THE THEORY OF ESTIMATION
IN ALGEBRAIC AND ANALYTIC EXPONENTIAL FAMILIES
WITH APPLICATIONS TO VARIANCE COMPONENTS MODELS

KRISHNAN UNNI

Stat-Math Division
INDIAN STATISTICAL INSTITUTE, CALCUTTA

A dissertation submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy.

March, 1978

ACKNOWLEDGEMENTS

With deep gratitude, I wish to acknowledge the guidance and encouragement I received from my thesis supervisor Prof. J.K. Ghosh.

I am indebted to Dr. Somesh Chandra Bagchi for critically reading an earlier version of the manuscript.

I am grateful to the staff and research scholars of the Stat-Math. division of the Indian Statistical Institute for their cooperation.

I wish to express my gratitude to Prof. A.M. Kagan for taking interest in my work.

I thank Mr. P. Sathisha Rao for typing the thesis neatly with great care and for his enthusiastic cooperation.

I thank my wife for her encouragement and for cheerfully bearing many inevitable inconveniences during the preparation of this work, and my parents for many, many things.

Krishnan Unni

PREFACE

There are many important statistical problems of the following kind. The family of probability measures $\mathcal{P}$ is parametrized by a vector parameter η varying in a q-dimensional domain. $\mathcal{P}$ can be represented as an exponential family of probability distributions with k canonical parameters where k is greater than q. The canonical parameters do not vary in a domain in R^k, but are restricted by polynomial or analytic equations. They vary on a curved surface defined by the polynomial or analytic equations within the natural parameter space of the exponential family. The present work is concerned with the problem of point estimation of parametric functions in such statistical problems. This work is done in the spirit of R.A. Wijsman, Ju. V. Linnik and A. M. Kagan.

In Chapter 1 we present the basic facts of the theory of exponential families. Several examples are given to indicate the importance of the kind of exponential families we study.

In Chapter 2 we prove a theorem of A.M. Kagan and V.P. Palamodov characterizing the class of uniformly minimum variance unbiased estimators in an exponential family dominated by the Lebesgue measure when the canonical parameters are restricted by polynomial equations. The proof we give brings about substantial simplifications in the original proof of Kagan and Palamodov. We use this theorem to prove two conjectures of J. K. Ghosh.

In Chapter 3 the variance components models, under the normality assumption, are treated as exponential families to characterize the uniformly minimum variance unbiased estimators. We consider this as a very important application of the theorem of Kagan and Palamodov. Explicit likelihood equations are also derived.

In Chapter 4 we extend and strengthen a result of A.M. Kagan on the inadmissibility of certain estimators which are functions of the minimal sufficient statistic. This result has an important application to a special type of location parameter family.

In Chapter 5 we prove an interesting theorem characterizing the uniformly minimum variance unbiased estimators in a family of normal distributions with an unknown integer mean.

As a corollary, the mean itself is shown to have no uniformly minimum variance unbiased estimator.

In Chapter 6 we discuss the problem of unbiased estimation in a censored gamma family of distributions.

CONTENTS

CHAPTER 0

PRELIMINARIES

A statistical model is a triplet $(X, \mathcal{B}, \mathcal{P})$ where X is a set, $\mathcal{B}$ a σ-algebra of subsets of X and $\mathcal{P}$ a family of probability measures on $(X, \mathcal{B})$. We shall be concerned only with Euclidean statistical models, i.e., X is a Borel subset of an N-dimensional real Euclidean space R^N and $\mathcal{B}$ is the Borel σ-algebra of X. A k-vector statistic $t = (t_1, \ldots, t_k)$ is a $\mathcal{B}$-measurable mapping $X \to R^k$. The mean or expectation (vector) and variance-covariance (matrix) of t with respect to $P \in \mathcal{P}$ will be denoted by $E_p t$ and $V_p t$ respectively.

If Θ is a subset of a Euclidean space R^k and if η is a one-one mapping from Θ onto $\mathcal{P}$ then η is said to be a parametrization of $\mathcal{P}$. The value of η at $\theta \in \Theta$ is denoted by P_θ and thus $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. In this case the mean and variance-covariance of the statistic t will be denoted by $E_\theta t$ and $V_\theta t$, rather than by $E_{P_\theta} t$. and $V_{P_\theta} t$.

Let $(X, \mathcal{B}, \mathcal{P})$ be a fixed statistical model with a fixed parametrization $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ The set Θ is called the set of parameters for the model (with respect to the given parametrization). A real valued function of the parameter (parametric function) g is said to be (unbiasedly) estimable if there is a statistic $G : X \to R$ such that $E_\Theta G = g(\Theta)$ for all Θ in Θ and in this case we call G an unbiased estimater of g. We shall denote by U_g the set of all unbiased estimators of the parametric function g. For the case $g(\Theta) \equiv 0$, U_0 will be the set of all unbiased estimators of zero.

$G \in U_g$ is a locally minimum variance unbiased (L.M.V.U.) estimator of $g(\Theta)$ at the point $\Theta_0 \in \Theta$ iff for all $G_1 \in U_g$ we have $V_{\Theta_0} G \leq V_{\Theta_0} G_1$, or equivalently, $E_{\Theta_0} G^2 \leq E_{\Theta_0} G_1^2$.

If $G \in U_g$ is an L.M.V.U. estimator of $g(\Theta)$ for each Θ in Θ, then G is called the uniformly minimum variance unbiased (U.M.V.U.) estimator of $g(\Theta)$.

The following well-known lemma will be used in the sequel. Its proof can be found in C.R. Rao (1965) p.257.

Lemma 0.1. An estimator $G \in U_g$ such that $E_{\Theta_0} G^2 < \infty$ is L.M.V.U. at $\Theta_0 \in \Theta$ iff $E_{\Theta_0}(G\,F) = 0$ for each $F \in U_0$ such that $E_{\Theta_0} F^2 < \infty$. //

Vectors are row or column vectors. The r-th coordinate of a vector x will be denoted by x_r. For a matrix A, A^T

is its transpose. $\langle x, y\rangle$ is the notation for the inner product of the vectors x and y. C is the complex field.

For two measures μ_1 and μ_2, $\mu_1 \ll \mu_2$ means μ_1 is dominated by μ_2. A relation holds except on a μ-null set is denoted by writing $[\mu]$ after the relation. Mostly the repetitious use of $[\mu]$ is omitted without mention.

The Laplace transform of a probability measure P on R^k is

$$(0.1) \qquad g(\theta) = \int_{R^k} e^{\langle\theta, t\rangle}\, dP$$

which is defined on the set $\text{dom } g = \{\theta \in R^k : g(\theta) < \infty\}$. If Θ is the interior of this set then g can be extended as an analytic function to the domain in C^k defined by

$$(0.2) \qquad \underset{\sim}{\Theta} = \{\xi : \xi = \theta + i\eta,\ \theta \in \Theta\ ,\quad \eta \in R^k\}$$

If $f(t)$ is a function on R^k, its Laplace transform is

$$(0.3) \qquad \hat{f}(\theta) = \int_{R^k} f(t)\, e^{\langle\theta, t\rangle}\, dt$$

defined on the set of $\theta \in R^k$ for which the integrol on the right hand side of (0.3) exists. If Θ is the interior of this set, then $f(\xi)$ is an analytic function on the domain $\underset{\sim}{\textcircled{H}}$ in C^k where $\underset{\sim}{\textcircled{H}}$ is as in (0.2).

If $P(z) = P(z_1, \ldots, z_k)$ is a polynomial with complex coefficients in the k variables $z_1, \ldots, z_k$ then the corresponding partial differential operator $P(D) = P(D_1, \ldots, D_k)$ is obtained by substituting D_j^i for z_j^i where $D_j^i f(t_1, \ldots, t_k) = (\delta^i/\delta t_j^i) f(t_1, \ldots, t_k)$. This gives a one-one correspondence between polynomials with complex coefficients and partial differential operators with complex constant coefficients.

The generalized Leibniz formula is given by

$$(0.4) \qquad P(D)\,[u(t).v(t)] = \sum_{\alpha} [D^{\alpha}u(t)]\,(\alpha!)^{-1}\,[P^{(\alpha)}(D)\,v(t)]$$

where $\alpha = (\alpha_1, \ldots, \alpha_k)$ is a multiindex and

$$D^{\alpha} = D_1^{\alpha_1} \cdots D_k^{\alpha_k}$$

$$P^{(\alpha)}(z) = D^{\alpha}P(z)$$

$$\alpha! = \alpha_1! \cdots \alpha_k!$$

For a proof, see Hormander (1963) p.10.

In (0.4) when $u(t) = e^{\langle \theta, t \rangle}$

$$P(D)\left[e^{\langle \Theta, t\rangle} v(t)\right] = \sum_{\alpha} D^{\alpha} e^{\langle \Theta, t\rangle} (\alpha!)^{-1} P^{(\alpha)} (D) v(t)$$

$$= e^{\langle \Theta, t\rangle} \sum_{\alpha} \Theta^{\alpha} P^{(\alpha)} (D) v(t) (\alpha!)^{-1}$$

From Taylor's expansion for polynomials

$$P(D + \Theta) = \sum_{\alpha} \Theta^{\alpha} P^{(\alpha)} (D) (\alpha!)^{-1}$$

Thus

(0.5) $$P(D)\left[e^{\langle \Theta, t\rangle} v(t)\right] = e^{\langle \Theta, t\rangle} \left[P(D + \Theta) v(t)\right] .$$

We shall need some elementary results from the theory of Schwartz distributions. See Donoghue, Jr. (1969).

The support of a function f is the closure of the set of points for which f is nonzero. For an open set U in R^k, $C_0^{\infty}(U)$ is the set of all functions on R^k supported by some compact set contained in U and whose partial derivatives of all orders exist. For $f(t) \in C_0^{\infty}(R^k)$,

(0.6) $$\int \left[P(-D) f(t)\right] e^{\langle \Theta, t\rangle} dt = P(\Theta) \hat{f}(\Theta)$$

The above result is easily proved by integrating by parts for $P(D) = D$ and then by induction for $P(D) = D^k$. Now, for a general $P(D)$, the result fullows from the linearity of an integral transformation.

We shall say that the sequence $f_n(t)$ in $C_o^\infty(R^k)$ converges in $C_o^\infty(R^k)$ to 0 iff (1) there is one compact set K such that $f_n(t) = 0$ for all $t \notin K$ and (2) $f_n(t)$ and its partial derivatives of all orders converge uniformly to 0.

A distribution G on R^k is a linear functional on the vector space $C_o^\infty(R^k)$ such that if $f_n \to 0$ in $C_o^\infty(R^k)$ then $G(f_n) \to 0$.

A locally integrable function $G(t)$ on R^k, i.e., a function $G(t)$ which is integrable on any compact subset of R^k, defines a distribution as

$$(0.7) \qquad G(f) = \int G(t)\, f(t)\, dt, \qquad f \in C_o^\infty(R^k).$$

When G is an arbitrary distribution it is customary to write $G(f)$ symbolically as in the right hand side of (0.7) even though there may not be a locally integrable function $G(t)$ corresponding to the distribution G. We shall follow this custom.

The Dirac's δ-function $\delta(x)$ and its translate $\delta_a(x) = \delta(x-a)$ are the symbolic functions corresponding to the distributions defined by

$$(0.8) \qquad \delta(f) = \int \delta(x)\, f(x)\, dx = f(0)$$

$$(0.9) \qquad \delta_a(f) = \int \delta(x-a)\, f(x)\, dx = f(a)$$

Notice that $\delta(x)$ and $\delta(x-a)$ are the symbolic densities of the one-point mass at 0 and a respectively with respect to the Lebesgue measure.

The partial differential operator $P(D)$ applied to the distribution G is defined by

$$(0.10) \qquad [P(D)\, G]\, f = G\, [P(-D)\, f], \qquad f \in C_0^\infty (R^k)$$

<u>Lemma</u> 0.2. Let G be a distribution on R such that $(d/dt)\, G = 0$. Then G is a constant.

<u>Proof</u>. Of course, the statement of the lemma means that G is the distribution defined by $G(f) = \int k\, f(t)\, dt$ for some constant k.

For each f in $C_0^\infty (R)$ we have

$$\int G\, (df/dt) = 0$$

Now, $h \in C_0^\infty (R)$ is the derivative of another function $f \in C_0^\infty (R)$ iff $\int h(t)\, dt = 0$.

Let $g \in C_0^\infty (R)$ such that $\int g(t)\, dt = 1$, for $f \in C_0^\infty(R)$

$$f(t) = [f(t) - \int f(x)\, dx.\, g(t)] + \int f(x)\, dx.\, g(t).$$

Since $f(t) - \int f(x)\,dx.\ g(t)$ is in $C_0^\infty(R)$ and its integral vanishes, it is the derivative of same function in $C_0^\infty(R)$. Therefore,

$$G(f) = \int f(x)\,dx.\ G(g).$$

Thus G corresponds to the constant $G(g)$.

The following lemma is immediate from the preceding result.

<u>Lemma</u> 0.3. If $(d^p/dt^p)G = 0$ for a distribution G on R, then G is a polynomial of degree less than p. //

<u>Lemma</u> 0.4. If $G(t_1, \ldots, t_k)$ is a locally integrable function on R^k and $(\delta/\delta t_1)G = 0$ in the sense of distributions on R^k, then G is independent of t_1.

<u>Proof</u>. Let $G(t_1, \ldots, t_n)$ be a locally integrable function and let for each $f(t) \in C_0^\infty(R^k)$

$$\int_{R^k} G(t_1, \ldots, t_k) \frac{\delta}{\delta t_1} f(t_1, \ldots, t_k) = 0$$

Now from the above, and also from Fubini's theorem,

$$\int_{R^{k-1}} \left[\int_R G(t_1, \ldots, t_k)\,(\delta/\delta t_1)\, f(t_1, \ldots, t_k)\,dt_1 \right] \times h(t_2, \ldots, t_k)\,dt_2 \ldots dt_k = 0$$

for any $h(t_2, \ldots, t_k) \in C_o^\infty (R^{k-1})$.

This implies that for any fixed $t_2^o, \ldots, t_k^o$ and for each $f(t) \in C_o^\infty (R^k)$

$$\int_R G(t_1, t_2^o, \ldots, t_k^o)(\delta/\delta t_1) f(t_1, t_2^o \ldots t_k^o)\, dt_1 = 0$$

Now from Lemma 0.2. we conclude $G(t_1, t_2^o, \ldots t_k^o)$ is a constant.

CHAPTER 1.

THE EXPONENTIAL FAMILIES

In this chapter we present the basic facts of exponential families. Information on exponential families can be found in Lehmann (1959), Linnik (1968), Chentsov (1966) and Barndorff-Nielsen (1970). We define algebraic and analytic exponential families and give some examples to indicate their importence. Apart from this, we follow Barndorff-Nielsen.

1.1. The Basic Facts

Let $(X, \mathcal{B}, \mathcal{P})$ be a fixed (Euclidean) statistical model.

Definition 1.1. $\mathcal{P}$ is said to be an exponential family provided there exists a σ-finite measure μ on $(X, \mathcal{B})$, a positive integer k, real valued functions c, $\theta_1, \dots, \theta_k$ on $\mathcal{P}$ and real valued measurable functions h, $t_1, \dots, t_k$ on X such that $h \geq 0$, $\mathcal{P} \ll \mu$ and for every $P \in \mathcal{P}$

$$(1.1) \qquad (dP/d\mu)(x) = c(P)\, h(x)\, e^{\langle \theta(P), t(x) \rangle} \quad [\mu]$$

where $\theta = (\theta_1, \dots, \theta_k)$ and $t = (t_1, \dots, t_k)$.

In this case (1.1) is called an exponential representation of the densities of $\mathcal{P}$ with respect to μ.

The probability measures in an exponential family are mutually equivalent. Let $P_0 \in \mathcal{P}$. Then for all $P \in \mathcal{P}$,

$$(1.2) \qquad dP/dP_0 = [c(P)/c(P_0)] \; e^{\langle \theta(P)-\theta(P_0), t \rangle} \qquad [\mathcal{P}]$$

When $\mathcal{P}$ is exponential then to any σ-finite measure μ dominating $\mathcal{P}$, there exists a representation of the form (1.1). For each dominating μ, let $k(\mu)$ denote the smallest integer such that the densities of the probability measures in $\mathcal{P}$ with respect to μ are representable as in (1.1). $k(\mu)$ is an integer which is independent of μ. This integer is called the order of the exponential family $\mathcal{P}$. An exponential family whose order is k is also called a k-variate exponential family. Any representation (1.1) where k is the order of $\mathcal{P}$ is said to be a minimal representation.

For proofs of the following theorems 1.2 through 1.5, see Barndorff-Nielsen (1970) or the references cited there.

<u>Theorem</u> 1.2. Let $\mathcal{P}$ be an exponential family of order k with (1.1) as a minimal representation. Let

$$(dP/d\mu)\,(x) = \overset{*}{c}(P)\,\overset{*}{h}(x)\; e^{\langle \overset{*}{\theta}(P),\; \overset{*}{t}(x)\rangle} \qquad [\mu]$$

be another representation. Then there exist two linear transformations A and $\bar{A}$ and vectors a and $\bar{a}$ such that

$$A\,t^{*} + a = t$$

$$\bar{A}\,\theta^{*} + \bar{a} = \theta$$

and

$$\bar{A}\,A^{T} = I_{k\times k}$$

<u>Theorem</u> 1.3. The representation (1.1) is minimal iff both of the following conditions are satisfied.

(1) $1, \theta_1, \ldots, \theta_k$ are linearly independent functions on $\mathcal{P}$.

(2) $1, t_1, \ldots, t_k$ are linearly independent functions on X.

<u>Theorem</u> 1.4. If (1.1) is a minimal representation then t is a minimal sufficient statistic.

<u>Theorem</u> 1.5. Let

$$(dP_\theta/dP_{\theta_0})\,(x) = c(\theta)\; e^{\langle \theta-\theta_0,\, t(x)\rangle}$$

Then the family of induced distributions of t, $\mathcal{P}^t = \{P_\theta^t\}$ is also an exponential family with order of $\mathcal{P}^t$ = order of $\mathcal{P}$, and

$$d\,P^t_\theta \,/\, d\,P^t_{\theta_0} = c(\theta)\, e^{\langle \theta-\theta_0, t\rangle}$$

For an exponential family $\mathcal{P}$, consider the representation (1.1) and let $\Theta_0 = \{\theta(P) : P \in \mathcal{P}\}$. The mapping $\theta : P \to \theta(P)$ from $\mathcal{P}$ into Θ_0 is one-one and onto, i.e., θ^{-1} is a parametrization of $\mathcal{P}$.

