{1}{2}\tilde{N}_2\tilde{\theta}_2 \\ B &= \sum_{r=1}^{N} x_r^2 = \frac{1}{3}\tilde{N}_1\tilde{\theta}_1^2 + \frac{1}{3}\tilde{N}_2\tilde{\theta}_2^2 \\ C &= \sum_{r=1}^{N} x_r^3 = \frac{1}{4}\tilde{N}_1\tilde{\theta}_1^3 + \frac{1}{4}\tilde{N}_2\tilde{\theta}_2^3 \end{aligned} \tag{1.2}$$

These equations are obtained by equating sample moments about the origin to their expected values and adding tildes to indicate the estimates so obtained. We have chosen to use the notation A,B,C for the first three moments rather than the more usual μ_1', μ_2', μ_3' in order to simplify the representation of the algebraic manipulations which follow. Equations (1.1) and (1.2) are to be regarded as a system of four equations to be solved simultaneously for the four unknowns $\tilde{\theta}_1$, $\tilde{\theta}_2$, $\tilde{N}_1$ and $\tilde{N}_2$.

Multiplying the first equation of (1.2) by $(\tilde{\theta}_1 + \tilde{\theta}_2)$, we obtain

$$2A(\tilde{\theta}_1+\tilde{\theta}_2) = \tilde{N}_1\tilde{\theta}_1^2 + \tilde{N}_2\tilde{\theta}_2^2 + \tilde{N}\tilde{\theta}_1\tilde{\theta}_2$$

and using the second equation of (1.2), this becomes

$$2A(\tilde{\theta}_1+\tilde{\theta}_2) = 3B + N\tilde{\theta}_1\tilde{\theta}_2. \tag{1.3}$$

Now multiply the second equation of (1.2) by $(\tilde{\theta}_1+\tilde{\theta}_2)$ so that

$$3B(\tilde{\theta}_1+\tilde{\theta}_2) = N_1\tilde{\theta}_1^3 + N_2\tilde{\theta}_2^3 + \tilde{\theta}_1\tilde{\theta}_2(\tilde{N}_1\tilde{\theta}_1+\tilde{N}_2\tilde{\theta}_2)$$

and we use the third equation of (1.2) to obtain

$$3B(\tilde{\theta}_1+\tilde{\theta}_2) = 4C + 2A\tilde{\theta}_1\tilde{\theta}_2. \tag{1.4}$$

Eliminating $(\tilde{\theta}_1+\tilde{\theta}_2)$ from equation (1.3) and (1.4), it follows that

$$3B(3B+N\tilde{\theta}_1\tilde{\theta}_2) = 2A(4C + 2A\tilde{\theta}_1\tilde{\theta}_2)$$

$$\text{i.e.}\quad \tilde{\theta}_1\tilde{\theta}_2(3NB-4A^2) = 8AC - 9B^2$$

$$\text{i.e.}\quad \tilde{\theta}_1\tilde{\theta}_2 = \frac{8AC - 9B^2}{3NB - 4A^2} \tag{1.5}$$

Note that $3NB - 4A^2 \neq 0$ with probability one.

Substituting the expression for $\tilde{\theta}_1\tilde{\theta}_2$ given by (1.5) into (1.3),

$$\tilde{\theta}_1+\tilde{\theta}_2 = \frac{1}{2A}\left\{3B + \frac{N(8AC-9B^2)}{3NB - 4A^2}\right\}$$

$$= \frac{4NC - 6AB}{3NB - 4A^2} \tag{1.6}$$

Now $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are solutions of the quadratic equation

$$y^2 - (\tilde{\theta}_1+\tilde{\theta}_2)y + \tilde{\theta}_1\tilde{\theta}_2 = 0$$

i.e. $$(3NB-4A^2)y^2 + (6AB-4NC)y + 8AC - 9B^2 = 0 \qquad (1.7)$$

If the method of moments is to succeed in the production of reasonable estimates for θ_1 and θ_2, it will be necessary for equation (1.7) to have roots which are both real and positive. In order to determine the conditions which must be satisfied by the observations for this to be so, the following relationships among the sample moments will be employed

$$NB = N^2m_2 + A^2$$
$$C = \frac{1}{N^2}(N^3m_3 + 3AN^2m_2 + A^3) \qquad (1.8)$$

where m_2 and m_3 are the second and third central sample moments respectively, i.e. $m_2 = \frac{1}{N}\sum_{r=1}^{N}(x_r-\bar{x})^2$ and $m_3 = \frac{1}{N}\sum_{r=1}^{N}(x_r-\bar{x})^3$ where $\bar{x}$ is the sample mean, $\bar{x} = \frac{1}{N}\sum_{r=1}^{N}x_r$.

Using the relationships (1.8) to rewrite each of the coefficients of our quadratic (1.7), we obtain

$$3NB - 4A^2 = 3N^2m_2 - A^2 \qquad (1.9)$$

$$\begin{aligned} 6AB-4NC &= \frac{6A}{N}(N^2m_2+A^2) - \frac{4}{N}(N^3m_3+3AN^2m_2+A^3) \\ &= 6ANm_2 + \frac{6A^3}{N} - 4N^2m_3 - 12ANm_2 - \frac{4A^3}{N} \\ &= \frac{2A^3}{N} - 4N^2m_3 - 6ANm_2 \\ &= -\frac{2A}{N}(3N^2m_2-A^2) - 4N^2m_3 \end{aligned} \qquad (1.10)$$

$$\begin{aligned} 8AC-9B^2 &= \frac{8A}{N^2}(N^3m_3+3AN^2m_2+A^3) - \frac{9}{N^2}(N^2m_2 + A^2)^2 \\ &= 8ANm_3+24A^2m_2 + \frac{8A^4}{N^2} - 9N^2m_2^2-18A^2m_2 - \frac{9A^4}{N^2} \\ &= 8ANm_3+6A^2m_2 - 9N^2m_2^2 - \frac{A^4}{N^2} \end{aligned}$$

$$= 8ANm_3 - \frac{1}{N^2}(3N^2m_2-A^2)^2. \qquad (1.11)$$

Now if we use the expressions (1.9), (1.10) and (1.11) for the coefficients, the discriminant of the quadratic (1.7) can be written as

$$\Delta = \{4N^2m_3+\frac{2A}{N}(3N^2m_2-A^2)\}^2-4(3N^2m_2-A^2)\{8ANm_3-\frac{1}{N^2}(3N^2m_2-A^2)^2\}$$

$$= 16N^4m_3^2 + 16ANm_3(3N^2m_2-A^2)+\frac{4A^2}{N^2}(3N^2m_2-A^2)^2$$

$$-32ANm_3(3N^2m_2-A^2) + \frac{4}{N^2}(3N^2m_2-A^2)^3$$

$$= 16N^4m_3^2-16ANm_3(3N^2m_2-A^2) + \frac{4}{N^2}(3N^2m_2-A^2)^2(A^2+3N^2m_2-A^2)$$

$$= 16N^4m_3^2-16ANm_3(3N^2m_2-A^2) + 12m_2(3N^2m_2-A^2)^2. \qquad (1.12)$$

To determine the appropriate relationship among the moments for the discriminant to be positive, we note that (1.12) can be regarded as a quadratic in m_3, not forgetting that as m_3 varies over different samples, so A and m_2 will also vary. Hence we can differentiate (1.12) with respect to m_3 for the purpose of finding the theoretical minimum value of (1.12). This minimum is reached when

$$32N^4m_3 - 16AN(3N^2m_2 - A^2) = 0$$

i.e. $$m_3 = \frac{A}{2N^3}(3N^2m_2 - A^2).$$

Hence the theoretical minimum value is given by

$$\frac{16N^4A^2}{4N^6}(3N^2m_2-A^2)^2 - \frac{16NA^2}{2N^3}(3N^2m_2-A^2)^2 + 12m_2(3N^2m_2-A^2)^2$$

$$= (3N^2m_2-A^2)\left(\frac{4A^2}{N^2} - \frac{8A^2}{N^2} + 12m_2\right)$$

$$= (3N^2m_2-A^2)4(3m_2 - \frac{A^2}{N^2})$$

$$= \frac{4}{N^2}(3N^2m_2-A^2)^3. \qquad (1.13)$$

Hence, if $3N^2m_2-A^2 > 0$, then $\Delta > 0$ and the roots are real.

Now consider the case $3N^2m_2 - A^2 < 0$. If it is also true that $m_3 > 0$, then Δ as given by (1.12) is positive since each of the three terms is positive (note that $A > 0$ follows from the fact that all the observations are assumed positive). On the other hand, if $m_3 < 0$ we examine the roots of (1.12). These occur at

$$\bar{m}_3, \underline{m}_3 = \frac{16AN(3N^2m_2-A^2) \pm \{256A^2N^2(3N^2m_2-A^2) - 3(256)N^4m_2(3N^2m_2-A^2)^2\}^{\frac{1}{2}}}{32N^4}$$

$$= \frac{1}{2N^3}\{A(3N^2m_2-A^2) \pm \left[-(3N^2m_2-A^2)^3\right]^{\frac{1}{2}}\}$$

$$= \frac{A^2-3N^2m_2}{2N^3}\{-A \pm (A^2-3N^2m_2)^{\frac{1}{2}}\}. \qquad (1.14)$$

To ensure that $\Delta > 0$ we require $m_3 > \bar{m}_3$ or $m_3 < \underline{m}_3$. Since $A > 0$ and $A^2 > A^2 - 3N^2m_2 > 0$ it is easily seen that both $\bar{m}_3$ and $\underline{m}_3$ are negative real numbers. Thus we can summarise by noting that (1.7) will have real roots if and only if

$$3N^2m_2 - A^2 > 0$$
$$\text{or} \quad 3N^2m_2 - A^2 < 0 \text{ and either } m_3 > \bar{m}_3 \text{ or } m_3 < \underline{m}_3 \qquad (1.15)$$

However, as we previously observed, it is not sufficient that the roots should be real - we also wish to determine the conditions under which they are both positive. To this end we shall adopt in turn each of the conditions (1.15) for the reality of the roots.

Case (1) : $3N^2m_2-A^2 > 0$

Since the first coefficient of (1.7) is positive in this case and since the roots are real, an application of Descartes' rule of signs shows that the roots are both positive if and only if the number of sign changes in the coefficients is two. Hence, reinvoking the expressions (1.9),(1.10) and (1.11) for the coefficients of our quadratic (1.7), it is required that

$$-\frac{2A}{N}(3N^2m_2-A^2) - 4N^2m_3 < 0$$

$$\text{i.e.}\quad m_3 > \frac{-A}{2N^3}(3N^2m_2-A^2) \tag{1.16}$$

and that

$$8ANm_3 - \frac{1}{N^2}(3N^2m_2-A^2)^2 > 0$$

$$\text{i.e.}\quad m_3 > \frac{1}{8AN^3}(3N^2m_2-A^2)^2. \tag{1.17}$$

The inequality (1.17) is more restrictive than (1.16) and so, in this case, (1.17) represents the required extra condition on the sample to ensure positive roots. For large sample size one would expect the right hand side of (1.17) to be fairly small in view of the factor of N^3 in the denominator, and so, heuristically, it seems likely that (1.17) would be satisfied in a reasonable proportion of large samples. There is no need to rely on this rather weak argument however; for, when we consider the asymptotic distribution of the moment estimators it will be shown that, for large sample sizes, the estimators of θ_1 and θ_2 obtained from the method of moments are both positive with probability close to unity.

<u>Case (2) : $3N^2m_2-A^2 < 0$ and either $m_3 > \bar{m}_3$ or $m_3 < \underline{m}_3$</u>

Now the first coefficient of (1.7) is negative and so, to apply Descartes' rule of signs it is necessary to multiply (1.7) by -1 so that

$$-(3N^2m_2-A^2)y^2+\{\frac{2A}{N}(3N^2m_2-A^2)+4m_3N^2\}y+\frac{1}{N^2}(3N^2m_2-A^2)^2-8ANm_3=0 \tag{1.18}$$

where we have again used (1.9),(1.10) and (1.11) for the coefficients. Hence, to ensure positive roots for equation (1.18) we require that

$$\frac{2A}{N}(3N^2m_2-A^2) + 4m_3N^2 < 0$$

i.e. $$m_3 < \frac{-A}{2N^3}(3N^2m_2-A^2) \tag{1.19}$$

and also $$\frac{1}{N^2}(3N^2m_2-A^2)^2 - 8ANm_3 > 0$$

i.e. $$m_3 < \frac{1}{8AN^3}(3N^2m_2-A^2)^2. \tag{1.20}$$

Both bounds (1.19) and (1.20) are positive, so to find the smallest we consider the difference

$$\begin{aligned}
&\frac{-A}{2N^3}(3N^2m_2-A^2) - \frac{1}{8AN^3}(3N^2m_2-A^2)^2 \\
&= \frac{-1}{8AN^3}\{(3N^2m_2-A^2)^2 + 4A^2(3N^2m_2-A^2)\} \\
&= \frac{-1}{8AN^3}\{9N^4m_2^2 - 6A^2N^2m_2 + A^4+12A^2N^2m_2 - 4A^4\} \\
&= \frac{-3}{8AN^3}(N^2m_2+A^2)(3N^2m_2-A^2) \\
&> 0 \quad \text{since } 3N^2m_2-A^2 < 0.
\end{aligned}$$

Hence the required bound in this case is given by the inequality (1.20). We can now summarise our results according to the following rule.

There will be real, positive roots for the quadratic (1.7) and hence real, positive estimates for θ_1 and θ_2 using the method of moments if and only if

$$\left.\begin{array}{ll}
 & 3N^2m_2-A^2 > 0 \text{ and } m_3 > \frac{1}{8AN^3}(3N^2m_2-A^2)^2 \\
\text{or} & 3N^2m_2-A^2 < 0 \text{ and either } \bar{m}_3 < m_3 < \frac{1}{8AN^3}(3N^2m_2-A^2)^2 \\
 & \qquad\qquad\qquad \text{or} \quad m_3 < \underline{m}_3
\end{array}\right\} \tag{1.21}$$

In any particular sample of size N under consideration the method of moments will succeed if and only if the conditions (1.21) hold. In §3.2 it will be shown that one can always expect real, positive solutions asymptotically and consequently, if (1.21) is not satisfied in a particular circumstance, then it is advisable to take further observations (if this is at all possible) so that the method will be sure to eventually succeed. If the method of moments fails on a sample which cannot be further supplemented, then of course one may still use the method of maximum likelihood to be discussed in the following section.

Assuming that the method of moments is successful inasmuch as equation (1.7) can be solved to yield two positive estimates $\tilde{\theta}_1$ and $\tilde{\theta}_2$ of θ_1 and θ_2 with $\tilde{\theta}_1 < \tilde{\theta}_2$, one then refers to equation (1.1) and the first equation of (1.2) to write

$$2A = \tilde{N}_1\tilde{\theta}_1 + (N-\tilde{N}_1)\tilde{\theta}_2$$

$$\text{i.e.} \quad \tilde{N}_1 = \frac{2A - N\tilde{\theta}_2}{\tilde{\theta}_1 - \tilde{\theta}_2} \quad . \tag{1.22}$$

Because the estimate for N_1 given by (1.22) is not generally integral, we choose those integers $\tilde{N}_1'$ and $\tilde{N}_2'$ which are between zero and N and closest to $\dfrac{2A-N\tilde{\theta}_2}{\tilde{\theta}_1-\tilde{\theta}_2}$ and $\dfrac{N\tilde{\theta}_1-2A}{\tilde{\theta}_1-\tilde{\theta}_2}$ respectively.

It is also of importance to classify the observations and to this end a reasonable procedure would be to allocate the first $\tilde{N}_1'$ smallest observations to the first population $U[0,\tilde{\theta}_1]$ and the remaining to the second population $U[\theta,\tilde{\theta}_2]$.

3.2 THE METHOD OF MAXIMUM LIKELIHOOD UNDER A MIXED SAMPLES MODEL

Consider a set of N observations $\{x_1,x_2,...,x_N\}$ which consists of an unknown number of N_1 observations on the first population $U[0,\theta_1]$ and N_2 observations on the second population $U[0,\theta_2]$, where $N_1+N_2 = N$. As in the method of moments, it can be assumed without loss of generality that $\theta_1 < \theta_2$. Let us also assume that θ_1,θ_2 and all the observations are positive. We wish to determine maximum likelihood estimates of the unknown parameters θ_1,θ_2, N_1 and N_2 based on the set of N observations, and further to attempt a classification of each observation according to the population from which it was taken.

Let ω_{ij} take the value one if the jth observation belongs to the ith population and zero otherwise for each i=1,2, and j=1,2,...,N. Then the likelihood of the sample is given by

$$L(x_1,...,x_N;\theta_1,\theta_2,N_1,N_2) = \frac{\prod_{j=1}^{N}\{u(\omega_{1j}x_j-\theta_1)u(\omega_{2j}x_j-\theta_2)\}}{\theta_1^{N_1}\ \theta_2^{N_2}} \qquad (2.1)$$

where u(.) is that function which is equal to one when the argument is nonpositive and zero otherwise.

Given θ_1,θ_2, N_1 and N_2, L is maximised by allocating an observation to the first population if it lies in the interval $[0,\theta_1]$ and to the second population if it lies in $(\theta_1,\theta_2]$, i.e. the maximum likelihood allocation in this case is given by

$$\hat{\omega}_{1j} = u(x_j-\theta_1)$$

$$\hat{\omega}_{2j} = 1 - \hat{\omega}_{1j} \qquad j=1,2,...,n. \qquad (2.2)$$

On the other hand, if the allocation ω_{ij} is given for all i and j the maximum likelihood estimates of θ_1 and θ_2

are given respectively by the largest observation in the first population and the largest observation in the second population,

$$\text{i.e.} \quad \hat{\theta}_1 = \max_{j=1,2,\ldots,N} \{x_j \omega_{1j}\}$$

$$\text{and} \quad \hat{\theta}_2 = \max_{j=1,2,\ldots,N} \{x_j \omega_{2j}\} . \tag{2.3}$$

Hence, it can be easily seen that to find the unconditional maximum likelihood estimates, one need only compare the N values of L generated in the following manner. Let $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, be the ordered set of observations and define $x_{(N+1)} = +\infty$. Let λ_c be the maximum value of L when the allocation is carried out according to $\omega_{1j} = u(x_j - c)$ and $\omega_{2j} = 1 - \omega_{1j}$. Then λ_c can take at most N distinct, non-zero values obtained by letting c take on a value between the jth and (j+1)th ordered observation, $x_{(j)}$ and $x_{(j+1)}$, for each j=1,2,...,N and then maximising L under this allocation by choosing estimates of θ_1 and θ_2 according to (2.3).

Let c* be that value of ĉ which yields the largest such "maximum" value of L and let j* be that integer in the set {1,2,...,N} which satisfies $j^* < \hat{c} < j^*+1$. Then the unconditional m.ℓ.e's of θ_1 and θ_2 are $\hat{\hat{\theta}}_1 = x_{(j^*)}$ and $\hat{\hat{\theta}}_2 = x_{(N)}$, and the m.ℓ. allocation puts $x_{(1)}, x_{(2)}, \ldots, x_{(j^*)}$ into the first population and $x_{(j^*+1)}, x_{(j^*+2)}, \ldots, x_{(N)}$ into the second population.

Note that after the observations have been ordered it is simply a matter of finding that integer j* in the set {1,2,...,N} which maximises

$$\frac{1}{x_{(j)}^{j} \, x_{(N)}^{N-j}} \tag{2.4}$$

and then allocating and estimating as above. Equivalently, one finds that integer j^* which minimises

$$j \log x_{(j)} + (N-j) \log x_{(N)}. \tag{2.5}$$

As the derivation of the distribution of these maximum likelihood estimators presents considerable difficulty, we do not attempt it in this thesis. A possible, though not very successful approach to this problem has been given by John (1970(a), p561). The distribution of the moment estimators is more straightforward and so we will move on to consider this in the following section.

3.3 THE ASYMPTOTIC DISTRIBUTION OF THE MOMENTS ESTIMATORS

The moment estimators are functions of the three sample sums $\sum_{r=1}^{N} x_r$, $\sum_{r=1}^{N} x_r^2$ and $\sum_{r=1}^{N} x_r^3$. Equivalently, they can be regarded as functions of $Z_t = \frac{1}{N^{\frac{1}{2}}} \left(\sum_{r=1}^{N} x_r^t - E \sum_{r=1}^{N} x_r^t \right)$, $t=1,2,3$.

If N is large these functions are approximately equal to their respective Taylor expansions in powers of Z_1, Z_2 and Z_3 with terms of order greater than one omitted. An application of the multivariate central limit theorem shows that asymptotically Z_1, Z_2 and Z_3 are jointly normally distributed. Thus the moment estimators are asymptotically distributed as linear functions of jointly normally distributed variables. It follows that the asymptotic distribution of the moment estimators is normal. This argument is well known and can be found in Cramér (1946, p218 and pp364-367).

Now $E\left(\sum_{r=1}^{N} x_r^t \right) = \frac{N_1 \theta_1^t}{t+1} + \frac{N_2 \theta_2^t}{t+1}$ for $t=1,2,3$ and so we can write

$$A = \sum_{r=1}^{N} x_r = N^{\frac{1}{2}}Z_1 + \frac{N_1\theta_1}{2} + \frac{N_2\theta_2}{2}$$

$$B = \sum_{r=1}^{N} x_r^2 = N^{\frac{1}{2}}Z_2 + \frac{N_1\theta_1^2}{3} + \frac{N_2\theta_2^2}{3} \qquad (3.1)$$

$$C = \sum_{r=1}^{N} x_r^3 = N^{\frac{1}{2}}Z_3 + \frac{N_1\theta_1^3}{4} + \frac{N_2\theta_2^3}{4}$$

Solving the quadratic (1.7), our estimates for θ_1 and θ_2 are given by

$$\tilde{\theta}_2, \tilde{\theta}_1 = (3NB-4A^2)^{-1}\{2NC-3AB \pm [(2NC-3AB)^2+(9B^2-8AC)(3NB-4A^2)]^{\frac{1}{2}}\} \qquad (3.2)$$

We now make use of the relations (3.1) to find the Taylor expansion of each of the terms in the right hand side of (3.2), neglecting terms of order greater than one. Firstly, we have

$$3NB-4A^2 = 3N^{3/2}Z_2 + (N_1\theta_1^2+N_2\theta_2^2)N - 4(N^{\frac{1}{2}}Z_1 + \frac{N_1\theta_1}{2} + \frac{N_2\theta_2}{2})^2$$

$$\simeq 3N^{3/2}Z_2 + N_1N_2(\theta_1^2+\theta_2^2) - 4N^{\frac{1}{2}}Z_1(N_1\theta_1+N_2\theta_2) - 2N_1N_2\theta_1\theta_2$$

$$= N_1N_2(\theta_1-\theta_2)^2 + 3N^{3/2}Z_2 - 4N^{\frac{1}{2}}Z_1(N_1\theta_1+N_2\theta_2). \qquad (3.3)$$

Hence $(3NB-4A^2)^{-1} \simeq [N_1N_2(\theta_1-\theta_2)^2]^{-1}\left\{1 - \dfrac{3N^{3/2}Z_2 - 4N^{\frac{1}{2}}Z_1(N_1\theta_1+N_2\theta_2)}{N_1N_2(\theta_1-\theta_2)^2}\right\}$

$$(3.4)$$

Also,

$$2NC-3AB = 2N^{3/2}Z_3 + N(\frac{N_1\theta_1^3}{2} + \frac{N_2\theta_2^3}{2}) - (N^{\frac{1}{2}}Z_1 + \frac{N_1\theta_1}{2} + \frac{N_2\theta_2}{2})(3N^{\frac{1}{2}}Z_2 + N_1\theta_1^2 + N_2\theta_2^2)$$

$$\simeq 2N^{3/2}Z_3 + N(\frac{N_1\theta_1^3}{2} + \frac{N_2\theta_2^3}{2}) - N^{\frac{1}{2}}Z_1(N_1\theta_1^2+N_2\theta_2^2)$$

$$-\frac{3}{2}N^{\frac{1}{2}}Z_2(N_1\theta_1+N_2\theta_2) - \tfrac{1}{2}(N_1\theta_1+N_2\theta_2)(N_1\theta_1^2+N_2\theta_2^2)$$

$$= 2N^{3/2}Z_3 - \frac{3}{2}N^{\frac{1}{2}}Z_2(N_1\theta_1+N_2\theta_2) - N^{\frac{1}{2}}Z_1(N_1\theta_1^2+N_2\theta_2^2)$$

$$+ \frac{N_1N_2}{2}(\theta_1-\theta_2)^2(\theta_1+\theta_2) \qquad (3.5)$$

Hence

$$(2NC-3AB)^2 \simeq N_1N_2(\theta_1-\theta_2)^2(\theta_1+\theta_2)\{2N^{\frac{3}{2}}Z_3 - \frac{3}{2}N^{\frac{1}{2}}Z_2(N_1\theta_1+N_2\theta_2)$$
$$-N^{\frac{1}{2}}Z_1(N_1\theta_1^2+N_2\theta_2^2)\}+\frac{N_1^2N_2^2}{4}(\theta_1-\theta_2)^4(\theta_1+\theta_2)^2 \quad (3.6)$$

Further,

$$9B^2-8AC = 9(N^{\frac{1}{2}}Z_2+\frac{N_1\theta_1^2}{3}+\frac{N_2\theta_2^2}{3})^2 - 8(N^{\frac{1}{2}}Z_1+\frac{N_1\theta_1}{2}+\frac{N_2\theta_2}{2})$$
$$\times(N^{\frac{1}{2}}Z_3+\frac{N_1\theta_1^3}{4}+\frac{N_2\theta_2^3}{4})$$
$$\simeq 6N^{\frac{1}{2}}Z_2(N_1\theta_1^2+N_2\theta_2^2) - 4N^{\frac{1}{2}}Z_3(N_1\theta_1+N_2\theta_2) - 2N^{\frac{1}{2}}Z_1(N_1\theta_1^3+N_2\theta_2^3)$$
$$+ (N_1\theta_1^2+N_2\theta_2^2)^2 - (N_1\theta_1+N_2\theta_2)(N_1\theta_1^3+N_2\theta_2^3)$$
$$= 6N^{\frac{1}{2}}Z_2(N_1\theta_1^2+N_2\theta_2^2)-4N^{\frac{1}{2}}Z_3(N_1\theta_1+N_2\theta_2)-2N^{\frac{1}{2}}Z_1(N_1\theta_1^3+N_2\theta_2^3)$$
$$-N_1N_2\theta_1\theta_2(\theta_1-\theta_2)^2 \quad (3.7)$$

Now combining (3.3) and (3.7) it follows that

$$(9B^2-8AC)(3NB-4A^2)$$
$$\simeq \{N_1N_2(\theta_1-\theta_2)^2+3N^{\frac{3}{2}}Z_2-4N^{\frac{1}{2}}Z_1(N_1\theta_1+N_2\theta_2)\}$$
$$\times \{6N^{\frac{1}{2}}Z_2(N_1\theta_1^2+N_2\theta_2^2)-4N^{\frac{1}{2}}Z_3(N_1\theta_1+N_2\theta_2)-2N^{\frac{1}{2}}Z_1(N_1\theta_1^3+N_2\theta_2^3)$$
$$-N_1N_2\theta_1\theta_2(\theta_1-\theta_2)^2\}$$
$$\simeq N_1N_2(\theta_1-\theta_2)^2\{6N^{\frac{1}{2}}Z_2(N_1\theta_1^2+N_2\theta_2^2)-4N^{\frac{1}{2}}Z_3(N_1\theta_1+N_2\theta_2)$$
$$-2N^{\frac{1}{2}}Z_1(N_1\theta_1^3+N_2\theta_2^3)\}-N_1N_2\theta_1\theta_2(\theta_1-\theta_2)^2$$
$$\times\{3N^{\frac{3}{2}}Z_2-4N^{\frac{1}{2}}Z_1(N_1\theta_1+N_2\theta_2)\}-N_1^2N_2^2\theta_1\theta_2(\theta_1-\theta_2)^4 \quad (3.8)$$

Adding this to expression (3.6) gives

$$(2NC-3AB)^2+(9B^2-8AC)(3NB-4A^2)$$
$$\simeq \frac{N_1^2N_2^2}{4}(\theta_1-\theta_2)^4(\theta_1+\theta_2)^2 - N_1^2N_2^2\theta_1\theta_2(\theta_1-\theta_2)^4$$

$$+N_1N_2(\theta_1-\theta_2)^2 \left\{ \begin{array}{l} 2N^{3/2}Z_3(\theta_1+\theta_2)-\frac{3}{2}N^{\frac{1}{2}}Z_2(N_1\theta_1+N_2\theta_2)(\theta_1+\theta_2) \\ -N^{\frac{1}{2}}Z_1(N_1\theta_1^2+N_2\theta_2^2)(\theta_1+\theta_2)+6N^{\frac{1}{2}}Z_2(N_1\theta_1^2+N_2\theta_2^2) \\ -4N^{\frac{1}{2}}Z_3(N_1\theta_1+N_2\theta_2)-2N^{\frac{1}{2}}Z_1(N_1\theta_1^3+N_2\theta_2^3) \\ -3N^{3/2}\theta_1\theta_2Z_2 + 4N^{\frac{1}{2}}Z_1\theta_1\theta_2(N_1\theta_1+N_2\theta_2) \end{array} \right\}$$

$$= \frac{N_1^2N_2^2}{4}(\theta_1-\theta_2)^6+N_1N_2(\theta_1-\theta_2)^2N^{\frac{1}{2}} \left\{ \begin{array}{l} Z_1(4N_1\theta_1^2\theta_2+4N_2\theta_1\theta_2^2-N_1\theta_1^3 \\ -N_1\theta_1^2\theta_2-N_2\theta_2^3-N_2\theta_1\theta_2^2-2N_1\theta_1^3-2N_2\theta_2^3) \\ +\frac{3}{2}Z_2(4N_1\theta_1^2+4N_2\theta_2^2-N_1\theta_1^2-N_1\theta_1\theta_2 \\ -N_2\theta_2^2-N_2\theta_1\theta_2-2N\theta_1\theta_2) \\ +2Z_3(N_1\theta_1+N_2\theta_2-2N_1\theta_1-2N_2\theta_2) \end{array} \right\}$$

$$= \frac{N_1^2N_2^2}{4}(\theta_1-\theta_2)^6+N_1N_2(\theta_1-\theta_2)^2N^{\frac{1}{2}} \left\{ \begin{array}{l} 3Z_1(\theta_2-\theta_1)(N_1\theta_1^2-N_2\theta_2^2) \\ +\frac{3}{2}Z_2(3N_1\theta_1^2+3N_2\theta_2^2-3N\theta_1\theta_2) \\ +2Z_3(N_2-N_1)(\theta_1-\theta_2) \end{array} \right\}$$

$$= \frac{N_1^2N_2^2}{4}(\theta_1-\theta_2)^6+N^{\frac{1}{2}}N_1N_2(\theta_1-\theta_2)^3 \left\{ \begin{array}{l} 3Z_1(N_2\theta_2^2-N_1\theta_1^2)+\frac{9}{2}Z_2(N_1\theta_1-N_2\theta_2) \\ -2Z_3(N_1-N_2) \end{array} \right\} \quad (3.9)$$

Hence $\{(2NC-3AB)^2+(9B^2-8AC)(3NB-4A^2)\}^{\frac{1}{2}}$

$$\simeq \frac{N_1N_2(\theta_2-\theta_1)^3}{2} \quad \{1+\frac{4N^{\frac{1}{2}}}{2(\theta_1-\theta_2)^3N_1N_2} \quad [3Z_1(N_2\theta_2^2-N_1\theta_1^2) + \frac{9}{2}Z_2(N_1\theta_1-N_2\theta_2)-2Z_3(N_1-N_2)]\}$$

Combining this with (3.5), we obtain

$$2NC-3AB+\{(2NC-3AB)^2+(9B^2-8AC)(3NB-4A^2)\}^{\frac{1}{2}}$$

$$\simeq 2N^{3/2}Z_3-\frac{3}{2}N^{\frac{1}{2}}Z_2(N_1\theta_1+N_2\theta_2)-N^{\frac{1}{2}}Z_1(N_1\theta_1^2+N_2\theta_2^2)+\frac{N_1N_2}{2}(\theta_1-\theta_2)^2(\theta_1+\theta_2)$$
$$+\frac{N_1N_2}{2}(\theta_2-\theta_1)^3-3N^{\frac{1}{2}}Z_1(N_2\theta_2^2-N_1\theta_1^2)-\frac{9}{2}N^{\frac{1}{2}}Z_2(N_1\theta_1-N_2\theta_2)+2N^{\frac{1}{2}}Z_3(N_1-N_2)$$

$$= N_1N_2\theta_2(\theta_2-\theta_1)^2-2N^{\frac{1}{2}}Z_1(2N_2\theta_2^2-N_1\theta_1^2)-3N^{\frac{1}{2}}Z_2(2N_1\theta_1-N_2\theta_2)+4N^{\frac{1}{2}}N_1Z_3 \quad (3.10)$$

Thus, from (3.2), (3.4) and (3.10) it follows that

$$\tilde{\theta}_2 \simeq \{N_1N_2(\theta_1-\theta_2)^2\}^{-2}\{N_1N_2(\theta_1-\theta_2)^2-3N^{3/2}Z_2+4N^{\frac{1}{2}}Z_1(N_1\theta_1+N_2\theta_2)\}$$
$$\times\{N_1N_2\theta_2(\theta_2-\theta_1)^2-2N^{\frac{1}{2}}Z_1(2N_2\theta_2^2-N_1\theta_1^2)-3N^{\frac{1}{2}}Z_2(2N_1\theta_1-N_2\theta_2)$$
$$+ 4N^{\frac{1}{2}}N_1Z_3\}$$

$$=\{N_1N_2(\theta_1-\theta_2)^2\}^{-2}\left\{\begin{array}{l}4N^{\frac{1}{2}}N_1^2N_2(\theta_1-\theta_2)^2Z_3+3N^{\frac{1}{2}}N_1N_2(\theta_1-\theta_2)^2 \\ \qquad\times Z_2(N_2\theta_2-2N_1\theta_1) \\ +2N^{\frac{1}{2}}N_1N_2(\theta_1-\theta_2)^2Z_1(N_1\theta_1^2-2N_2\theta_2^2) \\ +N_1^2N_2^2(\theta_1-\theta_2)^4\theta_2-3N^{3/2}N_1N_2(\theta_1-\theta_2)^2\theta_2Z_2 \\ +4N^{\frac{1}{2}}N_1N_2(\theta_1-\theta_2)^2\theta_2Z_1(N_1\theta_1+N_2\theta_2)\end{array}\right\}$$

$$=\{N_1N_2(\theta_1-\theta_2)^2\}^{-1}N^{\frac{1}{2}}\left\{\begin{array}{l}4N_1Z_3+3Z_2(N_2\theta_2-2N_1\theta_1) \\ +\ 2Z_1(N_1\theta_1^2-2N_2\theta_2^2 + \dfrac{N_1N_2}{N^{\frac{1}{2}}}(\theta_1-\theta_2)^2\theta_2 \\ -3N\theta_2Z_2+4\theta_2Z_1(N_1\theta_1+N_2\theta_2)\end{array}\right\}$$

$$= \theta_2 + \frac{N^{\frac{1}{2}}}{N_2(\theta_1-\theta_2)^2}\{2\theta_1Z_1(\theta_1+2\theta_2)-3Z_2(2\theta_1+\theta_2)+4Z_3\} \quad (3.11)$$

Hence the coefficients of $\tilde{\theta}_2$ with respect to Z_1, Z_2 and Z_3 are given by

$$\alpha_{21} = \frac{2N^{\frac{1}{2}}\theta_1(\theta_1+2\theta_2)}{N_2(\theta_1-\theta_2)^2}, \quad \alpha_{22} = \frac{-3N^{\frac{1}{2}}(2\theta_1+\theta_2)}{N_2(\theta_1-\theta_2)^2}$$

and $$\alpha_{23} = \frac{4N^{\frac{1}{2}}}{N_2(\theta_1-\theta_2)^2} . \quad (3.12)$$

By an argument of symmetry or by direct verification, the coefficients of $\tilde{\theta}_1$ with respect to Z_1, Z_2 and Z_3 are

$$\alpha_{11} = \frac{2N^{\frac{1}{2}}\theta_2(\theta_2+2\theta_1)}{N_1(\theta_1-\theta_2)^2}, \quad \alpha_{12} = \frac{-3N^{\frac{1}{2}}(\theta_1+2\theta_2)}{N_1(\theta_1-\theta_2)^2} \quad \text{and}$$

$$\alpha_{13} = \frac{4N^{\frac{1}{2}}}{N_1(\theta_1-\theta_2)^2}. \tag{3.13}$$

In §3.1 it was shown that the moment estimator $\tilde{N}_1'$ of N_1 should be taken as that integer between 0 and N which is closest to $\tilde{N}_1 = \dfrac{2A-N\tilde{\theta}_2}{\tilde{\theta}_1-\tilde{\theta}_2}$. We shall derive the expansion for $\tilde{N}_1$ and assume that its distribution is asymptotically a good approximation to that of the discrete random variable $\tilde{N}_1'$.

$$\text{By (3.1)}, \quad \frac{2A-N\tilde{\theta}_2}{\tilde{\theta}_1-\tilde{\theta}_2} = \frac{2N^{\frac{1}{2}}Z_1+N_1\theta_1+N_2\theta_2-N\tilde{\theta}_2}{\tilde{\theta}_1-\tilde{\theta}_2} \tag{3.14}$$

$$\text{Now } \tilde{\theta}_1-\tilde{\theta}_2 \simeq \theta_1+\theta_2+\frac{N^{\frac{1}{2}}}{N_1N_2(\theta_1-\theta_2)^2}\left\{\begin{array}{l}2N_2\theta_2(\theta_2+2\theta_1)Z_1-3N_2(\theta_1+2\theta_2)Z_2 \\ +4N_2Z_3-2N_1\theta_1(\theta_1+2\theta_2)Z_1+3N_1(2\theta_1+\theta_2)Z_2 \\ -4N_1Z_3\end{array}\right\}$$

$$=\theta_1-\theta_2+\frac{N^{\frac{1}{2}}}{N_1N_2(\theta_1-\theta_2)^2}\left\{\begin{array}{l}2Z_1\left[N_2\theta_2^2+2(N_2-N_1)\theta_1\theta_2-N_1\theta_1^2\right] \\ +3Z_2\left[N_1(2\theta_1+\theta_2)-N_2(\theta_1+2\theta_2)\right] \\ +4(N_2-N_1)Z_3\end{array}\right\}$$

$$\text{Hence } (\tilde{\theta}_1-\tilde{\theta}_2)^{-1} \simeq \frac{1}{\theta_1-\theta_2}\left\{1-\frac{N^{\frac{1}{2}}}{N_1N_2(\theta_1-\theta_2)^3}\left[\begin{array}{l}2Z_1(N_2\theta_2^2+2(N_2-N_1)\theta_1\theta_2-N_1\theta_1^2) \\ +3Z_2(N_1(2\theta_1+\theta_2) \\ -N_2(\theta_1+2\theta_2))+4(N_2-N_1)Z_3\end{array}\right]\right. \tag{3.15}$$

Combining (3.11) and (3.15) we obtain the following expansion of equation (3.14) and hence of $\tilde{N}_1$.

$$\tilde{N}_1 = \frac{2A-N\tilde{\theta}_2}{\tilde{\theta}_1-\tilde{\theta}_2}$$

$$=(\theta_1-\theta_2)^{-1}\left\{1-\frac{N^{\frac{1}{2}}}{N_1N_2(\theta_1-\theta_2)^3}\left[\begin{array}{l}2Z_1(N_2\theta_2^2+2(N_2-N_1)\theta_1\theta_2-N_1\theta_1^2) \\ +3Z_2(N_1(2\theta_1+\theta_2)-N_2(\theta_1+2\theta_2)) \\ +4(N_2-N_1)Z_3\end{array}\right]\right\}$$

$$\times\{2N^{\frac{1}{2}}Z_1+N_1\theta_1+N_2\theta_2-N\theta_2-\frac{N^{3/2}}{N_2(\theta_1-\theta_2)^2}[2\theta_1Z_1(\theta_1+2\theta_2)-3Z_2(2\theta_1+\theta_2)+4Z_3]\}$$

$$\simeq N_1+\frac{2N^{\frac{1}{2}}Z_1}{\theta_1-\theta_2}-\frac{N^{\frac{1}{2}}}{N_2(\theta_1-\theta_2)^3}\begin{Bmatrix}2Z_1[N_2\theta_2^2+2(N_2-N_1)\theta_1\theta_2-N_1\theta_1^2]\\+3Z_2[N_1(2\theta_1+\theta_2)-N_2(\theta_1+2\theta_2)]\\+4(N_2-N_1)Z_3\end{Bmatrix}$$

$$-\frac{N^{3/2}}{N_2(\theta_1-\theta_2)^3}\{2\theta_1Z_1(\theta_1+2\theta_2)-3Z_2(2\theta_1+\theta_2)+4Z_3\}$$

$$=N_1-\frac{N^{\frac{1}{2}}}{N_2(\theta_1-\theta_2)^3}\begin{Bmatrix}2Z_1[N_2\theta_2^2+2(N_2-N_1)\theta_1\theta_2-N_1\theta_1^2-N_2(\theta_1-\theta_2)^2\\+N\theta_1(\theta_1+2\theta_2)]\\+3Z_2[N_1(2\theta_1+\theta_2)-N_2(\theta_1+2\theta_2)-N(2\theta_1+\theta_2)]\\+4Z_3(N_2-N_1+N)\end{Bmatrix}$$

$$=N_1-\frac{N^{\frac{1}{2}}}{N_2(\theta_1-\theta_2)^3}\{12Z_1N_2\theta_1\theta_2-9N_2Z_2(\theta_1+\theta_2)+8N_2Z_3\}$$

Hence the coefficients of $\tilde{N}_1$ with respect to Z_1,Z_2 and Z_3 are

$$\beta_{11}=\frac{-12N^{\frac{1}{2}}\theta_1\theta_2}{(\theta_1-\theta_2)^3},\quad \beta_{12}=\frac{9N^{\frac{1}{2}}(\theta_1+\theta_2)}{(\theta_1-\theta_2)^3}\ \text{and}\ \beta_{13}=\frac{-8N^{\frac{1}{2}}}{(\theta_1-\theta_2)^3} \tag{3.16}$$

Thus the coefficients of the expansion of $\tilde{N}_2$ are given by

$$\beta_{21}=\frac{12N^{\frac{1}{2}}\theta_1\theta_2}{(\theta_1-\theta_2)^3},\quad \beta_{22}=\frac{-9N^{\frac{1}{2}}(\theta_1+\theta_2)}{(\theta_1-\theta_2)^3}\ \text{and}\ \beta_{23}=\frac{8N^{\frac{1}{2}}}{(\theta_1-\theta_2)^3} \tag{3.17}$$

Now that we have the expansion for each of the four estimators as linear functions of Z_1,Z_2 and Z_3, it will be possible to calculate the variances of these estimators as well as the various covariances between estimators, provided we first determine the variances and covariances between Z_1, Z_2 and Z_3. The covariance between any Z_i and Z_j is given by

$$\begin{aligned}
\operatorname{cov}(Z_i,Z_j) &= \operatorname{cov}\left\{N^{-\frac{1}{2}}\left(\sum_{r=1}^{N} x_r^i - E\sum_{r=1}^{N} x_r^i\right),\ N^{-\frac{1}{2}}\left(\sum_{r=1}^{N} x_r^j - E\sum_{r=1}^{N} x_r^j\right)\right\} \\
&= \frac{1}{N}\sum_{r=1}^{N} \operatorname{cov}(x_r^i,\ x_r^j) \\
&= \frac{1}{N}\sum_{r=1}^{N}\left\{E(x_r^{i+j}) - E(x_r^i)E(x_r^j)\right\} \\
&= \frac{1}{N}\left\{\frac{N_1\theta_1^{i+j}}{i+j+1} - \frac{N_1\theta_1^{i+j}}{(i+1)(j+1)} + \frac{N_2\theta_2^{i+j}}{i+j+1} - \frac{N_2\theta_2^{i+j}}{(i+1)(j+1)}\right\} \\
&= \frac{ij(N_1\theta_1^{i+j} + N_2\theta_2^{i+j})}{(i+1)(j+1)(i+j+1)} \qquad \begin{matrix} i=1,2,3 \\ j=1,2,3 \end{matrix} \qquad (3.18)
\end{aligned}$$

If one uses formula (3.18) with the fact that

$$\operatorname{cov}\left(\sum_{i=1}^{3}\lambda_{1i}Z_i,\ \sum_{j=1}^{3}\lambda_{2j}Z_j\right) = \sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_{1i}\lambda_{2j}\operatorname{cov}(Z_i,Z_j),$$

where the λ_{1i}'s and λ_{2j}'s are constants, then it is an easy matter to calculate the variance of a particular estimator or the covariance between two moment estimators. This completes the derivation of the asymptotic distribution of the moment estimators.

We shall now show that, for large sample sizes, the quadratic (1.7) has real, positive roots with probability close to one, and so, for sufficiently large samples, the method of moments can be expected to yield real, positive estimates for θ_1 and θ_2 with probability close to one. The discriminant of this quadratic is given by

$$\begin{aligned}
\Delta &= (6AB-4NC)^2 - 4(3NB-4A^2)(8AC-9B^2) \\
&= 4\{(2NC-3AB)^2 + (9B^2-8AC)(3NB-4A^2)\} \\
&\simeq N_1^2N_2^2(\theta_1-\theta_2)^6 + 4N^{\frac{1}{2}}N_1N_2(\theta_1-\theta_2)^3\left\{\begin{matrix} 3Z_1(N_2\theta_2^2) + \frac{9}{2}Z_2(N_1\theta_1-N_2\theta_2) \\ -2Z_3(N_1-N_2) \end{matrix}\right\} \qquad (3
\end{aligned}$$

where we have used equation (3.9). Since Z_1, Z_2 and Z_3 are all equal to $O(N^{-\frac{1}{2}})$ in the limit as N tends to $+\infty$, it can be seen that the terms containing Z_1, Z_2 and Z_3 in equation (3.19) are all negligible in comparison to $N_1^2 N_2^2 (\theta_1 - \theta_2)^6$ in the limit.

Hence $\Delta' \simeq N_1^2 N_2^2 (\theta_1 - \theta_2)^6 > 0$ asymptotically. Thus for large N the quadratic (1.7) will have real roots with probability close to one. In a similar manner it can be argued that, for large N, the coefficients of (1.7) are given by

$$3NB - 4A^2 \simeq N_1 N_2 (\theta_1 - \theta_2)^2$$
$$6AB - 4NC \simeq -N_1 N_2 (\theta_1 - \theta_2)^2 (\theta_1 + \theta_2)$$
$$8AC - 9B^2 \simeq N_1 N_2 \theta_1 \theta_2 (\theta_1 - \theta_2)^2$$

where we have referred to equations (3.3),(3.5) and (3.7). Thus the coefficients alternate in sign and so, by Descartes' rule of signs, both roots of (1.7) are positive. Hence, asymptotically, the method of moments can be relied upon to provide reasonable estimators for the parameters.

We complete this section by noting that we could have obtained the asymptotic distribution of our moment estimators directly from the distribution of the moment estimators in the case of a mixed sample from two gamma distributions as discussed by John (1970(b)). We shall make some slight changes in John's notation to permit an easy comparison with the work of this chapter. If the distribution in the i^{th} population (i=1,2) is supposed to have the density function

$$\left\{\frac{\theta_i^{-f}}{\Gamma(f)}\right\} x^{f-1} e^{-\frac{x}{\theta_i}} \qquad (x \geq 0) \qquad (3.20)$$

and if we put $f = \frac{1}{2}$, then the moment estimates $\tilde{\theta}_1, \tilde{\theta}_2, \tilde{N}_1$ and $\tilde{N}_2$ are those values which satisfy the equations $\tilde{N}_1 + \tilde{N}_2 = N$ and

$$\sum_{r=1}^{N} x_r = \frac{1}{2}\tilde{N}_1\tilde{\theta}_1 + \frac{1}{2}\tilde{N}_2\tilde{\theta}_2$$

$$\sum_{r=1}^{N} x_r^2 = \frac{3}{4}\tilde{N}_1\tilde{\theta}_1^2 + \frac{3}{4}\tilde{N}_2\tilde{\theta}_2^2 \qquad (3.21)$$

$$\sum_{r=1}^{N} x_r^3 = \frac{15}{8}\tilde{N}_1\tilde{\theta}_1^3 + \frac{15}{8}\tilde{N}_2\tilde{\theta}_2^3$$

If we multiply the second and third equations of (3.21) by $\frac{4}{9}$ and $\frac{2}{15}$ respectively we then have the three equations

$$\sum_{r=1}^{N} x_r = \frac{1}{2}\tilde{N}_1\tilde{\theta}_1 + \frac{1}{2}\tilde{N}_2\tilde{\theta}_2$$

$$\frac{4}{9}\sum_{r=1}^{N} x_r^2 = \frac{1}{3}\tilde{N}_1\tilde{\theta}_1^2 + \frac{1}{3}\tilde{N}_2\tilde{\theta}_2^2 \qquad (3.22)$$

$$\frac{2}{15}\sum_{r=1}^{N} x_r^3 = \frac{1}{4}\tilde{N}_1\tilde{\theta}_1^3 + \frac{1}{4}\tilde{N}_2\tilde{\theta}_2^3$$

The right hand sides of equations (3.22) are the same as the right hand sides of the equations (1.2) which express the moment equations in the case of a mixture of samples from two uniform distributions. John (1970(b)) gives the coefficients of $Z_t' = N^{-\frac{1}{2}}(\sum_{r=1}^{N} x_r^t - E\sum_{r=1}^{N} x_r^t)$ $t=1,2,3$. in the expansion of each of the estimators obtained by solving equations (3.21). From equations (3.22) we see that the corresponding coefficients in the case of a mixture of samples from two uniform distributions will be the coefficients of $Z_1 = Z_1', Z_2 = \frac{4}{9}Z_2'$ and $Z_3 = \frac{2}{15}Z_3'$.

For example, John gives the coefficients of the expansion of $\tilde{\theta}_1$ with respect to Z_1', Z_2' and Z_3' as

$$\alpha_{11}' = \frac{2\theta_2(2\theta_1+\theta_2)N^{\frac{1}{2}}}{N_1(\theta_1-\theta_2)^2}, \quad \alpha_{12}' = \frac{-4N^{\frac{1}{2}}(\theta_1+2\theta_2)}{3N_1(\theta_1-\theta_2)^2}, \quad \alpha_{13}' = \frac{8N^{\frac{1}{2}}}{15N_1(\theta_1-\theta_2)^2} \qquad (3.23)$$

which we multiply by 1, $\frac{9}{4}$ and $\frac{15}{2}$ respectively to obtain the coefficients with respect to Z_1, Z_2 and Z_3 as

$$\alpha_{11}=\frac{2\theta_2(2\theta_1+\theta_2)N^{\frac{1}{2}}}{N_1(\theta_1-\theta_2)^2},\quad \alpha_{12}=\frac{-3N^{\frac{1}{2}}(\theta_1+2\theta_2)}{N_1(\theta_1-\theta_2)^2},\quad \alpha_{13}=\frac{4N^{\frac{1}{2}}}{N_1(\theta_1-\theta_2)^2}\ .$$

These coefficients are exactly the same as those of (3.13) which we derived from first principles. Similarly, the coefficients in (3.12), (3.16) and (3.17) have been checked and shown to be accurate.

Of course one could simply have used this method to derive our coefficients in the first place, but we elected to use the longer, more detailed approach in order to illustrate the ideas and for convenience in showing that the method of moments is asymptotically reliable.

As a final comment we note that the method of moments as outlined in §3.1 and §3.3 applies equally well to the model of a compound uniform distribution $p(x,\theta)=\alpha U[0,\theta_1]+(1-\alpha)U[0,\theta_2]$, if we simply replace $\frac{N_1}{N}$ by α, $\frac{N_2}{N}$ by $1-\alpha$ and then estimate accordingly.

3.4 MAXIMUM LIKELIHOOD UNDER A MIXED DISTRIBUTION MODEL

While the method of moments is essentially the same whether we use a mixed samples or a mixed distributions model, the maximum likelihood method, on the other hand, tends to be fairly straightforward under the former model, but may involve complex iterative processes under a mixed distributions model, especially if there is a large number of parameters. We shall now consider the maximum likelihood method for a compound uniform distribution consisting of a mixture of two one-parameter uniform distributions $U[0,\theta_1]$ and $U[0,\theta_2]$. We shall show that the method is quite simple in this particular case and that the m.ℓ.e's obtained here should be similar to those of §3.2, especially if the sample size is large.

Suppose that the set $\{x_1, x_2, \ldots, x_N\}$ is a single sample of N observations on a nonnegative random variable X which has the compound density

$$p(x;\ \alpha,\theta_1,\theta_2) = \frac{\alpha}{\theta_1}\, u(x-\theta_1) + \frac{(1-\alpha)}{\theta_2}\, u(x-\theta_2) \qquad (4.1)$$

where u(.) is that function which takes the value one if the argument is nonpositive and zero otherwise, and where $x \geq 0$, $\theta_2 \geq \theta_1 > 0$ and $1 > \alpha > 0$. In this model the unknown parameters are α, θ_1 and θ_2.

The joint likelihood of the set of N observations on X is given by

$$L(x_1,\ldots,x_N;\alpha,\theta_1,\theta_2) = \prod_{r=1}^{N} \left\{\frac{\alpha}{\theta_1}\, u(x_r-\theta_1) + \frac{(1-\alpha)}{\theta_2}\, u(x_r-\theta_2)\right\} \qquad (4.2)$$

and we wish to maximise this expression with respect to α, θ_1 and θ_2.

If $\theta_1=\theta_2=\theta$, the likelihood reduces to

$$L(x_1,\ldots,x_N;\ \theta) = \prod_{r=1}^{N} \left\{ \frac{u(x_r-\theta)}{\theta} \right\}$$

independently of the value of α. In this case the m.ℓ.e. of θ is $\hat{\theta} = x_{(N)}$, the largest order statistic, and the maximum value of the likelihood is

$$L(x_1,\ldots,x_N;\ \hat{\theta}) = \frac{1}{x_{(N)}^{N}} . \tag{4.3}$$

Now assume $\theta_1 < \theta_2$. Let the set of ordered observations be $x_{(1)} < x_{(2)} < \ldots < x_{(N)}$. If $\tilde{\theta}_2 < x_{(N)}$ is an estimator of θ_2, the likelihood of the sample is given by $L(x_1,\ldots,x_N;\ \alpha,\theta_1,\tilde{\theta}_2) = 0$ and hence the m.ℓ.e. $\hat{\theta}_2$ of θ_2 satisfies the inequality $\hat{\theta}_2 \geq x_{(N)}$. But the likelihood is maximised for the smallest such estimate of θ_2, and so the m.ℓ.e. is given by

$$\hat{\theta}_2 = x_{(N)} . \tag{4.4}$$

Once it is maximised with respect to θ_2 the likelihood becomes

$$\begin{aligned} L(x_1,\ldots,x_N;\ \alpha,\theta_1,\hat{\theta}_2) &= \prod_{r=1}^{N} \left\{ \frac{\alpha}{\theta_1} u(x_r-\theta_1) + \frac{1-\alpha}{x_{(N)}} \right\} \\ &= \left\{ \frac{\alpha}{\theta_1} + \frac{1-\alpha}{x_{(N)}} \right\}^{N_1(\theta_1)} . \left\{ \frac{1-\alpha}{x_{(N)}} \right\}^{N-N_1(\theta_1)} \end{aligned} \tag{4.5}$$

where $N_1(\theta_1)$ is the number of observations less than or equal to θ_1. Now we can maximise (4.5) in the following way. Let $\tilde{\theta}_{1j}=x_{(j)}$ be an estimate of θ_1 for any $j \in \{1,2,\ldots,N-1\}$. Note that $N_1(x_{(j)}) = j$ is the number of observations less than or equal to $\tilde{\theta}_{1j} = x_{(j)}$. For this estimate of θ_1 the likelihood (4.5) becomes

$$L(x_1,\ldots,x_N;\ \alpha,\tilde{\theta}_{1j},\hat{\theta}_2) = \{\frac{\alpha}{x_{(j)}} + \frac{1-\alpha}{x_{(N)}}\}^j \{\frac{1-\alpha}{x_{(N)}}\}^{N-j} \quad (4.6)$$

which we shall maximise with respect to α, $0 < \alpha < 1$, for each $j=1,2,\ldots,N-1$. It should be noted that it is not necessary to consider a continuum of values for $\tilde{\theta}_1$ in the interval $[x_{(j)},x_{(j+1)})$, as the likelihood (4.6) is always greatest at the left hand end point of such an interval.

Now our method is to maximise (4.6) with respect to α for each $j \in \{1,2,\ldots,N-1\}$ and compare the N-1 values of the likelihood so obtained, as well as that value (4.3) derived under the assumption that $\theta_1 = \theta_2$. The largest of these N likelihood values will lead us to the required m.ℓ.e.'s of θ_1 and α. Note that it might also be thought necessary to maximise (4.6) for the case j=N, but this merely gives the estimates $\hat{\theta}_1 = \hat{\theta}_2 = x_{(N)}$ independently of α and the same value of the likelihood as obtained in (4.3) under the assumption that $\theta_1 = \theta_2$.

To maximise (4.6) with respect to α, $0 < \alpha < 1$, we first note that

$$\lim_{\alpha\to 0+} L(x_1,\ldots,x_N;\ \alpha,\tilde{\theta}_{1j},\hat{\theta}_2) = \frac{1}{x_{(N)}^N} \quad (4.7)$$

and $$\lim_{\alpha\to 1-} L(x_1,\ldots,x_N;\ \alpha,\tilde{\theta}_{1j},\hat{\theta}_2) = 0 \quad (4.8)$$

Hence, if there is in fact no local maximum for L as a function of $\alpha \in (0,1)$, the likelihood is maximised as $\alpha\to 0+$ and this maximum value is the value (4.3) obtained by maximising L under the assumption $\theta_1 = \theta_2$.

To find any local maxima we differentiate (4.6) with respect to α and equate to zero, thus

$$\frac{1}{x_{(j)}^{j}x_{(N)}^{N}}\left\{\begin{array}{l} j[\alpha x_{(N)}+(1-\alpha)x_{(j)}]^{j-1}(1-\alpha)^{N-j}(x_{(N)}-x_{(j)}) \\ -(N-j)(1-\alpha)^{N-j-1}[\alpha x_{(N)}+(1-\alpha)x_{(j)}] \end{array}\right\} = 0$$

Hence $j(1-\alpha)(x_{(N)}-x_{(j)}) = (N-j)\ \alpha x_{(N)}+(1-\alpha)x_{(j)}$ or $\alpha = 1$, but this last value can be discarded as shown by the limits (4.7) and (4.8). The first equation becomes

$$jx_{(N)}-jx_{(j)}-\alpha j(x_{(N)}-x_{(j)}) = \alpha(N-j)(x_{(N)}-x_{(j)}) + (N-j)x_{(j)}$$

i.e. $\alpha\{(N-j)(x_{(N)}-x_{(j)})+j(x_{(N)}-x_{(j)})\}= jx_{(N)}-jx_{(j)}-(N-j)x_{(j)}$

Hence $\alpha = \dfrac{jx_{(N)}-Nx_{(j)}}{N(x_{(N)}-x_{(j)})}$. (4.9)

Note that $\dfrac{jx_{(N)}-Nx_{(j)}}{N(x_{(N)}-x_{(j)})} > 0$ provided $jx_{(N)} > Nx_{(j)}$

and that $1 - \dfrac{jx_{(N)}-Nx_{(j)}}{N(x_{(N)}-x_{(j)})} = \dfrac{(N-j)x_{(N)}}{N(x_{(N)}-x_{(j)})} > 0$ for $j \varepsilon \{1,2,...,N-1\}$.

Given that $jx_{(N)} > Nx_{(j)}$ for some $j \varepsilon \{1,2,...,N-1\}$, we now show that (4.9) is in fact a local maximum for the likelihood (4.6). Since (4.9) is the only point in the interval (0,1) at which the derivative is zero, we need only examine the value of the derivative at $\alpha = 0+$ under the condition that $jx_{(N)} > Nx_{(j)}$:

$$\begin{aligned}
\frac{\partial L}{\partial \alpha}(x_1,...,x_N;\ \alpha=0+,\ \tilde{\theta}_{1j}= x_{(j)},\ \hat{\theta}_2 = x_{(N)}) \\
&= \frac{1}{x_{(j)}^{j}x_{(N)}^{N}}\{j\ x_{(j)}^{j-1}\ (x_{(N)}-x_{(j)}) - (N-j)x_{(j)}^{j}\} \\
&= \frac{1}{x_{(j)}x_{(N)}^{N}}\{j\ x_{(N)}-jx_{(j)}-(N-j)x_{(j)}\} \\
&= \frac{jx_{(N)}-Nx_{(j)}}{x_{(j)}x_{(N)}^{N}} \\
&> 0 \text{ if and only if } j\ x_{(N)} > N\ x_{(j)}
\end{aligned}$$

Hence for those values of $j \in \{1,2,\ldots,N-1\}$ which satisfy $jx_{(N)} > Nx_{(j)}$ the m.ℓ.e. $\hat{\alpha}_j$ of α is given by (4.9),

$$\text{i.e. } \hat{\alpha}_j = \frac{jx_{(N)}-Nx_{(j)}}{N(x_{(N)}-x_{(j)})}, \tag{4.10}$$

and the value of the likelihood (4.6) is

$$L(x_1,\ldots,x_N;\ \hat{\alpha}_j, \tilde{\theta}_{1j} = x_{(j)}, \tilde{\theta}_2 = x_{(N)})$$

$$= \frac{1}{x_{(j)}^j x_{(N)}^N} \ \{\hat{\alpha}_j x_{(N)} + (1-\hat{\alpha}_j)x_{(j)}\}^j (1-\hat{\alpha}_j)^{N-j}$$

$$= \frac{1}{x_{(j)}^j x_{(N)}^N} \cdot \frac{(jx_{(N)}^2 - Nx_{(j)}x_{(N)} + Nx_{(j)}x_{(N)} - jx_{(j)}x_{(N)})^j (Nx_{(N)} - jx_{(N)})^{N-j}}{\{N(x_{(N)}-x_{(j)})\}^N}$$

$$= \frac{j^j (x_{(N)}-x_{(j)})^j (N-j)^{N-j} x_{(N)}^{N-j}}{x_{(j)}^j \ x_{(N)}^{N-j} \ N^N (x_{(N)}-x_{(j)})^N}$$

$$= \frac{1}{N^N} \left(\frac{j}{x_{(j)}}\right)^j \left(\frac{N-j}{x_{(N)}-x_{(j)}}\right)^{N-j} \tag{4.11}$$

Thus the method reduces to the following simple process. The m.ℓ.e. of θ_2 is $\hat{\theta}_2 = x_{(N)}$. Then, for each $j \in \{0,1,\ldots,N-1\}$ which satisfies $jx_{(N)} > Nx_{(j)}$, one calculates the value of (4.11) and chooses that integer j^* for which (4.11) is maximum. Then the m.ℓ.e. of θ_1 is $\hat{\theta}_1 = \tilde{\theta}_{1j^*} = x_{(j^*)}$ and the m.ℓ.e. of α

$$\text{is } \hat{\alpha} = \hat{\alpha}_{j^*} = \frac{j^* x_{(N)} - Nx_{(j^*)}}{N(x_{(N)}-x_{(j^*)})} \tag{4.12}$$

If there is no $j \in \{1,2,\ldots,N-1\}$ for which $jx_{(N)} > Nx_{(j)}$, the maximum value of the likelihood is given by (4.3) and so the m.ℓ.e's are $\hat{\theta}_1 = \hat{\theta}_2 = x_{(N)}$ and the density (4.1) is independent of α for these estimates of θ_1 and θ_2.