<u>Definition</u> 1.6. A parametrization of the exponential family $\mathcal{P}$, represented as in (1.1), by θ^{-1} is called a canonical parametrization of $\mathcal{P}$ and θ is called the canonical parameter of the representation (1.1). This parametrization is called minimal canonical if (1.1) is minimal and in this case θ is called the minimal canonical parameter. The statistic t occuring in the various possible representations (1.1) are called canonical statistics. t is said to be minimal canonical if it occurs in a minimal representation.

Let $\mathcal{P}$ have an exponential representation with respect to a σ-finite measure μ. Consider a minimal canonical parametrization of $\mathcal{P} = \{P_\theta : \theta \in \Theta_0\}$ where $\Theta_0 \subset R^k$. We shall denote by p_θ a chosen version of the density of P_θ with respect to μ. We have

(1.3) $$p_\theta(x) = c(\theta)\, e^{\langle \theta, t(x)\rangle}\, h(x)$$

If the Laplace transform

$$g(\theta) = \int e^{\langle\theta,\ t(x)\rangle}\ h(x)\ d\mu$$

is finite for $\theta \in R^k$, then p_θ in (1.3) is a probability density with $c(\theta) = [g(\theta)]^{-1}$. Thus for $\theta \in Ⓗ = \text{dom } g$, p_θ as defined in (1.3) is a probability density.

Definition 1.7. The set Ⓗ as defined above is called the natural parameter space (of the minimal canonical parameter) in the representation (1.3) of the exponential family $\mathcal{P}$.

The natural parameter space Ⓗ of $\mathcal{P}$ should be distinguished from its set of minimal canonical parameters $Ⓗ_0$

Theorem 1.8. The natural parameter space of an exponential family is a convex set.

For a proof of the above theorem, see Lehmann (1959) p.51.

The statistical models of importance in exponential families are of the following three types.

(1) Canonical Models. The set of minimal canonical parameters $Ⓗ_0$ coincides with the natural parameter space Ⓗ.

(2) <u>Convex Models.</u> Θ_0 is a convex subset of Θ.

(3) <u>Curved Models.</u> Θ_0 is a smooth connected manifold contained in Θ.

If q is the dimension of the manifold Θ_0 then $\mathcal{P}$ is called a q-dimensional exponential family. Notice that in canonical and convex models the order and dimension of an exponential family coincide.

The curved models arise in the following way. Let a statistical model be specified as $(X, \mathcal{B}, P_\lambda)_{\lambda \varepsilon \Lambda}$ where the family of probability measures is parametrized by an open connected set Λ in R^q. Let the family $\mathcal{P} = \{P_\lambda : \lambda \varepsilon \Lambda\}$ be an exponential family. Suppose in a minimal canonical parametrization θ^{-1}, the functions $\theta_1(\lambda), \ldots, \theta_k(\lambda)$ are smooth with differentials of full rank at each point. Now the set of minimal canonical parameters $\Theta_0 = \theta(\mathcal{P})$ has the structure of a q-dimensional smooth manifold embedded in the k-dimensional Euclidean space.

We mention in passing that Effron (1970) has made use of one-dimensional curved exponential families in defining the curvature of a statistical problem.

Definition 1.9. A set $\underline{N} \subset C^k$ is called an algebraic set in C^k if it is the set of common zeroes of a finite number of polynomials with complex coefficients in k variables An analytic set in C^k is a set of the form .

$$U \cap \{z \in C^k : f_1(z) = 0, \ldots, f_r(z) = 0\}$$

where U is an open set in C^k and $f_1(z), \ldots, f_r(z)$ are analytic at least on U.

Definition 1.10. A k-variate exponential family $\mathcal{P}$ is called an algebraic (analytic) exponential family if the set of minimal canonical parameters of $\mathcal{P}$ is of the form $\Theta_0 = \Theta_1 \cap \underline{M}$ where $\underline{M}$ is an algebraic (analytic) set in C^k and Θ_1 is an open set in R^k contained in the natural parameter space of $\mathcal{P}$.

Lemma 1.11. Let t have an exponential family of densities

$$p_\theta(t) = c(\theta) \; e^{\langle \theta, t \rangle} \; h(t)$$

with respect to the Lebesgue measure on R^k. We assume that $h(t) > 0$ for all t in some open set S in R^k. Then if the canonical parameters satisfy a nontrivial polynomial equation $P(\theta_1, \ldots, \theta_k) = 0$ then t is not a complete statistic.

<u>Proof</u>. We can find $f(t) \epsilon C_0^\infty (S)$ such that

$$F(t) = \begin{cases} P(-D)\, f(t)/\, h(t)\,, & t \epsilon S \\ 0 \,, & t \not\epsilon S \end{cases}$$

is nontrivial. Now

$$E_\Theta F = C(\Theta) \int_S P(-D)\, f(t)\, e^{\langle\Theta,\, t\rangle}\, dt$$

$$= C(\Theta)\, P(\Theta)\, \hat{f}(\Theta)$$

which vanishes for all canonical parameters Θ.

It follows that t is not a complete statistic. //

The technique used to construct a nontrivial unbiased estimater of zero in the above lemma is known as Wijsman's D-method. See Wijsman (1958), where he uses this method to construct test functions satisfying the property of similarity in some important problems. Wijsman's D-method will find applications in the sequel.

1.2. Examples of Algebraic and Analytic Exponential Families

Example 1.12. Sampling from a normal distribution with a known coefficient of variation. Let $x_1, \ldots, x_N$ be independent and indentically distributed (i.i.d.) random variables and let $x_i \sim N(\mu, \sigma^2)$, i.e., x_i has a normal distribution with mean μ and variance σ^2. We assume that the coefficient of variation $r = \mu/\sigma$ is known but σ is unknown. The family of joint distributions of the sample $(x_1, \ldots, x_N)$, parametrized by $\sigma > 0$, is an exponential family. It has a minimal representation with respect to the Lebesgue measure μ^N on R^N, given by

$$(1.4) \qquad p_\sigma(x) = (2\pi\sigma^2)^{-N/2} \exp\left[-(1/2\,\sigma^2) \sum_{i=1}^{N} (x_i-\mu)^2\right]$$

$$= (2\pi\sigma^2)^{-N/2} \exp(-\mu^2/2\,\sigma^2) \exp\left[(-1/2\sigma^2) \sum_{i=1}^{N} x_i^2 + (\mu/\sigma^2) \sum_{i=1}^{N} x_i\right]$$

The minimal canonical parameters are $\theta_1 = -1/2\sigma^2$, $\theta_2 = \mu/\sigma^2$ and the minimal canonical statistics are $t_1 = \sum_{i=1}^{N} x_i^2$ and $t_2 = \sum_{i=1}^{N} x_i$. The exponential family (1.4) is a two-variate exponential family of dimension 1. The natural parameter space of (1.4) is $\Theta = \{(\theta_1, \theta_2) : \theta_1 < 0, \ \theta_2 \in R\}$ and the set of minimal canonical parameters is $\Theta_0 = \Theta \cap \underline{M}$ where $\underline{M} = \{(z_1, z_2) : z_1, z_2 \in C \text{ and } z_2^2 + 2r^2 z_1 = 0\}$. Thus the family (1.4) is an algebraic exponential family.

The minimal canonical statistic $t = (t_1, t_2)$ has the density

$$(1.5) \qquad p_\theta(t) = c(\theta) \ e^{\langle\theta, t\rangle} \ h(t)$$

with respect to the Lebesgue measure on R^2, where

$$(1.6) \qquad h(t_1, t_2) = \begin{cases} (t_1 - t_2^2)^{N/2 - 1} , & t_1 \geq t_2^2 \\ 0 , & t_1 < t_2^2 \end{cases}$$

<u>Example</u> 1.13. <u>The Behrens - Fisher Problem</u>.

Let $x_1, \ldots, x_{N_1}$ and $y_1, \ldots, y_{N_2}$ be two independent sets of repeated samples from $N(\mu, \sigma_1^2)$ and $N(\mu, \sigma_2^2)$ respectively where the unknown parameters are μ and σ^2, $\mu \in R$ and $\sigma^2 > 0$. The joint density of the sample (x, y), with respect to the Lebesgue measure on $R^{N_1+N_2}$, is

$$(1.7) \qquad p_{\mu,\sigma_1^2,\sigma_2^2}(x, y) = (2\pi)^{-(N_1+N_2)/2} \exp\Big[-(1/2\sigma_1^2) \sum_{i=1}^{N_1} (x_i-\mu)^2 - (1/2\sigma_2^2) \sum_{i=1}^{N_2} (y_i-\mu)^2\Big]$$

$$= (2\pi)^{-(N_1+N_2)/2} \exp\left[-(\mu^2/2\sigma_1^2) - (\mu^2/2\sigma_2^2)\right] \times$$

$$\exp\left[-(1/2\,\sigma_1^2) \sum_{i=1}^{N_1} x_i^2 + (\mu/\sigma_1^2) \sum_{i=1}^{N_1} x_i\right.$$

$$\left. -(1/2\,\sigma_2^2) \sum_{i=1}^{N_2} y_i^2 + (\mu/\sigma_2^2) \sum_{i=1}^{N_2} y_i\right]$$

(1.7) is a minimal representation with minimal canonical parameters $\theta_1 = -(1/2\,\sigma_1^2)$, $\theta_2 = (\mu/\,\sigma_1^2)$, $\theta_3 = -(1/2\,\sigma_2^2)$, $\theta_4 = (\mu/\sigma_2^2)$ and minimal canonical statistics $t_1 = \Sigma\, x_i^2$, $t_2 = \Sigma\, x_i$, $t_3 = \Sigma\, y_i^2$ and $t_4 = \Sigma\, y_i$. The natural parameter space of the family (1.7) is

$$\Theta = \{\theta : \theta_1 < 0,\ \theta_2 \in R,\ \theta_3 < 0,\ \theta_4 \in R\} .$$

The minimal canonical parameters of (1.7) satisfy the polynomial equation

$$(1.8) \qquad z_1\, z_4 - z_2\, z_3 = 0$$

so that (1.7) is an algebraic exponential family with parameter $\Theta_0 = \Theta \cap \underline{M}$ where $\underline{M}$ is the algebraic set defined by the equation (1.8). In this example we have a four-variate exponential family of dimension three.

<u>Example</u> 1.14. <u>Multinomial distributions with cell probabilities as functions cf a parameter</u>.

In genetics it is assumed that under random mating conditions an off-spring belongs to the genotypes AA, Aa or aa with probabilities p^2, $2pq$, q^2, where $0 < p < 1$ and $p + q = 1$.

In a fixed number N of individuals the probability that r_1 is AA, r_2 is Aa and r_3 is aa, is given by $[N! / r_1! \, r_2! \, r_3!] \, (p^2)^{r_1} \, (2pq)^{r_2} \, (q^2)^{r_3}$. Consider the family of probability distributions on the set of triplets of positive integers (r_1, r_2, r_3), $r_1 + r_2 + r_3 = N$, generated by varying p. The family has an exponential representation with respect to the counting measure λ on the set of triplets of positive integers. A minimal representation is given by the densities

$$(1.9) \qquad f_p = (N! / r_1! \, r_2! \, r_3!) \, e^{N \log 2pq} \, e^{r_1(\log p^2 - \log 2pq)} \times e^{r_3(\log q^2 - \log 2pq)}$$

with respect to λ, since the functions 1, $\log p^2 - \log 2pq$, $\log q^2 - \log 2pq$ are linearly independent and 1, r_1, r_3 are linearly independent. The family is of order 2 and dimension 1. The minimal canonical parameters are given by $\theta_1 = \log p^2 - \log 2pq$, $\theta_2 = \log q^2 - \log 2pq$ and the minimal canonical statistics are r_1 and r_2.

We have $e^{\theta_1} = p^2/\, 2pq = p/\, 2(1-p)$ and $e^{\theta_2} = (1-p)/\, 2p = (1/4)\, e^{\theta_1}$. The natural parameter space of the exponential family (1.9) is R^2 and the minimal canonical parameters form the set $(\theta_1, \theta_2) \in R^2 : e^{\theta_2} - (1/4)\, e^{-\theta_1} = 0$ so that the random mating model is an analytic exponential family

<u>Example</u> 1.15. <u>Location parameter exponential families</u>.

A location parameter exponential family of distributions on R is dominated by the Lebesgue measure μ, with respect to which, it has a density of the form

$$(1.10) \qquad f(x, \theta) = \exp \sum_{i=1}^{m} \exp\left[\alpha_i (x_i - \theta)\right] p_i(x-\theta)$$

where $\alpha_1, \ldots, \alpha_m$ are complex numbers and $p_1, \ldots, p_m$ are polynomials with complex coefficients. Naturally if $f(x)$ is to be a probability density then the complex constants in (1.10) must be chosen so that the function $f(x)$ is real, positive and satisfies the condition $\int f(x)\,dx = 1$. Under these conditions, conversely, formula (1.10) is the density of an exponential family of distributions in a location parameter θ. Soo Ferguson (1962). It is a univariate exponential family iff it is a family of the type $N(\theta, \sigma^2, \gamma)$, introduced by Ferguson, for some fixed positive σ^2 and real number γ. Otherwise, it is a multivariate exponential family of dimension 1 with the minimal canonical parameters varying over the curve

$$\theta \longrightarrow (\exp(-\alpha_1\theta)\theta, \quad \exp(-\alpha_1\theta)\theta^2, \quad \dots \quad , \quad \exp(-\alpha_2\theta)\theta, \quad \dots \quad)$$

When (1.10) is of the form $\exp[P(x-\theta)]$ for a polynomial P then we have an algebraic exponential family with the minimal canonical parameters satisfying the equations

$$\begin{aligned} \theta_2 - \theta_1^2 &= 0 \\ &\vdots \\ \theta_k - \theta_1^k &= 0 \end{aligned}$$

CHAPTER 2.

UNBIASED ESTIMATION IN ALGEBRAIC EXPONENTIAL FAMILIES

In this chapter we prove a theorem due to A.M. Kagan and V.P. Palamodov characterizing the class of uniformly minimum variance unbiased estimators in an algebraic exponential family dominated by the Lebesgue measure. The main arguments in our proof are based entirely on the ideas of A.M. Kagan and V.P. Palamodov but at the same time we bring about substantial simplifications in the details of the proof. See Kagan and Palamodov (1968). A point worth mentioning is that we completely avoid the use of some difficult results of H. Whitney and J. P. Serre in algebraic and analytic geometry which are used by Kagan and Palamodov. As a result, the proof given in this chapter should be more easily accessible to statisticians.

In 2.1. we prove the necessary results on polynomials of several variables which are needed for the proof of the main theorem. The results in this section also seem to be of independent interest. Given a set of polynomial constraints on C^k we show how to obtain the maximal subspace of

C^k where these constraints do not apply. We obtain the equations defining this subspace in terms of the ideal of all polynomials vanishing on the set of points satisfying the constraints.

In 2.2. we state and prove the main theorem. The statement is slightly changed from that of Kagan and Palamodov to make it suitable for direct applications.

In 2.3. the theorem is applied to prove two conjectures of J.K. Ghosh, namely, in the Behrens - Fisher problem and in the problem of normal samples with a known coefficient of variation the uniformly minimum variance unbiased estimators are trivial. An application of the theorem to variance components models is given in the next chapter.

2.1. Some Results on Polynomials of Several Variables

Polynomials in k variables with complex coefficients form a ring $\underline{\underline{P}}_k$ under the usual addition and multiplication of polynomials.

Definition 2.1. An ideal in the ring of polynomials $\underline{\underline{P}}_k$ is a set of polynomials $I \subset \underline{\underline{P}}_k$ which has the properties

(1) $P \in I, \quad Q \in I \Rightarrow P + Q \in I$

(2) $P \in I, \quad R \in \underline{\underline{P}}_k \Rightarrow RP \in I$

Definition 2.2. $A \subset \underline{\underline{P}}_k$ is a set of generators for the ideal I in $\underline{\underline{P}}_k$ if I is the smallest ideal in $\underline{\underline{P}}_k$ containing the set A. We denote this fact by $I = \text{id } A$.

Note that id A is the set of all finite sums of the form $\Sigma P_i Q_i$ where $P_i \in \underline{\underline{P}}_k$ and $Q_i \in A$.

The following two theorems in algebra are well-known.

Theorem 2.3. (Hilbert). Every ideal in $\underline{\underline{P}}_k$ has a finite set of generators. //

Let $\underline{\underline{N}}$ be an algebraic set in C^k defined as the set of common zeroes of the set of polynomials A. This is denoted by $\underline{N} = V(A)$. Notice $\underline{N} = V(\text{ id } A)$.

Theorem 2.4. (Hilbert's Nullstellensatz) Let $\underline{N} = V(A)$, $A \subset \underline{\underline{P}}_k$. Suppose $P \in \underline{\underline{P}}_k$ vanishes on $\underline{N}$. Then for some positive integer n, $P^n \in \text{ id } A$. //

The ideal of <u>all</u> polynomials vanishing on a set $\underline{N} \subset C^k$ will be denoted by id $\underline{N}$.

<u>Theorem</u> 2.5. Let $\underline{N}$ be an algebraic set in C^k and $I = \text{id } \underline{N}$. Let

$$(2.1) \qquad I_o = \text{id} \{P(z + \xi) : P(z) \in I, \quad \xi \in \underline{N}\}$$

and

$$(2.2) \qquad \underline{L} = V(I_o).$$

Then $\underline{L}$ is a linear subspace of C^k. It is the largest subspace of C^k such that

$$(2.3) \qquad \underline{L} + \underline{N} \subseteq \underline{N}.$$

<u>Proof</u>. Let

$$I_{\xi_o} = \{P(z + \xi_o) : P(z) \in I\}$$

Then $V(I_{\xi_o}) = \underline{N} - \xi_o$. Therefore

$$(2.4) \qquad \underline{L} = V(I_o) = \bigcap_{\xi \in \underline{N}} \underline{N} - \xi$$

To show $\underline{L}$ is closed under addition . Let ξ_1, ξ_2 be in $\underline{N}$. For arbitrary $\xi \in \underline{N}$, from (2.4), we can find $\eta_1, \eta_2 \in \underline{N}$ such that $\xi_1 = \eta_1 - \xi$ and $\eta_2 = \eta_2 - \eta_1$. Therefore, $\xi_1 + \xi_2 = \eta_2 - \xi \in \underline{L}$.

To show $\underline{L}$ is closed under scalar multiplication: Fix ξ in $\underline{L}$. From the previous part of the proof, for each positive integer n, $n\xi \in \underline{L}$. For $P \in I_0$, consider the polynomial in one complex variable α, $P_1(\alpha) = P(\alpha\xi)$. Since P_1 vanishes at all positive integers, $P_1 \equiv 0$, i.e., $P(\alpha\xi) = 0$ for each complex α. It follows that $\alpha\xi \in \underline{L}$ for each $\alpha \in C$.

Now, from (2.4) we have, for each ξ in $\underline{N}$

$$\underline{L} \subset \underline{N} - \xi$$

or

$$\underline{L} + \xi \subset \underline{N}.$$

Thus

$$\underline{L} + \underline{N} \subset \underline{N}.$$

On the other hand, if $\underline{M}$ is a subspace of C^k such that $\underline{M} + \underline{N} \subset \underline{N}$, then obviously $\underline{M}$ is contained in $\underline{N} - \xi$ for each ξ in $\underline{N}$. Thus $\underline{M} \subset \bigcap_{\xi \in \underline{N}} \underline{N} - \xi = \underline{L}$. Thus the theorem is proved. //

Let $\underline{N}$ be a fixed algebraic set in C^k. Let I, I_0, $\underline{L}$ be as before. We shall assume that we are working with an orthonormal basis $(e_1, \ldots, e_r, e_{r+1}, \ldots, e_k)$ for C^k so that $(e_1, \ldots, e_r)$ is a basis for $\underline{L}^{\perp}$ and $(e_{r+1}, \ldots, e_k)$ is a basis for $\underline{L}$. We shall identify $z = z_1 e_1 + \ldots + z_r e_r$

with the ordered r-tuple $(z_1, \ldots, z_r)$ and so on.

Let $\underline{V} = \underline{L}^{\perp} \cap \underline{N}$. If $(z_1^0, \ldots, z_k^0) \in \underline{N}$, then, from (2.3), $(z_1^0, \ldots, z_r^0) \in \underline{V}$ and for arbitrary $z_{r+1}, \ldots, z_k$, $(z_1^0, \ldots, z_r^0, z_{r+1}, \ldots, z_k) \in \underline{N}$. This shows that $\underline{N}$ can be identified with the cylinder $\underline{V} \times \underline{L}$.

let $P(z_1, \ldots, z_k) \in I$. Then the polynomial in $z_1, \ldots, z_r$, $P(z_1, \ldots, z_r, 0, \ldots, 0)$ will vanish on $\underline{V}$. It is also clear that such polynomials form the ideal I', in $\underline{\underline{P}}_r$, of all polynomials vanishing on $\underline{V}$. To see this, consider $P(z_1, \ldots, z_r) \in I'$ as a polynomial in $\underline{\underline{P}}_k$ with zero coefficients for $z_{r+1}, \ldots, z_k$. Then obviously P vanishes on $\underline{V} \times \underline{L}$ and consequently it is in I. In fact, I' is nothing but the set of all polynomials in $z_1, \ldots, z_r$ which are in I.

Lemma 2.6. I' generates I in $\underline{\underline{P}}_k$.

Proof. First, consider $P(z_1, \ldots, z_{r+1}) \in I$. We can put

$$(2.5) \qquad P(z_1, \ldots, z_{r+1}) = Q_0(z_1, \ldots, z_r) + z_{r+1} P_1(z_1, \ldots, z_{r+1})$$

Putting $z_{r+1} = 0$, we see that $Q_0 \in I'$. It follows that

$z_{r+1} P_1 \in I$. Now for any $z^o \in \underline{N}$ such that $z^o_{r+1} \neq 0$ P_1 must vanish. If $z^o_{r+1} = 0$, let $z^{(n)} = (z^o_1, \ldots, z^o_r, \frac{1}{n}, \ldots, z^o_k)$. From (2.3), $z^{(n)} \in \underline{N}$. Since $z^{(n)} \to z^o$ we have $P_1(z^o) = 0$. Thus $P_1 \in I$.

Decompose P_1 as in (2.5). Continuing this process, finally, we obtain

$$(2.6) \qquad P(z_1, \ldots, z_{r+1}) = Q_o(z_1, \ldots, z_r) + z_{r+1} Q_1(z_1, \ldots, z_r) + \ldots + z^n_{r+1} Q_n(z_1, \ldots, z_r)$$

where $Q_i \in I'$, $i = 1, \ldots, n$.

Thus P is in the ideal generated by I' in $\underline{\underline{P}}_k$.

Now, by induction on the number of variables, we obtain the result. //

Let

$$(2.7) \qquad I'_o = \text{id} \left\{ P(z + \xi) : P(z) \in I', \quad \xi \in \underline{V} \right\}.$$

In the above definition of I'_o, since $P(z)$ is a polynomial in $z_1, \ldots, z_r$ only, we can replace $\xi \in \underline{V}$ by $\xi \in \underline{V} \times \underline{L} = \underline{N}$. Arguing as in the proof of theorem 2.5., $V(I'_o)$ is a subspace of $\underline{L}$ which is contained in $\underline{V} - \xi$ for each $\xi \in \underline{N}$ and hence contained in $\underline{L}$. It follows that $V(I_o') = 0$. It is also clear that I_o' generates the ideal I_o in $\underline{P}_k$.

Lemma 2.7. The set of vectors

$$\{(\delta P/\delta z_1, \ldots, \delta P/\delta z_r)_o : P \in I_o'\}$$

spans $\underline{L}^{\perp}$.

Proof. Suppose $a \neq 0$ is in $\underline{L}$ and for each $P \in I_o'$, $a \perp (\delta P/\delta z_1, \ldots, \delta P/\delta z_r)_o$. Then for $P \in I'$ and $\xi \in \underline{V}$, the directional derivative $D_a P(\xi) = 0$. This shows $D_a P \in I'$. Repeating the above argument for $D_a P \in I'$, etc., finally we have for $\xi \in \underline{V}$, $D_a^n P(\xi) = 0$ for each positive integer n. This implies that I_o' vanishes on the vector space generated by the vector a. This is a contradiction. //

Theorem 2.8. I_o is the ideal of all polynomials vanishing on $\underline{L}$.

Proof. It is enough to show that the polynomials $z_1, \ldots, z_r$ are in I_o'.

From Lemma 2.7. we can choose $P_1, \ldots, P_r$ in I_o' so that the inverse function theorem can be applied to obtain an analytic isomorphism

$$(z_1,\dots,z_r) \quad \leftrightarrow \quad (P_1,\dots, P_r)$$

in a neighbourhood of zero. Let

(2.8) $$z_1 = F(P_1, \dots , P_r)$$

in a neighbourhood of zero, where

(2.9) $$F(P_1, \dots, P_r) = \Sigma\, a_{i_1\dots i_r} P_1^{i_1}\dots P_r^{i_r}.$$

If we expand the righthand side of (2.9) as a power series in $z_1, \dots, z_r$ then the coefficient of each form (monomial) $z_1^{i_1}, \dots , z_r^{i_r}$ must vanish, except that of z_1.

From Hilbert's Nullstellensatz, we can find m such that if a form in $z_1, \dots, z_r$ is of degree at least m then it is in I_o'. Also, except for a finite number of indices, all other $a_{i_1\dots i_r}\ P_1^{i_1} \dots P_r^{i_r}$ is a finite sum of forms in $z_1, \dots, z_r$ of degree at least equal to m. Let the sum $\Sigma'\, a_{i_1\dots i_r}\ P_1^{i_1} \dots P_r^{i_r}$ correspond to these finite number of indices. In this sum, the coefficients of each form, other than z_1, whose degree is less than m must vanish. So we have

$$\Sigma'\, a_{i_1 \dots i_r} P_1^{i_1} \dots P_r^{i_r} = z_1 + A(z_1,\dots,z_r)$$

where $A(z_1, \ldots, z_r)$ is a finite sum of forms in $z_1, \ldots, z_r$ of degree at least m. So $A(z_1, \ldots, z_r)$ is in I'_o. It follows that $z_1 \in I'_o$. //

2.2. Uniformly Minimum Variance Unbiased Estimation in Algebraic Exponential Families dominated by the Lebesgue Measure.

Let $(X, \mathcal{B}, \mathcal{P})$ be a statistical model where X is a Borel subset of R^N and $\mathcal{P}$ an algebraic exponential family dominated by the Lebesgue measure μ^N. In the canonical parametrization Θ^{-1} suppose $\mathcal{P}$ has a minimal representation

$$(2.10) \qquad p_\Theta(x) = c(\Theta) \; e^{\langle\Theta, \; t(x)\rangle} \; h_1(x), \qquad \Theta \in \Theta_o$$

with respect to μ^N.

Clearly, in the problem of unbiased estimation, we need consider only estimators which are functions of the sufficient statistic $t = (t_1, \ldots, t_k)$. Suppose the statistics $t_1, \ldots, t_k$ are continuously differentiable and functionally independent. Then the induced family of joint distributions of the statistic t is an exponential family, and has densities

(2.11) $$p_\Theta(t) = c(\Theta)\ e^{\langle\Theta,\ t\rangle}\ h(t), \quad \Theta \in \Theta_0$$

with respect to μ^k.

In the rest of this section we assume

(1) $t = (t_1, \ldots, t_k)$ has a family of densities of the form (2.11) where the set of canocical parameters $\Theta_0 = \Theta_1 \cap \underline{M}$. Θ_1 is an open set in R^k contained in the natural parameter space Θ and $\underline{M}$ is an algebraic set in C^k.

(2) $\mu^k \{t : h(t) > 0\} = \mu^k(S)$ where S is the interior of the set $\{t : h(t) > 0\}$.

(3) $h(t)$ is bounded away from zero on every compact subset of S.

If $h(t)$ is a continuous function then assumptions (2) and (3) are trivially satisfied.

The problem is to characterize the class of U.M.V.U. estimators in the family (2.11). Without loss of generality we assume all estimators vanish outside S.

Let K be any compact subset of S. From assumption (3), we can find $\delta > 0$ such that $h(t)\ e^{\langle\Theta,\ t\rangle}\ dt > \delta$ for all t

in K. Now

$$\infty > \int_K |G(t)|\ h(t)\ e^{\langle\Theta,\ t\rangle}\ dt > \delta \int_K |G(t)|\ dt.$$

Thus G is alocally integrable function and as in (0.7) G can be considered as a distribution.

Lemma 2.9. Let G(t) be an L.M.V.U. estimator for its expectation at the point $\Theta_o \varepsilon$ Θ_o. Then G(t) satisfies the partial differential equations (in the sense of distributions)

$$(2.12) \qquad P(D + \Theta_o)G = 0$$

for any polynomial $P(z_1, \ldots, z_k)$ (with complex coefficients) which vanishes on Θ_o.

Proof. Let P(z) e any polynomial vanishing on Θ_o. Take $f(t) \varepsilon C_o^\infty (S)$ and put

$$F(t) = P(-D)\ f(t)/\ h(t)\ , \qquad t \varepsilon S$$

Since

$$\int_S F(t)\ h(t)\ e^{\langle\Theta,\ t\rangle}\ dt = \int_S P(-D)\ f(t)\ e^{\langle\Theta, t\rangle}\ dt$$
$$= P(\Theta)\ \hat{f}(\Theta),$$

F(t) is an unbiased estimator of zero, i.e., $F(t) \varepsilon U_o$.

Since $h(t)$ is bounded away from zero on compact subsets of K, $E_\theta F^2 < \infty$ for each θ in H_0.

Now let $G(t)$ be L.M.V.U. at θ_0. From Lemma 0.1. we have for each $f(t) \in C_0^\infty(S)$

$$0 = \int G(t) \; [P(-D) f(t)] \; e^{\langle \theta_0, t \rangle} \; dt$$

$$= \int [P(D + \theta_0) G(t)] f(t) \; e^{\langle \theta_0, t \rangle} \; dt$$

The second part of the above equality follows from (0.5) and the definition of differential operator on distributions. Since $f(t) \to f(t) e^{\langle \theta_0, t \rangle}$ is one-one onto $C_0^\infty(S)$ and $G(t) = 0$ outside S,

$$P(D + \theta_0) G = 0 . \qquad //$$

<u>Theorem</u> 2.10. (A.M. Kagan and V.P. Palanodov). Let $\underline{N}$ be the smallest algebraic set in C^k containing Θ_0 and let L be the largest linear subspace of R^k such that $L + \underline{N} \subset \underline{N}$. An estimator $G(t)$, $E_\theta G^2 < \infty$ for all θ in Θ_0, is a U.M.V.U. estimator for its expectation iff $G(t)$ is a function of s only where s is the projection of t into L.

<u>Proof</u>. (1) <u>Sufficiency</u>. Let

$$t = r + s, \quad r \in L, \quad s \in L$$

$$\theta = \beta + \eta, \quad \beta \in L^{\perp}, \quad \eta \in L$$

Consider an estimator $G(s)$ such that $E_\theta G^2 < \infty$ for all θ in Θ_0. Let $F(t) \in U_0$ and $E_\theta F^2 < \infty$, $\theta \in \Theta_0$.

Fix $\theta_0 = \beta_0 + \eta_0$ in Θ_0. Let U_1 be a neighbourhood of β_0 in $L^{\perp}$ and U_2 a neighbourhood of η_0 in L such that $U_1 \times U_2 \subset \Theta_1$. Since

$$\beta_0 \times L \subset \underline{N}$$

we have

$$\beta_0 \times U_2 \subset \Theta_1 \cap \underline{N}.$$

Since $\underline{N}$ is the smallest algebraic set containing $\Theta_1 \cap \underline{M}$, where $\underline{M}$ is an algebraic set, it is clear that $\Theta_1 \cap \underline{N} = \Theta_1 \cap \underline{M} = \Theta_0$. Thus $\beta_0 \times U_2 \subset \Theta_0$.

Put

$$F_1(s) = \int_{L^{\perp}} F(r, s) \; e^{\langle \beta_0, r \rangle} \; h(r, s) \, dr$$

Since $F \in U_0$, we have

$$\int F_1(s) \; e^{\langle \eta, s \rangle} \; ds = 0$$

for all $\eta \in U_2$. This implies

$$F_1(s) = 0$$

Therefore, for all θ in Θ_o,

$$E_\theta [\, G(s)\ F(r) \,] = 0$$

which shows G is a U.M.V.U. estimator for its expection, from Lemma 0.1.

(2) <u>Necessity</u>. Suppose $G(t)$, $E_\theta G^2 < \infty$ for all θ in Θ_o, is a U.M.V.U. estimator for its expectation.

Consider a fixed polynomial P vanishing on $\underline{N}$ and a fixed $f(t) \in C_o^\infty (S)$. Now

$$\int [P(D + z)\ G(t)]\ f(t)\ dt = \int G(t)\ [P(-D + z)\ f(t)]\ dt$$

is a polynomial in z which, from Lemma 2.9., vanishes on Θ_o and consequently on $\underline{N}$. Thus, for all ξ in $\underline{N}$,

$$P(D + \xi)\ G = 0 \tag{2.13}$$

Now, let I be the ideal of all polynomials vanishing on $\underline{N}$ and

$$I_o = \text{id} \{P(z + \xi) : P(z) \in I, \quad \xi \in \underline{N}\}$$

Then, from (2.13), for all $P \in I_o$

$$P(D)\ G = 0 \tag{2.14}$$

Let

$$\underline{L} = V(I_o)$$

$$L = \underline{L} \cap R^k$$

where we assume we are working with an orthonormal basis $(e_1, \ldots, e_k)$ so that $(e_1, \ldots, e_r)$ forms a basis for $L^{\perp} = \underline{L}^{\perp} \cap R^k$ and $(e_{r+1}, \ldots, e_k)$ a basis for L. From Theorem 2.8. the polynomials $z_1, \ldots, z_r \in I_o$. Now from (2.14) $\delta G/ \delta t_1 = 0, \ldots, \quad \delta G/ \delta t_r = 0.$

This shows that G is independent of $t_1, \ldots, t_r$. The theorem is proved. //

<u>Remark</u> 2.11. If $L = \{0\}$, then the only U.M.V.U. estimators are constants.

<u>Remark</u> 2.12. The fact that the set of minimal canonical parameters Θ_o is of the form H_1 $\underline{M}$ is used in the sufficiency part of the proof. If Θ_o is any set and $\underline{N}$ the smallest algebraic set containing Θ_o then the necessity part of the theorem is true.

2.3. <u>Examples</u>

<u>Example</u> 2.13. In the problem of N independent samples from a normal distributions $N(\mu, \sigma^2)$ with a fixed standardized

mean, considered in Example 1.12., we have

$$\Theta_1 \cap \underline{M} = \{(\theta_1, \theta_2) : \theta_1 < 0, \ \theta_2 \in R, \ \theta_2^2 + 2r^2\theta_1 = 0\}$$

If $r = 0$, then $L = \{(\theta_1, \theta_2) : \theta_1 \in R, \ \theta_2 = 0\}$ and U.M.V.U. estimators are functions of t_1.

<u>Case</u> $r \neq 0$. For any fixed (β_1, β_2) such that $\beta_2^2 + 2r^2 \beta_1 = 0$, $\beta_1 < 0$, $\beta_2 \in R$ and for $(\theta_1, \theta_2) \in L$ we have $(\theta_2 + \beta_2)^2 + 2r^2(\theta_1 + \beta_1) = 0$, i.e., $\theta_2^2 + 2\theta_2\beta_2 + 2r^2\theta_1 = 0$.

Since $r \neq 0$, the above equality is true for any arbitrary $\beta_2 \in R$. But then if $\theta_2 \neq 0$ this equation determines β_2 uniquely. So $\theta_2 = 0$ and consequently $\theta_1 = 0$. It follows that $L = \{0\}$. So constants are the only U.M.V.U. estimators.

<u>Example</u> 2.14. <u>The Behrens-Fisher problem</u>.