Since one maximises

$$\frac{1}{N^N} \left(\frac{j}{x_{(j)}}\right)^j \left(\frac{N-j}{x_{(N)}-x_{(j)}}\right)^{N-j}$$

under a mixed distribution model and one maximises $\{x_{(j)}^{j} x_{(N)}^{N-j}\}$ under a mixed samples model (see §3.2), it is likely that the two models will not generally yield the same estimates of θ_1 and θ_2. A comparison of these expressions is not straightforward, mainly because the former is maximised under the additional constraint that $jx_{(N)} > Nx_{(j)}$ i.e. $\frac{j}{x_{(j)}} > \frac{N}{x_{(N)}}$, and it is not immediately obvious, even for large N, whether j should be small or large to satisfy this restriction. In any case, it is doubtful whether such a mathematical comparison would be very meaningful, seeing that the likelihoods are calculated under two different models. However, in view of the fact that a compound distribution model may be regarded as a random subsample from a mixed sample of size infinity, one might reasonably expect the models to yield the same results asymptotically.

For the purpose of comparing the relative performances of the m.ℓ.e.'s under each model one could consider a large variety of mixed samples or single samples from a compound uniform distribution and determine the estimates obtained by adopting each model in turn. Further comparison could be made with the moment estimators which are essentially the same under either model. We shall not attempt here such a laborious procedure, and shall rest content with a single example to illustrate the methods.

3.5 AN ILLUSTRATIVE EXAMPLE

Using random numbers, we construct a sample of size 10 on a random variable X which has a compound uniform distribution given by (4.1) and with the parametric values $\alpha=0.5$, $\theta_1=1$ and $\theta_2 = 3$. So we take 10 observations each of which has probability 0.5 of being taken from $U[0,\theta_1]$ and probability 0.5 of being taken from $U[0,\theta_2]$. In the example which we consider here these observations are

$\{2.50, 0.135, 2.22, 1.69\}$ from $U[0,\theta_2]$

and $\{0.682, 0.931, 0.870, 0.325, 0.287, 0.686\}$ from $U[0,\theta_1]$.

<u>The Method of Moments</u>

The sample moments are given by $A = \sum_{r=1}^{10} x_r = 10.326$, $B = \sum_{r=1}^{10} x_r^2 \equiv 16.80$ and $C = \sum_{r=1}^{10} x_r^3 \simeq 33.56$. Now we determine the coefficients of the quadratic (1.7) which is to be solved for the estimates $\tilde{\theta}_1, \tilde{\theta}_2$ of θ_1 and θ_2.

$$3NB-4A^2 \simeq 504.0-426.5 = 77.5$$

$$6AB-4NC \simeq 1040.8-1342.4 = -301.6$$

$$8AC-9B^2 \simeq 2772.3-2540.1 = 232.2$$

We note that conditions (1.21) are satisfied for the existence of real, positive estimates of θ_1 and θ_2, since $3N^2m_2-A^2 = 3NB-4A^2 > 0$, $m_3 = \frac{C}{N} - \frac{3AB}{N^2} + \frac{2A^3}{N^3} \simeq 0.354$ and $\frac{1}{8AN^3}(3N^2m_2-A^2) \simeq 0.0727 < m_3$.

In fact, the estimates of θ_1 and θ_2 are given by

$$\tilde{\theta}_2, \tilde{\theta}_1 \equiv \frac{301.6 \pm[(301.6)^2-4(77.5)(232.2)]^{\frac{1}{2}}}{155}$$

$$\simeq \frac{301.6\pm137.8}{155}$$

$$\simeq 2.83,\ 1.06.$$

Hence, by formula (1.22),

$$\tilde{N}_1 = \frac{2A-N\tilde{\theta}_2}{\tilde{\theta}_1-\tilde{\theta}_2} \simeq \frac{20.652-28.3}{-1.77} \simeq 4.32$$

Thus under a mixed samples model we estimate N_1 by $\tilde{N}_1'=4$ and N_2 by $\tilde{N}_2'=6$, which implies that we should allocate the four smallest observations to $U[0,\theta_1]$ and the remaining 6 to $U[0,\theta_2]$, a procedure which results in considerable misclassification. Under the model of a single sample from a compound uniform distribution we estimate the parameter α by $\hat{\alpha} \equiv 0.432$. Note that the estimates 1.06 and 2.83 of θ_1 and θ_2 are of course the same in either model.

Maximum Likelihood under a Mixed Distribution Model

The ordered set of observations is given by $x_{(1)}=0.135$, $x_{(2)} = 0.287$, $x_{(3)} = 0.325$, $x_{(4)} = 0.682$, $x_{(5)} = 0.686$, $x_{(6)} = 0.870$, $x_{(7)} = 0.931$, $x_{(8)} = 1.69$, $x_{(9)} = 2.22$, $x_{(10)} = 2.50$. It is easily verified that $jx_{(10)} > 10\, x_{(j)}$ for each $j=1,2,\ldots,9$, and so we must evaluate the likelihood

$\frac{1}{N^N}\left(\frac{j}{x_{(j)}}\right)^j\left(\frac{N-j}{x_{(N)}-x_{(j)}}\right)^{N-j}$ as given by (4.11) for each $j=1,2,\ldots,9$.

Notate these nine values $L_1,L_2,\ldots,L_9$ respectively. It follows that

$$L_1 = (10)^{-10}\left(\frac{1}{0.135}\right)\left(\frac{9}{2.365}\right)^9 \simeq 1.23\times10^{-4}$$

$$L_2 = (10)^{-10}\left(\frac{2}{0.287}\right)\left(\frac{8}{2.213}\right)^8 \simeq 1.42\times10^{-4}$$

$$L_3 = (10)^{-10}\left(\frac{3}{0.325}\right)\left(\frac{7}{2.175}\right)^7 \simeq 2.81\times10^{-4}$$

$$L_4 = (10)^{-10}\left(\frac{4}{0.682}\right)^4\left(\frac{6}{1.818}\right)^6 \simeq 1.53\times10^{-4}$$

$$L_5 = (10)^{-10}\left(\frac{5}{0.686}\right)^5\left(\frac{5}{1.814}\right)^5 \simeq 3.27\times10^{-4}$$

$$L_6 = (10)^{-10}\left(\frac{6}{0.870}\right)^6\left(\frac{4}{1.63}\right)^4 \simeq 3.90\times10^{-4}$$

$$L_7 = (10)^{-10}\left(\frac{7}{0.931}\right)^7\left(\frac{3}{1.569}\right)^3 \simeq 9.49\times10^{-4}$$

$$L_8 = (10)^{-10}\left(\frac{8}{1.69}\right)^8\left(\frac{2}{0.81}\right)^2 \simeq 1.53\ 10^{-4}$$

$$L_9 = (10)^{-10}\left(\frac{9}{2.22}\right)^9\left(\frac{1}{0.28}\right) \simeq 1.06\times10^{-4}$$

Note that all these values are greater than $\left(\frac{1}{x_{(10)}}\right)^{10} = 1.05\times10^{-4}$ which is the maximum value of the likelihood under the assumption that $\theta_1=\theta_2=\theta$ and that all the observations are in fact from the single population $U[0,\theta]$.

Now L_7 is the largest of the above likelihoods and hence $\hat{\theta}_1 = x_{(7)} = 0.931$, $\hat{\theta}_2 = x_{(10)} = 2.50$ and by equation (4.12), $\hat{\alpha} = \frac{7(2.50)-10(0.931)}{10(1.569)} \simeq 0.522$. These estimates are of similar accuracy to those obtained by the method of moments.

Maximum Likelihood under a Mixed Samples Model

Under the assumption that the set of 10 observations is a mixture of two samples, one of size N_1 from $U[0,\theta_1]$ and the other of size N_2 from $U[0,\theta_2]$, it was shown in §3.2 that we should maximise $\{x_{(j)}^{j}x_{(N)}^{N-j}\}^{-1}$ or equivalently minimise $j \log x_{(j)}+(N-j)\log x_{(N)}$ over $j=1,2,\ldots,10$. The values of the latter expression which need to be considered are

$$\log x_{(1)}+9\log x_{(10)} = -0.8697+3.5811 = 2.7114$$

$$2\log x_{(2)}+8\log x_{(10)} = -1.0842+3.1832 = 2.0990$$

$$3\log x_{(3)}+7\log x_{(10)} = -1.4643+2.7853 = 1.3210$$

$$4\log x_{(4)}+6\log x_{(10)} = -0.6480+2.3874 = 1.7394$$

$$5 \log x_{(5)} + 5 \log x_{(10)} = -0.8185 + 1.9895 = 1.1710$$
$$6 \log x_{(6)} + 4 \log x_{(10)} = -0.3630 + 1.1933 = 1.2286$$
$$7 \log x_{(7)} + 3 \log x_{(10)} = -0.3170 + 1.1937 = 0.8760$$
$$8 \log x_{(8)} + 2 \log x_{(10)} = 1.8232 + 0.7958 = 2.6190$$
$$9 \log x_{(9)} + \log x_{(10)} = 3.1176 + 0.3979 = 3.5155$$

The minimum of the above expressions occurs at $j=7$ and so, as in the mixed distributions model, the likelihood is maximised at $j=7$. Hence, under this model, the m.ℓ.e.'s are $\hat{\theta}_1 = x_{(7)} = 0.931$, $\hat{\theta}_2 = x_{(10)} = 2.50$, $\hat{N}_1=7$ and $\hat{N}_2=3$. We allocate $x_{(1)}, x_{(2)}, \ldots, x_{(7)}$ to the population $U[0,\hat{\theta}_1]$ and $x_{(8)}, x_{(9)}$ and $x_{(10)}$ to $U[0,\hat{\theta}_2]$. This results in the misclassification of the observation $x_{(1)} = 0.135$.

It is of interest that the method of moments does not yield as accurate estimates of N_1 and N_2 as the maximum likelihood method and one might speculate as to whether this is a common failing for small sample sizes.

CHAPTER FOUR

ESTIMATION IN MIXTURES OF TWO-PARAMETER UNIFORM DISTRIBUTIONS

4.1 A MIXED SAMPLE FROM $U[a_1-h,a_1+h]$ AND $U[a_2-h,a_2+h]$ ANALYSED BY THE METHOD OF MOMENTS

In this and following sections we wish to apply the methods discussed in the previous chapter to mixed samples and compound distributions of random variables possessing a uniform distribution with both scale and location parameters. By such a distribution, which we shall call a two-parameter uniform distribution, we mean that the random variable X should have a density function given by

$$p_X(x) = \begin{cases} \frac{1}{2h} & a-h \leq x \leq a+h \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

where $h > 0$. The introduction of another parameter will of course increase the difficulties of estimation in mixtures of samples and distributions. If, however, either end-point $a-h$ or $a+h$ is fixed and known to be the same for each of the mixing populations of type (1.1), then the appropriate simple transformation can be used to reduce the problem to one of a mixture of $U[0,\theta_1]$ and $U[0,\theta_2]$, which has already been discussed.

We shall be concerned in this section with the question of estimating the parameters of a mixed sample from $U[a_1-h,a_1+h]$ and $U[a_2-h,a_2+h]$ by the method of moments. The maximum likelihood method is discussed in the following section. As a practical situation requiring the type of analysis considered here, one might envisage the use of a measuring instrument which estimates some characteristic of true value "a" with the error distribution $U[a-h,a+h]$. If it is decided that the characteristic in fact took on two unknown values a_1 and a_2 during the accumulation of a body of measurements, while h remained constant, then one should regard the problem as either a mixture of samples from $U[a_1-h,a_1+h]$ and $U[a_2-h,a_2+h]$ or as a single sample from a mixture of the distributions $U[a_1-h,a_1+h]$ and $U[a_2-h,a_2+h]$. In view of the simplicity of the associated maximum likelihood method, the former model seems the more useful in this particular situation. Of course, the moment estimates of a_1,a_2 and h will be identical under either model.

Let us then adopt a mixed samples model. A set $\{x_1,x_2,\ldots,x_N\}$ of N observations is believed to consist of a sample of size N_1 from the population $U[a_1-h,a_1+h]$ and a sample of size N_2 from the population $U[a_2-h,a_2+h]$. It is necessary to estimate the unknown parameters N_1,N_2,a_1,a_2 and h. As in Chapter Three, to apply the method of moments, we equate sample moments about the origin to their expected values and solve for those estimates $\tilde{N}_1,\tilde{N}_2,\tilde{a}_1,\tilde{a}_2$ and $\tilde{h}$ of the parameters which simultaneously satisfy these equations. Thus, using the first four moments, we obtain

$$A = \sum_{r=1}^{N} x_r = \tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2$$

$$B = \sum_{r=1}^{N} x_r^2 = \tilde{N}_1(\tilde{a}_1^2+ \frac{1}{3}\,\tilde{h}^2)+\tilde{N}_2(\tilde{a}_2^2+ \frac{1}{3}\,\tilde{h}^2)$$

$$C = \sum_{r=1}^{N} x_r^3 = \tilde{N}_1\tilde{a}_1(\tilde{a}_1^2+\tilde{h}^2)+\tilde{N}_2\tilde{a}_2(\tilde{a}_2^2+\tilde{h}^2) \tag{1.2}$$

$$D = \sum_{r=1}^{N} x_r^4 = \tilde{N}_1(\tilde{a}_1^4+2\tilde{a}_1^2\tilde{h}^2+\frac{1}{5}\tilde{h}^4)+\tilde{N}_2(\tilde{a}_2^4+2\tilde{a}_2^2\tilde{h}^2+\frac{1}{5}\tilde{h}^4)$$

When we couple equations (1.2) with the further requirement that $\tilde{N}_1+\tilde{N}_2=N$, we obtain a system of five equations in the five estimates $\tilde{N}_1, \tilde{N}_2, \tilde{a}_1, \tilde{a}_2$ and $\tilde{h}$. Since the first three moments of the distribution $U[a-h,a+h]$ are the same as the first three moments of a normal distribution $N(a, \frac{1}{3}h^2)$, we expect considerable analogy with the method of moments applied to a mixed sample from $N(\mu_1,\sigma^2)$ and $N(\mu_2,\sigma^2)$ as discussed by John (1970(a)). While the analogy is of course useful in solving the equations, we shall show that it cannot be carried further to prove that there is a unique solution for each of the parameters.

Using the fact that $\tilde{N}_1+\tilde{N}_2=N$, equations (1.2) can be written as

$$\begin{aligned} A &= \tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2 \\ B &= \tilde{N}_1\tilde{a}_1^2+\tilde{N}_2\tilde{a}_2^2+\frac{1}{3}N\tilde{h}^2 \\ C &= \tilde{N}_1\tilde{a}_1^3+\tilde{N}_2\tilde{a}_2^3+\tilde{h}^2(\tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2) \\ D &= \tilde{N}_1\tilde{a}_1^4+\tilde{N}_2\tilde{a}_2^4+2\tilde{h}^2(\tilde{N}_1\tilde{a}_1^2+\tilde{N}_2\tilde{a}_2^2)+\frac{1}{5}N\tilde{h}^4 \end{aligned} \tag{1.3}$$

Multiplying the first of these equations by $(\tilde{a}_1+\tilde{a}_2)$ and substituting for B from the second equation, it follows that

$$\begin{aligned} A(\tilde{a}_1+\tilde{a}_2) &= \tilde{N}_1\tilde{a}_1^2+\tilde{N}_1\tilde{a}_1\tilde{a}_2+\tilde{N}_2\tilde{a}_2^2+\tilde{N}_2\tilde{a}_1\tilde{a}_2 \\ &= B - \frac{1}{3}N\tilde{h}^2 + N\tilde{a}_1\tilde{a}_2 \end{aligned}$$

Multiplying the second equation of the set (1.3) by $(\tilde{a}_1+\tilde{a}_2)$ and substituting for C from the third equation, it follows that

$$B(\tilde{a}_1+\tilde{a}_2) = (\tilde{N}_1\tilde{a}_1^2+\tilde{N}_2\tilde{a}_2^2)(\tilde{a}_1+\tilde{a}_2)+\frac{1}{3}N\tilde{h}^2(\tilde{a}_1+\tilde{a}_2)$$

$$= \tilde{N}_1\tilde{a}_1^3+\tilde{N}_2\tilde{a}_2^3+\tilde{N}_1\tilde{a}_1^2\tilde{a}_2+\tilde{N}_2\tilde{a}_1\tilde{a}_2^2+\frac{1}{3}N\tilde{h}^2(\tilde{a}_1+\tilde{a}_2)$$

$$= C-\tilde{h}^2(\tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2)+\tilde{a}_1\tilde{a}_2A+\frac{1}{3}N\tilde{h}^2(\tilde{a}_1+\tilde{a}_2)$$

Hence $(\tilde{a}_1+\tilde{a}_2)(B-\frac{1}{3}N\tilde{h}^2) = C-\tilde{h}^2(\tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2)+\tilde{a}_1\tilde{a}_2A$

Now manipulating the third equation of (1.3) in a similar manner,

$$C(\tilde{a}_1+\tilde{a}_2) = \tilde{N}_1\tilde{a}_1^3(\tilde{a}_1+\tilde{a}_2)+\tilde{N}_2\tilde{a}_2^3(\tilde{a}_1+\tilde{a}_2)+\tilde{h}^2(\tilde{a}_1+\tilde{a}_2)(\tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2)$$

$$= \tilde{N}_1\tilde{a}_1^4+\tilde{N}_2\tilde{a}_2^4+\tilde{N}_1\tilde{a}_1^3\tilde{a}_2+\tilde{N}_2\tilde{a}_1\tilde{a}_2^3+\tilde{h}^2(\tilde{a}_1+\tilde{a}_2)A$$

Hence $(C-A\tilde{h}^2)(\tilde{a}_1+\tilde{a}_2) = D-2\tilde{h}^2(\tilde{N}_1\tilde{a}_1^2+\tilde{N}_2\tilde{a}_2^2)-\frac{1}{5}N\tilde{h}^4+\tilde{a}_1\tilde{a}_2(\tilde{N}_1\tilde{a}_1^2+\tilde{N}_2\tilde{a}_2^2)$

$$= D-2\tilde{h}^2(B-\frac{1}{3}N\tilde{h}^2)-\frac{1}{5}N\tilde{h}^4+\tilde{a}_1\tilde{a}_2(B-\frac{1}{3}N\tilde{h}^2)$$

$$= D-2\tilde{h}^2B+\frac{7}{15}N\tilde{h}^4+\tilde{a}_1\tilde{a}_2(B-\frac{1}{3}N\tilde{h}^2)$$

Thus equations (1.3) have been reduced to the following system of three equations in the three unknowns $\tilde{a}_1, \tilde{a}_2$ and $\tilde{h}$.

$$\begin{aligned} A(\tilde{a}_1+\tilde{a}_2) &= B-\frac{1}{3}N\tilde{h}^2+N\tilde{a}_1\tilde{a}_2 \\ (B-\frac{1}{3}N\tilde{h}^2)(\tilde{a}_1+\tilde{a}_2) &= C-\tilde{h}^2A+\tilde{a}_1\tilde{a}_2A \\ (C-A\tilde{h}^2)(\tilde{a}_1+\tilde{a}_2) &= D-2\tilde{h}^2B+\frac{7}{15}N\tilde{h}^4+\tilde{a}_1\tilde{a}_2(B-\frac{1}{3}N\tilde{h}^2) \end{aligned} \tag{1.4}$$

Eliminating $(\tilde{a}_1+\tilde{a}_2)$ between the first and second equations of (1.4), one obtains

$$(B-\frac{1}{3}N\tilde{h}^2)(B-\frac{1}{3}N\tilde{h}^2+N\tilde{a}_1\tilde{a}_2) = A(C-A\tilde{h}^2)+\tilde{a}_1\tilde{a}_2A^2$$

i.e. $$[N(B-\frac{1}{3}N\tilde{h}^2)-A^2]\tilde{a}_1\tilde{a}_2 = A(C-A\tilde{h}^2)-(B-\frac{1}{3}N\tilde{h}^2)^2$$

Now eliminating $(\tilde{a}_1+\tilde{a}_2)$ from the first and third equations of (1.4), it follows that

$$(C-A\tilde{h}^2)(B-\frac{1}{3}N\tilde{h}^2+N\tilde{a}_1\tilde{a}_2)=A[D-2\tilde{h}^2B+\frac{7}{15}N\tilde{h}^4+\tilde{a}_1\tilde{a}_2(B-\frac{1}{3}N\tilde{h}^2)]$$

i.e. $$[N(C-A\tilde{h}^2)-A(B-\frac{1}{3}N\tilde{h}^2)]\tilde{a}_1\tilde{a}_2=A[D-2\tilde{h}^2B+\frac{7}{15}N\tilde{h}^4]-(C-A\tilde{h}^2)(B-\frac{1}{3}N\tilde{h}^2)$$

$$\text{i.e. } [NC-AB-\tfrac{2}{3}NA\tilde{h}^2]\tilde{a}_1\tilde{a}_2 = A(D-2\tilde{h}^2B)+\tfrac{2}{15}AN\tilde{h}^4-C(B-\tfrac{1}{3}N\tilde{h}^2)+AB\tilde{h}^2$$
$$= A(D-\tilde{h}^2B)+\tfrac{2}{15}AN\tilde{h}^4-C(B-\tfrac{1}{3}N\tilde{h}^2)$$

Thus we have reduced (1.4) to a system of two equations in the two unknowns $\tilde{a}_1\tilde{a}_2$ and $\tilde{h}^2$, given by

$$\left.\begin{aligned} [N(B-\tfrac{1}{3}N\tilde{h}^2)-A^2]\tilde{a}_1\tilde{a}_2 &= A(C-A\tilde{h}^2)-(B-\tfrac{1}{3}N\tilde{h}^2)^2 \\ [NC-AB-\tfrac{2}{3}NA\tilde{h}^2]\tilde{a}_1\tilde{a}_2 &= A(D-B\tilde{h}^2)+\tfrac{2}{15}AN\tilde{h}^4-C(B-\tfrac{1}{3}N\tilde{h}^2)\end{aligned}\right\}\quad(1.5)$$

Eliminating $\tilde{a}_1\tilde{a}_2$ between the two equations of (1.5) yields the following equation in $\tilde{h}^2$.

$$\{NC-AB-\tfrac{2}{3}NA\tilde{h}^2\}\{A(C-A\tilde{h}^2)-(B-\tfrac{1}{3}N\tilde{h}^2)^2\}$$
$$=\{N(B-\tfrac{1}{3}N\tilde{h}^2)-A^2\}\{A(D-B\tilde{h}^2)+\tfrac{2}{15}NA\tilde{h}^4-C(B-\tfrac{1}{3}N\tilde{h}^2)\}$$

$$\text{i.e. } \{NC-AB-\tfrac{2}{3}NA\tilde{h}^2\}\{AC-A^2\tilde{h}^2-B^2+\tfrac{2}{3}NB\tilde{h}^2-\tfrac{1}{9}N^2\tilde{h}^4\}$$
$$= \{NB-\tfrac{1}{3}N^2\tilde{h}^2-A^2\}\{AD-AB\tilde{h}^2+\tfrac{2}{15}NA\tilde{h}^4-BC+\tfrac{1}{3}NC\tilde{h}^2\}$$

$$\text{i.e. } NAC^2-NA^2C\tilde{h}^2-NB^2C+\tfrac{2}{3}N^2BC\tilde{h}^2-\tfrac{1}{9}N^3C\tilde{h}^4-A^2BC+A^3B\tilde{h}^2$$
$$+AB^3-\tfrac{2}{3}NAB^2\tilde{h}^2+\tfrac{1}{9}N^2AB\tilde{h}^4-\tfrac{2}{3}NA^2C\tilde{h}^2+\tfrac{2}{3}NA^3\tilde{h}^4+\tfrac{2}{3}NAB^2\tilde{h}^2$$
$$-\tfrac{4}{9}N^2AB\tilde{h}^4+\tfrac{2}{27}N^3A\tilde{h}^6$$
$$= NABD-NAB^2\tilde{h}^2+\tfrac{2}{15}N^2AB\tilde{h}^4-NB^2C+\tfrac{1}{3}N^2BC\tilde{h}^2-\tfrac{1}{3}N^2AD\tilde{h}^2$$
$$+\tfrac{1}{3}N^2AB\tilde{h}^4-\tfrac{2}{45}N^3A\tilde{h}^6+\tfrac{1}{3}N^2BC\tilde{h}^2-\tfrac{1}{9}N^3C\tilde{h}^4-A^3D+A^3B\tilde{h}^2$$
$$-\tfrac{2}{15}NA^3\tilde{h}^4+A^2BC-\tfrac{1}{3}NA^2C\tilde{h}^2$$

Hence, collecting terms with the same power of $\tilde{h}$, it follows that

$$(\tfrac{2}{27}+\tfrac{2}{45})N^3A\tilde{h}^6+\tilde{h}^4(\tfrac{1}{9}N^2AB+\tfrac{2}{3}NA^3-\tfrac{4}{9}N^2AB-\tfrac{2}{15}N^2AB-\tfrac{1}{3}N^2AB+\tfrac{2}{15}NA^3)$$
$$+\tilde{h}^2(-NA^2C-\tfrac{2}{3}NA^2C+NAB^2+\tfrac{1}{3}N^2AD+\tfrac{1}{3}NA^2C)$$
$$+NAC^2-A^2BC+AB^3-NABD+A^3D-A^2BC = 0$$

i.e. $\frac{16}{135}N^3A\tilde{h}^6+\tilde{h}^4(-\frac{4}{5}N^2AB+\frac{4}{5}NA^3)+\tilde{h}^2(-\frac{4}{3}NA^2C+NAB^2+\frac{1}{3}N^2AD)$

$$-2A^2BC+NAC^2+AB^3+A^3D-NABD = 0.$$

i.e. $\frac{16}{135}N^3\tilde{h}^6+\frac{4}{5}N\tilde{h}^4(A^2-NB)+\frac{1}{3}N\tilde{h}^2(3B^2+ND-4AC)$

$$+ NC^2-2ABC+B^3+A^2D-NBD = 0.$$

Thus $\tilde{h}^2$ is a root of the cubic

$$\frac{N^3}{135}f(y)= \frac{16}{135}N^3y^3+\frac{4}{5}N(A^2-BN)y^2+\frac{N}{3}(3B^2+ND-4AC)y+(NC^2-2ABC+B^3+A^2D-NBD)$$

which can be written in the form

$$f(y)=16y^3-108m_2y^2+45(m_4+3m_2^2)y-135(m_2m_4-m_3^2-m_2^3) \qquad (1.6)$$

where m_2, m_3 and m_4 are second, third and fourth central sample moments respectively,

$$m_2 = \frac{1}{N}\sum_{r=1}^{N}(x_r-\bar{x})^2 = \frac{B}{N} - (\frac{A}{N})^2$$

$$m_3 = \frac{1}{N}\sum_{r=1}^{N}(x_r-\bar{x})^3 = \frac{C}{N} - 3(\frac{A}{N})(\frac{B}{N}) + 2(\frac{A}{N})^3$$

$$m_4 = \frac{1}{N}\sum_{r=1}^{N}(x_r-\bar{x})^4 = \frac{D}{N} - 4(\frac{A}{N})(\frac{C}{N}) + 6(\frac{B}{N})(\frac{A}{N})^2 - 3(\frac{A}{N})^4$$

and where $\bar{x} = \frac{1}{N}\sum_{r=1}^{N}x_r$ is the sample mean.

It is obvious that the first three coefficients of the cubic f(y) are positive, negative and positive respectively. Since $m_2m_4-m_3^2-m_2^3 = m_2^3(\beta_2-\beta_1-1)$, where $\beta_2 = \frac{m_4}{m_2^2}$ is Pearson's measure of kurtosis and $\beta_1 = \frac{m_3^2}{m_2^2}$ is Pearson's measure of skewness, it follows that $m_2m_4 - m_3^2 - m_2^3 > 0$; for, by a well known inequality (Kendall and Stuart (1969), p92, Exercise 3.19), $\beta_2 > \beta_1 + 1$. Therefore, the coefficients of f(y) are alternatively positive and negative, and so, by

Descartes' rule of signs, f(y) cannot have any negative roots.

The derivative of (1.6) with respect to y is

$$f'(y) = 48y^2 - 216m_2y + 45(m_4 + 3m_2^2) \tag{1.7}$$

which takes its minimum value of $-108m_2^2 + 45m_4$ at $y = \frac{9}{4}m_2$. Hence, if $-108m_2^2 + 45m_4 > 0$ i.e. $12m_2^2 - 5m_4 < 0$, then $f'(y)$ is positive for all y and so f(y) can have at most one root. On the other hand, it is well known that every cubic with real coefficients has at least one real root. Thus it can be concluded that, if $12m_2^2 - 5m_4 < 0$, f(y) has exactly one positive root and so there will be a unique, positive solution for $\tilde{h}^2$ and hence for $\tilde{h} > 0$.

If one attempts to prove that there is also a unique, positive root when $12m_2^2 - 5m_4 > 0$ according to the method used by John (1970(a)) in the case of a mixture of normal samples, it is necessary to develop the following extra conditions. The first follows from the fact that $m_2 = \frac{B}{N} - (\frac{A}{N})^2$ and hence, using the first two equations of (1.3), one obtains

$$\begin{aligned} m_2 &= \frac{1}{N}(\tilde{N}_1\tilde{a}_1^2 + \tilde{N}_2\tilde{a}_2^2 + \tfrac{1}{3}N\tilde{h}^2) - \frac{1}{N^2}(\tilde{N}_1\tilde{a}_1 + \tilde{N}_2\tilde{a}_2)^2 \\ &= \frac{\tilde{h}^2}{3} + \frac{1}{N^2}(\tilde{N}_1\tilde{N}_2\tilde{a}_1^2 + \tilde{N}_1\tilde{N}_2\tilde{a}_2^2 - 2\tilde{N}_1\tilde{N}_2\tilde{a}_1\tilde{a}_2) \\ &= \frac{\tilde{h}^2}{3} + \frac{\tilde{N}_1\tilde{N}_2}{N^2}(\tilde{a}_1 - \tilde{a}_2)^2 \end{aligned}$$

Hence $3m_2 \geq \tilde{h}^2$. (1.8)

The second condition follows in a similar manner by developing an expression for m_4 using the equations (1.3):

$$m_4 = \frac{D}{N} - 4(\tfrac{A}{N})(\tfrac{C}{N}) + 6(\tfrac{B}{N})(\tfrac{A}{N})^2 - 3(\tfrac{A}{N})^4$$

$$= \frac{\tilde{N}_1}{N}\tilde{a}_1^4 + \frac{\tilde{N}_2}{N}\tilde{a}_2^4 + \frac{2\tilde{h}^2}{N}(B - \tfrac{1}{3}N\tilde{h}^2) + \tfrac{1}{5}\tilde{h}^4$$

$$-4(\tfrac{A}{N})(\frac{\tilde{N}_1}{N}\tilde{a}_1^3 + \frac{\tilde{N}_2}{N}\tilde{a}_2^3 + \frac{\tilde{h}^2}{N}A) + 6(\tfrac{B}{N})(\tfrac{A}{N})^2 - 3(\tfrac{A}{N})^4$$

i.e. $$m_4 - \frac{\tilde{h}^4}{5} = \frac{\tilde{N}_1}{N}\tilde{a}_1^4 + \frac{\tilde{N}_2}{N}\tilde{a}_2^4 + 2m_2\tilde{h}^2 - \tfrac{2}{3}\tilde{h}^4 - 4(\tfrac{A}{N})\frac{\tilde{N}_1\tilde{a}_1^3}{N} - 4(\tfrac{A}{N})\frac{\tilde{N}_2\tilde{a}_2^3}{N}$$

$$-2(\tfrac{A}{N})^2\tilde{h}^2 + 6(\tfrac{B}{N})(\tfrac{A}{N})^2 - 3(\tfrac{A}{N})^4$$

i.e. $$m_4 + \tfrac{7}{15}\tilde{h}^4 - 2m_2\tilde{h}^2 = \frac{\tilde{N}_1}{N}\tilde{a}_1^4 + \frac{\tilde{N}_2}{N}\tilde{a}_2^4 - \frac{4\tilde{N}_1}{N^2}\tilde{a}_1^3(\tilde{N}_1\tilde{a}_1 + \tilde{N}_2\tilde{a}_2)$$

$$- \frac{4\tilde{N}_2}{N^2}\tilde{a}_2^3(\tilde{N}_1\tilde{a}_1 + \tilde{N}_2\tilde{a}_2) - \frac{2\tilde{h}^2}{N^2}(\tilde{N}_1\tilde{a}_1 + \tilde{N}_2\tilde{a}_2)^2$$

$$+ \frac{6}{N^3}(\tilde{N}_1\tilde{a}_1 + \tilde{N}_2\tilde{a}_2)^2(\tilde{N}_1\tilde{a}_1^2 + \tilde{N}_2\tilde{a}_2^2 + \tfrac{1}{3}N\tilde{h}^2)$$

$$- \frac{3}{N^4}(\tilde{N}_1\tilde{a}_1 + \tilde{N}_2\tilde{a}_2)^4$$

After considerable algebraic manipulation, this reduces to

$$m_4 + \frac{7}{15}\tilde{h}^4 - 2m_2\tilde{h}^2 = (\tilde{a}_1 - \tilde{a}_2)^4\,\frac{\tilde{N}_1\tilde{N}_2}{N^5}\,(\tilde{N}_1^3 + \tilde{N}_2^3)$$

and hence $m_4 + \frac{7}{15}\tilde{h}^4 - 2m_2\tilde{h}^2 \geq 0$

i.e. $$15m_4 - 30m_2\tilde{h}^2 + 7\tilde{h}^4 \geq 0 \qquad (1.9)$$

This is our second required condition.