See Example 1.13. We have

$$\Theta_0 = \{\theta : \theta_1 < 0, \ \theta_3 < 0, \ \theta \in R, \ \theta_4 \in R, \ \theta_1\theta_4 - \theta_2\theta_3 = 0\}$$

For $\theta \in L$, we have

$$(\theta_1 + \beta_1)(\theta_4 + \beta_4) - (\theta_2 + \beta_2)(\theta_3 + \beta_3) = 0$$

where $\beta_1, \beta_2, \beta_3, \beta_4$ are any numbers such that $\beta_1 < 0, \beta_3 < 0,$ $\beta_1\beta_4 - \beta_2\beta_3 = 0$. Now, taking $\beta_2 = 0, \quad \beta_4 = 0,$ we have for arbitrary $\beta_1 < 0$ and $\beta_3 < 0$,

$$\theta_1\theta_4 + \theta_4\theta_1 - \theta_2\theta_3 - \theta_2\theta_3 = 0.$$

For a fixed $(\theta_1, \theta_2, \theta_3, \theta_4) \varepsilon L$, the equation holds for arbitrary $\beta_1 < 0$ and $\beta_3 < 0$.

Since $0 \varepsilon \underline{M}$, 0 is a limit point of Θ_0 and so $0 \varepsilon \underline{N}$. Thus, for $\theta \varepsilon L$,

$$\theta_1\theta_4 - \theta_2\theta_3 = 0$$

So, for arbitrary $\beta_1 < 0, \quad \beta_3 < 0,$ we have

$$\theta_4\theta_1 - \theta_2\theta_3 = 0.$$

This implies $\theta_4 = 0, \quad \theta_2 = 0.$

Now, arguing as before, we can show $\theta_1 = 0, \quad \theta_3 = 0.$

It follows that the only U.M.V.U. estimators in the Behrens-Fisher problem are constants.

The same result is true also for the multivariate analogue of the Behrens-Fisher problem which can be proved by similar methods.

CHAPTER 3.

ESTIMATION IN VARIANCE COMPONENTS MODELS

Let y be a random p-vector with a multivariate normal distribution $N_p(0, G)$ where G is unknown but is assumed to be in a fixed convex cone of positive definite (p.d.) matrices. Let $y_1, \ldots, y_N$ be N independent observations of y. N may be one. We give a representation of the joint densities of $(y_1, \ldots, y_N)$ as an algebraic exponential family. Making use of this fact, an elegant characterization of the class of U.M.V.U. estimators is obtained. The same point of view also leads to an interesting derivation of the explicit likelihood equation.

The models we consider arise from the variance component models $x \sim N_n(X\beta, \sum_i \eta_i V_i)$ when we are interested in estimators which are invariant under translation by $X\beta$. Let $y = Px$ where P is the projection of y on the orthogonal complement of the column space of X. Then y is a maximal invariant for the group of translations by $X\beta$ and $y \sim N_p(0, PGP^T)$. In traditional variance components models N is one, but replicated models are also studied under the name linear covariance models and they are of interest in some problems in psychometry and econometrics. These models occur also in the study of autoregressive processes with moving average residuals.

The literature on variance components models is extensive. C.R. Rao (1971) has characterized the class of quadratic estimators which are of uniform minimum variance in the class of quadratic estimators. It turns out that for quadratic estimators uniform minimum variance in the class of all estimators is equivalent to uniform minimum variance in the class of quadratic estimators. U.M.V.U. estimators are functions of U.M.V.U. quadratic estimators. For U.M.V.U. quadratic estimators our criterion reduces to the condition given by Rao. The likelihood equation we derive resembles the one given by Herbach(1959) for the two way classification random effects model. The likelihood equation for the model is also derived by Srivastava(1966) for the case when all the covariance matrices are simultaneously diagonalizable, and by Anderson (1969) in the more general case. Our likelihood equation is explicit in the sense that the equation does not contain the inverse of the unknown matrix. One of the advantages of this is that an iterative solution by substitution is possible in our case whereas one has to resort to more difficult techniques to solve Anderson's equation.

Our use of Jordan algebra in the problem is influenced by

Seely (1975). Also, our techniques have something in common with those of Srivastava (1966).

3.1. Jordan Algebras of Matrices

Lemma 3.1. Let $\underline{\underline{A}}$ be a real linear space of $p \times p$ matrices with real entries. The following conditions on $\underline{\underline{A}}$ are all equivalent.

(1) $G \in \underline{\underline{A}} \Rightarrow G^2 \in \underline{\underline{A}}$

(2) $G_1, G_2 \in \underline{\underline{A}} \Longrightarrow G_1G_2 + G_2G_1 \in \underline{\underline{A}}$

(3) For each positive integer n,

$$G \in \underline{\underline{A}} \Longrightarrow G^n \in \underline{\underline{A}}$$

Definition 3.2. A real linear space $\underline{\underline{A}}$ of matrices satisfying any of the conditions given in Lemma 3.1. is said to form a (real) Jordan algebra of matrices.

Remark 3.3. In fact, such a linear space of matrices forms a Jordan algebra under the usual definition of a Jordan algebra with the usual matrix addition and the multiplication

---

$\circ$ defined by $A \circ B = (AB + BA)/2$

In the rest of this section we shall denote by $\underline{A}$ a fixed Jordan algebra of $p \times p$ symmetric matrices containing the identity matrix I.

<u>Lemma</u> 3.4. If $G \varepsilon \underline{A}$ and G is p.d. then $G^{-1} \varepsilon \underline{A}$.

<u>Proof</u>. Let $G \varepsilon \underline{A}$ and G be p.d. For sufficiently small $\alpha > 0$

$$(3.1) \qquad G^{-1} = \alpha [I - (I - \alpha G)]^{-1}$$

$$= \alpha [I + (I - \alpha G) + (I - \alpha G)^2 + \ldots]$$

For each positive integer k, $(I - \alpha G)^k \varepsilon \underline{A}$. Since $\underline{A}$ is a finite dimensional linear space, it is closed. Therefore, $G^{-1} \varepsilon \underline{A}$. //

A linear space of symmetric matrices can be made into an inner product space by the inner product $\langle A,B \rangle = \text{tr } AB$. Let $(G_1, \ldots, G_n)$ be an orthonormal basis with respect to this inner product, Now we can identify R^n and $\underline{A}$ by the isometry

$$(3.2) \qquad \eta = (\eta_1, \ldots, \eta_n) \longleftrightarrow G_\eta = \sum_{i=1}^{n} \eta_i G_i$$

Let Θ be the convex cone in R^n defined by

$$(3.3) \qquad \Theta = \{\eta \varepsilon R^n : G_\eta \varepsilon \underline{A} , \; G_\eta \text{ is p.d.}\}$$

For $\eta \in \textcircled{H}$, from Lemma 3.4., we have a unique $\rho \in \textcircled{H}$ such that

$$(3.4) \qquad (G_\eta)^{-\dagger} = G_\rho$$

Let $U : \textcircled{H} \longrightarrow \textcircled{H}$ be the map

$$(3.5) \qquad U(\eta) = \rho$$

It is clear that U^2 is the identity map on $\textcircled{H}$.

<u>Theorem</u> 3.5. Let U be the map defined in (3.5). Then U is a rational map, i.e., each coordinate of ρ is given by

$$(3.6) \qquad \rho_i = U_i(\eta_1, \ldots, \eta_n)$$

where U_i is a well-defined rational function of $\eta_1, \ldots, \eta_n$ on $\textcircled{H}$.

<u>Proof</u>. We shall assume $(G_i)\ i = 1, \ldots, n$ is an orthonormal basis for $\underline{\underline{A}}$ and $G_1 = (1/\sqrt{p})I$. The modification in the form of U_i when this is not the case is clear.

We have

$$(3.7) \qquad I = (G_\rho G_\eta + G_\eta G_\rho)/2$$

For a p.d. $G_\eta \in \underline{\underline{A}}$, if $G_\rho \in \underline{\underline{A}}$ satisfies (3.7) then it can be easily shown that

$$G_{\theta} = (G_{\eta})^{-1}$$

Now, the equation (3.7) can be written as

$$(3.8) \qquad I = \sum_i \theta_i \left[\sum_j \eta_j (G_i G_j + G_j G_i)/2 \right]$$

In terms of the coordinates with respect to the basis (G_k) this becomes

$$(3.9) \qquad 1 = \sum_i \theta_i \left[\sum_j \eta_j \operatorname{tr} (G_i G_j + G_j G_i)/2p \right]$$

$$0 = \sum_i \theta_i \left[\sum_j \eta_j \operatorname{tr} (G_i G_j + G_j G_i) G_k/2 \right], \quad k = 2, \ldots, n.$$

The equation (3.9) can be put in matrix form as

$$(3.10) \qquad A(\eta).\theta = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the (k, i)th element of the matrix $A(\eta)$ is

$$(3.11) \qquad a_{ki}(\eta) = (1/2) \sum_j \eta_j \operatorname{tr} \left[(G_i G_j + G_j G_i) G_k \right], \quad k = 2, \ldots, n$$

$$a_{1i}(\eta) = (1/2p) \sum_j \eta_j \operatorname{tr} \left[(G_i G_j + G_j G_i) \right]$$

For $\eta \in \textcircled{H}$, from Lemma 3.4., equation (3.10) has a unique solution $\theta \in \textcircled{H}$. From Cramer's rule for the solution of linear equations

(3.12) $$\theta_i = U_i(\eta_1, \ldots, \eta_n) = |A_{1i}(\eta)|/|A(\eta)|.$$

Obviously $|A_{1i}(\eta)|$ and $|A(\eta)|$ are polynomials in $\eta_1, \ldots, \eta_n$ and for $\eta \in \textcircled{H}$, $|A(\eta)| \neq 0$.

This proves the theorem. //

3.2. <u>The Minimal Exponential Representation</u>

In the following $\underline{\underline{S}}$ denotes a convex cone of $p \times p$ p.d. matrices. $\underline{\underline{V}} = \{G^{-1} : G \in \underline{\underline{S}}\}$. Let $\underline{\underline{S}}_0$ and $\underline{\underline{V}}_0$ denote the affine spaces generated by $\underline{\underline{S}}$ and $\underline{\underline{V}}$ respectively. Since $\underline{\underline{S}}$ is a convex cone the affine spaces are actually seen to be linear spaces. Unless otherwise stated, we shall assume $\underline{\underline{S}}$ is open in $\underline{\underline{S}}_0$ and $\underline{\underline{S}}$ contains the identity matrix I. In the sequel $\underline{\underline{A}}$ will denote the smallest Jordan algebra of symmetric matrices which contains $\underline{\underline{S}}_0$.

<u>Lemma</u> 3.6. $\underline{\underline{S}}_0 \subset \underline{\underline{V}}_0 \subset \underline{\underline{A}}$. $\underline{\underline{S}}_0 = \underline{\underline{V}}_0$ iff $\underline{\underline{S}}_0 = \underline{\underline{A}}$

<u>Proof</u>. Clearly $\underline{\underline{V}}_0 \subset \underline{\underline{A}}$. For $\alpha > 0$ and $G \in \underline{\underline{S}}$,

$I + \alpha G$ is in $\underline{\underline{S}}$. Now, for $\alpha > 0$ small enough,

$$(I + \alpha G)^{-1} = I - \alpha G + \alpha^2 G^2 - \ldots$$

which is in $\underline{\underline{V}}_0$. Therefore,

$$-\alpha G + \alpha^2 G^2 \ldots \in \underline{\underline{V}}_0$$

or,

$$G - \alpha G^2 + \ldots \in \underline{\underline{V}}_0$$

Now letting $\alpha \longrightarrow 0$ we obtain $G \in \underline{\underline{V}}_0$. Thus $\underline{\underline{S}}_0 \subset \underline{\underline{V}}_0 \subset \underline{\underline{A}}$.

From the same argument we also have, for $G \in \underline{\underline{S}}$, $G^2 \in \underline{\underline{V}}_0$. For $G_1, G_2 \in \underline{\underline{S}}$ we have $G_1 + G_2 \in \underline{\underline{S}}$, and consequently, $(G_1 + G_2)^2 \in \underline{\underline{V}}_0$. Thus, for $G_1, G_2 \in \underline{\underline{S}}$, $G_1G_2 + G_2G_1 \in \underline{\underline{V}}_0$. This implies that for any $G \in \underline{\underline{S}}_0$, $G^2 \in \underline{\underline{V}}_0$. Thus if $\underline{\underline{S}}_0 = \underline{\underline{V}}_0$ then $\underline{\underline{S}}_0$ is a Jordan algebra.

On the other hand, if $\underline{\underline{S}}_0$ is a Jordan algebra, then from Lemma 3.2., $\underline{\underline{V}}_0 \subset \underline{\underline{S}}_0$ and consequently $\underline{\underline{V}}_0 = \underline{\underline{S}}_0$. //

We fix an orthonormal basis $(G_1, \ldots, G_m, \ldots, G_k, \ldots, G_n)$ for $\underline{\underline{A}}$ so that $(G_1, \ldots, G_m)$ is a basis for $\underline{\underline{S}}_0$ and $(G_1, \ldots, G_k)$ is a basis for $\underline{\underline{V}}_0$. We shall find it very convenient to identify G_η and η via the isometry (3.2) Under this identification $\underline{\underline{S}}_0 = R^m$, $\underline{\underline{V}}_0 = R^k$ and $\underline{\underline{A}} = R^n$ with, of course, the embeddings $R^m \subset R^k \subset R^n$.

Let $\underline{M}$ be the algebraic set defined by the polynomial equations

$$(3.13) \qquad |A_{1i}(z_1, \ldots, z_k, 0, 0 \ldots 0)| = 0, \qquad i = m+1, \ldots, n.$$

where A_{1i} is as in (3.12).

The set of $p \times p$ p.d. matrices is open in the set of all $p \times p$ matrices with the usual topology. The Euclidean topology on $\underline{\underline{A}}$ is the relative topology inherited from the topology on the set of $p \times p$ matrices. Therefore, we can conclude that $\textcircled{H}$, the set of p.d. matrices in $\underline{\underline{A}}$, is an open set in $\underline{\underline{A}}$.

Now, consider the map $\textcircled{H} \overset{U}{\longleftrightarrow} \textcircled{H}$ defined in (3.5). It is a rational map and hence bicontinuous. We also have

$$(3.14) \qquad \textcircled{H} \cap \underline{\underline{S}}_o \overset{U}{\longleftrightarrow} \textcircled{H} \cap \underline{M}$$

Since $\underline{\underline{S}}$ is open in $\underline{\underline{S}}_o$ and $\underline{\underline{S}} \subset \textcircled{H} \cap \underline{\underline{S}}_o$, $\underline{\underline{S}}$ is also open in $\textcircled{H} \cap \underline{\underline{S}}_o$. Therefore, $U(\underline{\underline{S}})$ is open in $\textcircled{H} \cap \underline{M}$. Now $\textcircled{H} \cap \underline{M} \subset R^k = \underline{\underline{V}}_o$. Thus $\textcircled{H}_o = U(\underline{\underline{S}})$ is of the form

$$(3.15) \qquad \textcircled{H}_o = \textcircled{H}_1 \cap \underline{M}$$

where $\textcircled{H}_1$ is an open set in R^k.

<u>Lemma</u> 3.7. The functions $1, U_1(\eta), \ldots, U_k(\eta)$ are linearly independent functions on $\underset{=}{S}$.

<u>Proof</u>. Suppose, for all $\eta \in \underset{=}{S}$ and some constant c, we have

$$\sum_{i=1}^{k} \alpha_i U_i(\eta) = c$$

Since $\underset{=}{S}$ is a convex cone, for $\eta_0 \in \underset{=}{S}$, $K \eta_0 \in \underset{=}{S}$ where K is any constant > 0. This implies that for arbitrary $a > 0$

$$\Sigma \alpha_i a U_i(\eta_0) = c$$

Thus $c = 0$. But the set of vectors

$$\{ (U_1(\eta), \ldots, U_k(\eta)) : \eta \in \underset{=}{S} \}$$

spans $\underset{=}{V}_0 = R^k$. This implies $\alpha_i = 0$, $i = 1, \ldots, k$. //

<u>Lemma</u> 3.8. $1, y^T G_1 y, \ldots, y^T G_k y$ are linearly independent functions on R^p.

<u>Proof</u>. Suppose, for all $y \in R^p$ and for some constant c

$$\sum_{i=1}^{k} \alpha_i y^T G_i y = c$$

or

$$y^T (\Sigma \alpha_i G_i) y = c$$

This implies

$$\Sigma \alpha_i G_i = 0$$

Since (G_i) $i = 1, \ldots, k$ are linearly independent we have $\alpha_i = 0$, $i = 1, \ldots, k$. //

<u>Theorem</u> 3.9. Let $y \sim N_p(0, G_\eta)$, $G_\eta \in \underline{\underline{S}}$. Let $y_1, \ldots, y_N$ be independent samples from y. Then the family of joint distributions of $(y_1, \ldots, y_N)$ has minimal exponential representation with respect to the Lebesgue measure in R^{pN}

(3.16) $$p_\Theta(y_1, \ldots, y_N) = c(\Theta)\, e^{\langle \Theta,\, t(y_1, \ldots, y_N)\rangle}$$

where

(3.17) $$t_i(y_1, \ldots, y_N) = -N\, tr(G_i C)/2, \quad i = 1, \ldots, k$$

$$C = \sum_{i=1}^{N} y_i y_i^T / N$$

and the canonical parametrization is by

(3.18) $$\Theta_i = U_i(\eta), \qquad G_\eta \in \underline{\underline{S}}$$

The canonical parameter Θ varies in the set $\Theta_0 = \Theta_1 \cap \underline{M}$ defined in (3.15).

t is a complete statistic iff $\underline{\underline{S}}_0 = \underline{\underline{A}}$.

Proof.

$$-(1/2) \sum_{i=1}^{N} y_i^T G_\eta^{-1} y_i = -N \operatorname{tr}(G_\eta^{-1} C)/2$$
$$= (-N/2) \sum_{i=1}^{k} \theta_i \operatorname{tr}(G_i C)$$

where $\theta_i = U_i(\eta)$.

The minimality of the exponential representation follows from Lemmas 3.7 and 3.8. Notice $\sum_{i=1}^{k} a_i \operatorname{tr}(G_i C) = c$ for all $y_1, \ldots, y_N$ implies, in particular, $\sum_{i=1}^{k} a_i y_1^T G_i y_1 = c$.

The set of minimal canonical parameters Θ_o is of the form (3.15) is clear.

If $\underline{\underline{S}}_o = \underline{\underline{A}}$ then Θ_o contains an open set in R^n and therefore t is a complete statistic.