Unfortunately, the inequality (1.9) may not always be effective in placing a restriction on the value of $\tilde{h}^2$. In fact, an examination of the minimum value of $15m_4 - 30m_2\tilde{h}^2 + 7\tilde{h}^4$ shows it to be nonnegative for all real $\tilde{h}^2$ if $15m_2^2 \leq 7m_4$, i.e. if $\frac{15}{7}m_2^2 \leq m_4 < \frac{12}{5}m_2^2$, where we have included the original restriction $12m_2^2 > 5m_4$ under which we are still working. Hence, the inequality (1.9) only succeeds in placing a restriction on the value of $\tilde{h}^2$ if

$$15m_2^2 - 7m_4 > 0 \qquad (1.10)$$

and then it can be seen from the roots of

$15m_4 - 30m_2\tilde{h}^2 + 7\tilde{h}^4$ that this restriction is

$$\tilde{h}^2 \geq \frac{15}{7}[m_2 + (m_2^2 - \frac{7}{15}m_4)^{\frac{1}{2}}] \equiv \alpha \tag{1.11}$$

or $$\tilde{h}^2 \leq \frac{15}{7}[m_2 - (m_2^2 - \frac{7}{15}m_4)^{\frac{1}{2}}] \equiv \beta \tag{1.12}$$

Combining these last two inequalities with (1.8), it can be said that, provided (1.10) is true, $\tilde{h}^2$ satisfies

$$\left.\begin{array}{ll} \text{either } 3m_2 \geq h^2 \geq \alpha \\ \text{or} \qquad\qquad h^2 \leq \beta \end{array}\right\} \tag{1.13}$$

Now we might hope to show that, provided (1.10) is true, $f'(\tilde{h}^2)$ is single-signed under inequalities (1.13). But consider the first of the inequalities (1.13). It reduces to $\tilde{h}^2 \geq \alpha$ unless

$$3m_2 \geq \alpha = \frac{15}{7}[m_2 + (m_2^2 - \frac{7}{15}m_4)^{\frac{1}{2}}]$$

i.e. $$\frac{6}{7}m_2 \geq \frac{15}{7}(m_2^2 - \frac{7}{15}m_4)^{\frac{1}{2}}$$

i.e. $$36m_2^2 \geq 225(m_2^2 - \frac{7}{15}m_4)$$

i.e. $$9m_2^2 \leq 5m_4 \tag{1.14}$$

Now it is easily seen that the derivative $f'(\tilde{h}^2)$ changes sign at a value of $\tilde{h}^2 \geq \alpha$; for, one of the zeroes of $f'(\tilde{h}^2)$ occurs at $\frac{9}{4}m_2 + \frac{3}{2}(m_2^2 - \frac{5}{12}m_4)^{\frac{1}{2}}$ which is certainly greater than α. To show that this zero is less than $3m_2$ in the event that $3m_2 \geq \alpha$, i.e. in the event that (1.14) is satisfied, we take the difference and apply inequality (1.14):

$$\begin{aligned} & 3m_2 - \frac{9}{4}m_2 - \frac{3}{2}(m_2^2 - \frac{5}{12}m_4)^{\frac{1}{2}} \\ & \geq \frac{3}{4}m_2 - \frac{3}{2}[m_2^2 - (\frac{5}{12})(\frac{9}{5})m_2^2]^{\frac{1}{2}} \\ & = 0 \end{aligned}$$

Hence the derivative $f'(\tilde{h}^2)$ changes sign in the interval specified by the first inequality of (1.13), and so these conditions cannot be used to establish the existence of a unique, positive root for the cubic (1.6). Instead, we shall derive the required relationships between the sample moments for the existence of such a root using the following straightforward method. By an application of Descartes' rule of signs, $f'(y)$ as given by (1.7) has two positive roots, say $y_1 > y_o > 0$, given by $y_1 = \frac{9}{4} m_2 + \frac{3}{2}(m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}}$ and $y_o = \frac{9}{4} m_2 - \frac{3}{2}(m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}}$. Note that we are still confining our interest to the case $12m_2^2 - 5m_4 > 0$, and so both y_o and y_1 are real. Since $f(y)$ has a local maximum at y_o and a local minimum at y_1, it readily follows that there is a unique, positive solution of $f(y) = 0$ if and only if

$$\left.\begin{array}{ll} \text{either} & f(y_1) > 0 \\ \text{or} & f(y_o) < 0 \end{array}\right\} \tag{1.15}$$

Note that, because y_o and y_1 are the only turning points and hence $f(y_o) > f(y_1)$, it is impossible that both inequalities of (1.15) should be simultaneously satisfied.

To convert inequalities (1.15) to a more explicit form, let us evaluate $f(y_1)$ and $f(y_o)$.

$$\begin{aligned} f(y_1) &= 16[\tfrac{9}{4} m_2 + \tfrac{3}{2}(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}}]^3 - 108m_2[\tfrac{9}{4}m_2 + \tfrac{3}{2}(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}}]^2 \\ &\quad + 45(m_4+3m_2^2)[\tfrac{9}{4}m_2 + \tfrac{3}{2}(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}}] - 135(m_2m_4 - m_3^2 - m_2^3) \\ &= \tfrac{729}{4} m_2^3 + \tfrac{729}{2} m_2(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}} + 243m_2(m_2^2 - \tfrac{5}{12} m_4) \\ &\quad + 54(m_2^2 - \tfrac{5}{12}m_4)^{3/2} - \tfrac{2187}{4} m_2^3 - 729m_2^2(m_2^2 - \tfrac{5}{12} m_4)^{\frac{1}{2}} \\ &\quad - 243m_2(m_2^2 - \tfrac{5}{12}m_4) + \tfrac{405}{4}m_2m_4 + \tfrac{135}{2}m_4(m_2^2 - \tfrac{5}{12} m_4)^{\frac{1}{2}} \\ &\quad + \tfrac{1215}{4} m_2^3 + \tfrac{405}{2} m_2^2(m_2^2 - \tfrac{5}{12} m_4)^{\frac{1}{2}} - 135m_2m_4 + 135m_3^2 + 135m_2^3 \\ &= \tfrac{297}{4} m_2^3 - \tfrac{135}{4} m_2m_4 + 135m_3^2 + (m_2^2 - \tfrac{5}{12} m_4)^{\frac{1}{2}}(-108m_2^2 + 45m_4) \\ &= \tfrac{27}{4} \{m_2(11m_2^2 - 5m_4) + 20m_3^2 - 16(m_2^2 - \tfrac{5}{12} m_4)^{3/2}\} \end{aligned}$$

Hence $f(y_1) > 0$ if and only if

$$20m_3^2 > 16(m_2^2 - \frac{5}{12} m_4)^{3/2} - m_2(11m_2^2 - 5m_4) \qquad (1.16)$$

Similarly,

$$\begin{aligned} f(y_o) &= 16[\tfrac{9}{4}m_2 - \tfrac{3}{2}(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}}]^3 - 108m_2[\tfrac{9}{4}m_2 - \tfrac{3}{2}(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}}]^2 \\ &\quad +45(m_4+3m_2^2)[\tfrac{9}{4}m_2^2 - \tfrac{3}{2}(m_2^2 - \tfrac{5}{12}m_4)^{\frac{1}{2}}] - 135(m_2m_4 - m_3^2 - m_2^3) \\ &= \frac{729}{4} m_2^3 - \frac{729}{2} m_2^2(m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}} + 243m_2(m_2^2 - \frac{5}{12} m_4) \\ &\quad + 54(m_2^2 - \frac{5}{12} m_4)^{3/2} - \frac{2187}{4} m_2^3 - 729m_2^2(m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}} \\ &\quad -243m_2(m_2^2 - \frac{5}{12} m_4) + \frac{405}{4} m_2m_4 + \frac{135}{2} m_4(m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}} \\ &\quad + \frac{1215}{4} m_2^3 - \frac{405}{2} m_2^2(m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}} - 135m_2m_4 + 135m_3^2 + 135m_2^3 \\ &= \frac{297}{4} m_2^3 - \frac{135}{4} m_2m_4 + 135m_3^2 + (m_2^2 - \frac{5}{12} m_4)^{\frac{1}{2}}(-108m_2^2 + 45m_4) \\ &= \frac{27}{4}\{m_2(11m_2^2 - 5m_4) + 20m_3^2 - 16(m_2^2 - \frac{5}{12} m_4)^{3/2}\} \end{aligned}$$

Hence $f(y_o) < 0$ if and only if

$$20m_3^2 < -16(m_2^2 - \frac{5}{12} m_4)^{3/2} - m_2(11m_2^2 - 5m_4) \qquad (1.17)$$

In summary, it can be said that the cubic $f(y)$, as given by (1.6) has a single, positive root if and only if

$$m_2^2 - \frac{5}{12} m_4 < 0$$

or $m_2^2 - \frac{5}{12} m_4 > 0$ and either $20m_3^2 > 16(m_2^2 - \frac{5}{12}m_4)^{3/2} - m_2(11m_2^2 - 5m_4)$

or $20m_3^2 < -16(m_2^2 - \frac{5}{12}m_4)^{3/2} - m_2(11m_2^2 - 5m_4)$

(1.18)

If conditions (1.18) are satisfied, $\tilde{a}_1\tilde{a}_2$ can be calculated from the first equation of (1.5) and hence $(\tilde{a}_1+\tilde{a}_2)$ can be determined from the first equation of (1.4). Then $\tilde{a}_1$ and $\tilde{a}_2$ are given by $\frac{1}{2}\{(\tilde{a}_1+\tilde{a}_2)\mp[(\tilde{a}_1+\tilde{a}_2)^2-4\tilde{a}_1\tilde{a}_2]^{\frac{1}{2}}\}$, provided these are real numbers. [If $(\tilde{a}_1+\tilde{a}_2)^2-4\tilde{a}_1\tilde{a}_2 < 0$, the method of moments fails.] Once $\tilde{a}_1$ and $\tilde{a}_2$ are known, $\tilde{N}_1$ and $\tilde{N}_2$ can be determined from the first equation of (1.3) and the equation $\tilde{N}_1+\tilde{N}_2=N$. In general, these values of $\tilde{N}_1$ and $\tilde{N}_2$ will not be integral, and so we should take the closest integer values between zero and N as our estimates $\tilde{N}_1'$ and $\tilde{N}_2'$ of N_1 and N_2. Once again, given these estimates $\tilde{N}_1', \tilde{N}_2', \tilde{a}_1, \tilde{a}_2$ and $\tilde{h}$, a reasonable allocation procedure is to assign the smallest $\tilde{N}_1'$ observations to the first population $U[\tilde{a}_1-\tilde{h}, \tilde{a}_1+\tilde{h}]$ and the remaining $\tilde{N}_2'$ observations to the second population $U[\tilde{a}_2-\tilde{h}, \tilde{a}_2+\tilde{h}]$.

We conclude this section by carrying out what should be a good check of our algebraic accuracy in the derivation of the cubic (1.6). If we have formulated (1.6) correctly, it should follow that $f(h^2)$ takes the value zero if we replace each of the coefficients by its expected value. Equivalently, we may consider the cubic given by

$$g(y) = 16N^3y^3+108N(A^2-BN)y^2+45N(3B^2+ND-4AC)y$$
$$+135(NC^2-2ABC+B^3+A^2D-NBD)$$

Again, we expect this cubic to vanish if we put $y=h^2$ and replace A,B,C and D by their expected values. The latter can be obtained from equations (1.3) if we simply discard all the tildes. It follows that $E\{g(h^2)\}$ is equal to

$$16N^3h^6+108Nh^4\{(N_1a_1+N_2a_2)^2-N(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)\}$$
$$+45Nh^2\left\{\begin{array}{l}3(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)^2+N[N_1a_1^4+N_2a_2^4+2h^2(N_1a_1^2+N_2a_2^2)+\tfrac{1}{5}Nh^4]\\ -4(N_1a_1+N_2a_2)[N_1a_1^3+N_2a_2^3+h^2(N_1a_1+N_2a_2)]\end{array}\right\}$$
$$+135\left\{\begin{array}{l}N[N_1a_1^3+N_2a_2^3+h^2(N_1a_1+N_2a_2)]^2-2(N_1a_1+N_2a_2)(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)\\ \quad[N_1a_1^3+N_2a_2^3+h^2(N_1a_1+N_2a_2)]+(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)^3\\ +(N_1a_1+N_2a_2)^2[N_1a_1^4+N_2a_2^4+2h^2(N_1a_1^2+N_2a_2^2)+\tfrac{1}{5}Nh^4]\\ -N(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)[N_1a_1^4+N_2a_2^4+2h^2(N_1a_1^2+N_2a_2^2)+\tfrac{1}{5}Nh^4]\end{array}\right\}$$
$$(1.19)$$

The term containing N^3h^6 in $E\{g(h^2)\}$ as given by (1.19) is $N^3h^6(16-36+15+9+5-9) = 0$. Similarly, if one goes through the tedious procedure of expanding all the expressions in (1.19), it is easily seen that the terms containing Nh^4, h^2 and the terms independent of h all vanish. Hence $E\{g(h^2)\}=0$ as required. Although this does not prove that the cubic (1.6) is correct, it does show that it satisfies quite a rigorous check.

So we have derived the moment equations, checked our results, and displayed the conditions (1.18) necessary for a unique positive value of $\tilde{h}^2$. It is a reasonable conjecture that (1.18) is always satisfied asymptotically, but we shall return to this question in §4.3 after having derived the asymptotic sampling distribution of the moment estimators. Firstly, however, let us examine the method of maximum likelihood under a mixed samples model.

4.2 ANALYSIS USING THE MAXIMUM LIKELIHOOD METHOD

Consider a sample of N observations $x_1, x_2, \ldots, x_N$ which is assumed to be a mixture of a sample of size N_1 from the first population $U[a_1-h, a_1+h]$ and a sample of size N_2 from the second population $U[a_2-h, a_2+h]$. Let ω_{ij} take the value one or zero according as the i^{th} observation does or does not belong to the j^{th} population for each $i=1,2,\ldots,n$ and $j=1,2$. Let $u(.)$ be that function which takes the value one if the argument is nonpositive and zero otherwise. Then the joint likelihood of $x_1, x_2, \ldots, x_N$ is given by

$$L(x_1,x_2,\ldots,x_N;\ a_1,a_2,h) = \frac{1}{(2h)^N} \prod_{i=1}^{N} \left. \begin{array}{l} \omega_{i1}u(x_i-a_1-h)u(a_1-h-x_i) \\ \quad + \omega_{i2}u(x_i-a_2-h)u(a_2-h-x_i) \end{array} \right\} (2\cdot 1)$$

Without loss of generality, we can assume $a_1 < a_2$. If a_1, a_2 and h are given, then the maximum likelihood allocation is arbitrary for those observations (if any) which fall in $[a_1-h, a_1+h] \cap [a_2-h, a_2+h]$ and we allocate those observations in $[a_1-h, a_1+h] \setminus [a_2-h, a_2+h]$ to the first population and those in $[a_2-h, a_2+h] \setminus [a_1-h, a_1+h]$ to the second population.

On the other hand, if the allocation is given, the m.ℓ.e. of h is half the maximum of the ranges of the two samples and so the m.ℓ.e. of "a" for whichever population has the largest range is given by the mid-point of this range, the so-called mid-range. There is no m.ℓ.e. for the parameter "a" of that population with the smaller sample range; however, if ξ_1 is the smallest observation of this sample and ξ_2 the largest, then we are able to bound the value of "a" for this population according to

$$\xi_2 - h \leq a \leq \xi_1 + h \tag{2.2}$$

Consideration of the above arguments and the form of the likelihood (2.1) shows that the m.ℓ.e. of h has the following simple form. Let the ordered set of observations be $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$. Assume that $N \geq 3$. Then there exists $m \varepsilon \{2,3,\ldots,N-1\}$ such that

$$x_{(m)} \leq \frac{x_{(1)} + x_{(N)}}{2} \leq x_{(m+1)} \tag{2.3}$$

and it follows that the m.ℓ.e. of h is given by

$$\hat{h} = \frac{1}{2} \max \{x_{(m)} - x_{(1)},\ x_{(N)} - x_{(m+1)}\} \tag{2.4}$$

since this is the smallest possible estimate of h which gives a nonzero value to the likelihood (2.1). The maximum likelihood allocation is $x_{(1)}, x_{(2)}, \ldots, x_{(m)}$ to the first population and $x_{(m+1)}, x_{(m+2)}, \ldots, x_{(N)}$ to the second. Whichever population has the largest sample range (i.e. sample range equal to $2\hat{h}$) under this allocation, will also have its sample mid-range as an m.ℓ.e. for its location parameter. Bounds for the location parameter of the other population are given by (2.2).

Though the method suffers from the weakness that not all the parameters are estimable, it does have the virtue that an estimate of h can be very readily obtained. This could be of use in the initiation of iterative processes, if one wished to obtain m.ℓ.e.'s of a_1, a_2, h and the mixing parameter α under the assumption that the observations form a single sample from a compound uniform distribution.

4.3 THE ASYMPTOTIC DISTRIBUTION OF THE MOMENT ESTIMATORS

This section is devoted to the presentation of the asymptotic distribution of the moment estimators derived in §4.1. We note that the moment estimators are functions of the four sample sums $\sum_{r=1}^{N} x_r^t$, $t=1,2,3,4$. Equivalently, they can be regarded as functions of Z_t, $t=1,2,3,4$, where $Z_t = N^{-\frac{1}{2}}(\sum_{r=1}^{N} x_r^t - E\sum_{r=1}^{N} x_r^t)$. If N is large, these functions are approximately equal to their respective Taylor expansions in powers of Z_1, Z_2, Z_3 and Z_4 with terms of order greater than one omitted. An application of the central limit theorem shows that asymptotically Z_1, Z_2, Z_3 and Z_4 are jointly normally distributed. Hence the moment estimators are asymptotically distributed as linear functions of jointly normally distributed random variables. It follows that the asymptotic distribution of the moment estimators is normal.

It is easily verified that equations (1.2) remain true after replacing every x_r by $x_r - c$, where c is an arbitrary constant, if simultaneously $\tilde{a}_i$ is replaced by $\tilde{a}_i - c$, $i=1,2$. Therefore, in calculating the standard errors of $\tilde{h}^2, \tilde{a}_1, \tilde{a}_2$, $\frac{\tilde{N}_1}{N}$ and $\frac{\tilde{N}_2}{N}$ we may without loss of generality make the simplifying assumption

$$N_1 a_1 + N_2 a_2 = 0 \tag{3.1}$$

It follows that

$$\begin{aligned}
A &= N^{\frac{1}{2}} Z_1 \\
B &= N^{\frac{1}{2}} Z_2 + N_1 a_1^2 + N_2 a_2^2 + \tfrac{1}{3} N h^2 \\
C &= N^{\frac{1}{2}} Z_3 + N_1 a_1^3 + N_2 a_2^3 \\
D &= N^{\frac{1}{2}} Z_4 + N_1 a_1^2 (a_1^2 + 2h^2) + N_2 a_2^2 (a_2^2 + 2h^2) + \tfrac{1}{5} N h^4
\end{aligned} \tag{3.2}$$

From §4.1, $\tilde{h}^2$ is a root of the cubic

$$g(y) = 16N^3y^3+108N(A^2-BN)y^2+45N(3B^2+ND-4AC)y$$
$$+135(NC^2-2ABC+B^3+A^2D-NBD) \qquad (3.3)$$

Let us now use the relations (3.2) to evaluate each of the coefficients of the cubic (3.3) in turn, retaining only terms of order at most one in Z_t, $t=1,2,3,4$.

$$A^2-BN \simeq -N^{3/2}Z_2-NN_1a_1^2-NN_2a_2^2 - \tfrac{1}{3}N^2h^2$$
$$= -N^{3/2}Z_2-N_1N_2(a_1^2+a_2^2)-N_1^2a_1^2-N_2^2a_2^2- \tfrac{1}{3}N^2h^2$$
$$= -N^{3/2}Z_2-N_1N_2(a_1^2+a_2^2)-(N_1a_1+N_2a_2)^2+2N_1N_2a_1a_2- \tfrac{1}{3}N^2h^2$$
$$= -N^{3/2}Z_2-N_1N_2(a_1-a_2)^2- \tfrac{1}{3}N^2h^2 \qquad (3.4)$$

To obtain the last equality we have used (3.1).

Also, $$3B^2+ND-4AC \simeq 2N^{1/2}Z_2(3N_1a_1^2+3N_2a_2^2+Nh^2)+3(N_1a_1^2+N_2a_2^2+ \tfrac{1}{3}Nh^2)^2$$
$$+N^{3/2}Z_4+NN_1a_1^2(a_1^2+2h^2)+NN_2a_2^2(a_2^2+2h^2)$$
$$+ \tfrac{1}{5}N^2h^4-4N^{1/2}Z_1(N_1a_1^3+N_2a_2^3)$$

The terms independent of Z_t $(t=1,2,3,4)$ in the above are

$$3N_1^2a_1^4+3N_2^2a_2^4+ \tfrac{1}{3}N^2h^4+6N_1N_2a_1^2a_2^2+2NN_1a_1^2h^2+2NN_2a_2^2h^2$$
$$+NN_1a_1^4+2NN_1a_1^2h^2+NN_2a_2^4+2NN_2a_2^2h^2+ \tfrac{1}{5}N^2h^4$$
$$= 4N_1^2a_1^4+N_1N_2a_1^4+4N_2^2a_2^4+N_1N_2a_2^4+4NN_1a_1^2h^2+4NN_2a_2^2h^2$$
$$+6N_1N_2a_1^2a_2^2+ \tfrac{8}{15}N^2h^4$$

Now using $N_1a_1 = -N_2a_2$ to obtain a common factor of N_1N_2 in as many terms as possible, these constant terms become

$$-4N_1N_2a_1^3a_2+N_1N_2a_1^4-4N_1N_2a_1a_2^3+N_1N_2a_2^4+4N_1N_2a_1^2h^2-4N_1N_2a_1a_2h^2$$
$$+4N_1N_2a_2^2h^2-4N_1N_2a_1a_2h^2+6N_1N_2a_1^2a_2^2+ \tfrac{8}{15}N^2h^4$$
$$= N_1N_2(a_1-a_2)^4+4N_1N_2h^2(a_1^2+a_2^2-2a_1a_2)+ \tfrac{8}{15}N^2h^4$$
$$= N_1N_2(a_1-a_2)^4+4N_1N_2(a_1-a_2)^2h^2+ \tfrac{8}{15}N^2h^4$$

Hence $3B^2+ND-4AC \simeq -4N^{\frac{1}{2}}Z_1(N_1a_1^3+N_2a_2^3)+2N^{\frac{1}{2}}Z_2(3N_1a_1^2+3N_2a_2^2+Nh^2)$

$$+N^{3/2}Z_4+N_1N_2(a_1-a_2)^4+4N_1N_2(a_1-a_2)^2h^2+\frac{8}{15}N^2h^4 \qquad (3.5)$$

Finally, the last coefficient of the cubic (3.3) which needs to be considered is

$$NC^2-2ABC+B^3+A^2D-NBD$$

$$\begin{aligned}
&\simeq 2N^{3/2}Z_3(N_1a_1^3+N_2a_2^3)+N(N_1a_1^3+N_2a_2^3)^2-2N^{\frac{1}{2}}Z_1(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)\\
&\quad\times(N_1a_1^3+N_2a_2^3)+3N^{\frac{1}{2}}Z_2(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)^2+(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)^3\\
&\quad-N^{3/2}Z_2[N_1a_1^2(a_1^2+2h^2)+N_2a_2^2(a_2^2+2h^2)+\tfrac{1}{5}Nh^4]-N^{3/2}Z_4(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)\\
&\quad-N(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)[N_1a_1^2(a_1^2+2h^2)+N_2a_2^2(a_2^2+2h^2)+\tfrac{1}{5}Nh^4] \qquad (3.6)
\end{aligned}$$

We shall use (3.1) to simplify these terms wherever possible. The term containing Z_1 in (3.6) is given by

$$\begin{aligned}
&-2N^{\frac{1}{2}}Z_1\{N_1^2a_1^5+N_2^2a_2^5+N_1N_2a_1^2a_2^3+N_1N_2a_1^3a_2^2+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^3)\}\\
&= -2N^{\frac{1}{2}}Z_1\{-N_1N_2a_1^4a_2-N_1N_2a_1a_2^4+N_1N_2a_1^2a_2^3+N_1N_2a_1^3a_2^2+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^3)\}\\
&= -2N^{\frac{1}{2}}Z_1\{-N_1N_2a_1a_2(a_1^3-a_1a_2^2-a_1^2a_2+a_2^3)+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^3)\}\\
&= -2N^{\frac{1}{2}}Z_1\{-N_1N_2a_1a_2(a_1^2-a_2^2)(a_1-a_2)+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^3)\}\\
&= -2N^{\frac{1}{2}}Z_1\{-N_1N_2a_1a_2(a_1-a_2)^2(a_1+a_2)+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^3)\}
\end{aligned}$$

The term containing Z_2 in (3.6) is given by

$$\begin{aligned}
&N^{\frac{1}{2}}Z_2\{3(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)^2-N[N_1a_1^2(a_1^2+2h^2)+N_2a_2^2(a_2^2+2h)^2+\tfrac{1}{5}Nh^4]\}\\
&= N^{\frac{1}{2}}Z_2\left\{\begin{matrix}3N_1^2a_1^4+3N_2^2a_2^4+\tfrac{1}{3}N^2h^4+6N_1N_2a_1^2a_2^2+2NN_1a_1^2h^2+2NN_2a_2^2h^2\\ -NN_1a_1^4-2NN_1a_1^2h^2-NN_2a_2^4-2NN_2a_2^2h^2-\tfrac{1}{5}N^2h^4\end{matrix}\right\}\\
&= N^{\frac{1}{2}}Z_2(2N_1^2a_1^4-N_1N_2a_1^4+2N_2^2a_2^4-N_1N_2a_2^4+\tfrac{2}{15}N^2h^4+6N_1N_2a_1^2a_2^2)\\
&= N^{\frac{1}{2}}Z_2(-2N_1N_2a_1^3a_2-N_1N_2a_1^4-2N_1N_2a_1a_2^3-N_1N_2a_2^4+6N_1N_2a_1^2a_2^2+\tfrac{2}{15}N^2h^4)\\
&= N^{\frac{1}{2}}Z_2[-N_1N_2(a_1^4+a_2^4+2a_1^3a_2+2a_1a_2^3-6a_1^2a_2^2)+\tfrac{2}{15}N^2h^4]\\
&= N^{\frac{1}{2}}Z_2\{-N_1N_2[a_1^2(a_1^2-a_2^2)+a_2^2(a_2^2-a_1^2)+2a_1a_2(a_1^2+a_2^2-2a_1a_2)]+\tfrac{2}{15}N^2h^4\}
\end{aligned}$$

$$= N^{\frac{1}{2}}Z_2\{-N_1N_2[(a_1^2-a_2^2)(a_1^2-a_2^2)+2a_1a_2(a_1-a_2)^2]+\frac{2}{15}N^2h^4\}$$

$$= N^{\frac{1}{2}}Z_2\{-N_1N_2(a_1-a_2)^2[(a_1+a_2)^2+2a_1a_2]+\frac{2}{15}N^2h^4\}$$

Also, the term independent of $Z_t(t=1,2,3,4)$ in (3.6) is given by

$$N(N_1a_1^3+N_2a_2^3)^2+(N_1a_1^2+N_2a_2^2+\frac{1}{3}Nh^2)^3-N(N_1a_1^2+N_2a_2^2+\frac{1}{3}Nh^2)[N_1a_1^2(a_1^2+2h^2)+N_2a_2^2(a_2^2+h^2)+\frac{1}{5}Nh^4]$$

$$= NN_1^2a_1^6+NN_2^2a_2^6+2NN_1N_2a_1^3a_2^3+N_1^3a_1^6+N_2^3a_2^6+\frac{1}{27}N^3h^6+3N_1^2N_2a_1^4a_2^2$$
$$+3N_1N_2^2a_1^2a_2^4+Nh^2(N_1^2a_1^4+N_2^2a_2^4+2N_1N_2a_1^2a_2^2)+\frac{1}{3}N^2h^4(N_1a_1^2+N_2a_2^2)$$
$$-NN_1^2a_1^6-2NN_1^2a_1^4h^2-NN_1N_2a_1^2a_2^4-2NN_1N_2a_1^2a_2^2h^2-\frac{1}{5}N^2N_1a_1^2h^4$$
$$-NN_1N_2a_1^4a_2^2-2NN_1N_2a_1^2a_2^2h^2-NN_2^2a_2^6-2NN_2^2a_2^4h^2-\frac{1}{5}N^2N_2a_2^2h^4$$
$$-\frac{1}{3}N^2N_1a_1^4h^2-\frac{2}{3}N^2N_1a_1^2h^4-\frac{1}{3}N^2N_2a_2^4h^2-\frac{2}{3}N^2N_2a_2^2h^4-\frac{1}{15}N^3h^6$$

$$= N_1^3a_1^6+N_2^3a_2^6+2NN_1N_2a_1^3a_2^3+3N_1^2N_2a_1^4a_2^2-NN_1N_2a_1^4a_2^2+3N_1N_2^2a_1^2a_2^4$$
$$-NN_1N_2a_1^2a_2^4+Nh^2(N_1^2a_1^4+N_2^2a_2^4-2N_1^2a_1^4-2N_1N_2a_1^2a_2^2-2N_2^2a_2^4-\frac{1}{3}NN_1a_1^4-\frac{1}{3}NN_2a_2^4)$$
$$+N^2h^4(\frac{1}{3}N_1a_1^2+\frac{1}{3}N_2a_2^2-\frac{1}{5}N_1a_1^2-\frac{1}{5}N_2a_2^2-\frac{2}{3}N_1a_1^2-\frac{2}{3}N_2a_2^2)-\frac{4}{135}N^3h^6$$

$$= N_1N_2^2a_1^4a_2^2+N_1^2N_2a_1^2a_2^4+2NN_1N_2a_1^3a_2^3+2N_1^2N_2a_1^4a_2^2-N_1N_2^2a_1^4a_2^2$$
$$+2N_1N_2^2a_1^2a_2^4-N_1^2N_2a_1^2a_2^4+Nh^2(-N_1^2a_1^4-\frac{1}{3}NN_1a_1^4-N_2^2a_2^4-\frac{1}{3}NN_2a_2^4-2N_1N_2a_1^2a_2^2)$$
$$+N^2h^4(-\frac{8}{15}N_1a_1^2-\frac{8}{15}N_2a_2^2)-\frac{4}{135}N^3h^6$$

$$= 2(N_1^2N_2+N_1N_2^2)a_1^3a_2^3-2N_1N_2^2a_1^3a_2^3-2N_1^2N_2a_1^3a_2^3$$
$$+Nh^2(\frac{4}{3}N_1N_2a_1^3a_2-\frac{1}{3}N_1N_2a_1^4+\frac{4}{3}N_1N_2a_1a_2^3-\frac{1}{3}N_1N_2a_2^4-2N_1N_2a_1^2a_2^2)$$
$$-\frac{8}{15}N^2h^4(N_1a_1^2+N_2a_2^2)-\frac{4}{135}N^3h^6$$

$$=-\frac{1}{3}Nh^2N_1N_2(a_1-a_2)^4-\frac{8}{15}N^2h^4(N_1a_1^2+N_2a_2^2)-\frac{4}{135}N^3h^6$$

$$=-\frac{1}{3}Nh^2N_1N_2(a_1-a_2)^4-\frac{8}{15}NN_1N_2h^4(a_1-a_2)^2-\frac{4}{135}N^3h^6$$