If $\underline{\underline{S}}_o \neq \underline{\underline{A}}$, then $\underline{\underline{S}}_o \neq \underline{\underline{V}}_o$. In this case the minimal canonical parameters span R^k but they are zeros of the non-trivial polynomials in (3.13). Now, from Lemma 1.11, it follows that the minimal sufficient statistic t is incomplete.

Remark 3.10. Suppose $y \sim N_p(0, G)$, $G \in \underline{\underline{S}}^*$ where $I \notin \underline{\underline{S}}^*$. Fix $V \in \underline{\underline{S}}^*$ and look at

$$\underline{\underline{S}} = \left\{ V^{-\frac{1}{2}} G V^{-\frac{1}{2}} : G \varepsilon \underline{\underline{S}}^* \right\}$$

Then it is easily seen that the above theorem is true with the set of minimal canonical parameters $V^{-1/2} \; \Theta_o \; V^{-1/2}$ where Θ_o is as before.

Example 3.11. We shall obtain the minimal canonical exponential representation of the two way balanced random effects model as an illustration of the method of this section. Herbach (1959) has obtained the same by simultaneously diagonalizing the covariance matrices. The unbalanced case can also be treated by our method but the Jordan algebra generated by the covariance matrices will depend upon the unbalanced model we consider.

The following model is assumed for a two way classification with K_o observations per cell.

(3.19) $$y_{ijk} = \mu + e_i^A + e_j^B + e_{ij}^{AB} + e_{ijk} ,$$

$i = 1, \ldots, I_o$; $j = 1, \ldots, J_o$; $k = 1, \ldots, K_o$ where y_{ijk} is the k th observation on the (i, j) th cell. The main effect μ is assumed to be a constant and the components e_i^A, e_j^B, e_{ij}^{AB}, e_{ijk} are assumed to be independent normal variables with mean zero and variances σ_a^2, σ_b^2, σ_{ab}^2, σ_e^2

respectively. In a self-explanatory matrix notation, (3.19) can be written as

$$(3.20) \qquad Y = J\mu + Ua + Vb + Wc + Ie$$

Put $G_1 = UU^T$, $G_2 = VV^T$, $G_3 = WW^T$. The matrices I, G_1, G_2, G_3 do not generate a Jordan algebra. We have

$$G_1G_2 = G_2G_1 = 2G_4$$

where G_4 is the matrix with all its entries unity. We have

$$G_1^2 = (J_0 - K_0)G_1, \quad G_2^2 = (I_0 - K_0)G_2, \quad G_3^2 = K_0G_3,$$

$$G_1G_2 = G_2G_1 = K_0G_4, \quad G_1G_3 = G_3G_1 = K_0G_1,$$

$$G_1G_4 = G_4G_1 = (J_0 - K_0)G_4, \quad G_2G_3 = G_3G_2 = K_0G_2$$

$$G_2G_4 = G_4G_2 = (I_0 - K_0)G_4, \quad G_3G_4 = G_4G_3 = K_0G_4$$

so that G_1, G_2, G_3, G_4, I span a Jordan algebra. Let

$$\Sigma = \sigma_e^2 I + \sigma_a^2 G_1 + \sigma_b^2 G_2 + \sigma_{ab}^2 G_3$$

$$\Sigma^{-1} = c_0 I + c_1 G_1 + c_2 G_2 + c_3 G_3 + c_4 G_4$$

Then we have

$$(3.21) \qquad I = \sigma_e^2 \theta_o I + G_1(\theta_1 [\sigma_e^2 + J_o.K_o\sigma_a^2 + K_o\sigma_{ab}^2] + \sigma_a^2 [\theta_o + K_o\theta_3]) + G_2(\theta_2 [\sigma_e^2 + I_o.K_o\sigma_b^2 + K_o\sigma_{ab}^2] + \sigma_b^2 [\theta_o + K_o\theta_3]) + G_3(\theta_3 [\sigma_e^2 + K_o\sigma_{ab}^2] + \sigma_{ab}^2\theta_o) + G_4(\theta_4 [\sigma_e^2 + J_o.K_o\sigma_a^2 + I_o.K_o\sigma_b^2 + K_o\sigma_{ab}^2] + K_o [\sigma_a^2\theta_2 + \theta_1\sigma_b^2])$$

Put

$$\eta_o = \sigma_e^2$$

$$\eta_1 = \sigma_e^2 + J_o.K_o\ \sigma_a^2 + K_o\ \sigma_{ab}^2$$

$$\eta_2 = \sigma_e^2 + I_o.K_o\ \sigma_b^2 + K_o\ \sigma_{ab}^2$$

$$\eta_3 = \sigma_e^2 + K_o\ \sigma_{ab}^2$$

$$\eta_4 = \sigma_e^2 + J_o.K_o\ \sigma_a^2 + I_o.K_o\ \sigma_b^2 + K_o\ \sigma_{ab}^2$$

Then we have the following linear relation

$$\eta_4 = \eta_1 + \eta_2 - \eta_3$$

Also we have

$$\sigma_e^2 = \eta_o$$

$$\sigma_{ab}^2 = (\eta_3 - \eta_o)/K_o$$

$$(3.21)\qquad I = \sigma_e^2\theta_o I + G_1(\theta_1[\sigma_e^2 + J_o.K_o\sigma_a^2 + K_o\sigma_{ab}^2] + \sigma_a^2[\theta_o + K_o\theta_3]) + G_2(\theta_2[\sigma_e^2 + I_o.K_o\sigma_b^2 + K_o\sigma_{ab}^2] + \sigma_b^2[\theta_o + K_o\theta_3]) + G_3(\theta_3[\sigma_e^2 + K_o\sigma_{ab}^2] + \sigma_{ab}^2\theta_o) + G_4(\theta_4[\sigma_e^2 + J_o.K_o\sigma_a^2 + I_o.K_o\sigma_b^2 + K_o\sigma_{ab}^2] + K_o[\sigma_a^2\theta_2 + \theta_1\sigma_b^2])$$

Put

$$\eta_o = \sigma_e^2$$

$$\eta_1 = \sigma_e^2 + J_o.K_o\,\sigma_a^2 + K_o\,\sigma_{ab}^2$$

$$\eta_2 = \sigma_e^2 + I_o.K_o\,\sigma_b^2 + K_o\,\sigma_{ab}^2$$

$$\eta_3 = \sigma_e^2 + K_o\,\sigma_{ab}^2$$

$$\eta_4 = \sigma_e^2 + J_o.K_o\,\sigma_a^2 + I_o.K_o\,\sigma_b^2 + K_o\,\sigma_{ab}^2$$

Then we have the following linear relation

$$\eta_4 = \eta_1 + \eta_2 - \eta_3$$

Also we have

$$\sigma_e^2 = \eta_o$$

$$\sigma_{ab}^2 = (\eta_3 - \eta_o)/K_o$$

$$\sigma_a^2 = (\eta_1 - \eta_3)/J_o.K_o = (\eta_4 - \eta_2)/J_o.K_o$$
$$\sigma_b^2 = (\eta_2 - \eta_3)/I_o.K_o = (\eta_4 - \eta_1)/I_o.K_o$$

Now; from the equation (3.21) we obtain

$$\theta_o = 1/\eta_o$$

$$\theta_3\eta_3 + \sigma_{ab}^2\,\theta_o = 0$$

or

$$\theta_3 = -\sigma_{ab}^2\,\theta_o/\eta_3 = -(\eta_3 - \eta_o)/K_o\eta_o\eta_3$$
$$= K_o^{-1}(\eta_3^{-1} - \eta_o^{-1}).$$

$$\theta_o + K_o\,\theta_3 = \eta_o^{-1} + (\eta_3^{-1} - \eta_o^{-1}) = \eta_3^{-1}$$

$$\sigma_a^2(\theta_o + K_o\,\theta_3) = (\eta_1 - \eta_3)/J_o.K_o\,\eta_3$$

Therefore,

$$\theta_1 = -(\eta_1 - \eta_3)/J_o.K_o\,\eta_1\,\eta_3 = (J_o.K_o)^{-1}(\eta_1^{-1} - \eta_3^{-1})$$

Similarly,

$$\theta_2 = (I_o.K_o)^{-1}(\eta_2^{-1} - \eta_3^{-1}).$$

$$\theta_4\eta_4 + K_o(\sigma_a^2\theta_2 + \theta_1\sigma_b^2) = 0.$$

Solving the above equation,

$$\theta_4 = (I_o.J_o.K_o)^{-1}(\eta_4^{-1} + \eta_3^{-1} - \eta_2^{-1} - \eta_1^{-1})$$

For the time being let us assume $\mu = 0$. We shall evaluate the quadratic form $y^T \Sigma^{-1} y$. First observe

$$y^T G_1 y = \sum_i (y_{i..})^2$$

$$y^T G_2 y = \sum_j (y_{.j.})^2$$

$$y^T G_3 y = \sum_{i,j} (y_{ij.})^2$$

$$y^T G_4 y = (y_{...})^2$$

Now,

$$y^T \Sigma^{-1} y = y^T \Big[\sum_{r=0}^{4} \theta_r G_r\Big] y$$

$$= y^T \Big[\eta_0^{-1}(I_o - K_o^{-1} G_3) + \eta_1^{-1}((J_o.K_o)^{-1} G_1 - (I_o.J_o.K_o)^{-1} G_4)$$

$$+ \eta_2^{-1}((I_o.K_o)^{-1} G_2 - (I_o.J_o.K_o)^{-1} G_4)$$

$$+ \eta_3^{-1}((K_o)^{-1} G_3 - (J_o.K_o)^{-1} G_1 - (I_o.K_o)^{-1} G_2$$

$$+ (I_o.K_o.J_o)^{-1} G_4) + \eta_4^{-1}((I_o.J_o.K_o)^{-1} G_4)\Big] y$$

$$= (S_o/\eta_o) + (S_1/\eta_1) + (S_2/\eta_2) + (S_3/\eta_3) + (S_4/\eta_4)$$

where

$$S_o = y^T y - \Sigma (y_{ij.})^2 / K_o$$

$$S_1 = (J_o.K_o)^{-1} \Sigma (y_{i..})^2 - (I_o.J_o.K_o)^{-1} (y_{...})^2$$

$$S_2 = (I_o.K_o)^{-1} \Sigma (y_{.j.})^2 - (I_o.J_o.K_o)^{-1} (y_{...})^2$$

$$S_3 = K_o^{-1} \Sigma (y_{ij.})^2 - (J_o.K_o)^{-1} \Sigma (y_{i..})^2$$

$$- (I_o.K_o)^{-1} \Sigma (y_{.j.})^2 + (I_o.J_o.K_o)^{-1} (y_{...})^2$$

$$S_4 = (I_o.J_o.K_o)^{-1} (y_{...})^2$$

Thus we have an exponential family with the minimal canonical statistics S_r and minimal canonical parameters η_r^{-1} with the non-linear relation (for η_r^{-1})

$$\eta_4 = \eta_1 + \eta_2 - \eta_3$$

restricting the parameters. The family is incomplete.

Now let μ be unknown in $(-\infty, \infty)$. Then replacing y_{ijk} by $y_{ijk} - \mu$, obviously S_i $(i = 0,1,2,3)$ remain the same. $S_4 = (I_o.J_o.K_o)^{-1} (y_{...} - I_o.J_o.K_o.\mu)^2$. The minimal canonical parameters are η_o^{-1}, η_1^{-1}, η_2^{-1}, η_3^{-1}, η_4^{-1}, μ/η_4 and the minimal canonical statistics are S_o, S_1, S_2, S_3, $(y_{...})^2$, $y_{...}$ The minimal canonical statistic is not a 'proper' minimal sufficient statistic since $y_{...}^{\;2}$ is a function of $y_{...}$ The minimal sufficient statistic can be shown to be complete in spite of the non-linear relation restricting the parameters. See Gautschi (1959).

3.3. Uniformly Minimum Variance Unbiased Estimation

Theorem 3.12. Let $y \sim N_p(0, G)$, $G \in \underset{\sim}{S}$ where $\underset{\sim}{S}$ is a convex cone of $p \times p$ p.d. matrices containing the identity matrix. Suppose $y_1, \ldots, y_N$ are N independent observations of y. Let

$$(3.22) \qquad L_o = \{G \in \underline{S}_o : AGA \in \underline{S}_o \text{ for all } A \in \underline{S}_o\}$$

$$= \{G \in \underline{S}_o : G_i G G_j + G_j G G_i \in \underline{S}_o, \ i=1, \ldots, m\}$$

Suppose $R_1, \ldots, R_r$ is basis for the vector space L_o. Then the class of U.M.V.U. estimators, with finite variance, is the same as the class of estimators, with finite variance, which are functions of $(\text{tr}(R_1C), \ldots, \text{tr}(R_rC))$.

Proof. Since $\underset{\sim}{S}$ is a convex cone the interior of $\underset{\sim}{S}$ in $\underline{S}_o$ is dense in $\underset{\sim}{S}$. Therefore we assume, without loss of generality, that $\underset{\sim}{S}$ is open in $\underline{S}_o$. In the minimal exponential representation (3.17) the minimal canonical statistics $t_1, \ldots, t_k$ has an induced family of distributions which has an exponential representation with respect to the Lebesgue measure μ^k

$$p_\theta(t) = c(\theta)\, e^{\langle\theta, t\rangle} h(t), \quad \theta \in \Theta_o = \Theta_1 \cap \underline{M}$$

$h(t)$ is continuous.

Suppose $\underline{N}$ is the smallest algebraic set containing $\textcircled{H}_o$ and let L be the largest subspace of R^k such that $L + \underline{N} = \underline{N}$.

Now, appealing to the theorem of Kagan and Palamodov, we are done if we show $L = L_o$.

<u>To show</u> $L \subset L_o$. Suppose $G \varepsilon L$. Take $F \varepsilon \textcircled{H}_o$. Now, from the property of L, $F + \alpha G \varepsilon \underline{N}$. But since $\textcircled{H}_1$ is open, for small enough α we have, $F + \alpha G \varepsilon \textcircled{H}_1$. Therefore,

$$F + \alpha G \varepsilon \textcircled{H}_1 \cap \underline{N} = \textcircled{H}_1 \cap \underline{M} = \textcircled{H}_o$$

Again, for small enough α,

$$\begin{aligned}(F + \alpha G)^{-1} &= [F(I + \alpha F^{-1}G)]^{-1} \\ &= (I - \alpha F^{-1}G + \alpha^2 F^{-1}GF^{-1}G - \ldots)\, F^{-1} \\ &= F^{-1} - \alpha F^{-1}GF^{-1} + \alpha^2 F^{-1}GF^{-1}GF^{-1} - \ldots\end{aligned}$$

Since $F + \alpha G \varepsilon \textcircled{H}_o$, $(F + \alpha G)^{-1} \varepsilon \underline{\underline{S}}$. Also $F^{-1} \varepsilon \underline{\underline{S}}$. It follows that

$$F^{-1}GF^{-1} - \alpha F^{-1}GF^{-1}GF^{-1} + \ldots \varepsilon \underline{\underline{S}}_o$$

Now, letting $\alpha \longrightarrow 0$,

$$F^{-1}GF^{-1} \in \underline{\underline{S}}_o.$$

Thus, for each $A \in \underline{\underline{S}}$

$$G \in L \implies AGA \in \underline{\underline{S}}_o$$

Since $\underline{\underline{S}}$ is a convex cone, the above implies that $AGA \in \underline{\underline{S}}_o$ for all $A \in \underline{\underline{S}}_o$.

<u>To show</u> $L_o \subset L$. Let $G \in L_o$. Take $F \in \textcircled{H}_1 \cap \underline{M}$. Now, we can find α_F such that $|\alpha| < \alpha_F$ implies $F + \alpha G \in \textcircled{H}_1$ and

$$(F + \alpha G)^{-1} = F^{-1} - \alpha F^{-1}GF^{-1} + \alpha^2 F^{-1}GF^{-1}GF^{-1} - \dots$$

$$F \in \textcircled{H}_o \implies F^{-1} \in \underline{\underline{S}}_o$$

$$G \in L_o, \quad F^{-1} \in \underline{\underline{S}}_o \implies F^{-1}GF^{-1} \in \underline{\underline{S}}_o.$$

But we also have

$$F^{-1} \in \underline{\underline{S}}_o, \quad F^{-1}GF^{-1} , \underline{\underline{S}}_o, \quad G \in L_o \implies F^{-1}GF^{-1}GF^{-1} \in \underline{\underline{S}}_o.$$

and so on. To see the above, notice

$$A, B \in \underline{\underline{S}}_o, \quad G \in L_o \implies (A + B)\, G(A + B) \in \underline{\underline{S}}_o$$

$$\implies AGB + BGA \in \underline{\underline{S}}_o$$

Thus we have

$$(F + \alpha G)^{-1} \in \underline{\underline{S}}_o$$

Also, for $|\alpha| < \alpha_F$

$$(F + \alpha G)^{-1} \in Ⓗ$$

where Ⓗ is, as before, the set of p.d. matrices in the Jordan algebra generated by $\underline{\underline{S}}_o$. Thus for $|\alpha| < \alpha_F$

$$(F + \alpha G)^{-1} \in Ⓗ \cap \underline{\underline{S}}_o.$$

It follows that

$$F + \alpha G \in Ⓗ \cap \underline{M}$$

and, since for $|\alpha| < \alpha_F$, $F + \alpha G$ is in $Ⓗ_1$,

$$F + \alpha G \in Ⓗ_1 \cap \underline{M}, \qquad |\alpha| < \alpha_F$$

or

$$F + \alpha G \in \underline{N}, \qquad |\alpha| < \alpha_F$$

But, since $\underline{N}$ is an algebraic set, the above implies that for each real α

$$F + \alpha G \in \underline{N}$$

Thus we have shown that for each real α

$$\text{Ⓗ}_1 \cap \underline{M} + \alpha G \subset \underline{N}$$

or

$$\text{Ⓗ}_1 \cap \underline{M} \subset \underline{N} - \alpha G$$

and since $\underline{N} - \alpha G$ is also an algebraic set

$$\underline{N} \subset \underline{N} - \alpha G$$

or

$$\underline{N} + \alpha G \subset \underline{N}$$

Thus $G \varepsilon L$ or $L_o \subset L$. //

Remark 3.13. If $I \notin \underline{\underline{S}}$, then from Remark 3.10, it is easily seen that the change in the definition of L_o should be the following.

$$L_o = \{G \varepsilon \underline{\underline{S}}_o : AV^{-1}GV^{-1}A \varepsilon \underline{\underline{S}}_o \quad \text{for all} \quad A \varepsilon \underline{\underline{S}}_o\}$$

$$= \{G \varepsilon \underline{\underline{S}}_o : G_iV^{-1}GV^{-1}G_j + G_jV^{-1}GV^{-1}G_i \varepsilon \underline{\underline{S}}_o\}$$

where V is any fixed matrix in $\underline{\underline{S}}$. //

The parametric functions of most interest are the linear functions on $\underline{\underline{S}}_o$. Notice the linear functions on $\underline{\underline{S}}_o$ are of the form $tr(AG)$, $A \varepsilon \underline{\underline{S}}_o$ or in terms of the coordinates $\eta_1, \ldots, \eta_m$, $\sum_i p_i\eta_i$ where $p_i = tr\, AG_i$. The simplest estimators for the linear parametric functions are the linear functions of the sufficient statistic C which are of the form tr RC. Since they are the quadratic functions of the observations they are called quadratic estimators.

Corollary 3.14. tr RC is a U.M.V.U. estimator for tr AG, $A \in \mathbf{S}_0$ iff $R = A$ and $BAB \in \mathbf{S}_0$ for all $B \in \mathbf{S}_0$. tr AG, $A \in \mathbf{S}_0$ is U.M.V.U. estimable by a quadratic estimator iff $BAB \in S_0$ for all $B \in \mathbf{S}_0$.

Proof. tr RC is a U.M.V.U. estimator implies, from Theorem 3.12, $R \in \mathbf{S}_0$ and $BRB \in \mathbf{S}_0$ for all $B \in \mathbf{S}_0$. Since $R \in \mathbf{S}_0$ and tr RG = tr AG for all $G \in \mathbf{S}_0$, we have $R = A$. //

It is interesting to note that a quadratic estimator is U.M.V.U. iff it is U.M.V.U. within the class of quadratic estimators.

Theorem 3.15. tr RC is U.M.V.U. for its expectation within the class of quadratic estimators iff $BRB \in \mathbf{S}_0$ for all $B \in \mathbf{S}_0$.

Proof. tr FC is an unbiased estimator of zero iff tr FG = 0 for all $G \in \mathbf{S}_0$, i.e., $F \in \mathbf{S}_0^{\perp}$. Now tr RC is U.M.V.U. within the class of quadratic estimators iff cov(tr RC, tr FC) = 0 for all $F \in \mathbf{S}_0^{\perp}$. But Cov(tr RC, tr FC)

$= 2 \mathrm{tr}\, FGRG$ when the covariance matrix of the observation vector is assumed to be G. Thus $\mathrm{tr}\, RG$ is U.M.V.U. within the class of quadratic estimators iff $\mathrm{tr}\, FGRG = 0$ for all $G \in \underline{\underline{S}}_0$ and for all $F \in \underline{\underline{S}}_0^{\perp}$. Or for all $G \in \underline{\underline{S}}_0$, $GRG \in \underline{\underline{S}}_0$.

The above result is obtained by C.R. Rao (1971) by directly minimizing the variance at a fixed $G \in \underline{\underline{S}}_0$ and then noting the condition for the estimator to be independent of G.

3.4. Maximum Likelihood Estimation

Let us assume that $\underline{\underline{S}}$ is the convex cone of all p.d. matrices in $\underline{\underline{S}}_0$ or $\underline{\underline{S}} = \text{Ⓗ} \cap \underline{\underline{S}}_0$. Then from the shape of the likelihood function it is clear that it attains a maximum in $\underline{\underline{S}}$. A necessary condition for $\tilde{G}$ to be a maximum is that it satisfies the likelihood equation derived below.

The logarithm of the likelihood function is proportional to

$$(3.23) \qquad (2/N) \log L(G) = -p \log 2\pi + \log|G^{-1}| - \operatorname{tr} G^{-1}C$$

where G varies in the set $\text{Ⓗ} \cap \underline{\underline{S}}_0$. Consider the function as a function of G^{-1} (or as a function of $\sigma_1, \ldots, \sigma_k$) where $G^{-1} \varepsilon \text{ Ⓗ} \cap \underline{M}$. The necessary condition for a point $\tilde{\sigma}$ to be a maximum of the above function is that

$$(3.24) \qquad \delta \left[(\log L(G) + \sum_{i=m+1}^{n} \beta_i U_i \right]_\sigma / \delta\sigma_j = 0, \qquad j = 1, \ldots, k$$

$$U_i(\sigma) = 0, \quad i = m+1, \ldots, n$$

where U_i is the function $\eta_i = U_i(\sigma)$ defined in (3.12) and $\beta_{m+1}, \ldots, \beta_n$ are the Lagrangian multipliers for the conditions $\eta_{m+1} = 0, \ldots, \eta_n = 0$.

We make use of the following two well-known formulae.

(3.25) $\delta \log |G^{-1}|/\delta\Theta_j = \text{tr } GG_j$

(3.26) $\delta G/\delta\Theta_j = - GG_jG.$

Now,

(3.27) $\delta G/\delta\Theta_j = \delta(\sum_i \eta_i G_i)/\delta\Theta_j = \sum_i [\delta U_i(\Theta)/\delta\Theta_j] G_i.$

Therefore,

(3.28) $\delta U_i(\Theta)/\delta\Theta_j = - \text{tr } GG_jGG_i.$

Let A_{ij} be the matrix whose (r, k)th element is given by $\text{tr} [(G_rG_jG_k + G_kG_jG_r)G_i]$. Then (3.28) can explicitly be written as a quadratic form

(3.29) $\delta U_i(\Theta)/\delta\Theta_j = - \eta^T A_{ij}\eta.$

Now we can write down the likelihood equation (3.24) explicitly as

(3.30) $\sum_i \eta_i \text{ tr } G_iG_j - \text{tr } G_jC - \sum_{i=m+1}^{n} \beta_i \eta^T A_{ij}\eta = 0, \quad j=1,\ldots, n$

$$\eta_i = 0, \quad i = m+1,\ldots, n.$$

Or in terms of an orthonormal basis G_i

$$(3.31)\qquad \eta_j - \operatorname{tr} G_jC - \Sigma\, \beta_i\eta^T A_{ij}\eta = 0, \qquad j = 1,\dots, m$$

$$- \operatorname{tr} G_jC - \Sigma\, \beta_i\eta^T A_{ij}\eta = 0, \qquad j = m+1, \dots, n.$$

We can attempt a solution of (3.31) by iteration by substitution. Get initial values of η_j (which may be taken as $\operatorname{tr} G_jC$), $j = 1,\dots, m$. Then from the second set of equations in (3.31) compute $\beta_{m+1},\dots, \beta_n$. Now we can obtain new values of η_j from the first set of equations in (3.31).

In an actual computation of the likelihood equation the following points may be noted. (1) It is not necessary to compute the matrices A_{ij}. At each stage we can compute $\operatorname{tr} GG_iGG_j$ directly. (2) It is not necessary to obtain an orthonormal basis. (3.30) can be used. (3) If the Jordan algebra $\underline{\underline{A}}$ is not easily recognized, it is not necessary to find it. We can consider the Jordan algebra of all $p \times p$ symmetric matrices with its natural basis G_{ij} where G_{ij} has its (i,j)th element and (j,i)th element 1 and all other elements zero. In this case $\underline{\underline{S}}_0$ is the kernel (or null space) of certain linear equations $\sum_{i,j} \ell_{ij}^{(r)} \eta_{ij} = 0$ in the algebra of $p \times p$ symmetric matrices. So the constraints $\eta_{m+1} = 0, \dots, \eta_n = 0$ will be replaced by $\sum_{i,j} \ell_{ij}^{(r)} \eta_{ij} = 0$.

Let

$$(3.32) \qquad F = -\sum_{m+1}^{n} \beta_i G_i .$$

Then the equation (3.31) can be put in the matrix form as

$$(3.33) \qquad G + GFG = C_{|\underline{\underline{A}}}$$

where $G \in \underline{\underline{S}}_0$, $F \in \underline{\underline{S}}_0^{\perp}$ and $C_{|\underline{\underline{A}}}$ denotes the projection of C into $\underline{\underline{A}}$. (3.33) can be written as an orthogonal sum

$$(3.34) \qquad I + FG = G^{-1} C_{|\underline{\underline{A}}} .$$

From (3.34) we immediately get

$$(3.35) \qquad \operatorname{tr} \tilde{G}^{-1} C = p$$

for any solution $\tilde{G}$ of (3.34), a result obtained by Anderson (1969).

Also, it is clear that the likelihood ratio criterion for the hypothesis that $G \in \underline{\underline{S}}_0$ versus the unrestricted alternative is the $\frac{1}{2} N$ th power of

$$(3.36) \qquad |C|/|\tilde{G}| = |I + \tilde{G}\tilde{F}|$$

where $\tilde{G}$ is the maximum likelihood estimator.

It is interesting to compare our equation (3.31) with the likelihood equation obtained by Anderson. Anderson's equation is

$$(3.37) \qquad \mathrm{tr}\,[(\Sigma\eta_i G_i)^{-1}G_j] = \mathrm{tr}\,[(\Sigma\eta_i G_i)^{-1}C(\Sigma\eta_i G_i)^{-1}G_j],$$

$$j = 1, \ldots, m,$$

or in matrix form as

$$(3.38) \qquad G^{-1}{}_{|\underline{\underline{S}}_0} = (G^{-1}CG^{-1})_{|\underline{\underline{S}}_0}.$$

Now putting

$$F_1 = G^{-1}\,C\,G^{-1} - G^{-1}$$

we have

$$G + G\,F_1\,G = C$$

for F_1 in the orthogonal complement of $\underline{\underline{S}}_0$ with respect to all symmetric $p \times p$ matrices whereas we have from our equation

$$G + G\,F\,G = C_{|\underline{\underline{A}}}$$

for F in the orthogonal complement of $\underline{\underline{S}}_0$ with respect to $\underline{\underline{A}}$. Thus our equation is of smaller dimension than Anderson's.

Finally the question of uniqueness of the solution of the likelihood equation leads to the following interesting question. Let $\underline{\underline{S}}_0$ be a subspace of a Jordan algebra of matrices $\underline{\underline{A}}$.

Under what conditions we have a unique dicomposition for the (p.d.) matrices $C \varepsilon \underline{\underline{A}}$ as

$$C = G + GFG, \quad G \varepsilon \underline{\underline{S}}_0, \quad F \varepsilon \underline{\underline{S}}_0^{\perp}.$$

CHAPTER 4.

INADMISSIBILITY OF CERTAIN ESTIMATORS WHICH ARE FUNCTIONS OF THE MINIMAL SUFFICIENT STATISTICS

In this chapter we extend and strengthen a result of A.M. Kagan on the inadmissibility (in the class of unbiased estimators) of certain estimators which are functions of the minimal sufficient statistic. This result has an important application to a special type of location parameter family.

4.1. Introduction

Consider the family of exponential densities on R^k, with respect to the Lebesgue measure μ^k on R^k

$$(4.1) \qquad p_\theta(t) = c(\theta)\, e^{\langle\theta, t\rangle}\, h(t), \qquad \theta \in \Theta_0$$

We assume

(1) $h(t)$ is a continuous function of t.

(2) The set of canonical parameters Θ_0 is a subset of the algebraic set $\underline{M}$ in C^k, defined by the polynomial equations

$$(4.2)\qquad \begin{aligned} P_1(z) &= z_2 - z_1^2 = 0 \\ &\vdots \\ P_{k-1}(z) &= z_k - z_1^k = 0 \end{aligned}$$

Under the assumption that Θ_o is a bounded set, A.M. Kagan (1968)-p. 86-89 shows that any nonconstant polynomial estimator $Q(t_2, \ldots, t_k)$, which is independent of t_1, is inadmissible in the class of unbiased estimators of $E_\Theta Q$.

We extend the result of Kagan from polynomial estimators to continuously differentiable estimators. But what is more important is that we discard the unnatural assumption of the boundedness of the parameter space Θ_o.

An exponential family with the canonical parameters satisfying the equations (4.2) appears, for example, when a sample $x_1, x_2, \ldots, x_N$ is taken from an m-dimensional population whose probability density with respect to the Lebesgue measure in R^m is of the form

$$(4.3)\qquad f(x, \eta) = c(\eta)\, e^{\eta r_1(x) + \eta^2 r_2(x) + \ldots + \eta^k r_k(x) + r_o(x)}, \quad \eta \in R.$$

In the canonical parametrization, we have $\Theta_1 = \eta, \ldots, \Theta_k = \eta^k$.

Our main theorem has a very interesting consequence to a special case of the family (4.3). Let

$$(4.4) \qquad f(x, \eta) = c \; e^{-(x-\eta)^{2k}}$$

where $x \in R$, $\eta \in R$ is a location parameter, and, $k \geq 2$ is an integer. For a sample of size N, where N is at least 2k, the theorem implies that none of the sample moments is an admissible estimator for the corresponding population moment, except one (which is in fact admissible).

4.2. <u>An Elementary Lemma on Admissibility</u>

Let $(X, \mathcal{B}, \mathcal{P})$ be a statistical model and

$$\mathcal{P} = \{P_\theta : \theta \in \Theta_0\}$$

<u>Definition</u> 4.1. We say G is an inadmissible unbiased estimator of $g(\theta)$ (in the class of unbiased estimators U_g) if $G \in U_g$ and there is some $G_1 \in U_g$ such that $E_\theta G_1^2 \leq E_\theta G^2$ for all θ in Θ_0. with strict inequality for some $\theta_0 \in \Theta_0$. G is admissible if it is not inadmissible.

<u>Lemma</u> 4.2. $G \in U_g$ is inadmissible in the class of unbiased estimators of $g(\theta)$ iff there exists an unbiased estimator of zero F and a constant $r > 0$ such that

$$(4.5)\quad \begin{aligned} &E_\theta GF \geq r\, E_\theta F^2 \quad \text{for all } \theta \text{ in } \Theta_o \\ &E_{\theta_o} GF > r\, E_{\theta_o} F^2 \quad \text{for some } \theta_o \text{ in } \Theta_o. \end{aligned}$$

Proof. (1) Suppose $G \varepsilon U_g$ is inadmissible. Then we have some

$$G_1 = G - F$$

where $F \varepsilon U_o$ and for all θ in Θ_o

$$E_\theta (G - F)^2 \leq E_\theta G^2$$

i.e.,

$$-2\, E_\theta\, GF + E_\theta\, F^2 \leq 0$$

or

$$\frac{1}{2} E_\theta\, F^2 \leq E_\theta\, GF.$$

For some $\theta_o \varepsilon \Theta_o$ all the above inequalities must be strict.

(2) Suppose that for some $r > 0$, $E_\theta GF \geq r\, E_\theta F^2$ for all $\theta \varepsilon \Theta_o$ with strict inequality at $\theta_o \varepsilon \Theta_o$. Notice that this implies $E_\theta F^2 > 0$.

For $\alpha > 0$, put $f = F/\alpha$. Then we have

$$E_\theta Gf \geq \alpha r\, E_\theta f^2$$

Now shoosing $\alpha > 1/2r$, $E_\theta Gf \geq \frac{1}{2} E_\theta f^2$, and hence

$$E_\Theta(G - f)^2 \leq E_\Theta G^2$$

with strict inequality, of course, at $\Theta = \Theta_0$. //

4.3. The Main Theorem

Theorem 4.2. In the exponential family (4.1) satisfying the assumptions (1) and (2), a continuously differentiable estimator $G(t_2, \ldots, t_k)$, where $G \in U_g$ and G is a function of only the statistics $t_2, \ldots, t_k$, is inadmissible in the class of unbiased estimators of $g(\Theta)$ unless G is a constant.

Proof. Suppose $G(t_2, \ldots, t_k) \in U_g$ is admissible in the class of unbiased estimators of $g(\Theta)$. Let S be the open set

$$S = \left\{ t \in R^k : h(t) > 0 \right\}.$$

Without loss of generality, we assume $G(t) = 0$ for $t \notin S$. To show

$$\delta G(t)/\delta t_2 \equiv 0.$$

If possible, let $(\delta G/\delta t_2)(t^*) > 0$ for some $t^* \in S$. We can find an open neighbourhood U of t^* contained in S

such that for all $t \in U$, $\delta G(t)/\delta t_2 > 0$.

Let V be an open sphere of radius α with centre at t^* such that the closure of V, $\bar{V} \subset U$. Let for all $t \in \bar{V}$,

(4.6)
$$(\delta G/\delta t_2)(t) \geq \gamma > 0$$

$$h(t) \geq \rho > 0$$

Now, for the polynomial P_1 as in (4.2) and for a fixed $\theta \in Ⓗ_0$, we have

$$P_1(D + \theta)G = [(\delta/\delta t_2 + \theta_2) - (\delta/\delta t_1 + \theta_1)^2]\, G.$$

Since $(\delta/\delta t_1)G \equiv 0$ and $P_1(\theta_1, \theta_2) = 0$, we have

(4.7)
$$P_1\,(D + \theta)\; G = \delta G/\delta t_2.$$

Now, we shall find a function $F \in U_0$ and a constant $r > 0$ such that $E_\theta GF > r\, E_\theta F^2$ for all θ in $Ⓗ_0$, which will be a contradiction to the assumption that G is admissible

For $f \in C_0^\infty(R^k)$, support of $f \subset \bar{V}$, we have

$$F(t) = P_1(-D)\, f(t)/h(t)$$

is in U_0 and

$$(4.8) \qquad E_Q GF = \int G(t)\,[P_1(-D)\,f(t)]\,e^{\langle Q,t\rangle}\,dt$$

$$= \int [P_1(D+Q)G(t)]\,f(t)\,e^{\langle Q,t\rangle}\,dt \qquad \text{[from (0.5)]}$$

$$= \int (\delta G/\delta t_2) f(t)\,e^{\langle Q,t\rangle}\,dt \qquad \text{[from (4.7)]}$$

Since $E_Q F^2 = \int_V \{[P_1(-D)\,f(t)]^2/h(t)\}\,e^{\langle Q,t\rangle}\,dt$, it is enough to choose $f(t)$ and $r > 0$ such that

$$(4.9) \qquad (\delta G/\delta t_2)(t)\,f(t)\,h(t) > r\,[P_1(-D)\,f(t)]^2 \quad \text{for } t \in V,$$

or because of (4.6), for some constant $r_1 > 0$

$$(4.10) \qquad f(t) > r_1\,[P_1(-D)\,f(t)]^2 \qquad t \in V.$$

Now choose

$$f(t) = \begin{cases} e^{-\alpha^2/(\alpha^2 - \|t-t^*\|^2)}, & t \in V \\ 0, & t \notin V. \end{cases}$$

For any polynomial $P(Q)$, clearly, $P(-D)\,f(t)$ is of the form $f(t)\,Q(t)$ where $Q(t)$ is a rational function. Also, $f(t)\,Q^2(t) \longrightarrow 0$ as $\|t-t^*\| \longrightarrow \alpha$, i.e., we can choose $r_2 > 0$ such that

$$(4.11) \qquad 1 > r_2\,f(t)\,Q^2(t)$$

so that r_1 can be chosen to satisfy (4.10).