To establish the last equality it should be noted that

$$N^2(N_1a_1^2+N_2a_2^2) = N(N_1N_2a_1^2+N_1^2a_1^2+N_1N_2a_2^2+N_2^2a_2^2)$$
$$= N(N_1N_2a_1^2-N_1N_2a_1a_2+N_1N_2a_2^2-N_1N_2a_1a_2)$$
$$= NN_1N_2(a_1-a_2)^2$$

Now we are in a position to write equation (3.6) as

$$NC^2 - 2ABC + B^3 + A^2D - NBD$$
$$\simeq -2N^{\frac{1}{2}}Z_1\{-N_1N_2a_1a_2(a_1-a_2)^2(a_1+a_2)+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^3)\}$$
$$+N^{\frac{1}{2}}Z_2\{-N_1N_2(a_1-a_2)^2[(a_1+a_2)^2+2a_1a_2]+\tfrac{2}{15}N^2h^4\}$$
$$+2N^{3/2}Z_3(N_1a_1^3+N_2a_2^3)-N^{3/2}Z_4(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)$$
$$-\tfrac{1}{3}Nh^2N_1N_2(a_1-a_2)^4 - \tfrac{8}{15}NN_1N_2h^4(a_1-a_2)^2-\tfrac{4}{135}N^3h^6 \qquad (3.7)$$

Thus all the coefficients of the cubic (3.3) have been written as linear functions of Z_1,Z_2,Z_3 and Z_4. Let us assume that, asymptotically, conditions (1.18) for the existence of a unique, positive estimator $\tilde{h}^2$ are satisfied with probability one. From the asymptotic sampling theory of such an estimator we know that, for large sample size,

$$\tilde{h}^2 = h^2 \text{ some linear function of } Z_1,Z_2,Z_3 \text{ and } Z_4$$
$$\simeq h^2 + \Delta\tilde{h}^2 \qquad (3.8)$$

So if we substitute $y = h^2+\Delta\tilde{h}^2$ into the cubic (3.3) and ignore powers of $\Delta\tilde{h}^2$ greater than one, we shall be able to solve for $\Delta\tilde{h}^2$ as a linear function of Z_1,Z_2,Z_3 and Z_4:

$$16N^3(h^2+\Delta\tilde{h}^2)^3+108N(A^2-BN)(h^2+\Delta\tilde{h}^2)^2+45N(3B^2+ND-4AC)(h^2+\Delta\tilde{h}^2)$$
$$+135(NC^2-2ABC+B^3+A^2D-NBD) = 0$$

i.e. $$16N^3(h^6+3h^4\Delta\tilde{h}^2)+108N(A^2-BN)(h^4+2h^2\Delta\tilde{h}^2)+45N(3B^2+ND-4AC)(h^2+\Delta\tilde{h}^2)$$
$$+135(NC^2-2ABC+B^3+A^2D-NBD) = 0$$

i.e. $$\Delta\tilde{h}^2\{48N^3h^4+216Nh^2(A^2-BN)+45N(3B^2+ND-4AC)\}$$
$$= -\{16N^3h^6+108Nh^4(A^2-BN)+45Nh^2(3B^2+ND-4AC)$$
$$+135(NC^2-2ABC+B^3+A^2D-NBD)\}$$

$$\text{i.e. } \Delta\tilde{h}^2 = \frac{-\{16N^3h^6+108Nh^4(A^2-BN)+45Nh^2(3B^2+ND-4AC) + 135(NC^2-2ABC+B^3+A^2D-NBD)\}}{\{48N^3h^4+216Nh^2(A^2-BN)+45N(3B^2+ND-4AC)\}} \tag{3.9}$$

Substituting the expressions (3.4), (3.5) and (3.7) into (3.9), it follows that the numerator of (3.9) can be written as

$$\begin{aligned}
&16N^3h^6+108Nh^4\left\{-N^{3/2}Z_2-N_1N_2(a_1-a_2)^2-\tfrac{1}{3}N^2h^2\right\} \\
&+45Nh^2\left\{\begin{aligned}&-4N^{\frac12}Z_1(N_1a_1^3+N_2a_2^3)+2N^{\frac12}Z_2(3N_1a_1^2+3N_2a_2^2+Nh^2)+N^{3/2}Z_4 \\ &+N_1N_2(a_1-a_2)^4+4N_1N_2(a_1-a_2)^2h^2+\tfrac{8}{15}N^2h^4\end{aligned}\right\} \\
&+135\left\{\begin{aligned}&-2N^{\frac12}Z_1[-N_1N_2a_1a_2(a_1-a_2)^2(a_1+a_2)+\tfrac{1}{3}Nh^2(N_1a_1^3+N_2a_2^2)] \\ &+N^{\frac12}Z_2[-N_1N_2(a_1-a_2)^2((a_1+a_2)^2+2a_1a_2)+\tfrac{2}{15}N^2h^4] \\ &+2N^{3/2}Z_3(N_1a_1^3+N_2a_2^3)-N^{3/2}Z_4(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2) \\ &-\tfrac{1}{3}Nh^2N_1N_2(a_1-a_2)^4-\tfrac{8}{15}NN_1N_2h^4(a_1-a_2)^2-\tfrac{4}{135}N^3h^6\end{aligned}\right\}
\end{aligned} \tag{3.10}$$

We expect the term independent of Z_1, Z_2, Z_3 and Z_4 in (3.10) to be zero. In fact, it is

$$\begin{aligned}
&16N^3h^6-108NN_1N_2h^4(a_1-a_2)^2-36N^3h^6+45NN_1N_2h^2(a_1-a_2)^4 \\
&+180NN_1N_2h^4(a_1-a_2)^2+24N^3h^6-45NN_1N_2h^2(a_1-a_2)^4-72NN_1N_2h^4(a_1-a_2) \\
&-4N^3h^6
\end{aligned}$$

which is easily seen to vanish, as required. Hence the numerator (3.10) can be written as

$$\begin{aligned}
&N^{\frac12}Z_1\{-180Nh^2(N_1a_1^3+N_2a_2^3)+270N_1N_2a_1a_2(a_1-a_2)^2(a_1+a_2) \\
&-90Nh^2(N_1a_1^3+N_2a_2^3)\}+N^{\frac12}Z_2\{-108N^2h^4+90Nh^2(3N_1a_1^2+3N_2a_2^2+Nh^2) \\
&+135[-N_1N_2(a_1-a_2)^2((a_1+a_2)^2+2a_1a_2)+\tfrac{2}{15}N^2h^4]\} \\
&+270N^{3/2}Z_3(N_1a_1^3+N_2a_2^3)+N^{3/2}Z_4\{45Nh^2-135(N_1a_1^2+N_2a_2^2+\tfrac{1}{3}Nh^2)\} \\
&= 270N^{\frac12}Z_1\{N_1N_2a_1a_2(a_1-a_2)^2(a_1+a_2)-Nh^2(N_1a_1^3+N_2a_2^3)\} \\
&\quad +135N^{\frac12}Z_2\{2Nh^2(N_1a_1^2+N_2a_2^2)-N_1N_2(a_1-a_2)^2[(a_1+a_2)^2+2a_1a_2]\}
\end{aligned}$$

$$+270N^{3/2}Z_3(N_1a_1^3+N_2a_2^3)-135N^{3/2}Z_4(N_1a_1^2+N_2a_2^2) \qquad (3.11)$$

Since the numerator of (3.9) as given by (3.11) contains no constant term, we need only evaluate the term independent of Z_1, Z_2, Z_3 and Z_4 in the denominator of (3.9). It is given by

$$-\{48N^3h^6+216Nh^2[-N_1N_2(a_1-a_2)^2-\tfrac{1}{3}N^2h^2]+45N[N_1N_2(a_1-a_2)^4$$
$$+4N_1N_2(a_1-a_2)^2h^2+\frac{8}{15}N^2h^4]\}$$

$$=36NN_1N_2h^2(a_1-a_2)^2-45NN_1N_2(a_1-a_2)^4$$
$$=9NN_1N_2(a_1-a_2)^2[4h^2-5(a_1-a_2)^2] \qquad (3.12)$$

We have previously noted that $N(N_1a_1^2+N_2a_2^2)=N_1N_2(a_1-a_2)^2$. It also follows that

$$N_1a_1^3+N_2a_2^3 = N_1\left\{a_1-\frac{N_1a_1+N_2a_2}{N}\right\}^3+N_2\left\{a_2-\frac{N_1a_1+N_2a_2}{N}\right\}^3$$
$$=\frac{N_1}{N^3}(N_2a_1-N_2a_2)^3+\frac{N_2}{N^3}(N_1a_2-N_1a_1)^3$$
$$=\frac{N_1N_2}{N^3}(N_2^2-N_1^2)(a_1-a_2)^3$$
$$=\frac{N_1N_2}{N^2}(N_2-N_1)(a_1-a_2)^3 \qquad (3.13)$$

Using these results, it is now possible to write down the coefficients β_i of Z_i $(i=1,2,3,4)$ in the expansion of $\Delta\tilde{h}^2$. Indeed, dividing (3.11) by (3.12), it follows that

$$\beta_1=\frac{270N^{\frac{1}{2}}\{N_1N_2a_1a_2(a_1-a_2)^2(a_1+a_2)-N^{-1}h^2N_1N_2(N_2-N_1)(a_1-a_2)^3\}}{9NN_1N_2(a_1-a_2)^2[4h^2-5(a_1-a_2)^2]}$$

$$=\frac{30N^{-3/2}\{Na_1a_2(a_1+a_2)-(N_2-N_1)(a_1-a_2)h^2\}}{4h^2-5(a_1-a_2)^2}$$

Now $Na_1a_2(a_1+a_2)=N_1a_1^2a_2+N_2a_1^2a_2+N_1a_1a_2^2+N_2a_1a_2^2$

$$=-N_2a_1a_2^2-N_1a_1^3-N_2a_2^3+N_2a_1a_2^2$$
$$=-(N_1a_1^3+N_2a_2^3)$$

Hence we can use (3.13) to write

$$\beta_1 = \frac{30N^{-5/2}(N_1-N_2)(a_1-a_2)\{N_1N_2(a_1-a_2)^2+h^2\}}{4h^2 - 5(a_1-a_2)^2} \qquad (3.14)$$

Similarly, the coefficient of Z_2 is given by

$$\beta_2 = \frac{135N^{\frac{1}{2}}\{2h^2N_1N_2(a_1-a_2)^2-N_1N_2(a_1-a_2)^2[(a_1+a_2)^2+2a_1a_2]\}}{9NN_1N_2(a_1-a_2)^2[4h^2-5(a_1-a_2)^2]}$$

$$= \frac{15N^{-\frac{1}{2}}\{2h^2-(a_1-a_2)^2-6a_1a_2\}}{4h^2-5(a_1-a_2)^2}$$

$$\text{Now } N^2a_1a_2 = (N_1^2+2N_1N_2+N_2^2)a_1a_2$$

$$= -N_1N_2a_2^2+2N_1N_2a_1a_2-N_1N_2a_1^2$$

$$= -N_1N_2(a_1-a_2)^2$$

$$\text{Hence } \beta_2 = \frac{15N^{-5/2}\{2N^2h^2-(N^2-6N_1N_2)(a_1-a_2)^2\}}{4h^2-5(a_1-a_2)^2} \qquad (3.15)$$

$$\text{Similarly, } \beta_3 = \frac{270N^{\frac{1}{2}}N_1N_2(a_1-a_2)^2(a_1+a_2)}{9NN_1N_2(a_1-a_2)^2[4h^2-5(a_1-a_2)^2]}$$

$$= \frac{30N^{-\frac{1}{2}}(a_1+a_2)}{4h^2-5(a_1-a_2)^2}$$

and since we can write $N(a_1+a_2) = N_1a_1+N_2a_1+N_1a_2+N_2a_2 = -N_2a_2+N_2a_1+N_1a_2-N_1a_1 = (N_2-N_1)(a_1-a_2)$ it follows that

$$\beta_3 = \frac{30N^{-3/2}(N_2-N_1)(a_1-a_2)}{4h^2-5(a_1-a_2)^2} \qquad (3.16)$$

$$\text{Finally, } \beta_4 = \frac{-135N^{\frac{1}{2}}N_1N_2(a_1-a_2)^2}{9NN_1N_2(a_1-a_2)^2[4h^2-5(a_1-a_2)^2]}$$

$$= \frac{-15N^{-\frac{1}{2}}}{4h^2-5(a_1-a_2)^2} \qquad (3.17)$$

The asymptotic distribution of $\tilde{h}=h+\Delta\tilde{h}$ can be obtained in the following way. Since $\Delta\tilde{h}^2 = 2\tilde{h}\Delta\tilde{h} = 2(h+\Delta\tilde{h})\Delta\tilde{h} \simeq 2h\Delta\tilde{h}$, it follows that the coefficient of Z_i in the expansion of $\Delta\tilde{h}$ is given by β_i divided by $2h$ for each $i=1,2,3,4$.

Now let us write

$$\tilde{a}_1 = a_1 + \Delta\tilde{a}_1 = a_1 + \sum_{i=1}^{4} \gamma_i Z_i$$

$$\tilde{a}_2 = a_2 + \Delta\tilde{a}_2 = a_2 + \sum_{i=1}^{4} \delta_i Z_i$$

$$\frac{\tilde{N}_1}{N} = \frac{N_1}{N} + \frac{\Delta\tilde{N}_1}{N} = \frac{N_1}{N} + \sum_{i=1}^{4} \varepsilon_i Z_i$$

$$\frac{\tilde{N}_2}{N} = \frac{N_2}{N} + \frac{\Delta\tilde{N}_2}{N} = \frac{N-\tilde{N}_1}{N} = \frac{N_2}{N} - \sum_{i=1}^{4} \varepsilon_i Z_i$$

We firstly wish to determine $\Delta\tilde{a}_1 = \sum_{i=1}^{4} \gamma_i Z_i$ from the known value of $\Delta\tilde{h}^2 = \sum_{i=1}^{4} \beta_i Z_i$. Consider the first equation of (1.4),

$$A(\tilde{a}_1+\tilde{a}_2) = B - \tfrac{1}{3}N\tilde{h}^2+N\tilde{a}_1\tilde{a}_2$$

Putting $\tilde{a}_1 = a_1+\Delta\tilde{a}_1$ etc. and substituting for the expected values of A and B as given by equations (3.2), we obtain

$$N^{\frac{1}{2}}Z_1(a_1+\Delta\tilde{a}_1+a_2+\Delta\tilde{a}_2) = N^{\frac{1}{2}}Z_2+N_1a_1^2+N_2a_2^2-\tfrac{1}{3}N\Delta\tilde{h}^2+N(a_1+\Delta\tilde{a}_1)(a_2+\Delta\tilde{a}_2).$$

Retaining only terms of order less than or equal to one, this becomes

$$N^{\frac{1}{2}}Z_1(a_1+a_2) = N^{\frac{1}{2}}Z_2-\tfrac{1}{3}N\Delta\tilde{h}^2+N(a_1\Delta\tilde{a}_2+a_2\Delta\tilde{a}_1)+N_1a_1^2+N_2a_2^2+(N_1+N_2)a_1a_2$$
$$= N^{\frac{1}{2}}Z_2-\tfrac{1}{3}N\Delta\tilde{h}^2+N(a_1\Delta\tilde{a}_2+a_2\Delta\tilde{a}_1)$$

Hence $a_1\Delta\tilde{a}_2 = -a_2\Delta a_1+ \tfrac{1}{3}\Delta\tilde{h}^2-N^{-\frac{1}{2}}Z_2+N^{-\frac{1}{2}}Z_1(a_1+a_2)$ (3.18)

Treating the second equation of (1.4) in a similar manner, it follows from

$$(B-\tfrac{1}{3}N\tilde{h}^2)(\tilde{a}_1+\tilde{a}_2) = C - \tilde{h}^2A+\tilde{a}_1\tilde{a}_2A$$

that we obtain

$$(N^{\frac{1}{2}}Z_2+N_1a_1^2+N_2a_2^2-\tfrac{1}{3}N\Delta\tilde{h}^2)(a_1+\Delta\tilde{a}_1+a_2+\Delta\tilde{a}_2)$$

$$= N^{\frac{1}{2}}Z_3+N_1a_1^3+N_2a_2^3-h^2N^{\frac{1}{2}}Z_1+a_1a_2N^{\frac{1}{2}}Z_1$$

i.e. $N^{\frac{1}{2}}Z_2(a_1+a_2)+(N_1a_1^2+N_2a_2^2)(\Delta\tilde{a}_1+\Delta\tilde{a}_2)-\frac{1}{3}N\Delta\tilde{h}^2(a_1+a_2)$

$$= N^{\frac{1}{2}}Z_3-h^2N^{\frac{1}{2}}Z_1+a_1a_2N^{\frac{1}{2}}Z_1 \qquad (3.19)$$

We now substitute the expression given for $a_1\Delta\tilde{a}_2$ in (3.18) into equation (3.19) so that we shall be able to solve for $\Delta\tilde{a}_1$ in terms of Z_1,Z_2,Z_3,Z_4 and $\Delta\tilde{h}^2$:

$$N^{\frac{1}{2}}Z_2(a_1+a_2)+(N_1a_1^2+N_2a_2^2)\Delta\tilde{a}_1+(N_1a_1-N_1a_2)\{-a_2\Delta\tilde{a}_1+\tfrac{1}{3}\Delta\tilde{h}^2-N^{-\frac{1}{2}}Z_2$$
$$+N^{-\frac{1}{2}}Z_1(a_1+a_2)\}-\tfrac{1}{3}N\Delta\tilde{h}^2(a_1+a_2)$$
$$= N^{\frac{1}{2}}Z_3-h^2N^{\frac{1}{2}}Z_1+a_1a_2N^{\frac{1}{2}}Z_1$$

i.e. $(N_1a_1^2+N_2a_2^2-N_1a_1a_2+N_1a_2^2)\Delta\tilde{a}_1= -N^{\frac{1}{2}}Z_2(a_1+a_2)-N_1(a_1-a_2)\{\frac{1}{3}\Delta\tilde{h}^2$

$$-N^{-\frac{1}{2}}Z_2+N^{-\frac{1}{2}}Z_1(a_1+a_2)\}+\tfrac{1}{3}N\Delta\tilde{h}^2(a_1+a_2)$$
$$+N^{\frac{1}{2}}Z_3-h^2N^{\frac{1}{2}}Z_1+a_1a_2N^{\frac{1}{2}}Z_1$$

i.e. $N_1(a_1-a_2)^2\Delta\tilde{a}_1=-N^{-\frac{1}{2}}Z_2(Na_1+Na_2-N_1a_1+N_1a_2)+N^{\frac{1}{2}}Z_3$

$$+\tfrac{1}{3}\Delta\tilde{h}^2(Na_1+Na_2-N_1a_1+N_1a_2)-N^{-\frac{1}{2}}Z_1[N_1(a_1-a_2)(a_1+a_2)$$
$$+Nh^2-Na_1a_2]$$
$$=N^{-\frac{1}{2}}Z_2(N_2a_1+2N_1a_2+N_2a_2)+N^{\frac{1}{2}}Z_3+\tfrac{1}{3}\Delta\tilde{h}^2(N_2a_1+2N_1a_2+N_2a$$
$$-N^{-\frac{1}{2}}Z_1(N_1a_1^2-N_1a_2^2-Na_1a_2+Nh^2)$$
$$=N^{-\frac{1}{2}}Z_2a_2\frac{N_2}{N_1}(2N_1-N_2)+N^{\frac{1}{2}}Z_3+\tfrac{1}{3}\Delta\tilde{h}^2a_2\,\frac{N}{N_1}(2N_1-N_2)$$
$$-N^{-\frac{1}{2}}Z_1[(a_1-a_2)(2a_1+a_2)N_1+Nh^2]$$
$$=N^{-\frac{1}{2}}Z_2N_1(a_2-a_1)(2-\frac{N_2}{N_1})+N^{\frac{1}{2}}Z_3+\tfrac{1}{3}\Delta\tilde{h}^2N_1(a_2-a_1)(2-\frac{N_2}{N_1})$$
$$-N^{-\frac{1}{2}}Z_1[N_1(a_1-a_2)(2a_1+a_2)+Nh^2] \qquad (3.20)$$

From equation (3.20) and the fact that $N(2a_1+a_2)=(2N_2-N_1)(a_1-a_2)$, it follows that

$$\gamma_1 = \frac{-N^{-\frac{1}{2}}[N_1(a_1-a_2)(2a_1+a_2)+Nh^2]}{N_1(a_1-a_2)^2} + \frac{N_1(a_2-a_1)(2-\frac{N_2}{N_1})\beta_1}{3N_1(a_1-a_2)^2}$$

$$= \frac{-N^{-\frac{1}{2}}(2a_1+a_2)}{a_1-a_2} - \frac{N^{\frac{1}{2}}h^2}{N_1(a_1-a_2)^2} + \frac{(2-\frac{N_2}{N_1})\beta_1}{3(a_2-a_1)}$$

$$= (a_2-a_1)^{-1}\{N^{-\frac{1}{2}}(2a_1+a_2) - \frac{N^{\frac{1}{2}}}{N_1}h^2(a_2-a_1)^{-1} + \frac{1}{3}(2-\frac{N_2}{N_1})\beta_1\}$$

$$= (a_2-a_1)^{-1}\{N^{-3/2}(2N_2-N_1)(a_1-a_2) - \frac{N^{\frac{1}{2}}}{N_1}h^2(a_2-a_1)^{-1} + \frac{1}{3}(2-\frac{N_2}{N_1})\beta_1\} \tag{3.21}$$

$$\gamma_2 = \frac{-N^{-\frac{1}{2}}N_1(a_2-a_1)(2-\frac{N_2}{N_1})}{N_1(a_1-a_2)^2} + \frac{(2-\frac{N_2}{N_1})\beta_2}{3(a_2-a_1)}$$

$$= \frac{(2-\frac{N_2}{N_1})}{a_2-a_1}(\frac{1}{3}\beta_2 - N^{-\frac{1}{2}}) \tag{3.22}$$

$$\gamma_3 = \frac{N^{\frac{1}{2}}}{N_1(a_1-a_2)^2} + \frac{(2-\frac{N_2}{N_1})\beta_3}{3(a_2-a_1)} \tag{3.23}$$

$$\gamma_4 = \frac{(2-\frac{N_2}{N_1})\beta_4}{3(a_2-a_1)} \tag{3.24}$$

The values of δ_i can be obtained from those for γ_i ($i=1,2,3,4$) in equations (3.20) to (3.23) simply by interchanging N_1 with N_2 and a_1 with a_2. To obtain $\frac{\Delta N_1}{N}$ we note that the first equation in (1.3) is

$$A = \tilde{N}_1\tilde{a}_1+\tilde{N}_2\tilde{a}_2$$

$$= \tilde{N}_1\tilde{a}_1+(N-\tilde{N}_1)\tilde{a}_2$$

Hence, asymptotically,

$$N^{\frac{1}{2}}Z_1 = (N_1+\Delta\tilde{N}_1)(a_1+\Delta\tilde{a}_1)+(N-N_1-\Delta\tilde{N}_1)(a_2+\Delta\tilde{a}_2)$$

$$= N_1a_1+N_1\Delta\tilde{a}_1+a_1\Delta\tilde{N}_1+N_2a_2+N_2\Delta\tilde{a}_2-a_2\Delta\tilde{N}_1$$
$$= \Delta\tilde{N}_1(a_1-a_2)+N_1\Delta\tilde{a}_1+N_2\Delta\tilde{a}_2$$

Hence $\Delta\tilde{N}_1(a_1-a_2) = N^{\frac{1}{2}}Z_1-N_1\Delta\tilde{a}_1-N_2\Delta\tilde{a}_2$ and so the coefficients of $\frac{\Delta\tilde{N}_1}{N}$ are given by

$$\begin{aligned}
\varepsilon_1 &= (a_1-a_2)^{-1}(N^{\frac{1}{2}}-N_1\gamma_1-N_2\delta_1)N^{-1}\\
\varepsilon_2 &= -(a_1-a_2)^{-1}(N_1\gamma_2+N_2\delta_2)N^{-1}\\
\varepsilon_3 &= -(a_1-a_2)^{-1}(N_1\gamma_3+N_2\delta_3)N^{-1}\\
\varepsilon_4 &= -(a_1-a_2)^{-1}(N_1\gamma_4+N_2\delta_4)N^{-1}
\end{aligned} \tag{3.25}$$

Since $\frac{\Delta\tilde{N}_2}{N} = -\frac{\Delta\tilde{N}_1}{N}$, the corresponding coefficients of $\Delta\tilde{N}_2$ are given by $-\varepsilon_1$, $-\varepsilon_2$, $-\varepsilon_3$ and $-\varepsilon_4$.

Before determining the various covariances between the Z_i's, it is necessary to obtain the expected value of X^k $(k=1,2,3...)$ where X has a uniform distribution on the interval $[a-h,a+h]$ with density function given by (1.1).

$$\begin{aligned}
E_a(X^k) &= \left.\frac{x^{k+1}}{2h(k+1)}\right]_{a-h}^{a+h}\\
&= \frac{(a+h)^{k+1} - (a-h)^{k+1}}{2h(k+1)}
\end{aligned} \tag{3.26}$$

The covariance between Z_i and Z_j $(i,j=1,2,3,4)$ is given by

$$\begin{aligned}
\text{cov}(Z_i,Z_j) &= \text{cov}\{N^{-\frac{1}{2}}(\sum_{r=1}^{N} x_r^i-E\sum_{r=1}^{N} x_r^i),N^{-\frac{1}{2}}(\sum_{r=1}^{N} x_r^j-E\sum_{r=1}^{N} x_r^j)\}\\
&= \frac{1}{N}\sum_{r=1}^{N}\text{cov}(x_r^i,x_r^j)\\
&= \frac{1}{N}\sum_{r=1}^{N}[E(x_r^{i+j}) - E(x_r^i)E(x_r^j)]\\
&= \frac{1}{N}\left\{\begin{aligned}&N_1E_{a_1}(X^{i+j})-N_1E_{a_1}(X^i)E_{a_1}(X^j)\\&+N_2E_{a_2}(X^{i+j})-N_2E_{a_2}(X^i)E_{a_2}(X^j)\end{aligned}\right\}
\end{aligned} \tag{3.27}$$

Thus from formulae (3.26) and (3.27) it is possible to calculate the covariance between any Z_i and Z_j. This, together with the fact that

$$\operatorname{cov}\left(\sum_{i=1}^{4} \lambda_{1i} Z_i, \sum_{j=1}^{4} \lambda_{2j} Z_j\right) = \sum_{i=1}^{4} \sum_{j=1}^{4} \lambda_{1i}\lambda_{2j} \operatorname{cov}(Z_i, Z_j),$$

for any scalars λ_{1i} and λ_{2j}, enables one to determine the variances of $\tilde{h}$, $\tilde{a}_1$, $\tilde{a}_2$, $\frac{\tilde{N}_1}{N}$ and $\frac{\tilde{N}_2}{N}$, as well as the co-variances between any two of these estimators, simply by referring to their expansions as linear functions of Z_1, Z_2, Z_3 and Z_4. This completes our derivation of the asymptotic distribution of the moment estimators.

We shall complete this section with some remarks on a conjecture which we put forward in the last paragraph of §4.1, that, asymptotically, the cubic given by (3.3) should have a unique positive solution for $\tilde{h}^2$. We may still assume that $N_1 a_1 + N_2 a_2 = 0$. For large sample sizes the coefficients of (3.3) can be replaced by their expected values as given by equations (3.4), (3.5) and (3.7). Thus we wish to examine the roots of the cubic

$$\begin{aligned} g^*(y) = {} & 16N^3y^3 + 108Ny^2\{-N_1N_2(a_1-a_2)^2 - \tfrac{1}{3}N^2h^2\} \\ & + 45Ny\{\tfrac{8}{15}N^2h^4 + 4N_1N_2h^2(a_1-a_2)^2 + N_1N_2(a_1-a_2)^4\} \\ & + 135\{-\tfrac{8}{15}NN_1N_2h^4(a_1-a_2)^2 - \tfrac{1}{3}NN_1N_2h^2(a_1-a_2)^4 - \tfrac{4}{135}N^3h^6\} \end{aligned} \tag{3.28}$$

We know that this will have at least one root given by $y = h^2$, and so we take out a factor of $(y-h^2)$:

$$\begin{aligned} g^*(y) = {} & 16N^3y^2(y-h^2) - 20N^3h^2y(y-h^2) + 4N^3h^4(y-h^2) \\ & - 108NN_1N_2(a_1-a_2)^2y(y-h^2) + 72NN_1N_2(a_1-a_2)^2h^2(y-h^2) \\ & + 45NN_1N_2(a_1-a_2)^4(y-h^2) \\ = {} & (y-h^2)\left\{\begin{aligned} & 16N^3y^2 - 20N^3h^2y + 4N^3h^4 - 108NN_1N_2(a_1-a_2)^2y \\ & + 72NN_1N_2(a_1-a_2)^2h^2 + 45NN_1N_2(a_1-a_2)^4 \end{aligned}\right\} \end{aligned} \tag{3.29}$$

The second term in the product (3.29) is a quadratic in y, and an application of Descartes' rule of signs shows that either both its roots are positive or both unreal. We note that the cubic (3.28) will have a single positive root $y = h^2$ if and only if the quadratic term in (3.29) has unreal roots. To establish the conditions for such roots we consider the discriminant of the quadratic term,

$$\begin{aligned}
\Delta &= \{-20N^3h^2-108NN_1N_2(a_1-a_2)^2\}^2 \\
&\quad -64N^3\{4N^3h^4+72NN_1N_2(a_1-a_2)^2h^2+45NN_1N_2(a_1-a_2)^4\} \\
&= 16N^2\{5N^2h^2+27N_1N_2(a_1-a_2)^2\}^2 \\
&\quad -64N^3\{4N^3h^4+72NN_1N_2(a_1-a_2)^2h^2+45NN_1N_2(a_1-a_2)^4\} \\
&= 16N^2 \left\{ \begin{array}{l} 25N^4h^4+270N^2N_1N_2h^2(a_1-a_2)^2+729N_1^2N_2^2(a_1-a_2)^4-16N^4h^4 \\ -288N^2N_1N_2(a_1-a_2)^2h^2-180N^2N_1N_2(a_1-a_2)^4 \end{array} \right\} \\
&= 16N^2\{9N^4h^4-18N^2N_1N_2(a_1-a_2)^2h^2+729N_1^2N_2^2(a_1-a_2)^4 \\
&\qquad -180N^2N_1N_2(a_1-a_2)^4\} \\
&= 144N^2\{N^4h^4-2N^2N_1N_2(a_1-a_2)^2h^2+81N_1^2N_2^2(a_1-a_2)^4-20N^2N_1N_2(a_1-a_2)^4\} \\
&= 144N^2\{[N^2h^2-N_1N_2(a_1-a_2)^2]^2-20N_1N_2(N^2-4N_1N_2)(a_1-a_2)^4\} \\
&= 144N^2\{[N^2h^2-N_1N_2(a_1-a_2)^2]^2-20N_1N_2(N_1-N_2)^2(a_1-a_2)^4\}
\end{aligned} \tag{3.30}$$

Hence $\Delta < 0$ if and only if

$$|N^2h^2-N_1N_2(a_1-a_2)^2| < (20N_1N_2)^{\frac{1}{2}}|N_1-N_2|(a_1-a_2)^2. \tag{3.31}$$

Thus (3.31) is the condition necessary for the cubic (3.28) to have a single, positive root $y = h^2$ i.e. for the cubic (3.3) to have a unique, positive solution $y = \tilde{h}^2$ with probability close to one for large sample sizes.

Of course, in a practical situation it will not be known whether condition (3.31) is satisfied for large sample sizes and hence one cannot be certain whether any non-uniqueness of solutions for $\tilde{h}^2$ is likely to be removed by taking further observations. It should be noted that (3.31) is likely to be satisfied if h^2 is not larger than $(a_1-a_2)^2$, especially if N_1 and N_2 are not too close to one another. If $\frac{N_1}{N} \to \alpha$ and $\frac{N_2}{N} \to 1-\alpha$ in the limit as $N \to \infty$, where $0 < \alpha < 1$, then the condition (3.31) for an asymptotically unique solution becomes

$$|h^2-\alpha(1-\alpha)(a_1-a_2)^2|<[20\alpha(1-\alpha)]^{\frac{1}{2}}|2\alpha-1|(a_1-a_2)^2. \tag{3.32}$$

Thus an experimenter is probably best advised to take as many observations as is conveniently possible and check conditions (1.18) for the uniqueness of a solution for $\tilde{h}^2$. If these are not satisfied he might try taking some further observations, though this cannot in general be guaranteed to provide a unique solution to the problem. In certain cases it may be possible to rule out some estimates of h^2 if they provide unreasonable or unreal estimates for the other parameters. If all other considerations fail one might choose that estimate of h^2 which gives the best fit to the fifth sample moment, $\sum_{r=1}^{N} x_r^5$, as the best available estimate from the method of moments.

4.4 ESTIMATION IN A MIXTURE OF THE DISTRIBUTIONS $U[a_1-h_1,a_1+h_1]$ and $U[a_2-h_2,a_2+h_2]$

Consider a set of observations $\{x_1,x_2,\ldots,x_N\}$ which we choose to regard as a single sample of size N on the random variable X which has a compound uniform distribution given by

$$p(x;\alpha,a_1,a_2,h_1,h_2)=\frac{\alpha u(x-a_1-h_1)u(a_1-h_1-x)}{2h_1}+\frac{(1-\alpha)u(x-a_2h_2)u(a_2-h_2-x)}{2h_2} \tag{4.1}$$

where $h_1,h_2 > 0$, $0 < \alpha < 1$ and $u(.)$ is that function which take the value unity if the argument is nonpositive and zero otherwise. We wish to employ the method of moments to estimate the five parameters α,a_1,a_2,h_1 and h_2.

Let $\mu_k = E(x-\mu_1')^k$ $k=1,2,3,4,5$ be the first five central moments, where $\mu_1' = E(x) = \alpha a_1+(1-\alpha)a_2$ is the first moment about the origin. Let $m_1=a_1-\mu_1'$ and $m_2=a_2-\mu_1'$. Then the moments defined above follow easily as

$$\begin{aligned}
\mu_1 &= \alpha m_1+(1-\alpha)m_2\\
\mu_2 &= \alpha(\tfrac{1}{3}h_1^2+m_1^2)+(1-\alpha)(\tfrac{1}{3}h_2^2+m_2^2)\\
\mu_3 &= \alpha m_1(h_1^2+m_1^2)+(1-\alpha)m_2(h_2^2+m_2^2)\\
\mu_4 &= \alpha(\tfrac{1}{5}h_1^4+2h_1^2m_1^2+m_1^4)+(1-\alpha)(\tfrac{1}{5}h_2^4+2h_2^2m_2^2+m_2^4)\\
\mu_5 &= \alpha(m_1h_1^4+\tfrac{10}{3}m_1^3h_1^2+m_1^5)+(1-\alpha)(m_2h_2^4+\tfrac{10}{3}m_2^3h_2^2+m_2^5)
\end{aligned} \tag{4.2}$$

Without loss of generality we suppose that $a_1 < a_2$, and hence $a_1 < \mu_1' < a_2$, $m_1 < 0 < m_2$. The case $a_1=a_2=a$ must be discussed separately. Under the model of a mixture of samples, we have already in §4.1 offered a possible approach to the case $h_1 = h_2 = h$ using moments about the origin.

Upon equating population moments to their corresponding sample moments, it follows from equations (4.2) that

$$
\begin{aligned}
&\alpha m_1+(1-\alpha)m_2 = 0\\
&\alpha(\tfrac{1}{3}h_1^2+m_1^2-\nu_2)+(1-\alpha)(\tfrac{1}{3}h_2^2+m_2^2-\nu_2) = 0\\
&\alpha(h_1^2m_1^2+m_1^3-\nu_3)+(1-\alpha)(h_2^2m_2+m_2^3-\nu_3) = 0 \qquad (4.3)\\
&\alpha(\tfrac{1}{5}h_1^4+2h_1^2m_1^2+m_1^4-\nu_4)+(1-\alpha)(\tfrac{1}{5}h_2^2+2h_2^2m_2^2+m_2^4-\nu_4) = 0\\
&\alpha(m_1h_1^4+ \tfrac{10}{3}m_1^3h_1^2+m_1^5-\nu_5)+(1-\alpha)(m_2h_2^4+ \tfrac{10}{3} m_2^3h_2^2+m_2^5-\nu_5) = 0
\end{aligned}
$$

where $\nu_i(1=2,3,4,5)$ is the i^{th} central moment of the sample, $\nu_i = \frac{1}{N}\sum_{r=1}^{N}(x_r-\bar{x})^i$. Equations (4.3) constitute a system of five equations in the five unknowns a_1,a_2,h_1,h_2 and α to be solved simultaneously for the estimates of these parameters. We have set the problem in the above manner so that it may be easily resolved by following the method of Cohen (1967) in the case of a mixture of normal distributions. We shall not display the laborious details here, but we note that the final estimates depend on a negative solution for a nonic in the product m_1m_2. For some combinations of sample data it may well be possible that such a nonic has more than one negative root, and as a possible solution to this problem one might follow the suggestion of Pearson (1894) and choose that set of estimates which results in the closest agreement between the sixth central moment of the sample and the corresponding moment of the fitted compound curve. Another disadvantage of the method of moments is that, due to certain differences in the coefficients, the method suggested by Cohen (1967) for circumventing the nonic in the case of a compound normal distribution cannot be adopted to the uniform case.

It is important to note that the method outlined above applies equally well to the model wherein $\{x_1,x_2,...,x_N\}$ is assumed to be a mixture of samples of size N_1 and N_2 from $U[a_1-h_1,a_1+h_1]$ and $U[a_2-h_2,a_2+h_2]$, if we simply replace α by

$\frac{N_1}{N}$ and $1-\alpha$ by $\frac{N_2}{N}$. However, because of the disadvantages of the method of moments mentioned above, the mixed samples model is better suited to the maximum likelihood method which we shall now briefly discuss.

Consider a sample of N observations $\{x_1, x_2, \ldots, x_N\}$ which is assumed to be a mixture of a sample of size N_1 from the first population $U[a_1-h_1, a_1+h_1]$ and a sample of size N_2 from the second population $U[a_2-h_2, a_2+h_2]$. Let ω_{ij} take the value one or zero according as the i^{th} observation does or does not belong to the j^{th} population for each $i=1,2,\ldots,n$ and $j=1,2$. Let $u(.)$ be as defined at the beginning of this section. Then the joint likelihood of $x_1, x_2, \ldots, x_N$ is given by $L(x_1, \ldots, x_N;\ a_1, a_2, h_1, h_2, N_1, N_2)$

$$= \frac{1}{(2h_1)^{N_1}} \cdot \frac{1}{(2h_2)^{N_2}} \prod_{i=1}^{N} \left\{ \begin{array}{l} \omega_{i1} u(x_i-a_1-h_1) u(a_1-h_1-x_i) \\ +\omega_{i2} u(x_i-a_2-h_2) u(a_2-h_2-x_i) \end{array} \right\} \tag{4.4}$$

We assume without loss of generality that $h_1 < h_2$. Given a_1, a_2, h_1 and h_2, the likelihood (4.4) is maximised if one allocates all observations falling in $[a_1-h_1, a_1+h_1]$ to the first population and the remaining to the second population. On the other hand, given the allocation ω_{ij}, the m.ℓ.e. of h_i is equal to one half of the range of the observations from population i and the m.ℓ.e. of a_i is the mid-range of the observations from population i for each $i=1,2$. Here it is necessary to make the additional assumption that $N_1 \neq 1$ and $N_2 \neq 1$; for, if an allocation of size one is made to either of the populations, the maximum value of the likelihood is infinite. This assumption is not particularly restrictive for large samples unless we are dealing with a "contaminated" sample where there is quite likely to be only one observation on the contaminating population. The method considered here does not apply to such a situation and a different approach would be required. A recent paper discussing contamination by outlying observations is that of Grubbs (1969).

Assuming $N_1, N_2 \geq 2$, all the possible values of the likelihood which need to be considered in the search for a maximum can be generated in the following manner. Let $j \in \{2,3,4,\ldots,N-2\}$. One must firstly determine the minimum m_j of the quantities

$$(x_{(j+1)}-x_{(2)}),\ (x_{(j+2)}-x_{(3)}),\ldots,(x_{(N-1)}-x_{(N-j)}) \quad (4.5)$$

where, as usual, $x_{(1)}, x_{(2)},\ldots,x_{(N)}$ are the order statistics of the sample. Having found m_j, one calculates that value of the likelihood given by

$$\lambda_{oj} = \frac{1}{(x_{(N)}-x_{(1)})^{N-j}} \cdot \frac{1}{m_j^j} \quad . \qquad (4.6)$$

This value of the likelihood corresponds to the maximum obtained by allocating any j observations to population 1 except for $x_{(1)}$ and $x_{(N)}$ which are allocated to population 2 along with the remaining N-j-2 observations. It is also necessary to consider the allocation of the smallest j observations including $x_{(1)}$ to the first population and the remaining to the second, as well as the allocation of the largest j observations including $x_{(N)}$ to the first population and the remaining to the second. The two corresponding values of the likelihood which need to be considered are given respectively by

$$\lambda_{1j} = \frac{1}{(x_{(j)}-x_{(1)})^j} \cdot \frac{1}{(x_{(N)}-x_{(j+1)})^{N-j}} \qquad (4.7)$$

$$\lambda_{2j} = \frac{1}{(x_{(N-j)}-x_{(1)})^{N-j}} \cdot \frac{1}{(x_{(N)}-x_{(N-j+1)})^{j}} \qquad (4.8)$$

Thus one compares λ_{oj}, λ_{1j} and λ_{2j} for each $j=2,3,\ldots,N-2$ in order to determine which such value of the likelihood is greatest. The corresponding allocation gives the maximum likelihood allocation and hence the m.ℓ.e.'s of a_1, a_2, h_1 and h_2 as discussed previously. It can be seen that such a method is ideally suited to a computer, especially if sample sizes are large. In fact, the total number of numerical comparisons required is given by

$$2(N-1) + \sum_{j=2}^{N-2} (N-j-1)$$

$$=2(N-1)+N(N-1)-\tfrac{1}{2}N(N-1)-(N-1)$$

$$= \tfrac{1}{2}(N-1)(N+2).$$

It is worth noting that it might sometimes be known a priori that $h_2 > h_1 > \delta > 0$, where δ is a known quantity perhaps determined from the range of a previous sample on the population $U[a_1-h_1, a_1+h_1]$. If this is so, allocations of size one may thenbe considered to contribute the factor $\frac{1}{\delta}$ to the likelihood of the sample and hence the resulting finite values of the likelihood may also be considered in the search for a maximum.

We conclude this section by remarking that, because of the appealing virtue of their uniqueness, the m.ℓ.e.'s would usually be preferable to the moment estimates, if we are using a mixed samples model. On the other hand, under a compound distribution model, the m.ℓ.e.'s can only be derived using complex iterative procedures and so the moment estimates might be of more use here, if only as first approximations to the parameters.

CHAPTER FIVE

A GENERAL FAMILY OF DISTRIBUTIONS POSSESSING A MONOTONE LIKELIHOOD RATIO

5.1 INTRODUCTION AND PRELIMINARIES

Consider an observed random variable x defined on a subset X of the real line R and depending on a continuous parameter θ taking values in a subset Θ of the real line. Without loss of generality, we need only consider the two cases where either X is an interval on which the random variable has a probability density $p(x,\theta)$ for each $\theta \in \Theta$, or X is a set of integers, in which case we let $p(x,\theta)$ be the probability of the observation x under the parametric value $\theta \in \Theta$.

Definition (1.1) If $p(x,\theta)$ satisfies

$$p(x_1,\theta_1)p(x_2,\theta_2) - p(x_1,\theta_2)p(x_2,\theta_1) \geq 0 \qquad (1.1)$$

for any $x_1,x_2 \in X$ with $x_1 > x_2$ and any $\theta_1,\theta_2 \in \Theta$ with $\theta_1 > \theta_2$, then $p(x,\theta)$ is said to possess a monotone likelihood ratio (M.L.R.) in x. If the inequality in (1.1) is strict, $p(x,\theta)$ will be said to possess a strict M.L.R. in x.

The above is the definition according to Karlin and Rubin (1956) and under certain regularity conditions it can be regarded as essentially equivalent to the more intuitively appealing definition given by Lehmann (1959, p.68):

Defintion (1.2) The real-parameter family of densities $p(x,\theta)$ is said to have a monotone likelihood ratio in a statistic T(x), a real-valued function of x, if for any $\theta_1 > \theta_2$ the distributions P_{θ_1} and P_{θ_2} are distinct, and the ratio

$\dfrac{p(x,\theta_1)}{p(x,\theta_o)}$ is a nondecreasing function of T(x).

Almost all the principal distributions occuring in statistical practice possess the property of monotone likelihood ratio. Some examples are the exponential family of densities given by $p(x,\theta) = \exp\{A(x)B(\theta)+C(x)+D(\theta)\}$ where $B(\theta)$ is nondecreasing in θ, the non-central t, the non-central F, the non-central chi-square and the power series distributions such as the binominal, Poisson and hypergeometric. Proofs that these distributions do indeed possess an M.L.R., as well as some further examples, are to be found in Karlin and Rubin (1956), Karlin (1956) and Lehmann (1959).

The work cited in the above references, plus that of Karlin (1957(a), 1957(b), 1958), has shown the M.L.R. property and the extension to Pòlya-type densities to be of great use in the unification and strengthening of a wide variety of statistical decision problems. The most well known results are the existence of uniformly most powerful (U.M.P.) one-sided tests of simple hypotheses, the uniformly most accurate (U.M.A.) confidence limits that accompany such tests, and the existence of U.M.P. two-sided tests in certain cases. The M.L.R. property has also been of use in the testing of hypotheses by sequential methods; for example, to show that a sequential probability ratio test of a simple hypothesis against a simple alternative has a nonincreasing operating characteristic (O.C.) function, given that the sample under consideration is from an M.L.R. family [see Ghosh (1970, p.100-103)]

The connection between the M.L.R. property and the classification and estimation problems of the type considered in the first two chapters of this thesis has been pointed out by John (1970(c)). Suppose that a sample $x_1,x_2,...,x_N$ is postulated to be a mixture of one sample from $p(x,\theta_1)$ and another from $p(x,\theta_2)$, where $\theta_2 < \theta_1$. We wish to use the method of maximum likelihood to classify the observations and estimate the parameters. The likelihood is

$$\prod_{r=1}^{N}\left[p(x_r\theta_1)\right]^{\omega_r}\left[p(x_r,\theta_2)\right]^{1-\omega_r} \tag{1.2}$$

where ω_r takes the value one if x_r is from $p(x,\theta_1)$ and zero otherwise. Given θ_1 and θ_2, the likelihood is maximised by taking ω_r one or zero according as the ratio $\rho(x_r) = \dfrac{p(x_r,\theta_1)}{p(x_r,\theta_2)}$ is greater or less than one; if the ratio is equal to one, it does not matter whether ω_r is given the value one or zero. If $\rho(x)$ can take values both less and greater than one as x varies over its range space and if $p(x,\theta)$ possesses an M.L.R. in x, then there exists a point η in the range space of x such that $\rho(x)$ is greater or not greater than one according as x is greater or not greater than η. This permits the simple allocation rule of classifying x_r as an observation on $p(x,\theta_1)$ or $p(x,\theta_2)$ according as x_r is greater or not greater than η. As the uniform distribution $U[0,\theta]$ possesses an M.L.R. in x, such an allocation procedure was applied in the third section of Chapter One when discussing a mixture of samples from $U[0,\theta_1]$ and $U[0,\theta_2]$.

We now define the range over which $p(x,\theta)$ is positive and quote without proof three useful properties of this range due to Karlin and Rubin (1956). These properties are also to be found in the book by Karlin (1968), wherein the term "totally positive, type two" is used for a function of two real variables possessing an M.L.R.

<u>Definition (1.3)</u> We define the spectrum of $p(x,\theta)$ by the set $\sigma_\theta = \{x|p(x,\theta) > 0\}$, for each $\theta \varepsilon \Theta$.

<u>Lemma (1.1)</u> If $x < y < z$ and $x \varepsilon \sigma_{\theta_1}$, $y \not\varepsilon \sigma_{\theta_1}$ and $y \varepsilon \sigma_{\theta_2}$, then $\theta_1 < \theta_2$ and $z \varepsilon \sigma_{\theta_1}$. A similar result holds true if we merely reverse all the inequalities.

<u>Lemma (1.2)</u> The set σ_θ for any $\theta \varepsilon \Theta$ is a relative interval in X, i.e. the intersection of an interval and X.

<u>Lemma (1.3)</u> If $\theta_1 < \theta_2$, then $\sigma_{\theta_1} < \sigma_{\theta_2}$ in the sense that, if $x_1 \varepsilon \sigma_{\theta_1}$ and $x_2 \varepsilon \sigma_{\theta_2}$, then either $x_1 \varepsilon \sigma_{\theta_1} \cap \sigma_{\theta_2}$ and $x_2 \varepsilon \sigma_{\theta_1} \cap \sigma_{\theta_2}$ or $x_1 < x_2$.

In the case of x being a continuous random variable, these lemmas permit us to specify σ_θ according to

$$\sigma_\theta = (\ell_1(\theta), \ell_2(\theta)), \tag{1.2}$$

where $\ell_1(\theta)$ and $\ell_2(\theta)$ are nondecreasing, real-valued functions of the parameter $\theta \varepsilon \Theta$; and similarly, for x a discrete-valued random variable,

$$\sigma_\theta = \{\ell_1(\theta), \ell_1(\theta)+1, \ldots, \ell_2(\theta)\} \tag{1.3}$$

where $\ell_1(\theta)$ and $\ell_2(\theta)$ are integer-valued functions of the parameter. We allow the possibility that $\ell_1(\theta) = -\infty$ and/or $\ell_2(\theta) = +\infty$ for some or all $\theta \varepsilon \Theta$.

Using the concept of statistical connectedness, Karlin ane Rubin (1956) have also shown that, without any loss of generality, Θ can be taken as an interval in the real line, not necessarily finite. We assume that σ_θ is non-empty for each $\theta \varepsilon \Theta$.

<u>Definition (1.4)</u> Let $S_x = \{\theta|p(x,\theta) > 0\}$ for each $x \varepsilon \bigcup_{\theta\varepsilon\Theta} \sigma_\theta$ be the selection statistic.

Then every such S_x is a nonempty interval in Θ, either open or closed. In fact,

$$\bar{S}_x = [m_1(x), m_2(x)], \tag{1.4}$$

where $$m_1(x) = \inf\{\theta \in \Theta \mid \ell_2(\theta) \geq x\} \tag{1.5}$$

and $$m_2(x) = \sup\{\theta \in \Theta \mid \ell_1(\theta) \leq x\} \tag{1.6}$$

for each value of the discrete random variable in $\bigcup_{\theta \in \Theta} \sigma_\theta$, and similarly for the continuous random variable except that the inequalities must be strict. If we take $X = \bigcup_{\theta \in \Theta} \sigma_\theta$, then $m_1(x)$ and $m_2(x)$ are both nondecreasing on X since $\ell_1(\theta)$ and $\ell_2(\theta)$ are both nondecreasing on Θ.

The following result due to Karlin (1957(a)) allows one to identify and characterise those density and probability functions which possess an M.L.R. in x, provided that the spectrum is independent of θ.

<u>Lemma (1.4)</u> Suppose $p(x,\theta) > 0 \ \forall \ x \in X$, $\theta \in \Theta$ and that $\frac{\partial^2 \log p(x,\theta)}{\partial x \partial \theta}$ exists on $X \times \Theta$. Then $p(x,\theta)$ has an M.L.R. in x if and only if $\frac{\partial^2 \log p(x,\theta)}{\partial x \partial \theta} \geq 0$ on $X \times \Theta$.

In the next section we shall derive a characterisation of some M.L.R. families which is suggested by Lemma (1.4), but which requires only the existence of $\frac{\partial}{\partial \theta} \log p(x,\theta) \ \forall \ \theta \in S_x$, $x \in X$. Having derived this general family of density or probability functions, we shall then examine some of its properties and consider how best to estimate the parameter and to construct confidence and tolerance intervals.

5.2 A GENERAL FAMILY POSSESSING A MONOTONE LIKELIHOOD RATIO

<u>Theorem (2.1)</u> Assume that $\frac{\partial}{\partial\theta} \log p(x,\theta)$ exists $\forall\ \theta \in S_x$, $x \in X$. Then a necessary and sufficient condition for $p(x,\theta)$ to possess an M.L.R. according to (1.1) above is that it can be written in the form

$$p(x,\theta) = \begin{cases} h(x) \exp\{ \int_a^\theta g(x,u)du \} & x \in \sigma_\theta \\ 0 & \notin \sigma_\theta \end{cases} \qquad (2.1)$$

for each $\theta \in \Theta$, where $g(x,\theta)$ is a nondecreasing function of $x \in \sigma_\theta$ for each $\theta \in \Theta$, $g(x,\theta)$ is zero for $x \notin \sigma_\theta$ for each $\theta \in \Theta$ and is integrable with respect to $\theta \in \Theta$ for each $x \in X$; $a \geq -\infty$ is the infinuum of Θ and $h(x)$ is a real-valued positive function of $x \in X$.

<u>Proof</u> *Necessity:*

If $p(x,\theta)$ has an M.L.R. as given by (1.1) and if $\theta_1 > \theta_2$ are two points in Θ, then

$$\eta(x,\theta_1,\theta_2) = \frac{p(x,\theta_1)}{p(x,\theta_2)} \qquad (2.2)$$

is a nondecreasing function of $x \in \sigma_{\theta_2}$. Taking logarithms and dividing by $\theta_1-\theta_2$,

$$\frac{\log p(x,\theta_1)-\log p(x,\theta_2)}{\theta_1 - \theta_2} = \frac{\log \eta(x,\theta_1,\theta_2)}{\theta_1 - \theta_2} \qquad x \in \sigma_{\theta_2}$$

$$= \frac{\log \eta(x,\theta_1,\theta_2)-\log \eta(x,\theta_2,\theta_2)}{\theta_1 - \theta_2} \quad x \in \sigma_{\theta_2}$$

$$(2.3)$$

since $\log \eta(x,\theta_2,\theta_2) = 0$. The right hand side of (2.3) is still a nondecreasing function of $x \in \sigma_{\theta_2}$. Taking the limit as θ_1 tends to θ_2 from the right,

$$\frac{\partial}{\partial\theta} \log p(x,\theta_2) = \frac{\partial}{\partial\theta} \log \eta(x,\theta,\theta_2)\Big|_{\theta=\theta_2} \qquad x \in \sigma_{\theta_2}$$

$$\equiv g(x,\theta_2) \qquad x \in \sigma_{\theta_2} \qquad (2.4)$$

where $g(x,\theta_2)$ is a nondecreasing function of $x \varepsilon \sigma_{\theta_2}$.
We see that θ_2 in equation (2.4) may take any value in Θ except perhaps the supremum θ_1^o of Θ if this is finite and belongs to Θ. However, if we put $\theta_1 = \theta_1^o$ in equation (2.3) and take the limit as θ_2 tends to θ_1^o from the left, then

$$\frac{\partial}{\partial\theta} \log p(x,\theta_1^o) = \frac{\partial}{\partial\theta} \log \eta(x,\theta_1^o,\theta)\Big|_{\theta=\theta_1^o} \qquad x \varepsilon \sigma_{\theta_1^o}$$

$$\equiv g(x,\theta_1^o) \qquad x \varepsilon \sigma_{\theta_1^o} \qquad (2.5)$$

Now for any $x \varepsilon X$ we can combine equations (2.4) and (2.5) so that

$$\frac{\partial}{\partial\theta} \ell n\ p(x,\theta) = g(x,\theta) \qquad \forall \theta \varepsilon S_x, \qquad (2.6)$$

and if we define $g(x,\theta) = 0 \qquad \forall \theta \varepsilon \Theta - S_x, \qquad (2.7)$

then we see that $g(x,\theta)$ is integrable with respect to θ on Θ.
In fact,

$$\log p(x,\theta) = \int_a^\theta g(x,u)\,du + h_1(x) \qquad \forall \theta \varepsilon S_x \qquad (2.8)$$

where $a = \inf \Theta$ and $h_1(x)$ is a real-valued function of $x \varepsilon X$.
Since equation (2.8) holds for any $x \varepsilon X$ we see that $p(x,\theta)$ can be written in the form

$$p(x,\theta) = \begin{cases} h(x) \exp\{\int_a^\theta g(x,u)\,du\} & x \varepsilon \sigma_\theta \\ 0 & x \not\varepsilon \sigma_\theta \end{cases}$$

for each $\theta \varepsilon \Theta$, where $h(x) > 0 \ \forall x \varepsilon X$, as required.

Sufficiency:

Given $p(x,\theta)$ is of the form (2.1), let $x_1 > x_2$ be points of X and $\theta_1 > \theta_2$ points of Θ.
Let $L = p(x_1,\theta_1)p(x_2,\theta_2) - p(x_1,\theta_2)p(x_2,\theta_1)$.
We shall show that $L \geq 0$ for all such possible choices of x_1, x_2, θ_1 and θ_2. For example, in the discrete case,

$x_1 \not\varepsilon \sigma_{\theta_1} \Rightarrow \{x_1 > \ell_2(\theta_1) \text{ or } x_1 < \ell_1(\theta_1)\} \Rightarrow \{x_1 > \ell_2(\theta_2) \text{ or } x_2 < \ell_1(\theta_1)\} \Rightarrow L=0$

$x_1 \not\varepsilon \sigma_{\theta_2} \Rightarrow p(x_1\theta_2) = 0 \Rightarrow L \geq 0$

$x_2 \notin \sigma_{\theta_1} \Rightarrow p(x_2,\theta_1) = 0 \Rightarrow L \geq 0$

$x_2 \notin \sigma_{\theta_2} \Rightarrow \{x_2 > \ell_2(\theta_2) \text{ or } x_2 < \ell_1(\theta_2)\} \Rightarrow \{x_1 > \ell_2(\theta_2) \text{ or } x_2 < \ell_1(\theta_1)\} \Rightarrow L = 0$

Now let $x_1, x_2 \in \sigma_{\theta_1} \cap \sigma_{\theta_2}$ with $x_1 > x_2$

From (2.1) we have that

$$L = h(x_1)h(x_2)\left\{ \exp\left[\int_a^{\theta_1} g(x_1,u)\,du + \int_a^{\theta_2} g(x_2,u)\,du\right] - \exp\left[\int_a^{\theta_2} g(x_1,u)\,du + \int_a^{\theta_1} g(x_2,u)\,du\right]\right\}$$

$$= h(x_1)h(x_2) \exp\left[\int_a^{\theta_2} g(x_1,u)\,dn + \int_a^{\theta_2} g(x_2,u)\,du\right] \left\{\exp\left[\int_{\theta_2}^{\theta_1} g(x_1,u)\,du\right] - \exp\left[\int_{\theta_2}^{\theta_1} g(x_2,u)\,du\right]\right\}$$

≥ 0 since $h(x_1)$, $h(x_2) \geq 0$ and $g(x_1,u) \geq g(x_2,u)\ \forall\, u \in (\theta_2,\theta_1)$

Q.E.D.

Given a particular $p(x,\theta)$, the above theorem can be used to determine whether $p(x,\theta)$ possesses an M.L.R. simply by taking $\frac{\partial}{\partial\theta} \log p(x,\theta)$ (if it exists) and examining it to see whether it is a nondecreasing function of $x \in \sigma_\theta$ for any $\theta \in \Theta$ The theorem will also be of use in the construction of families with the M.L.R. property, but before considering this we shall first note some results following from the fact that $p(x,\theta)$ as given by (2.1) must be a density in the continuous case and a probability distribution in the discrete case. If the random variable is continuous, we require that

$$\int_{\ell_1(\theta)}^{\ell_2(\theta)} h(x)\exp\left\{\int_a^\theta g(x,u)\,du\right\}dx = 1 \ \forall\, \theta \in \Theta \qquad (2.9)$$

Assuming $\ell_1(\theta)$ and $\ell_2(\theta)$ are differentiable on Θ we can differentiate (2.7) with respect to θ to obtain

$$\int_{\ell_1(\theta)}^{\ell_2(\theta)} h(x)g(x,\theta)\exp\left\{\int_a^\theta g(x,u)du\right\}dx + \ell_2'(\theta)h(\ell_2(\theta)-0)$$
$$\times \exp\left\{\int_a^\theta g(\ell_2(\theta)-0,u)du\right\}-\ell_1'(\theta)h(\ell_1(\theta)+0)\exp\left\{\int_a^\theta g(\ell_1(\theta)+0,u)du\right\}=0$$
$$\forall\, \theta \in \Theta \qquad (2.10)$$

So, if we add the extra conditions that

$$\ell_1'(\theta) \text{ and } \ell_2'(\theta) \text{ exist } \quad \forall\, \theta \in \Theta \qquad (2.11)$$

and that

$$\lim_{x\to\ell_1(\theta)+} \int_a^\theta g(x,u)du = -\infty = \lim_{x\to\ell_2(\theta)-} \int_a^\theta g(x,u)du \quad \forall\, \theta \in \Theta, \qquad (2.12)$$

then we have
$$\left.\begin{aligned} &E_\theta\{g(x,\theta)\} = 0 \quad \forall\, \theta \in \Theta \\ &E_\theta\{g^2(x,\theta)+ \tfrac{\partial}{\partial\theta}\, g(x,\theta)\} = 0 \quad \forall\, \theta \in \Theta \\ &\qquad \text{etc.} \end{aligned}\right\} \qquad (2.13)$$

where the series of equations can be obtained by repeated differentiation of equation (2.10).

If the r.v. is discrete, then

$$\sum_{x=\ell_1(\theta)}^{\ell_2(\theta)} p(x,\theta) = 1 \quad \forall\, \theta \in \Theta, \qquad (2.14)$$

where $p(x,\theta)$ is given by (2.1) and $\ell_1(\theta)$ and $\ell_2(\theta)$ are non-decreasing integer-valued functions of $\theta \in \Theta$. If $\ell_1(\theta)$ and $\ell_2(\theta)$ are both continuous from the right or both continuous from the left, then we claim that the results of (2.13) also hold in the discrete case. For example, if $\ell_1(\theta)$ and $\ell_2(\theta)$ are both right continuous on Θ, it follows from (2.14) that

$$0 = \lim_{\delta\to 0+} \frac{1}{\delta}\left\{\sum_{x=\ell_1(\theta+\delta)}^{\ell_2(\theta+\delta)} p(x,\theta+\delta) - \sum_{x=\ell_1(\theta)}^{\ell_2(\theta)} p(x,\theta)\right\} \quad \forall\, \theta \in \Theta$$

$$= \lim_{\delta\to 0+} \frac{1}{\delta}\left\{\sum_{x=\ell_1(\theta)}^{\ell_2(\theta)} [p(x,\theta+\delta)-p(x,\theta)]\right\} \text{ by right continuity}$$

$$= \sum_{x=\ell_1(\theta)}^{\ell_2(\theta)} \frac{\partial p(x,\theta)}{\partial\theta} \qquad \text{since the partial derivative exists } \forall\, x \in \sigma_\theta,\ \forall\, \theta \in \Theta$$

$$= \sum_{x=\ell_1(\theta)}^{\ell_2(\theta)} h(x)g(x,\theta)\exp\left\{\int_a^\theta g(x,u)du\right\} \qquad \forall\, \theta \in \Theta$$

$$= E_\theta(g(x,\theta)) \qquad \forall\, \theta \in \Theta \tag{2.15}$$

Similarly, by taking the second derivative from the right we have $E_\theta\{g^2(x,\theta) + \frac{\partial}{\partial\theta} g(x,\theta)\} = 0 \quad \forall\, \theta \in \Theta$, etc.