If $\delta G/\delta t_2 < 0$ for some $t^* \varepsilon S$, take

$$f(t) = -e^{-\alpha^2/(\alpha^2 - ||t-t^*||^2)}$$

in the above argument.

Thus, it follows that $\delta G/\delta t_2 \equiv 0$.

As before, we can show that

$$P_i(D + \Theta)G = \delta G/\delta t_i, \quad i = 2, \ldots, k.$$

The same kind of argument as before will show

$$\delta G/\delta t_i = 0, \qquad i = 2, \ldots, k.$$

This shows G is a constant. The theorem is proved. //

Corollary 4.3. For the location parameter exponential family (4.4), for a sample of size $N > 2k$ the sample moments $m_1, m_2, \ldots, m_{2k-2}, m_{2k}, m_{2k+1}, \ldots$ are inadmissible estimators of the corresponding moments $\mu_1, \mu_2, \ldots, \mu_{2k-2}, \mu_{2k}, \mu_{2k+1}, \ldots$ in the class of unbiased estimators of these population moments.

Proof. For a sample of size $N > 2k$, the canonical statistics of the family

$$t_1 = \Sigma x_i^{2k-1}, \quad t_2 = \Sigma x_i^{2k-2}, \quad \ldots, \quad t_{2k-1} = \Sigma x_i$$

have a joint distribution of the type (4.1), satisfying the assumptions. So from our theorem, the sample moments, $m_q = \Sigma x_i^q$, $q = 1, \ldots, 2k-2$ are inadmissible. The sample moments $m_{2k}, m_{2k+1}, \ldots$ are inadmissible because they are not functions of the sufficient statistics $t_1, \ldots, t_{2k-1}$.

CHAPTER 5.

UNBIASED ESTIMATION IN A NORMAL DISTRIBUTION WITH AN UNKNOWN INTEGER MEAN AND A KNOWN VARIANCE

In Chapter 2 we have seen that in an algebriac exponential family, dominated by the Lebesgue measure, the U.M.V.U. estimators form a σ-algebra in the sense that there is a sub-σ-algebra of the Borel σ-algebra of the sample space with the property that the estimators measurable with respect to this σ-algebra are exactly the U.M.V.U. estimators. In an analytic exponential family, dominated by the Lebesgue measure, the U.M.V.U. estimators may not form a σ-algebra. However, one may expect, that the U.M.V.U. estimators will be a mathematically interesting class of functions.

In this chapter we characterize the U.M.V.U. estimators in the case of a normal distribution $N(\mu, 1)$ on R with an unknown integer mean μ and variance 1. As a corollary, we show that the parametric function μ has no U.M.V.U. estimator.

Notation. By a $2\pi i$-periodic function we shall mean a periodic entire function of period $2\pi i$. The set of integers will be denoted by Z.

<u>Theorem</u> 5.1. For the family $N(\mu, 1)$, $\mu \varepsilon Z$, the class of U.M.V.U. estimators coincide with the class of $2\pi i$-periodic functions $G(x)$, $E_\mu G^2 < \infty$ for $\mu \varepsilon Z$, of the form $G(x) = \sum_{-\infty}^{\infty} c_n e^{-n^2/2} e^{nx}$ where the function of the complex variable μ defined by $E(\mu) = \sum_{-\infty}^{\infty} c_n e^{n\mu}$ is uniformly convergent on compacts. We have $E_\mu G = E(\mu)$.

<u>Proof.</u> Let $G(x)$ be a fixed U.M.V.U. estimator. Let $E(\mu) = E_\mu G$. Then

$$(5.1) \qquad E(\mu) = (1/\sqrt{2\pi})\, e^{-\mu^2/2} \int G(x)\, e^{-x^2/2}\, e^{\mu x}\, dx$$

$E(\mu)$ is an entire function. We shall show that it is a $2\pi i$-periodic function.

We know

$$(5.2) \qquad \int e^{-x^2/2}\, e^{\mu x} = \sqrt{2\pi}\, e^{\mu^2/2}.$$

Therefore,

$$(5.3) \qquad \int e^{2\pi i x}\, e^{-x^2/2}\, e^{\mu x} = \sqrt{2\pi}\, e^{(\mu+2\pi i)^2/2}.$$

Thus, we have for $\mu \varepsilon Z$

$$(5.4) \qquad E_\mu [\, e^{2\pi^2}\, e^{2\pi i x} - 1 \,] = e^{2\pi i \mu} - 1 = 0$$

This shows

$$(5.5) \qquad (e^{2\pi^2}\, e^{2\pi i x} - 1) \varepsilon U_0.$$

Therefore, since G is U.M.V.U., we have for $\mu \varepsilon Z$

(5.6) $$E_{\mu}[G(x).\ (e^{2\pi^2}\ e^{2\pi ix} - 1)] = 0$$

It follows that

(5.7) $$E_{\mu}G = e^{2\pi i\mu}\ E_{\mu+2\pi i}\ G = E(\mu + 2\pi i), \quad \mu \varepsilon Z.$$

Now, to show that $E(\mu) = E(\mu + 2\pi i)$ for all complex μ, it is enough to show that for each positive integer n

(5.8) $$d^n\ E(\mu)/\ d\mu^n = d^n\ E(\mu + 2\pi i)/d\mu^n, \qquad \mu \varepsilon Z$$

Let us suppose that we have shown (5.8) for positive integers less than or equal to n-1. From (0.5) we have

(5.9) $$D^n(e^{\mu^2/2 + 2\pi i\mu}) = [(D + 2\pi i)^n\ e^{\mu^2/2}]\ e^{2\pi i\mu}$$

Now,

(5.10) $$\int [\ e^{2\pi^2}\ x^n\ e^{2\pi ix} - (x+2\pi i)^n]\ e^{-x^2/2}\ e^{\mu x}\ dx$$

$$= D^n[\ \int e^{2\pi^2}\ e^{2\pi ix}\ e^{-x^2/2}\ e^{\mu x}\ dx\ ]$$

$$-(D + 2\pi i)^n[\ \int e^{-x^2/2}\ e^{\mu x}\ dx\ ]$$

$$= D^n[\ \sqrt{2\pi}\ e^{\mu^2/2 + 2\pi i\mu}]$$

$$- (D + 2\pi i)^n\ (\sqrt{2\pi}\ e^{\mu^2/2}\ )$$

$$= [(D + 2\pi i)^n (\sqrt{2\pi} \; e^{\mu^2/2})] \; e^{2\pi i \mu}$$

$$- (D + 2\pi i)^n (\sqrt{2\pi}\, e^{\mu^2/2})$$

$$= (e^{2\pi i\mu} - 1)\sqrt{2\pi}(D + 2\pi i)^n (e^{\mu^2/2}).$$

Thus we have,

(5.11) $$[e^{2\pi^2} x^n e^{2\pi i x} - (x + 2\pi i)^n] \varepsilon U_o$$

Therefore, again we have for $\mu \varepsilon Z$,

(5.12) $$\int G(x) [e^{2\pi^2} x^n e^{2\pi i x} - (x + 2\pi i)^n] e^{-x^2/2} e^{\mu x} dx$$

$$= 0$$

That is,

(5.13) $$D^n [\int G(x) \; e^{2\pi^2} \; e^{2\pi i x} \; e^{-x^2/2} \; e^{\mu x} \; dx]$$

$$= (D + 2\pi i)^n \int G(x) e^{-x^2/2} \; e^{\mu x} \, dx, \quad \mu \varepsilon Z$$

(5.14) $$D^n [\sqrt{2\pi}\, e^{\mu^2/2 + 2\pi i \mu} \; E(\mu + 2\pi i)]$$

$$= (D + 2\pi i)^n [\sqrt{2\pi} \; e^{\mu^2/2} \, E(\mu)], \quad \mu \varepsilon Z.$$

But we have

(5.15) $$D^n [e^{\mu^2/2 + 2\pi i \mu} . E(\mu + 2\pi i)]$$

$$= \sum_r \binom{n}{r} [D^r E(\mu + 2\pi i)][D^{n-r}(e^{\mu^2/2 + 2\pi i \mu})]$$

$$= \sum_r \binom{n}{r} [D^r E(\mu + 2\pi i)][(D + 2\pi i)^{n-r} e^{\mu^2/2}] e^{2\pi i\mu}$$

(5.16) $$(D + 2\pi i)^n [e^{\mu^2/2} E(\mu)]$$

$$= \sum \binom{n}{r} [D^r E(\mu)][(D + 2\pi i)^{n-r} e^{\mu^2/2}]$$

Now from the induction hypothesis that (5.8) is true for positive integers less than n, it follows that (5.8) is true for n also.

Thus we have shown that $E(\mu)$ is a $2\pi i$-periodic function.

Since $E(\mu)$ is an entire periodic function of period $2\pi i$, it has the (complex) Fourier expansion

(5.17) $$E(\mu) = \sum c_n e^{n\mu},$$

where the series converges uniformly on compact sets.

(5.18) $$\int [G(x) e^{-x^2/2}] e^{i\mu x} dx = \sqrt{2\pi}\, e^{-\mu^2/2} E(i\mu).$$

Since $E(i\mu)$ is bounded,

(5.19) $$G(x) e^{-x^2/2} = 1/\sqrt{2\pi} \,.\int e^{-\mu^2/2} E(i\mu) e^{-i\mu x} d\mu$$

$$= 1/\sqrt{2\pi}. \sum_{-\infty}^{\infty} \int e^{-\mu^2/2} c_n e^{in\mu} e^{-i\mu x} d\mu.$$

It follows that

(5.20) $$G(x) = \sum_{-\infty}^{\infty} c_n e^{-n^2/2} e^{nx}$$

and G(x) satisfies the conditions stated in the theorem.

On the other hand if G(x) satisfies the conditions stated in the theorem, it is clearly U.M.V.U. //

Corollary*5.2. In the family $N(\mu, 1)$, $\mu \varepsilon Z$, the parametric function $g(\mu) = \mu$ has no U.M.V.U. estimator.

Proof. If possible, let G(x) be a U.M.V.U. estimator for μ.

It is easy to see that in $N(\mu, 1)$, for any estimator R(x),

(5.21) $$E_\mu R(x+1) = E_{\mu+1} R(x).$$

Thus we have

$$G_1(x) = G(x + 1) - 1$$

and

$$G_2(x) = G(x - 1) + 1$$

both are unbiased for μ and therefore for $\mu \varepsilon Z$

(5.22) $$E_\mu G_1^2 = E_{\mu+1} G^2 - 2\mu - 1 \leq E_\mu G^2$$

(5.23) $$E_\mu G_2^2 = E_{\mu-1} G^2 + 2\mu - 1 \leq E_\mu G^2$$

---

* The same result has also been obtained recently by Professor Morimoto and a student of his, Y. Kojima.

But (5.23) can also be written as

$$E_{\mu} G^2 + 2\mu + 1 \leq E_{\mu+1} G^2 \tag{5.24}$$

Now, from (5.22), for $\mu \varepsilon Z$

$$E_{\mu} G_1^2 = E_{\mu} G^2$$

Since U.M.V.U. estimator is essentially unique we have $G_1 = G$, or

$$G(x + 1) = G(x) + 1. \tag{5.25}$$

However, since $G(z)$ is a periodic entire function, let

$$|G(z)| \leq A$$

for $0 \leq \text{re } z \leq 1$, $0 \leq \text{im } z \leq 2\pi i$. Then for $z \varepsilon C$

$$|G(z)| \leq A + |z|. \tag{5.26}$$

But (5.26) will force $G(z)$ to be a polynomial of degree at most one and we have a contradiction. //

CHAPTER 6.

ESTIMATION IN CENSORED GAMMA MODELS

Consider a censored sample from a gamma distribution $\Gamma(p,\lambda)$ with a known form parameter $p > 0$ and an unknown scale parameter $\lambda > 0$. In this chapter we treat the problem of unbiased estimation in this family. This generalizes some results of E.N. Torgersen.

6.1. Introduction

Suppose our experiment is to observe the time of death where the observation has a gamma distribution $\Gamma(p,\lambda)$ whose density is

$$(6.1) \qquad (\lambda^p/\Gamma p)\, x^{p-1}\, e^{-\lambda x}, \qquad x > 0$$

We assume $p > 0$ is known and $\lambda > 0$ is unknown. Suppose our observation is limited to a fixed time interval $(0, t)$. If death does not occur before time t then the observation is taken as t. In this case the observation is said to be censored at time t. Without loss of generality t can be taken as 1. Such experiments occur in the study of life testing models. A minimal sufficient statistic for this model is the total number of deaths recorded together with the sum of lifelengths of individuals dying before time t.

For the case $p = 1$, and the number of observations $N \geq 2$ Torgersen (1973) showed that the minimal sufficient statistic is incomplete. Generalizing this result Unni(1976) proved that this holds for any integer $p \geq 1$. Unni also showed that when p is irrational the minimal sufficient statistic is complete. Finally Torgersen (1977) completed this result by showing when p is rational r/s where r and s are relatively prime the minimal sufficient statistic is complete iff the number of observations $N \leq s$. See also Torgersen and Unni (1978).

For the case of integer p, Unni has also characterized the class of U.M.V.U. estimators generalizing similar results of Torgersen.

In this chapter we present the results of Unni (1976).

<u>Notation.</u> In this chapter μ and λ will denote the Lebesgue measure on R and the counting measure on the set of positive integers respectively. δ_a will denote the one-

point mass at a or the distribution defined by $\delta(x-a)$ where $\delta(x)$ is the Dirac's δ-function, f_*g denotes the convolution of the functions f and g and f^{*k} denotes the k-fold convolution of f with itself.

6.2. The Exponential Representation

For a censored observation x, the distribution function $F(x)$ is given by

$$(6.2) \qquad F(x) = \begin{cases} (\lambda^p / \Gamma p) \int_0^x u^{p-1} e^{-\lambda u}\, du\,, & 0 < x < 1 \\ 1\,, & x = 1 \end{cases}$$

Let

$$(6.3) \qquad d(x) = \begin{cases} 1, & x < 1 \\ 0, & x = 1 \end{cases}$$

Then the censored observation x has a density with respect to $\mu + \delta_1$ given by

$$(6.4) \qquad p_\lambda(x) = (\lambda^p / \Gamma p)\, x^{(p-1)d(x)}\, e^{-\lambda x\, d(x)}\, [A(\lambda)]^{1-d(x)}$$

where

$$(6.5) \qquad A(\lambda) = \int_1^\infty u^{p-1} e^{-\lambda u}\, du$$

Note that $A(\lambda)$ is the Laplace transform of $v_1(x)-v_2(x)$ where

$$(6.6) \qquad v_1(x) = \begin{cases} x^{p-1}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$(6.7) \qquad v_2(x) = \begin{cases} x^{p-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Clearly $A(z)$ is an analytic function for $\text{re } z > \theta$.

A censored sample of size N is a repeated sample $x_1, \dots, x_N$ from F. The sample has joint density with respect to the measure $(\mu + \delta_1)^N$ on R^N given by

$$(6.8) \qquad p_\lambda(x_1,\dots,x_N) = (\lambda^p/\Gamma(p))^N x_1^{(p-1)d(x_1)} \dots x_N^{(p-1)d(x_N)} \times e^{-\lambda T(x_1,\dots,x_N)} [A(\lambda)]^{N-D(x_1,\dots x_N)}$$

where

$$(6.9) \qquad D(x_1,\dots, x_N) = \sum_{i=1}^{N} d(x_i)$$

$$T(x_1,\dots, x_N) = \sum_{i=1}^{N} x_i\, d(x_i)$$

Since

$$(6.10) \qquad [A(\lambda)]^{N-D} = e^{(N-D) \log A(\lambda)},$$

the formula (6.8) gives a canonical exponential representation of the joint distributions of the censored sample with respect to the measure $(\mu + \delta_1)^N$ on R^N, Since 1, $\theta_1 = -\lambda$ and $\theta_2 = - \log A(\lambda)$ are linearly independent and 1, D, T are linearly independent, the representation (6.8) is minimal canonical. It is easily seen that the natural parameter space of the family is $\Theta = \{(\theta_1, \theta_2) : \theta_1 < 0, \theta_2 \in R\}$ and the minimal canonical parameters are θ in Θ which satisfy the analytic equation

$$(6.11) \qquad e^{-z_2} - A(-z_1) = 0$$

defined on the set re $z_1 < 0$, $z_2 \in C$ so that the family (6.8) is an analytic exponential family.

6.2 The induced family of distributions of the minimal canonical statistic

For any statistic $R(x_1, \ldots, x_N)$, let $P_\lambda(R)$ denote the induced family of probability distributions of the statistic R.

Since the probability of the event $d(x_1) = 1$ is $[1 - (\lambda^p/\Gamma p)A(\lambda)]$, D is binomially distributed with the success parameter $1 - (\lambda^p/\Gamma p) A(\lambda)$, i.e., (6.9) $$P_\lambda(D = d) = \binom{N}{d} \left[1-(\lambda^p/\Gamma p)A(\lambda)\right]^d \left[(\lambda^p/\Gamma p)A(\lambda)\right]^{N-d}$$

Now, for $D = 1, \ldots, N$, $P_\lambda(T < t, D = d)$ is given by

$$\Sigma\, P_\lambda(x_{i_1} + \ldots + x_{i_d} < t, \quad x_{i_1} < 1, \ \ldots, \ x_{i_d} < 1, \ \text{all other } x_i = 1)$$

where the summation is over all combinations of d elements $i_1, \ldots, i_d$ from $1, \ldots, N$.