To construct a member of the family (2.1) in the case of a continuous r.v., one firstly chooses a parametric interval Θ in the real line with infinium $a \geq -\infty$. Then define two real-valued, nondecreasing, differentiable functions $\ell_1(\theta) < \ell_2(\theta)$ on Θ. Let $\sigma_\theta = (\ell_1(\theta), \ell_2(\theta)) \ \forall\, \theta \in \Theta$ and $\bar{S}_x$ be as defined in (1.4) for each $x \in X = \bigcup_{\theta\in\Theta} \sigma_\theta$. Next one chooses $g(x,\theta)$, a nondecreasing function of $x \in \sigma_\theta$ and vanishing if $x \notin \sigma_\theta$ for each $\theta \in \Theta$, integrable with respect to $\theta \in \Theta$ for any $x \in X$, and such that

$$\int_{\ell_1(\theta)}^{\ell_2(\theta)} h(x) \exp\left\{\int_a^\theta g(x,u)\,du\right\} dx \qquad \text{is finite,}$$

where $h(x)$ is a positive integrable function of $x \in X$.

$$\text{Let } \phi(\theta) = \left\{\int_{\ell_1(\theta)}^{\ell_2(\theta)} h(x) \exp\left[\int_a^\theta g(x,u)\,du\right] dx\right\}^{-1} \forall\theta \in \Theta. \tag{2.16}$$

Since $\phi(\theta)$ is differentiable on Θ and satisfies $0 < \phi(\theta) < \infty$ $\forall\, \theta \in \Theta$ we have that

$$\int_{\ell_1(\theta)}^{\ell_2(\theta)} h(x) \exp\left\{\int_a^\theta \left[g(x,u) + \frac{\phi'(u)}{\phi(u)}\right] du\right\} dx = \frac{1}{\phi(a)} \quad \forall\theta \in \Theta$$

$$\Rightarrow \int_{\ell_1(\theta)}^{\ell_2(\theta)} h(x)\phi(a) \exp\left\{\int_a^\theta \left[g(x,u) + \frac{\phi'(u)}{\phi(u)}\right] du\right\} dx = 1 \quad \forall\theta \in \Theta \tag{2.17}$$

Hence, if we put $h^*(x) = h(x)\,\phi(a) \ \forall\, x \in X$ and

$g^*(x,\theta) = g(x,\theta) + \frac{\phi'(\theta)}{\phi(\theta)} \quad \forall x \in \sigma_\theta, \quad \theta \in \Theta$ and

$g^*(x,\theta) = 0 \ \forall x \notin \sigma_\theta, \theta \in \Theta$,

then we have constructed a density of the form (2.1):

$$p(x,\theta) = \begin{cases} h^*(x) \exp\left\{\int_a^\theta g^*(x,u)\,du\right\} & x \in \sigma_\theta \\ 0 & x \notin \sigma_\theta \end{cases}$$

where $g^*(x,\theta)$ is a nondecreasing function of $x \in \sigma_\theta$ for each $\theta \in \Theta$ and so $p(x,\theta)$ possesses an M.L.R.

The construction for the discrete case is analogous if we put $\sigma_\theta = \{\ell_1(\theta), \ell_1(\theta)+1\ldots,\ell_2(\theta)\}$ where ℓ_1 and ℓ_2 are integer-valued functions of $\theta \in \Theta$ and are either both left continuous or both right continuous, so that $\phi(\theta)$ defined by

$$\phi(\theta) = \left\{\sum_{x=\ell_1(\theta)}^{\ell_2(\theta)} h(x)\exp\left[\int_a^\theta g(x,u)\,du\right]\right\}^{-1} \quad \theta \in \Theta \quad (2.18)$$

will be differentiable with respect to θ on Θ if it is finite on Θ.

We remark that in the discrete case the above construction will be sufficient for the properties (2.13) to hold, but for the case of a continuous r.v. the extra conditions given by (2.12) will be needed if (2.13) is to hold.

5.3 ESTIMATION OF THE PARAMETER

In this section we treat only the case of a continuous r.v. satisfying (2.11), (2.12) and hence (2.13), the analogy for a discrete r.v. being obvious. From a sample of n independent observations $(x_1,\ldots,x_n)$ on the continuous r.v. with density given by (2.1) we wish to estimate the value of θ.

Firstly, we consider the maximum likelihood estimator. The joint likelihood of the sample is given by

$$L = L(x_1\ldots,x_n;\theta) = \left\{\prod_{r=1}^{n} h(x_r)\right\}\exp\left\{\int_a^\theta \sum_{r=1}^{n} g(x_r,u)\,du\right\}$$

for all $\theta \in \bigcap_{r=1}^{n} S_{x_r} = \{\theta \mid \ell_1(\theta) < x_{(1)} \text{ and } \ell_2(\theta) > x_{(n)}\}$

$$= (m_1(x_{(n)}),\ m_2(x_{(1)})) \quad (3.1)$$

since $\ell_1(\theta)$ and $\ell_2(\theta)$ are differentiable on Θ, where $x_{(1)},\ldots,x_{(n)}$

are the order statistics formed from the sample. So the m.ℓ.e. will be given by $\hat{\theta}$, that value of θ which satisfies

$$\sum_{r=1}^{n} g(x_r, \theta) = 0 \qquad (3.2)$$

provided that $\hat{\theta} \in (m_1(x_{(n)}), m_2(x_{(1)}))$.
The solution of (3.2) may have to be obtained by iterative methods such as the Newton-Raphson process or Fisher's "the method of scoring for parameters", both discussed in Kendall and Stuart (1967).

From (2.13), $E_\theta\{g^2(x,\theta) + \frac{\partial}{\partial\theta} g(x,\theta)\} = 0 \quad \forall \theta \in \Theta$,

and hence $\quad E_\theta\{\frac{\partial}{\partial\theta} g(x,\theta)\} < 0 \quad \forall \theta \in \Theta, \qquad (3.3)$

assuming that for each $\theta \in \Theta \; \exists \; x \in \sigma_\theta$ s.t. $g(x,\theta) \neq 0$.
Thus (3.2) will have exactly one solution with probability close to one for large sample size. So, provided we can start with an initial estimate sufficiently close to $\hat{\theta}$, we expect a method such as the Newton-Raphson process to converge to $\hat{\theta}$ with probability close to one for large sample size. If the value of $\hat{\theta}$ so determined falls outside the interval $(m_1(x_{(n)}), m_2(x_{(1)}))$, then we take the closest boundary point of this interval as the m.ℓ.e.

As is well known, the asymptotic distribution of $\hat{\theta}$ will be normal with mean θ (the true value of the parameter) and variance given by

$$-\{E_\theta(\frac{\partial^2 \log L}{\partial\theta^2})\}^{-1} = -\{E_\theta\{\sum_{r=1}^{n} \frac{\partial}{\partial\theta} g(x_r,\theta)\}\}^{-1}$$

$$= \{\sum_{r=1}^{n} E_\theta(g^2(x_r,\theta))\}^{-1} \quad \text{by (2.13)}$$

$$= \{n\, E_\theta(g^2(x,\theta))\}^{-1} \qquad (3.4)$$

We can also use equations (2.13) to obtain the estimates of θ given by the solutions to the equations

$$\sum_{r=1}^{n} g(x_r,\theta) = 0 \qquad (3.5)$$

$$\sum_{r=1}^{n} \{g^2(x_r,\theta) + \frac{\partial}{\partial\theta} g(x_r,\theta)\} = 0 \qquad (3.6)$$

$$\sum_{r=1}^{n} \{g^3(x_r\theta) + 3g(x_r,\theta) \frac{\partial}{\partial\theta} g(x_r,\theta) + \frac{\partial^2 g(x,\theta)}{\partial\theta^2}\} = 0 \qquad (3.7)$$

etc.

These estimates will also be subject to the condition that they belong to $(m_1(x_{(n)}), m_2(x_{(1)}))$ and we note that (3.5) yields the m.ℓ.e. The above three equations may be considered as convenient analogies to the moment equations when these may in fact be too complicated to be useful. Hence it may be possible in some cases that they could be more easily applied to the problem of solving for several unknown parameters in the case of a mixture of populations. In general, however, the solution of these equations will require methods of iteration.

Since $E_\theta(g(x,\theta)) = 0 \;\forall\; \theta \in \Theta$ we might consider another estimator $\tilde{\theta}$ of θ based on minimising the sum of squares $S = \sum_{r=1}^{n} g^2(x_r,\theta)$. Then $\tilde{\theta}$ will satisfy the equation

$$\sum_{r=1}^{n} g(x_r,\theta) \frac{\partial}{\partial\theta} g(x_r,\theta) = 0 \qquad (3.8)$$

provided this does in fact minimise S and subject to $\tilde{\theta} \in (m_1(x_{(n)}), m_2(x_{(1)}))$. If $\tilde{\theta}$ falls outside this interval we take that boundary point which minimises S as our least squares estimate of θ.

We shall now illustrate these various estimates of the parameter by applying the methods outlined above to some examples of M.L.R. families to be found in Karlin (1956, p.125) and Karlin and Rubin (1956).

Example (3.1) Let $p(x,\theta) = \frac{1}{\pi}\,\mathrm{sech}(x-\theta)$ for $x,\theta \in R$.

Hence $g(x,\theta) = \frac{\partial}{\partial\theta}\log p(x,\theta) = \frac{\mathrm{sech}(x-\theta)\tanh(x-\theta)}{\mathrm{sech}(x-\theta)} = \tanh(x-\theta)$

and so the m.ℓ.e. of θ is given by $\sum_{r=1}^{n} g(x_r,\hat{\theta}) = 0$,

i.e. $\sum_{r=1}^{n} \tanh(x_r-\hat{\theta}) = 0$.

The least squares estimate according to (3.8) is given by

$$\sum_{r=1}^{n} \tanh(x_r-\tilde{\theta})\,\mathrm{sech}^2(x_r-\tilde{\theta}) = 0$$

i.e. $\sum_{r=1}^{n} \tanh(x_r-\tilde{\theta})[1 - \tanh^2(x_r-\tilde{\theta})] = 0$,

and from equation (3.6) the second moment-like estimator is given by that value of θ which satisfies the equation

$$\sum_{r=1}^{n} [\tanh^2(x_r-\theta)-\mathrm{sech}^2(x_r-\theta)] = 0$$

i.e. $\sum_{r=1}^{n} [2\tanh^2(x_r-\theta)-1] = 0$

i.e. $\sum_{r=1}^{n} \tanh^2(x_r-\theta) = \frac{N}{2}$

It can be seen that iterative methods will be required for all three of these estimators, but that such methods will be simplest for the maximum likelihood estimator.

Example 3.2 Let $p(x,\theta) = e^{x-\theta-e^{x-\theta}}$ for $x,\theta \in R$.

Then $g(x,\theta) = \frac{\partial}{\partial\theta}\log p(x,\theta) = -1 + e^{x-\theta}$ and so the m.ℓ.e. of θ is given by $\sum_{r=1}^{n} (-1+e^{x_r-\hat{\theta}}) = 0$, i.e. $\hat{\theta} = \log(\frac{1}{n}\sum_{r=1}^{n} e^{x_r})$.

The least squares estimate is given by

$$\tilde{\theta} = \log\left(\sum_{r=1}^{n} e^{2x_r}\right) - \log\left(\sum_{r=1}^{n} e^{x_r}\right),$$

and the second moment-like estimate satisfies the quadratic in e^{θ} given by

$$ne^{2\theta} - 3e^{\theta}\sum_{r=1}^{n} e^{x_r} + \sum_{r=1}^{n} e^{2x_r} = 0.$$

Example (3.3) If $p(x,\theta) = \begin{cases} \frac{1}{n!}(x-\theta)^m e^{-(x-\theta)} & ; x \geq \theta \\ 0 & ; x < \theta \end{cases}$

where $m \geq 0$ is an integer, then the m.ℓ.e., least squares estimate and second moment-like estimate are given by the solutions to the equations

$$\sum_{r=1}^{n}(x_r-\theta)^{-1} = \frac{n}{m}, \quad \sum_{r=1}^{n}\frac{m(x_r-\theta-m)}{(x_r-\theta)^3} = 0 \quad \text{and}$$

$$\sum_{r=1}^{n}\left\{\frac{m(m-1)-2m(x_r-\theta)+(x_r-\theta)^2}{(x_r-\theta)}\right\} = 0$$ respectively. Hence it is apparent in these and other examples that the m.ℓ.e. is usually the simplest and hence the most useful of the estimates discussed above.

5.4 CONFIDENCE AND TOLERANCE INTERVALS

Lehmann (1959) has shown how the theory of U.M.P. one-sided tests for M.L.R. families depending on one parameter θ can be used to construct U.M.A. one-sided γ confidence intervals for θ and Zacks (1970, 1971) has shown how these in turn can be used to derive U.M.A. one-sided (β,γ) tolerance intervals. We shall not define the various concepts here, but refer the reader to Zacks (1971) or Guttman (1970). However, to expedite the discussion, we quote the following three main results.

Theorem (4.1) If $\mathcal{P} = \{p(x,\theta);\ \theta \in \Theta\}$ is a family of density functions which is M.L.R. in T(x), and if the distribution function $H(t;\theta)$ of T(x) is continuous in t for each θ, then

(i) there exists a U.M.A. γ lower confidence limit $\underline{\theta}(T)$ for all γ, $0 < \gamma < 1$;

(ii) if the root $\hat{\theta}_\gamma(T)$ of the equation $H(T;\theta) = \gamma$ belongs to Θ, then $\underline{\theta}(T) = \hat{\theta}_\gamma(T)$.

Theorem 4.2 Let $\mathcal{P} = \{p(x,\theta); \theta \in \Theta\}$ be an M.L.R. family of density functions of an integer-valued discrete r.v. x. Let U be an independent r.v. having a uniform distribution on (0,1). Then the U.M.A.γ lower confidence limit for θ is the root $\underline{\theta}$ of the equation

$$UF(x;\theta) + (1-U)F(x-1;\theta) = \gamma$$

where $F(x;\theta)$ is the distribution function of x.

Theorem 4.3 If $\mathcal{P}$ is an M.L.R. family of one parameter distribution functions and $K_\gamma(x)$ is a U.M.A.γ upper (resp. lower) confidence limit for θ, then $\bar{L}_{\beta,\gamma}(x) = F^{-1}(\beta; K_\gamma(x))$ is a U.M.A. (β,γ) upper (resp. lower) tolerance limit.

Now one can see that to apply these theorems to the construction of confidence and tolerance limits based on a random sample $x_1,\ldots,x_n$ of independent observations on an M.L.R. family $\{p(x,\theta);\ \theta \in \Theta\}$, it will be necessary for the joint likelihood $L(x_1,\ldots,x_n;\theta) = \prod_{r=1}^{n} p(x_r,\theta)$ to possess an M.L.R. in some statistic $T(x_1,\ldots,x_n)$, in the sense that the ratio $\frac{L(x_1,\ldots,x_n;\theta_1)}{L(x_1,\ldots,x_n;\theta_2)}$ should be a nondecreasing function of $T(x_1,\ldots,x_n)$ for any $\theta_1 > \theta_2$, say

$$\frac{L(x_1,\ldots,x_n;\theta_1)}{L(x_1,\ldots,x_n;\theta_2)} = \rho(T(x_1,\ldots,x_n);\ \theta_1,\theta_2) \qquad (4.1)$$

If we fix $\theta_2 = \theta_o$ in equation (4.1), we have that

$L(x_1,\ldots,x_n;\theta) = \rho(T(x_1,\ldots,x_n);\theta,\theta_o)L(x_1,\ldots,x_n;\theta_o)$ for $\theta \geq \theta_o$,

and if we fix $\theta_1 = \theta_o$,

$L(x_1,\ldots,x_n;\theta) = \rho^{-1}(T(x_1,\ldots,x_n);\theta_o,\theta)L(x_1,\ldots,x_n;\theta_o)$ for $\theta < \theta_o$. Hence, by the well known factorisation theorem (Halmos and Savage (1949)), $T(x_1,\ldots,x_n)$ is a sufficient statistic for θ. Hence, theorems (4.1), (4.2) and (4.3) will not be applicable unless the M.L.R. family$\{p(x,\theta), \theta \in \Theta\}$ admits a sufficient statistic. Moreover, $T(x_1,\ldots,x_n)$ must be a single, one-dimensional, sufficient statistic and so, assuming certain regularity conditions are satisfied, $p(x,\theta)$ must be of the well known Darmois (1935) - Koopman (1936) - Pitman (1936) form,

$$p(x,\theta) = \exp\{A(x)B(\theta) + C(\theta) + D(x)\} \quad ; \quad x \in X, \; \theta \in \Theta \tag{4.2}$$

given that its spectrum σ_θ is independent of θ. This result was proved under various regularity conditions on $p(x,\theta)$ and/or $T(x_1,\ldots,x_n)$ by the above three authors. The work has been extended by Dynkin (1961) with emphasis on a dimensionality analysis of the log-likelihood function. Brown (1964) has given a more rigorous statement of the regularity conditions under which these extended results hold true. A very rigorous treatment of the dimensionality problem has been carried out by Barankin and Maitra (1963) who obtained necessary and sufficient conditions for results like (4.2) to be both locally and globally valid. We also need the form (4.2) in the case of a discrete-valued random variable and for this we can refer to Fraser (1963). This author displays a simple proof based on a method of likelihood, that fixed dimensionality of the sufficient statistic under increasing sample size is restricted to the exponential family, the analysis not being confined to continuous densities having a piecewise smooth derivative with respect to x, as in Dynkin's work. A recent paper which derives from (4.2) for a discrete-valued random variable possessing a sufficient statistic is that of Andersen (1970), who places all the regularity assumptions on the structure of the statistic.

In the case where σ_θ depends on θ, $\sigma_\theta = (\ell_1(\theta), \ell_2(\theta))$ for a continuous r.v. and $\sigma_\theta = \{\ell_1(\theta), \ell_1(\theta)+1, \ldots, \ell_2(\theta)\}$ for a discrete-valued r.v., the likelihood of a sample of n observations $x_1, x_2, \ldots, x_n$ can be written as

$$L(x_1,\ldots,x_n;\theta) = u(\ell_1(\theta)-x_{(1)})u(x_{(n)}-\ell_2(\theta)) \prod_{r=1}^{n} p(x_r,\theta) \qquad (4.3)$$

where $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are the order statistics from the sample and where $u(.)$ is that function which takes the value one if the argument is nonpositive and zero otherwise. It has been pointed out by Pitman (1936) and Davis (1951) that a factorisation of the likelihood (4.3) appropriate for the existence of a single, one-dimensional sufficient statistic is possible if and only if one of $\ell_1(\theta)$ or $\ell_2(\theta)$ is independent of θ and fixed, and $p(x,\theta)$ is of the form

$$p(x,\theta) = \frac{j(x)}{k(\theta)} \quad x \in \sigma_\theta. \qquad (4.4)$$

If $\ell_2(\theta)$ is fixed and $\ell_1(\theta)$ varies with $\theta \in \Theta$, the single sufficient statistic will be $T(x_1,\ldots,x_n) = x_{(1)}$; and if $\ell_1(\theta)$ is fixed and the upper limit $\ell_2(\theta)$ varies with θ_1, than the sufficient statistic is given by $T(x_1,\ldots,x_n) = x_{(n)}$. It is easily seen that, for each of these cases, the likelihood as given by (4.3) and (4.4) will possess an M.L.R. in the appropriate statistic $T(x_1,\ldots,x_n)$.

Thus we have shown that requiring the likelihood of a sample to possess an M.L.R. in a statistic T puts the very narrow constraint (4.4) on the form of our density or probability function $p(x,\theta)$ if σ_θ depends on θ, and it is easily seen that densities of type (4.4) possess an M.L.R. in x. If σ_θ does not depend on θ, then we are restricted to an exponential family of type (4.2) and so the likelihood of a sample,

$$L(x_1,\ldots,x_n;\theta) = \exp\{B(\theta)\sum_{r=1}^{n} A(x_r)+nC(\theta)+\sum_{r=1}^{n} D(x_r)\}, \tag{4.5}$$

possesses an M.L.R. in the statistic $T(x_1,\ldots,x_n)=\sum_{r=1}^{n} A(x_r)$ provided $B(\theta)$ is nondecreasing in $\theta \in \Theta$, or in the statistic $T(x_1,\ldots,x_n) = -\sum_{r=1}^{n} A(x_r)$ if $B(\theta)$ is nonincreasing for $\theta \in \Theta$. So we can apply theorems (4.1), (4.2) and (4.3) to any $p(x,\theta)$ of type (4.2), provided that $B(\theta)$ is a monotonic function of $\theta \in \Theta$. It should be noted that this does not imply that $p(x,\theta)$ has an M.L.R. in x (though it does of course have an M.L.R. in $A(x)$) unless the function $A(x)$ is monotonic in the same manner as $B(\theta)$.

Some examples of families possessing an M.L.R. in x, but for which one cannot apply theorems (4.1) to (4.3) to the likelihood of a sample are

$$p_1(x,\theta) = \begin{cases} mx^{m-1}/\theta & 0 < x < \theta,\ \theta > 0,\ m \text{ a positive integer} \\ 0 & \text{elsewhere} \end{cases}$$

$$p_2(x,\theta) = \begin{cases} \frac{1}{m!}(x-\theta)^m e^{-(x-\theta)} & x \geq \theta,\ \theta \in R,\ m \text{ a nonnegative integer} \\ 0 & \text{elsewhere} \end{cases}$$

$$p_3(x,\theta) = \frac{1}{\pi} \text{sech } (x-\theta) \qquad x,\ \theta \in R.$$

For M.L.R. families such as these we seek other confidence and tolerance intervals which are easily calculated and which we hope possess some optimum properties.

Following Rao (1965, p.384) let $x_1,\ldots,x_n$ be n observations on the M.L.R. family $\{p(x,\theta);\ \theta \in \Theta\}$ of type (2.1). Then the locally most powerful one-sided test of H_o: $\theta = \theta_o$ against H_1: $\theta > \theta_o$ at level α is

$$\frac{\sum\limits_{r=1}^{n} \frac{\partial}{\partial\theta} p(x_r, \theta_o)}{p(x_r, \theta_o)} \geq k, \tag{4.6}$$

i.e. $$\sum_{r=1}^{n} g(x_r, \theta_o) \geq k \tag{4.7}$$

where k is such that the type I error is α.

Under regularity conditions discussed in Section 2,

$$E_{\theta_o} \left\{ \frac{\frac{\partial}{\partial\theta} p(x, \theta_o)}{p(x, \theta_o)} \right\} = E_{\theta_o} \{g(x, \theta_o)\} = 0$$

and $$V_{\theta_o} \left\{ \frac{\frac{\partial}{\partial\theta} p(x, \theta_o)}{p(x, \theta_o)} \right\} = E_{\theta_o} \{g^2(x, \theta_o)\} = i(\theta_o),$$

where $i(\theta_o)$ is Fisher's information at θ_o.

If n is large, by an application of the central limit theorem,

$$\frac{1}{\sqrt{ni(\theta_o)}} \sum_{r=1}^{n} g(x_r, \theta_o) \sim N(0,1).$$

Hence if d_α is the upper α-point of an N(0,1) distribution, the approximate value of k in (4.7) is

$$k \simeq d_\alpha \sqrt{ni(\theta_o)}.$$

Hence the approximate one-sided "locally accurate" $(1-\alpha)$ confidence interval for θ based on the test (4.7) will be

$$\left\{ \theta \,\middle|\, \frac{\sum\limits_{r=1}^{n} g(x_r, \theta)}{\{E_\theta(g^2(x, \theta))\}^{\frac{1}{2}}} \leq d_\alpha \sqrt{n} \right\} \cap \{\theta \mid \theta \varepsilon \bigcap_{r=1}^{n} S_{x_r}\} \tag{4.8}$$

This can be taken as an implicit confidence interval for θ which will be easily calculated provided

$$\frac{\sum\limits_{r=1}^{n} g(x_r, \theta)}{\{E_\theta(g^2(x, \theta))\}^{\frac{1}{2}}} \tag{4.9}$$

is invertible as a function of θ. We have that

$$\frac{\partial}{\partial\theta}\left\{\frac{g(x,\theta)}{[E_\theta(g^2(x,\theta))]^{\frac{1}{2}}}\right\} = \frac{[E_\theta(g^2(x,\theta))]^{\frac{1}{2}}\frac{\partial}{\partial\theta}g(x,\theta) - \frac{1}{2}g(x,\theta)[\frac{\partial}{\partial\theta}E_\theta(g^2(x,\theta))][E_\theta(g^2(x,\theta))]^{-\frac{1}{2}}}{E_\theta(g^2(x,\theta))} \tag{4.10}$$

and the expectation of (4.10) taken for given θ will be strictly negative since

$$E_\theta(g(x,\theta)) = 0$$

and $E_\theta(\frac{\partial}{\partial\theta}\, g(x,\theta)) = -\, E_\theta(g^2(x,\theta)) < 0$,

provided we add the regularity condition that, for each $\theta \varepsilon \Theta$, there exists at least one value of $x \varepsilon X$ such that $g(x,\theta) > 0$. Hence for large sample size (4.9) will be invertible as a function of θ with probability close to one, and so the confidence interval (4.8) will be of the form

$$(\underline{\theta},\ \sup\Theta) \cap \bigcap_{r=1}^{n} S_{x_r} \tag{4.11}$$

where $\underline{\theta}$ is the asymptotically unique solution for $\theta \varepsilon \Theta$ to the equation

$$\frac{\sum_{r=1}^{n} g(x_r,\theta)}{\{E_\theta(g^2(x,\theta))\}^{\frac{1}{2}}} = d_\alpha\sqrt{n}$$

Upper confidence limits $\bar{\theta}$ can of course be obtained in a similar manner. It is to be noted that since we have taken the intersection of the ordinary upper confidence interval with the so-called selection statistic $\bigcap_{r=1}^{n} S_{x_r}$, the confidence coefficient will be at least as great as $1-\alpha$.

Also, if we test $H_o:\theta = \theta_o$ against $H_1:\theta \neq \theta_o$ we can set up the two-sided $(1-\alpha)$ confidence interval given by

$$\left\{\theta: \left|\frac{\sum_{r=1}^{n} g(x_r,\theta)}{[E_\theta(g^2(x,\theta))]^{\frac{1}{2}}}\right| \leq d_{\frac{\alpha}{2}}\sqrt{n}\right\} \cap \{\theta:\ \theta\ \varepsilon \bigcap_{r=1}^{n} S_{x_r}\} \tag{4.12}$$

The test on which this confidence interval is based is not locally most powerful unbiased but has been shown to have some optimum properties in large samples - see Rao (1948(b)), Rao and Poti (1946) and Wald (1941).

Assuming that we have a $\gamma = 1-\alpha$ two-sided confidence interval for θ, we now wish to find a (β,γ) tolerance interval, i.e. we wish to find $\underline{L}_{\beta,\gamma}(x_1,\ldots,x_n)$ and $\bar{L}_{\beta,\gamma}(x_1,\ldots,x_n)$ such that

$$P_\theta\{P_\theta[\underline{L}_{\beta,\gamma}(x_1,\ldots,x_n) \leq x \leq \bar{L}_{\beta,\gamma}(x_1,\ldots,x_n)] \geq \beta\} \geq \gamma \tag{4.13}$$

Let the two-sided γ confidence interval be $[\theta_o,\theta_1]$ where θ_o and θ_1 are both functions of $x_1,\ldots,x_n$ derived from (4.12). Let $F(x,\theta)$ be the c.d.f. of $p(x,\theta)$. Since $F(x,\theta)$ is a non-decreasing function of x for each $\theta \in \Theta$, we can define $\hat{b}$ to be the smallest point in X such that

$$F(\hat{b},\theta_1) = Pr(x \leq \hat{b}|\theta_1) \geq \beta_1 \tag{4.14}$$

and $\hat{a}$ to be the largest point in X such that

$$F(\hat{a},\theta_o) = Pr(x \leq \hat{a}|\theta_o) \leq \beta_o < \beta_1. \tag{4.15}$$

Lehmann (1959, p74) has shown that if $p(x,\theta)$ has an M.L.R. in x, then $F(x,\theta)$ is a nonincreasing function of the parameter for each $x \in X$.

Hence $\beta_o < \beta_1 \leq F(\hat{b},\theta_1) \leq F(\hat{b},\theta_o)$

and so, by (4.15), $F(\hat{a},\theta_o) < F(\hat{b},\theta_o)$

$$\Rightarrow \hat{a} < \hat{b}$$

Hence $P_\theta(\hat{a} \leq X \leq \hat{b}) \geq \beta_1 - \beta_o$ for $\theta \in [\theta_o,\theta_1]$

$= \beta$ if we put $\beta_1 = \frac{1}{2}(1+\beta)$

and $\beta_o = \frac{1}{2}(1-\beta)$, which is only one possibility, not necessarily optimal.

Hence $P_\theta\{P_\theta(\hat{a} \leq X \leq \hat{b}) \geq \beta\} \geq \gamma$, as required. One-sided tolerance intervals can be constructed from the "locally accurate" one-sided confidence intervals in the same manner.

5.5 A RELATED PROBLEM

In this section we shall investigate the conditions under which a joint density function given by

$$p(\underset{\sim}{x},\theta) = \frac{1}{(2\pi)^{n/2}|\Lambda^{-1}(\theta)|^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}\underset{\sim}{x}'\Lambda(\theta)\underset{\sim}{x}\} \quad \theta \in \Theta \tag{5.1}$$

has a monotone likelihood ratio in real-valued statistic $T(\underset{\sim}{x})$. Here $\underset{\sim}{x}'=(x_1,x_2,\ldots,x_n)$ is a vector of an observations with a joint normal distribution, the parametric space Θ is an open interval in the real line R and $\Lambda^{-1}(\theta)$ is the positive definite variance-covariance matrix associated with $(x_1,x_2,\ldots,x_n)$.

Such a problem arises for example in the following context. Let V be a random variable with distribution function $F(v)$ and characteristic function given by

$$\psi(t) = E(e^{itV}) = \int e^{itv}\, dF(v). \tag{5.2}$$

On the basis of m independent, identically distributed observations $V_1,V_2,\ldots,V_m$ we wish to test the goodness of fit hypothesis

$$H_o: F(v) = F_o(v) \tag{5.3}$$

where $F_o(v)$ is completely specified and V is symmetric about the origin. This is equivalent to the hypothesis

$$H_o': \psi(t) = \psi_o(t)$$

where $\psi_o(t)$ is the completely specified characteristic function of a random variable V symmetric about the origin. Such questions have been considered by Heathcote (1972), who proposed the test statistic

$$W_m(t) = \left(\frac{2}{m}\right)^{\frac{1}{2}} \sum_{j=1}^{m} \sin\left\{\frac{tV_j}{2}\right\} \tag{5.4}$$

which has the property that the vector $(W_m(t_1),\ W_m(t_2),\ldots,W_m(t_n))$ is asymptotically multivariate normal with zero mean vector.

Hence, for tests using the test statistic $(x_1, x_2, \ldots, x_n) = (W_m(t_1), W_m(t_2), \ldots, W_m(t_n))$ it would be of interest to establish the conditions under which a multivariate normal density of the type (5.1) possesses an M.L.R. in some statistic $T(\underset{\sim}{x}) = T(x_1, x_2, \ldots, x_n)$.

If such a statistic $T(\underset{\sim}{x})$ exists it will be a sufficient statistic for the family $\{p(x,\theta);\ \theta \in \Theta\}$, and so we expect some analogy to the Koopman-Pitman-Darmois form for $p(x,\theta)$. However, we cannot apply this well known result directly, as it is based on n independent identically distributed observations on a real-valued random variable. Since the random variable $\underset{\sim}{x}$ in the above problem takes values in the Euclidean space R^n, it will be necessary to develop a new method of characterisation.

We shall need the following lemma which is a generalisation to the multivariate case of Theorem (2.1).

Lemma (5.1) Let $p(\underset{\sim}{x},\theta)$ be the joint density of a random variable $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ taking values in R^n under the one-dimensional parameter $\theta \in \Theta$, an open interval in R. Assume that $\frac{\partial}{\partial\theta} \log p(\underset{\sim}{x},\theta)$ exists for all $\underset{\sim}{x}$ and all θ. Then a necessary and sufficient condition for $p(\underset{\sim}{x},\theta)$ to possess an M.L.R. in a real-valued statistic $T(\underset{\sim}{x})$ is that it can be written in the form

$$p(\underset{\sim}{x},\theta) = H(\underset{\sim}{x}) \exp\left\{\int_a^\theta G(T(\underset{\sim}{x}),u)\,du\right\} \qquad (5.5)$$

where $H(\underset{\sim}{x})$ is a positive, real-valued function of $\underset{\sim}{x}$, $G(T(\underset{\sim}{x}),\theta)$ is a real-valued, nondecreasing function of $T(\underset{\sim}{x})$ for each $\theta \in \Theta$ and $a = \inf \Theta$.

<u>Proof</u> The proof is a natural generalisation of that for theorem (2.1). In this case one considers the ratio $\frac{p(\underset{\sim}{x},\theta_1)}{p(\underset{\sim}{x},\theta_2)}$ for $\theta_1 > \theta_2$, which is known to be a nondecreasing function of the one-dimensional statistic $T(\underset{\sim}{x})$. Taking logarithms and limits as in theorem (2.1), it is easily shown that (5.5) is the necessary form for $p(x,\theta)$, where

$$G(T(\underset{\sim}{x}),\theta) = \frac{\partial}{\partial\theta} \log p(\underset{\sim}{x},\theta) \quad \forall \underset{\sim}{x} \in R^n, \theta \in \Theta \qquad (5.6)$$

and $G(T(\underset{\sim}{x}),\theta)$ is a nondecreasing function of $T(\underset{\sim}{x})$ for each $\theta \in \Theta$. In this case the additional complication of the spectrum depending on θ is absent.

Q.E.D.

Using this lemma, we can now state and verify a necessary and sufficient condition on the form of $\Lambda(\theta)$ for $p(\underset{\sim}{x},\theta)$ as given by (5.1) to have an M.L.R. in a statistic $T(\underset{\sim}{x})$, provided certain regularity assumptions are satisfied.

Theorem (5.1) Suppose that $p(\underset{\sim}{x},\theta)$ is given by (5.1) and that $\frac{d\Lambda(\theta)}{d\theta}$ exists on Θ. Consider a statistic $T(\underset{\sim}{x})$, a function from R^n into R, such that there exists $t_o \in R$ such that there $\{\underset{\sim}{x}: T(\underset{\sim}{x}) = t_o\}$ contains amongst its elements some $\frac{n(n+1)}{2}$ distinct points $\underset{\sim}{x}_r = (x_{r_1}, x_{r_2}, \ldots, x_{r_n})$ belonging to R^n which possess the property that the matrix Z has full rank, where the r^{th} row of Z, written in vector form, is given by

$$Z_r = (x_{r_1}^2, x_{r_2}^2, \ldots, x_{r_n}^2, x_{r_2}x_{r_1}, x_{r_3}x_{r_1}, x_{r_3}x_{r_2}, x_{r_4}x_{r_1}, x_{r_4}x_{r_2},$$
$$x_{r_4}x_{r_3}, \ldots, x_{r_n}x_{r(n-2)}, x_{r_n}x_{r(n-1)}\} \qquad (5.7)$$

for each $r=1,2,\ldots,\frac{n}{2}(n+1)$.

Assume also that $\exists\, \underset{\sim}{x}_o \in \{x: T(\underset{\sim}{x}) = t_o\}$ such that

$$\underset{\sim}{x}_o' \frac{d\Lambda(\theta)}{d\theta} \underset{\sim}{x}_o \neq 0 \qquad \theta \in \Theta. \qquad (5.8)$$

Then a necessary and sufficient condition for $p(\underset{\sim}{x},\theta)$ to possess an M.L.R. in the statistic $T(\underset{\sim}{x})$ is that $\Lambda(\theta)$ should be of the form

$$\Lambda(\theta) = \phi(\theta)M+N \qquad \theta \in \Theta \qquad (5.9)$$

where $\phi(\theta)$ is a real-valued, differentiable and monotonic function of θ, and where M and N are $n\times n$ matrices independent of the value of θ in Θ.

Proof If the form (5.9) for $\Lambda(\theta)$ is satisfied, it follows that

$$\frac{\partial}{\partial\theta}\log p(\underset{\sim}{x},\theta) = -\tfrac{1}{2}\frac{d}{d\theta}\log|\Lambda^{-1}(\theta)| - \tfrac{1}{2}\,\underset{\sim}{x}'\frac{d\Lambda(\theta)}{d\theta}\underset{\sim}{x}$$

$$= -\tfrac{1}{2}\frac{d}{d\theta}\log|\Lambda^{-1}(\theta)| - \tfrac{1}{2}\frac{d\phi(\theta)}{d\theta}\underset{\sim}{x}'M\underset{\sim}{x} \qquad (5.10)$$

The right hand side of (5.10) is a nondecreasing function of the real-valued statistic $T(\underset{\sim}{x}) = \pm\,\underset{\sim}{x}'M\,\underset{\sim}{x}$ according as $\frac{d\phi(\theta)}{d\theta}$ is nonpositive for all θ in Θ or nonnegative for all θ in Θ. Hence, by an application of lemma (5.1), the joint density $p(\underset{\sim}{x},\theta)$, as given by the equation (5.1), possesses an M.L.R. in $T(\underset{\sim}{x})$.

To prove the necessity of the form (5.9) for $\Lambda(\theta)$, we note that $p(\underset{\sim}{x},\theta)$, as given by (5.1), must also be of the form

$$p(\underset{\sim}{x},\theta) = H(\underset{\sim}{x})\exp\left\{\int_a^\theta G(T(\underset{\sim}{x}),u)\,du\right\} \quad \underset{\sim}{x}\in R^n,\ \theta\in\Theta \qquad (5.11)$$

This follows from Lemma (5.1), where $H(\underset{\sim}{x})$ is a positive, real-valued function of $\underset{\sim}{x}$, $a = \inf\Theta$ and $G(T(\underset{\sim}{x}),\theta)$ is a non-decreasing function of $T(\underset{\sim}{x})$ for each θ in Θ. After equating the two forms (5.1) and (5.11) for $p(\underset{\sim}{x},\theta)$, taking logarithms and differentiating with respect to θ, it follows that $\Lambda(\theta)$ satisfies the differential equation

$$G(T(\underset{\sim}{x}),\theta) = \tfrac{1}{2}\frac{d}{d\theta}\log|(\Lambda(\theta))| - \tfrac{1}{2}\,\underset{\sim}{x}'\frac{d\Lambda(\theta)}{d\theta}\underset{\sim}{x} \qquad \theta\in\Theta \quad (5.12)$$

We do not attempt to solve equation (5.12), but we wish to make use of it in order to derive the form of $\Lambda(\theta)$. If we define a new function $K(T(\underset{\sim}{x}),\theta)$ of $T(\underset{\sim}{x})$ and θ according to

$$K(T(\underset{\sim}{x}),\theta) = \frac{d}{d\theta} \log|\Lambda(\theta)| - 2\, G(T(\underset{\sim}{x}),\theta), \qquad (5.13)$$

it follows that $K(T(\underset{\sim}{x}),\theta)$ is a nonincreasing function of $T(\underset{\sim}{x})$ for each $\theta \in \Theta$. Also, equation (5.12) can then be rearranged to yield

$$K(T(\underset{\sim}{x}),\theta) = \underset{\sim}{x}' \frac{d\Lambda(\theta)}{d\theta} \underset{\sim}{x} \quad \underset{\sim}{x} \in R^n, \quad \theta \in \Theta. \qquad (5.14)$$

It follows from equation (5.14) that, for each $\theta \in \Theta$,

$$\{\underset{\sim}{x}: T(\underset{\sim}{x}) = t_o\} \subset \{\underset{\sim}{x}: \underset{\sim}{x}' \frac{d\Lambda(\theta)}{d\theta} \underset{\sim}{x} = K(t_o,\theta)\} \qquad (5.15)$$

By the assumption (5.8) we have that $T(\underset{\sim}{x}_o) = t_o$ and $\underset{\sim}{x}_o' \frac{d\Lambda(\theta)}{d\theta} \underset{\sim}{x}_o \neq 0 \;\forall\; \theta \in \Theta$. Hence we can conclude from equation (5.14) that

$$K(t_o,\theta) \neq 0 \;\forall \theta \in \Theta. \qquad (5.16)$$

Consider $\{\underset{\sim}{x}: \underset{\sim}{x}' \frac{d\Lambda(\theta)}{d(\theta)} \underset{\sim}{x} = K(t_o,\theta)\}$ for any value of $\theta \in \Theta$. It can be written in the form

$$\{\underset{\sim}{x}: \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\theta)\, x_i x_j = 1\}, \qquad (5.17)$$

where x_i is the i^{th} component of x for $i=1,2,\ldots,n$ and $a_{ij}(\theta)$ is the $(i,j)^{th}$ entry of the matrix $\frac{1}{K(t_o,\theta)} \frac{d\Lambda(\theta)}{d\theta}$ for $i,j=1,2,\ldots,n$. Furthermore, since $\underset{\sim}{x}_r \in \{\underset{\sim}{x}: T(\underset{\sim}{x}) = t_o\}$ for each $r=1,2,\ldots,\frac{n}{2}(n+1)$ it follows from the inclusion (5.15) that

$\underset{\sim}{x}_r \in \{\underset{\sim}{x}: \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\theta)x_i x_j = 1\}$ for each $r=1,2,\ldots,\frac{n}{2}(n+1)$ and for any $\theta \in \Theta$.

Hence $$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\theta)x_{ri}x_{rj} = 1 \qquad r=1,2,\ldots,\frac{n}{2}(n+1) \qquad (5.18)$$

Equations (5.18) can be regarded as a set of $\frac{n}{2}(n+1)$ equations in the $\frac{n}{2}(n+1)$ unknown coefficients a_{ij}, where $i \geq j$ and $i=1,2,\ldots,n$. [Note that from the symmetry of $\Lambda(\theta)$ and hence of $\frac{d\Lambda(\theta)}{d\theta}$, we have that $a_{ij} = a_{ji}$ for $i < j$ and $j=1,2,\ldots,n$. Therefore there are only $\frac{n}{2}(n+1)$ unknowns as given above, and not n^2 unknowns as it might at first appear.] Now, the assumption that the matrix Z is of full rank implies that the $\frac{n}{2}(n+1)$ equations of (5.18) are linearly independent and so there is a unique solution for each of the coefficients $a_{ij}(\theta)$; and moreover, this unique solution will be independent of the value of θ since it depends only on the fixed points $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_{\frac{n}{2}(n+1)}$. This implies that, for any $\theta \in \Theta$, $\frac{1}{K(t_o,\theta)} \cdot \frac{d\Lambda(\theta)}{d\theta} = M$ where M is an $n\times n$ matrix independent of θ and hence $\Lambda(\theta)$ can be written in the form

$$\Lambda(\theta) = \phi(\theta)M+N, \qquad \theta \in \Theta, \tag{5.19}$$

where N is also an $n\times n$ matrix independent of θ, and $\phi(\theta)$ is differentiable on Θ. Equation (5.14) now becomes

$$K(T(\underset{\sim}{x}),\theta) = \frac{d\phi(\theta)}{d\theta}\ \underset{\sim}{x}'M\underset{\sim}{x} \qquad \underset{\sim}{x} \in R^n, \quad \theta \in \Theta,$$

the left hand side of which we know to be a nonincreasing function of $T(\underset{\sim}{x})$ for each $\theta \in \Theta$. This implies that $\phi(\theta)$ is monotonic on Θ because $T(\underset{\sim}{x})$ can be taken, without loss of generality, to be a scalar multiple of $\underset{\sim}{x}'M\underset{\sim}{x}$(Such a scalar must be negative if $\phi(\theta)$ is nondecreasing on Θ and positive if $\phi(\theta)$ is nonincreasing on Θ). Equation (5.19) is therefore the required result. Q.E.D.

The awkward regularity conditions on $T(\underset{\sim}{x})$ in the hypothesis of Theorem (5.1) may seem rather artificial and so we shall now present a similar result derived under somewhat more natural conditions.

Theorem (5.2) Let $p(\underset{\sim}{x},\theta)$ be given by (5.1) and let $T(\underset{\sim}{x})$ be a mapping from R^n into R. Assume that $\frac{\partial}{\partial\theta}\log p(x,\theta)$ exists and is differentiable with respect to $T(\underset{\sim}{x})$ for all $(x,\theta)\in R^n\times\Theta$. Let $\theta_o \in \Theta$. Then a necessary and sufficient condition for $p(\underset{\sim}{x},\theta)$ to possess an M.L.R. in $T(\underset{\sim}{x})$ such that

$G(T(\underset{\sim}{x}),\theta_o) = \frac{1}{2}\frac{d}{d\theta}\log|\Lambda(\theta_o)| - \frac{1}{2}\underset{\sim}{x}'\frac{d\Lambda(\theta_o)}{d\theta}\underset{\sim}{x}$ is a strictly increasing function of $T(\underset{\sim}{x})$, is that $\Lambda(\theta)$ can be written in the form

$$\Lambda(\theta) = \phi(\theta)\ M+N \qquad \theta \in \Theta \qquad (5.20)$$

where $\phi(\theta)$, M and N are as given in Theorem (5.1) except that $\phi(\theta)$ possesses the additional property that $\frac{d\phi(\theta_o)}{d\theta} \neq 0$.

<u>Proof</u> The sufficiency of the form (5.20) follows as in Theorem (5.1) with the additional remark that since $\frac{d\phi(\theta_o)}{d\theta} \neq 0$ it can be seen from equation (5.10) that $G(T(\underset{\sim}{x}),\theta_o) = \frac{1}{2}\frac{d}{d\theta}\log|\Lambda(\theta_o)| - \frac{1}{2}\underset{\sim}{x}'\frac{d\Lambda(\theta_o)}{d\theta}\underset{\sim}{x}$ is a strictly increasing function of $T(\underset{\sim}{x})$.

To prove the necessity of the form (5.20) we follow Theorem (5.1) to the point where we are able to write

$$K(T(\underset{\sim}{x}),\theta) = \underset{\sim}{x}'\frac{d\Lambda(\theta)}{d\theta}\underset{\sim}{x} \qquad \theta \in \Theta \qquad (5.21)$$

where $K(T(\underset{\sim}{x}),\theta)$ is a nonincreasing function of $T(\underset{\sim}{x})$ for each $\theta \in \Theta$ and is strictly decreasing in $T(\underset{\sim}{x})$ for $\theta = \theta_o$. Since $G(T(\underset{\sim}{x}),\theta) = \frac{\partial}{\partial\theta}\log p(\underset{\sim}{x},\theta)$ is differentiable with respect to $T(\underset{\sim}{x})$ for each θ by assumption, so is $K(T(\underset{\sim}{x}),\theta)$.

For $\theta = \theta_o$ we may invert equation (5.21) to obtain

$$T(\underset{\sim}{x}) = K_{\theta_o}^{-1}\left[\underset{\sim}{x}'\frac{d\Lambda(\theta_o)}{d\theta}\underset{\sim}{x}\right]. \qquad (5.22)$$

Now let us define $M = \frac{d\Lambda(\theta_o)}{d\theta}$, a mapping T_o from R^n into R given by

$$T_o(\underset{\sim}{x}) = \underset{\sim}{x}'M\underset{\sim}{x} \qquad (5.23)$$

and a bivariate function on $R\times\Theta$ given by

$$K_o(\cdot,\theta) = K\left\{K_{\theta_o}^{-1}(\cdot),\theta\right\} \qquad (5.24)$$

From (5.22), (5.23) and (5.24) it follows that $T_o(\underset{\sim}{x}) = \underset{\sim}{x}'M\underset{\sim}{x}$

is a well-defined statistic satisfying

$$K_o(T_o(\underset{\sim}{x}),\theta) = \underset{\sim}{x}' \frac{d\Lambda(\theta)}{d\theta} \underset{\sim}{x} \qquad \theta \in \Theta \tag{5.25}$$

where $K_o(T_o(\underset{\sim}{x}),\theta)$ is a nonincreasing function of $T_o(\underset{\sim}{x})$ for each $\theta \in \Theta$ and differentiable with respect to $T_o(\underset{\sim}{x})$ for each $\theta \in \Theta$.

Now equation (5.25) can be written as

$$K_o(T_o(\underset{\sim}{x}),\theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \lambda'_{ij}(\theta) \tag{5.26}$$

where $\Lambda(\theta) = (\lambda_{ij}(\theta))$ and $\lambda'_{ij}(\theta) = \frac{d}{d\theta} \lambda_{ij}(\theta)$.

Differentiating (5.26) with respect to x_k we have

$$\frac{\partial K_o(T_o(\underset{\sim}{x}),\theta)}{\partial T_o(\theta)} \cdot \frac{\partial T_o(\underset{\sim}{x})}{\partial x_k} = 2 \sum_{i=1}^{n} \lambda'_{ik}(\theta)\, x_i \tag{5.27}$$

$$k = 1,2,\ldots,n$$

where we have used the fact that $\Lambda(\theta)$ (and hence $\frac{d\Lambda(\theta)}{d\theta}$) is a symmetric matrix, being the inverse of a variance - covariance matrix. Let $M = (m_{ij})$. Then $T_o(\underset{\sim}{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} x_i x_j$ and so equation (5.27) becomes, since M is also symmetric,

$$\frac{\partial K_o(T_o(\underset{\sim}{x}),\theta)}{\partial T_o(\underset{\sim}{x})} \cdot 2 \sum_{i=1}^{n} m_{ik} x_i = 2 \sum_{i=1}^{n} \lambda'_{ik}(\theta) x_i \tag{5.28}$$

Hence

$$\frac{\sum_{i=1}^{n} \lambda'_{ik}(\theta) x_i}{\sum_{i=1}^{n} m_{ik}\, x_i} = \frac{\partial K_o(T_o(\underset{\sim}{x}),\theta)}{\partial T_o(\underset{\sim}{x})} \qquad k=1,2,\ldots,n \quad \underset{\sim}{x} \neq \underset{\sim}{0} \tag{5.29}$$

The right hand side of (5.29) does not vary with k and so, if we put all the components of x equal to zero except the j-th which we put equal to one, then we can write, for $m_{jk} \neq 0$,

$$\frac{\lambda'_{jk}(\theta)}{m_{jk}} = \phi'_j(\theta) \qquad j=1,2,\ldots,n \quad k=1,2,\ldots,n \tag{5.30}$$

where $\phi_j'(\theta)$ is an integrable function of θ, negative for all $\theta \in \Theta$ and independent of the value of k. If $m_{jk} = 0$, then, by (5.28), $\lambda_{jk} = 0$. Hence

$$\lambda_{jk}'(\theta) = \phi_j'(\theta) m_{jk} \qquad j=1,2,\ldots,n \text{ and } k=1,2,\ldots,n \tag{5.31}$$

However, both $\frac{d\Lambda(\theta)}{d\theta}$ and M are symmetric matrices, i.e. $\lambda_{jk}'(\theta) = \lambda_{kj}'(\theta)$ and $m_{jk} = m_{kj}$ for all j and k, and so it follows that

$$\phi_1'(\theta) = \phi_2'(\theta) = \ldots = \phi_n'(\theta) \equiv \phi'(\theta).$$

Hence $\frac{d\Lambda(\theta)}{d\theta} = \phi'(\theta)M$ and so $\Lambda(\theta) = \phi(\theta)M+N$ where $\phi(\theta)$ has nonpositive derivative on Θ and M,N are $n\times n$ matrices of constants such that $\Lambda(\theta)$ is positive definite. We could have chosen $T_o(\underset{\sim}{x}) = -\underset{\sim}{x}'M\underset{\sim}{x}$, so that in general we may conclude that $\phi(\theta)$ is monotonic and differentiable with respect to θ. Also, since $G(T(\underset{\sim}{x}),\theta_o)$ is strictly increasing in $T(\underset{\sim}{x})$, it follows that $\frac{d\phi(\theta_o)}{d\theta} \neq 0$.

Q.E.D.

It is interesting to note that Theorem (5.1) can be extended to a characterisation of $\Lambda(\theta)$ such that $p(\underset{\sim}{x},\theta)$ possesses a one-dimensional sufficient statistic $T(\underset{\sim}{x})$ rather than an M.L.R. in $T(\underset{\sim}{x})$, provided the same regularity assumptions on $T(\underset{\sim}{x})$ are made and provided as before that $\frac{\partial}{\partial\theta}\log p(\underset{\sim}{x},\theta)$ exists on $R^n\times\Theta$. This last assumption means that $p(\underset{\sim}{x},\theta)$ will have the form (5.5) given by Lemma (5.1),

$$p(\underset{\sim}{x},\theta) = H(\underset{\sim}{x}) \exp\left\{\int_a^\theta G(T(\underset{\sim}{x}),u)\,du\right\},$$

except that $G(T(\underset{\sim}{x}),\theta)$ will not necessarily be a monotonic function of $T(\underset{\sim}{x})$ in this case. Thus we are able to develop the same form (5.9) for $\Lambda(\theta)$ by following the proof of Theorem (5.1) and omitting the questions of monotonicity of $K(T(\underset{\sim}{x}),\theta)$ and $\phi(\theta)$. We state this result as a corollary to

Theorem (5.1).

Corollary (5.1) Suppose all the assumptions of Theorem (5.1) are satisfied. Then a necessary and sufficient condition for $T(\underset{\sim}{x})$, a mapping from R^n into R, to be a sufficient statistic for $p(\underset{\sim}{x},\theta)$ is that $\Lambda(\theta)$ should be of the form $\Lambda(\theta) = \phi(\theta)M+N$, where M and N are as given in Theorem (5.1) and $\phi(\theta)$ is a differentiable function of $\theta \in \Theta$

Similarly, one can extend Theorem (5.2) to a characterisation of $\Lambda(\theta)$ such that $T(\underset{\sim}{x})$ is a sufficient statistic for $p(\underset{\sim}{x},\theta)$.

Corollary (5.2) Suppose all the assumptions of Theorem (5.2) are satisfied. Then a necessary and sufficient condition for $T(\underset{\sim}{x})$ to be a sufficient statistic for $p(\underset{\sim}{x},\theta)$ is that $\Lambda(\theta)$ should be of the form $\Lambda(\theta) = \phi(\theta)M+N$, where M and N are as given in Theorem (5.1) and $\phi(\theta)$ is a differentiable function of $\theta \in \Theta$ satisfying the further condition $\frac{d\phi(\theta_o)}{d\theta} \neq 0$.

So we have seen that requiring $p(\underset{\sim}{x},\theta) = \frac{1}{(2\pi)^{n/2}|\Lambda^{-1}(\theta)|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\, \underset{\sim}{x}'\Lambda(\theta)\underset{\sim}{x}\}$ to possess an M.L.R. in $T(\underset{\sim}{x})$, or merely to possess a one-dimensional sufficient statistic $T(\underset{\sim}{x})$, places a rather narrow constraint on the form of $\Lambda(\theta)$ and hence on the variance-covariance matrix $\Lambda^{-1}(\theta)$ of the observations. In fact, this form for $\Lambda^{-1}(\theta)$ is not generally enjoyed by the test statistic $(x_1,\ldots,x_n) = (W_m(t_1),\ldots,W_m(t_n))$ discussed at the beginning of this section. Nevertheless, it is possible that the results derived in this section, as well as being of some interest in themselves, may be of more use in reference to some other application.

REFERENCES

ANDERSEN, E.B. (1970): Sufficiency and exponential families for discrete sample spaces. *J. Amer. Stat. Assoc., 65*, 1248-55.

ANSCOMBE, F.J. (1961): Estimating a mixed-exponential response law. *J. Amer. Stat. Assoc., 56*, 493-502.

BARANKIN, E.W. and MAITRA, A. (1963): Generalisations of the Fisher-Darmois-Koopman-Pitman theorems on sufficient statistics. *Sankhya A25*, 217-44.

BARNARD, G.A. (1954): Sampling inspection and statistical decisions. *J. Roy.Stat.Soc. 16B*, 151-174.

BHATTACHARYA, C.G. (1967): A simple method of resolution of a distribution into Gaussian components. *Biometrics 23*, 115-35.

BLISCHKE, W.R. (1962): Calculation of C.E.P. for a mixed population. (Unpublished Memo).

____________ (1963): Mixtures of discrete distributions. *Proceedings of the International Symposium on the Classical and Contagious Distributions.* Montreal, Pergamon Press, 351-72.

BOES, D.C. (1966): On the estimation of mixing distributions. *Ann. Math. Stat. 37*, 177-88.

BROWN, L.D. (1964): Sufficient statistics in the case of independent random variables. *Ann. Math.Stat. 35*, 1456-74.

BUCHANAN-WOLLASTON H.G. and HODGESON, W.C. (1929): A new method of treating frequency curves in fishery statistics, with some results. *J. Conservation 4*, 207-25.

BURRAU, C. (1934): The half-invariants of the sum of two typical laws of errors, with an application to the problem of dissecting a frequency curve into components. *Skandinavisk Aktuarietidskrift 17*, 1-6.

CASSIE, R.M. (1954): Some uses of probability paper in the analysis of size frequency distributions. *Australian J. Marine and Freshwater Research 5*, 513-23.

CHARLIER C.V.L. (1906): Researches into the theory of probability. *Acta.Univ. Lund.* (Neue Folge. Abt. 2) 1, 33-38.

CHARLIER, C.V.L. and WICKSELL, S.D. (1924): On the dissection of frequency functions. *Arkiv. för Matematik, Astronomi och Fysik*, Bd 18, No. 6.

COHEN, A.C. (1967): Estimation in mixtures of two normal distributions. *Technometrics* 9, 15-28.

COX, D.R. (1959): The analysis of exponentially distributed life-times with two types of failure. *J. Roy. Stat. Soc.* 21B, 411-21.

CRAMÈR, H. (1946): Mathematical Methods of Statistics. Princeton; Princeton University Press.

DARMOIS, G. (1935): Sur les lois de probabilité à estimation exhaustive. *Comptes Rendues Acad. Sci.*, Paris 200, 1265.

DAVIS, R.C. (1951): On minimum variance in nonregular estimation. *Ann. Math. Stat.* 22, 43.

DYNKIN, E.B. (1961): Necessary and sufficient statistics for a family of probability distributions. *Uspeh Matem. Nauk 6*. English translation appeared in *Math. Stat. & Prob., Vol 1.*

FELLER, W. (1943): On a general class of "contagious" distributions. *Ann. Math. Stat.* 14, 389-400.

FRASER, D.A.S. (1963): On sufficiency and the exponential family. *J. Roy. Stat. Soc.* 25, 115-23.

GHOSH, B, K. (1970): *Sequential Tests of Statistical Hypotheses.* Addison Wesley.

GRUBBS, F.E. (1969): Procedures for detecting outlying observations in samples. *Technometrics* 11, 1-22.

GUTTMAN, I. (1970): *Statistical Tolerance Regions: Classical and Bayesian.* Griffin.

HALD, A. (1949): Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuarietidskrift* 32, 119-134.

_______ (1952): *Statistical Theory with Engineering Applications.* Wiley, New York.

HALMOS, P.R. and SAVAGE, L.J. (1949): Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* 20, 225-41.

HARDING, J.P. (1949): The use of probability paper for the graphical analysis of polymodal frequency distributions. *J. Marine Biological Assoc.* 28, 141-153.

HASSELBLAD, V. (1966): Estimation of parameters for a mixture of normal distributions. *Technometrics* 8, 431-444.

HEATHCOTE, C.R. (1972): A test of goodness of fit for symmetric random variables. *Australian J. Stat.* 14, 172-181.

JOHN, S. (1970(a)): On identifying the population of origin of each observation in a mixture of observations from two normal populations. *Technometrics 12*, 553-63.

________ (1970(b)): On identifying the population of origin of each observation in a mixture of observations from two gamma populations. *Technometrics 12*, 565-68.

________ (1970(c)): On analysing mixed samples. *J. Amer. Stat. Assoc. 65*, 755-62.

KAO, J.H.K. (1959): A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. *Technometrics 1*, 389-407.

KARLIN, S. (1956): Decision theory for Pólya-type distributions. Case of two actions I. *Proc 3rd Symp. Berkeley Math. Stat. Prob. 1*, 115-128.

________ (1957(a)): Pólya-type distributions II. *Ann. Math. Stat. 28*, 281-308.

________ (1957(b)): Pólya-type distributions III: Admissability for multi-action problems. *Ann. Math. Stat. 28*, 839-60.

________ (1958): Pólya-type distributions IV: some principles for selecting a single procedure from a complete class. *Ann. Math. Stat. 29*, 406-36.

________ (1968): *Total Positivity.* Stanford U.P.

KARLIN, S. and RUBIN, H. (1956): The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Stat. 27*, 272-300.

KENDALL, M.G. and STUART, A. (1967): *Advanced Theory of Statistics, Vol II*, 2nd Ed. Griffin.

________________ (1969): *Advanced Theory of Statistics, Vol I*, 3rd Ed. Griffin.

KOOPMAN, B.O. (1936): On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc. 39*, 399.

LEHMANN, E.L. (1959): *Testing Statistical Hypotheses.* Wiley.

MALLOWS, C.L. (1962): Medgyessy, P. "Decomposition of Superpositions of Distribution Functions. (Book Review) *Biometrics 18*, 617.

MEDGYESSY, P. (1961): *Decomposition of Superpositions of Distribution Functions.* Publishing House of the Hungarian Academy of Sciences, Budapest.

MENDENHALL,W. and HADER,R.J. (1958): Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika* *45*, 504-520.

OKA,M. (1954): Ecological studies on the kidai by the statistical method II. On the growth of kidai (taires trumifrons). *Bull. Fac. Fish. Nagasaki* *2*, 8-25.

PEARSON,K. (1894): Contributions to the mathematical theory of evolution. *Philosoph. Trans.Roy.Soc.* *185*, 71-110.

PITMAN,E.J.G. (1936): Sufficient statistics and intrinsic accuracy. *Proc.Camb. Philosoph. Soc.* *32*, 567.

POLLARD,H.S. (1934): On the relative stability of the median and arithmetic mean, with particular reference to certain frequency distributions which can be dissected into normal distributions. *Ann. Math. Stat.* *5*, 227-262.

PRESTON,E.J. (1953): A graphical method for the analysis of statistical distributions into two normal components. *Biometrika* *40*, 460-464.

RAO,C.R. (1948(a)): The utilisation of multiple measurements in problems of biological classification. *J. Roy. Stat. Soc.* *10B*, 159-203.

________ (1948(b)): Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Camb. Philosoph. Soc.* *44*, 50-57.

________ (1952): *Advanced Statistical Methods in Biometric Research.* Wiley.

________ (1965): *Linear Statistical Inference and its Applications.* Wiley.

RAO,C.R. and POTI,S.J. (1946): On locally most powerful tests when alternatives are one-sided. *Sankhya* *7*, 439.

SITTIG,J. (1948): Superpositie van twee frequentierverdelingen *Statistica* *2*, 206-227.

STRÖMGREN, B. (1934): Tables and diagrams for dissecting a frequency curve into components by the half-invariant method. *Skandinavisk Aktuarietidskrift* *17*, 7-54.

TANAKA, S. (1962): A method of analysing of polymodal frequency distributions and its application to the length distribution of the Porgy, *Taires Tumifrons*. *J. Fish. Res. Bd. Can.* *19*, 1143-59.

TEICHER,H. (1960): On the mixture of distributions. *Ann. Math. Stat.* *31*, 55-73.

TEICHER, H. (1961): Identifiability of mixtures. *Ann. Math. Stat.* *32*, 244-48.

_________ (1963): Identifiability of finite mixtures. *Ann. Math. Stat.* *34*, 1265-1269.

VAGHOLKAR, M.K. (1959): The process curve and the equivalent mixed binomial with two components. *J. Roy. Stat. Soc.* *21B*, 63-66.

WALD, A. (1941): Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Stat.* *12*, 1-19.

WEICHSELBERGER, K. (1951): Über ein graphisches Verfahren zur Frennung von Mischverteilungen und zur Identifikation kupierter Normalverteilungen bei grossem Stichprobenumfang. *Metrika* *4*, 178-229.

ZACKS, S. (1970): Uniformly most accurate upper tolerance limits for monotone likelihood ratio families of discrete distributions. *Amer. Stat. Assoc.* *64*, 307-316.

_________ (1971): *The Theory of Statistical Inference.* Wiley.