It follows that

$$P_\lambda(T < t, D = d) = \binom{N}{d} P_\lambda(x_1 + \ldots + x_d < t, \ x_1 < 1, \ \ldots \ x_d < 1) . P_\lambda(x_{d+1} = 1, \ldots, x_N = 1)$$

$$= P_\lambda(x_1 + \ldots + x_d < t \mid x_1 < 1, \ldots, x_d < 1) . \binom{N}{d} \left[P_\lambda(x_1 < 1)\right]^d \left[P_\lambda(x_{d+1} = 1)\right]^{N-d}$$

or

(6.10) $$P_\lambda(T < t, D = d) = P_\lambda(x_1 + \ldots + x_d < t \mid x_1 < 1, \ldots, x_d < 1) . \binom{N}{d} \left[1-(\lambda^p/\Gamma p)A(\lambda)\right]^d \left[(\lambda^p/\Gamma p)A(\lambda)\right]^{N-d}$$

Now, the conditional probability distribution of

$x_1 + \ldots + x_d$ given $x_1 < 1, \ldots, x_d < 1$ is the d-fold convolution of the conditional probability distribution of x_1 given $x_1 < 1$. $P_\lambda(x_1 | x_1 < 1)$ has the density

(6.11) $$(\lambda^p/\Gamma p)(1-A(\lambda)\lambda^p/\Gamma p)^{-1} e^{-\lambda x_1}$$

with respect to the ~~uniform~~ measure $x_1^{p-1}dx_1$ on $[0, 1]$. Therefore, $P_\lambda(x_1 + \ldots + x_d | x_1 < 1, \ldots, x_d < 1)$ has the density

(6.12) $$(\lambda^p/\Gamma p)^d [1-\lambda^p A(\lambda)/\Gamma p]^{-d} e^{-\lambda t} h'(d,t)$$

with respect to the Lebesgue measure μ on R, where $h'(d,t)$ is the density of the d-fold convolution of the ~~uniform~~ measure $x^{p-1}dx$ on $[0, 1]$ with respect to μ. For an explicit expression of $h'(d, t)$, se Feller (1966). We shall only need the fact that $h'(d, t)$ is supported by the closed interval $[0, d]$.

The conditional distribution of T given D = 0 is concentrated at the point 0.

Now, it is clear, the induced family of distributions of (D, T) has a density with respect to $\delta(0,0) + \nu \times \mu$ where $\delta(0, 0)$ is the one-point mass at (0, 0). This density can be formally written with respect to $\nu \times \mu$ as

(6.13) $$p_\lambda(d, t) = (\lambda^p/\Gamma p)^N h(d,t) [A(\lambda)]^{N-d} e^{-\lambda t}$$

where $h(0, t)$ is the Dirac's δ-function and for $d = 1, \ldots, N$

$$h(d, t) = \binom{N}{d} h'(d, t) \tag{6.14}$$

which is supported by $[0, d]$

6.4 Unbiased Estimators of Zero

Let $g(d,t)$ be a statistic and for all $\lambda > 0$ let $E_\lambda |g(d,t)| < \infty$. Then we have

$$(6.15) \quad E_\lambda [g(d,t)] = (\lambda^p / \Gamma p)^N \sum_{d=0}^{N} [A(\lambda)]^{N-d} \int_0^d g(d,t) e^{-\lambda t} h(d,t) dt$$

Let us put

$$(6.16) \quad G_d(t) = g(d,t)\, h(d,t).$$

Notice $E_\lambda |g(d,t)| < \infty$ iff for each d $G_d(t)$ is integrable

From (6.5), (6.6), (6.7) and from the familiar properties of the Laplace transform

$$(6.17) \quad E_\lambda [g(d,t)] = (\lambda^p / \Gamma p)^N \int [\sum_{d=0}^{N} (v_1 - v_2)^{*N-d} * G_d(t)] e^{-\lambda t} \, dt$$

Expanding $(v_1-v_2)^{*N-d}$ by the binomial theorem (which is easily seen to be applicable looking at the Laplace transform) and collecting the same convolution powers of v_1 in the sum in (6.17) we obtain

$$(6.18) \quad E_\lambda [g(d,t)] = (\lambda^p / \Gamma p)^N \sum_{d=0}^{N} \int [v_1^{*N-d} * R_d(t)] \, e^{-\lambda t} \, dt$$

where

$$(6.19) \quad R_d(t) = \sum_{i=0}^{d} (-1)^{d-i} \binom{N-i}{d-i} v_2^{*d-i} * G_i(t).$$

Notice $R_d(t)$ is supported by $[0, d]$.

Since we know $\int v_1 e^{-\lambda t} dt = \Gamma p / \lambda^p$, we have

(6.20) $$E_\lambda g = (\lambda^p / \Gamma p)^N \sum_{d=0}^{N} (\Gamma p / \lambda^p)^{N-d} \hat{R}_d(\lambda)$$

where

$$\hat{R}_d(\lambda) = \int R_d(t) e^{-\lambda t} dt$$

Thus we obtain

(6.21) $$E_\lambda g = g(0,0) + \sum_{d=1}^{N} \lambda^{pd} \hat{R}_d(\lambda) / (\Gamma p)^d.$$

Let $g(d, t)$ be an unbiased estimator of zero.

As $\lambda \longrightarrow 0$, $\Sigma [\lambda^{pd} / (\Gamma p)^d] \hat{R}_d(\lambda) \longrightarrow 0$.

Therefore, from (6.21),

(6.22) $$g(0, 0) = 0.$$

Thus it is clear that for the case $N = 1$ (D, T) is complete.

Since $R_d(t)$, $d = 1, \ldots, N$ are integrable functions supported by $[0, d]$, $\hat{R}_d(\lambda)$ are entire functions.

Now suppose p is an integer. Then it is well-known that $\lambda^{pd} \hat{R}_d(\lambda) / (\Gamma p)^d$ is the Laplace transform of the compactly supported distribution $(\Gamma p)^{-d} (d^{pd}/dt^{pd}) [R_d(t)]$.

To find the most general unbiased estimators of zero, take integrable functions $R_d(t)$ supported by $[0, d]$ satisfying the differential equation (in the sense of distributions)

$$\sum_d (\Gamma p)^{-d}(d^{pd}/dt^{pd})[R_d(t)] = 0 \tag{6.23}$$

That there exist nontrivial $R_d(t)$, $d = 1,\ldots,N$; $N \geq 2$, satisfying (6.23) is obvious and in any case examples are given in the proof of Theorem 6.2.

The question remains about the case of a noninteger p. A complete answer to the question of completeness of (D,T) is given by the following theorem.

<u>Theorem</u> 6.1. Let $x_1, \ldots, x_N$ be a repeated sample from a distribution function F_λ where

$$F_\lambda(x) = \begin{cases} \int_0^x (\lambda^p/\Gamma p)\, u^{p-1} e^{-\lambda u}\, du, & 0 < x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

with an unknown parameter $\lambda > 0$ and a known parameter $p > 0$. Then the family of joint distributions of $x_1,\ldots, x_N$ admits a complete and sufficient statistic iff it admits a boundedly complete and sufficient statistic and this is the case iff either

(1) p is irrational

or

(2) p is rational $= r/s$ where the integers r and s are relatively prime and the number of observations $N \leq s$.

For a proof of the above theorem, see Torgersen and Unni (1978).

Notice the fragility of the property of completeness. Although the distributions are strongly continuous in p, the situation when p is rational is entirely different from the situation when p is irrational.

6.5. Unbiased Estimation

Next theorem characterizes the class of U.M.V.U. estimators in the censored gamma families when p is an integer. A similar theorem is true also in the incomplete rational case. This is proved in Torgersen and Unni.

Theorem 6.2. Let $x_1, \ldots, x_N$, $N > 1$ be a repeated sample from a censored gamma family with an integer p. Then an estimator ϕ with everywhere finite variance is a U.M.V.U. estimator for its expectation iff ϕ is a function of the minimal sufficient statistic (D, T) of the form

$$\begin{aligned} \phi(0, 0) &= c_0 \\ \phi(d, t) &= c, \qquad d = 1, \ldots, N-1 \\ \phi(N, t) &= c, \qquad 1 \leq t \leq N \end{aligned} \tag{6.24}$$

where c and c_0 are arbitrary constants and $\phi(N, t)$, $0 \leq t \leq 1$, is an arbitrary function satisfying the square integrability condition

$$\int_0^1 [\phi(N, t)]^2 \, t^{Np-1} \, e^{-\lambda t} \, dt < \infty \tag{6.25}$$

Proof. (1) Sufficiency. Note that for a statistic ϕ of the given form, the existence of variance for each $\lambda > 0$ is equivalent to (6.25) from the fact that $\phi(N, t)$ is defined on $x_1 + \ldots + x_N < 1$ and from the additivity property of the gamma distributions.

If $g(d, t)$ is an unbiased estimator of zero then from (6.17)

$$\sum_{d=0}^{N} (v_1 - v_2)^{*N-d} * G_d(t) \equiv 0 \tag{6.26}$$

Now, $(v_1 - v_2)^{*N-d} * G_d(t)$ has support in $[N - d, \infty)$.

Therefore, we have

(6.27) $$g(N, t) = 0 , \quad 0 \leq t < 1$$

Also, from

$$g(0, 0) = 0.$$

Now, let $\phi(d, t)$ be a function of the form (6.24). For any unbiased estimator of zero $g(d,t)$, $E_\lambda g^2 < \infty$ for $\lambda > 0$,

$$E_\lambda \phi g = cE_\lambda g = 0.$$

Thus, ϕ is U.M.V.U.

(2) <u>Necessity</u>. From (6.17), $g(d, t)$ is an unbiased estimator of zero iff $g(0, 0) = 0$ and

$$\sum_{d=1}^{N} (v_1 - v_2)^{*N-d} * G_d(t) \equiv 0$$

Suppose $g(d, t)$ is an unbiased estimator of zero (U.E.Z.) such that $g(d, t) = 0$ for $d \neq 1,2$. Then we have from (6.27) $(v_1-v_2)^{*N-1} * G_1 + (v_1-v_2)^{*N-2} * G_2 = 0$ and looking at the Laplace transform we see

$$(6.25) \qquad G_2(t) = -(v_1 - v_2) * G_1$$

$$= \begin{cases} -\int_0^{t-1} (t-u)^{p-1} G_1(u)\, du & t < 2 \\ -\int_0^{1} (t-u)^{p-1} G_1(u)\, du & t \geq 2 \end{cases}$$

since G_1 is supported by $[0, 1]$. Now, the fact that G_2 is supported by $[0, 2]$ implies

$$(6.26) \qquad \int_0^1 (t-u)^{p-1} G_1(u)\, du = 0 \quad \text{for all} \quad t \geq 2$$

or

$$(6.27) \qquad \int_0^1 u^k G_1(u)\, du = 0 \quad \text{for} \quad k = 0, \ldots, p-1.$$

Conversely, for G_1 satisfying (6.27) if we define $G_2(t)$ by (6.25) and $g(d, t) \equiv 0$ for $d \neq 1, 2$ then $g(d, t)$ is a U.E.Z.

Now, let us suppose that $\phi(d, t)$ is a U.M.V.U. estimator. Again, $\phi(d, t)\, g(d, t)$ is a U.E.Z. and it vanishes for $d \neq 1, 2$, where $g(d, t)$ is any U.E.Z. vanishing for $d \neq 1,2$.

This implies

$$(6.28) \qquad \int_0^1 u^k G_1(u)\, \phi(1, u)\, du = 0 \quad \text{for} \quad k = 0,\ldots,p-1$$

for any G_1 satisfying (6.27).

Take $F \varepsilon C_0^\infty (0, 1)$ and put $G_1 = (d^p/dt^p) F(t)$.
Then we have

$$(6.29) \qquad \int_0^1 u^k \left[(d^p/du^p)F(u)\right]du = \int_0^1 \left[(-1)^p d^p u^k/du^p\right]F(u)\, du$$

$$= 0 \quad \text{for} \quad k = 0, \ldots, p-1.$$

Therefore, from (6.28),

$$(6.30) \qquad \int_0^1 u^k\phi(1, u) \left[d^p F(u)/du^p\right] du = 0, \quad k = 0,\ldots,p-1.$$

But (6.30) implies, in the sense of distributions,

$$(6.31) \qquad (d^p/du^p) \left[u^k\phi(1, u)\right] = 0, \quad k = 0,\ldots, p-1$$

The differential equation (6.31) shows that $u^k\phi(1, u)$ is a polynomial of degree less than p in the variable u for $k = 0, \ldots, p-1$. It follows that $\phi(1, u)$ is a constant, say c.

Again, from (6.25), since $g(d, t)\ \phi(d, t)$ is a U.E.Z., for a U.E.Z. $g(d, t)$ of the above type

$$(6.32) \qquad G_2(t)\, \phi(2, t) = -\int_0^{t-1} (1-u)^{p-1} G_1(u)\, \phi(1,u)\, du$$

$$= -c \int_0^{t-1} (t-u)^{p-1}\, G_1(u)\, du$$

$$= -c\, [-G_2(t)].$$

For any $t \in (1, 2)$ we can take G_1 and G_2 so that $G_2 \neq 0$. Therefore,

$$\phi(2, t) = c, \qquad t \in (1, 2).$$

Now use the same argument replacing G_1 by G_2 and G_2 by G_3. Then we obtain that $\phi(2, t)$ is a constant. This constant must be the same as $\phi(1, t)$ because for $t \in (1, 2)$, $\phi(1, t) = \phi(2, t)$.

Proceeding in the same manner, the necessity part of the theorem is proved. //

The next theorem characterizes the U.M.V.U. estimable functions.

<u>Theorem</u> 6.3. In the censored gamma family with an integer p, α parametric function $g(\lambda)$ is U.M.V.U. estimable iff it has the form

$$g(\lambda) = a + b(\lambda^p / p)^N [A(\lambda)]^N + (\lambda^p / p)^N \int_0^1 f(t) t^{Np-1} e^{-\lambda t} \, dt \tag{6.33}$$

where

$$\int_0^1 f^2(t) t^{Np-1} e^{-\lambda t} \, dt < \infty \tag{6.34}$$

<u>Proof</u>. Suppose $g(\lambda)$ is a U.M.V.U. estimable function and let $\phi(d, t)$ be a U.M.V.U. estimator for $g(\lambda)$. Now ϕ is necessarily of the form (6.22). So,

$$\begin{aligned} E_\lambda [\phi(d, t)] &= c_0(\lambda^p / \Gamma p)^N [A(\lambda)]^N \\ &+ c \;\; 1 - (\lambda^p / \Gamma p) [A(\lambda)]^N - (\lambda^p / \Gamma p)^N \int_0^1 t^{Np-1} e^{-\lambda t} \\ &+ (\lambda^p / \Gamma p)^N \int_0^1 \phi(N, t) t^{Np-1} e^{-\lambda t} \, dt. \end{aligned} \tag{6.35}$$

Put $a = c$, $b = c_0 - c$ and $f(t) = \phi(N, t) - c$ for $0 < t < 1$, to obtain the form (6.33)

The sufficiency part is, now, clear.

BIBLIOGRAPHY

T.W. Anderson (1969). Statistical inference for covariance matrices with linear structure. In Multivariate Analysis, Vol.II (edited by P.R. Krishnaiah) 55-66.

T.W. Anderson (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. Ann. Stat. 3, 1283-1304.

O. Barndorff-Nielsen (1970). Exponential Families-Exact Theory. Various Publication Series No.19. Matematisk Institut, Aarhus Universitet.

N.N. Chentsov (1966). A systematic theory of exponential families of distributions. Theory Prob. App. 11, 425-435.

W.F. Donoghue, Jr. (1969). Distributions and Fourier Transforms. Academic Press, London.

B. Effron (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). Ann. Stat. 3, 1189-1243.

W. Feller (1966). Introduction to Probability Theory and Applications, Vol.II. Wiley, New York.

T. Ferguson (1962). Location and scale parameters in exponential families of distributions. Ann. Math. Stat. 33, 986-1001.

W. Gautschi (1959). Some Remarks on Herbach's paper 'Optimum Nature of the F-test for Model II in the Balanced Case' Ann. Math. Stat. 30, 960-963.

L.H. Herbach (1959). Properties of Model II type analysis of variance tests. Ann. Math. Stat. 30, 939-959.

L. Hormander (1963). Linear Partial Differential Operators. Springer-Verlag, Berlin.

S. T. Jensen (1975). Covariance hypotheses which are linear in both the covariance and the inverse covariance. Pre-print of the Institute of Mathematical Statistics, University of Copenhagen.

A.M. Kagan (1968). Estimation theory for families with location and scale parameters, and for exponential families. Proceedings of the Steklov Institute of Mathematics No. 104.

A.M. Kagan and V.P. Palamodov (1967). Conditions of optimal unbiased estimation of parametric functions for incomplete exponential families with polynomial ties. Dokl. Akad.Nauk SSSR

A.M. Kagan and V.P. Palamodov (1968). New results in theory of estimation and testing hypotheses for problems with nuisance parameters. Supplement to Linnik's Statistical Problems with Nuisance Parameters. 232-240.

E.L. Lehmann (1959). Testing Statistical Hypotheses. Wiley, New York

Ju. V. Linnik (1968). Statistical Problems with Nuisance Parameters. American Mathematical Society, Providence.

C. R. Rao (1965). Linear Statistical Inference and its Applications. John Wiley, New York.

C. R. Rao (1971). Minimum variance quadratic unbiased estimation of variance components. J. Mult. Anal. 1, 445-456.

J. Seely (1971). Quadratic subspaces and completeness. Ann. Math. Stat. 42, 710-721.

J. Seely (1977). Minimal sufficient statistics and completeness for multivariate normal families. Sankhya 39A(2), 170-185.

J. N. Srivastava (1966). On testing hypotheses regarding a class of covariance structures. Psychometrika 31, 147-164.

E. N. Torgersen (1973). Uniformly minimum variance unbiased estimators based on a sample from right truncated and right accumulated exponential distributions. Statistical Research Report, Inst. Math., University of Oslo.

E. N. Torgersen (1977). Fixed point censoring of gamma distributions. Statistical Research Report, Inst. Math., University of Oslo.

E. N. Torgersen and K. Unni (1978). Censored exponential models. To appear.

K. Unni (1976). Unbiased estimation in type I censored gamma distributions. Research Report. Indian Statistical Institute, Calcutta.

R.A. Wijsman (1958). Incomplete sufficient statistics and similar tests. Ann. Math. Stat. 29, 1028-1045.

ps: