NIETO DE PASCUAL, Jose. THEORY OF MINIMUM VARIANCE ESTIMATION WITH APPLICATIONS.

Iowa State University of Science and Technology
Ph.D., 1961
Mathematics

THEORY OF MINIMUM VARIANCE ESTIMATION WITH APPLICATIONS

by

Jose Nieto de Pascual

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1961

TABLE OF CONTENTS

I. INTRODUCTION

The estimation of the parameters of a probability distribution is a problem of basic importance in modern statistics. A unified solution to this problem is not yet known. Several theories of estimation are available, however, which apply under rather general conditions and which may be placed into two broad classes, the theory of interval estimation and the theory of point estimation. The present study shall be concerned with a particular type of estimation that belongs to the latter class.

In general, the problem of point estimation may be stated as follows:

Given a population with a probability density function

$$\phi\,(x;\, \theta_1,\, \theta_2,\, \ldots,\, \theta_k)$$

which would be completely determined if the value of the parameters $\theta_1, \theta_2, \ldots, \theta_k$ were known. By means of a random sample of observations $x_1, x_2, \ldots, x_n$ from this population, it is required to determine some function $t(x_1, \ldots, x_n)$ such that the distribution of $t(x_1, \ldots, x_n)=t(x)$ upon repeated sampling from the same population, with samples of equal size n, will be 'concentrated around' the values of the parameters of the population. The properties of the estimator $t(x)$ are assessed according to the properties of the population of estimates thus generated. The function $t(x)$ of the random sample $x_1, \ldots, x_n$ is called the Estimator (or the Statistic) and the values it assumes with each particular random sample are called the Estimates of the parameters of the population. It should be clear that $t(x)$ is a random variable.

Several properties have been suggested that $t(x)$ should satisfy in order to qualify as a "good" estimator -- that is, an estimator with a distribution 'closely concentrated' around the value of $(\theta_1, \ldots, \theta_k)$. The properties of unbiasedness, consistency, efficiency and sufficiency are well known. The first two are limiting properties; the efficiency of an estimator may or may not be defined as an asymptotic property; the concept of sufficiency applies for any sample size.

For simplicity, consider a population with one parameter θ, with probability density function

$$\phi(x_1; \theta).$$

Let the joint density of the sample $x_1, \ldots, x_n$ be denoted by

$$\phi(x_1, x_2, \ldots, x_n; \theta) = \phi(x; \theta)$$

The estimator $t(x)$ is unbiased if its expectation is equal to θ, that is, if

$$\int t(x)\, \phi(x; \theta)\, dx = \theta$$

where integration is over the sample space and $dx = dx_1 \ldots dx_n$.

The estimator $t(x)$ is consistent if, for any positive ϵ and δ, there exists an n_o such that

$$P\left\{ \left| t(x) - \theta \right| < \epsilon \right\} > 1 - \delta$$

for $n > n_o$, or in words, the probability that $t(x)$ differs from θ by less than ϵ is greater than $1 - \delta$ for n sufficiently large. Thus, $t(x)$ is a consistent estimator of θ if it converges in probability to θ.

Suppose that $u(x)$ is any other estimator of θ. Then, if the variance of $t(x)$ is less than the variance of $u(x)$, $t(x)$ is more efficient than $u(x)$. Another definition of efficiency, due to Cramér, shall be given in Chapter II.

Suppose that the joint probability density of $t(x)$ and $u(x)$ is of the form

$$\phi \left\{ t(x),\ u(x) \right\} = \phi_1 \left\{ t(x),\ \theta \right\} \cdot \phi_2 \left\{ u(x) \mid t(x) \right\}$$

where ϕ_1 is the probability density for $t(x)$ and ϕ_2 is the conditional density of $u(x)$ given $t(x)$, which is independent of θ. Then $t(x)$ is said to be a sufficient estimator, or a sufficient statistic, for the parameter θ. It can be shown from this definition that the necessary and sufficient condition for $\phi(x); \theta)$ to admit a sufficient estimator is that $\phi(x); \theta)$ may be factored as

$$\phi(x;\ \theta) = \phi_1 \left\{ t(x),\ \theta \right\} \cdot \phi_o(x)$$

where $\phi_o(x)$ is a function of $x_1, \ldots, x_n$ only. [See 12, p. 120]. It is essential, for this condition to apply, that the range of the x's be independent of θ.

Several criteria are known that yield techniques for finding estimators with desirable properties such as the ones given above. Among these criteria, the methods of estimation by Least Squares and by Maximum Likelihood are widely employed; other criteria such as the method of

Minimum Chi-Square or the method of Moments are also used.

Least Squares estimation is most successful when applied to linear observational equations affected by error, as in regression problems. The normal equations for the optimal values of the unknown parameters can be deduced [2] by two different sets of postulates. The first set, A, proceeds by

i) assuming a normal distribution of errors of observation,

ii) accepting as optimal values for the unknown parameters those values which make the joint density of the sample a maximum.

The second set of postulates, B, proceeds by

i) assuming that the optimal values are unbiased linear combinations of the observations,

ii) accepting those particular linear combinations for which the error variance is a minimum.

For example [16, p. 267], consider a sample of n independent values $x_1, \ldots, x_n$ drawn from a population π_i with mean μ_i and variance σ_i^2. Suppose a function θ is defined by

$$\theta = \sum_{j=1}^{s} b_j \, p_j$$

where the b_j's are unknown constants and the parameters p_j depend on μ_i according to the equation

$$\mu_i = \sum_{j=1}^{s} a_{ij} \, p_j ; \; s \leq n \; .$$

The a_{ij}'s are also known. Then an unbiased estimator $t(x)$ of θ, with minimum variance, may be written as

$$t(x) = \sum_{j=1}^{n} \lambda_j x_j$$

where $t(x)$ is found by substituting for the p_j's in the expression for θ the functions q_j obtained by minimizing

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left\{ x_i - \sum_{j=1}^{s} a_{ij} q_j \right\}^2$$

with respect to the q_j's. This is often referred to as Markoff's theorem.

Maximum Likelihood estimation resembles postulate ii) of the set A above in that it accepts, as the optimal estimator of the unknown parameter θ, a function $t(x)$ of the sample values $x_1, \ldots, x_n$ which maximizes the joint density of the x_j's. The evaluation of the merits of the method of Maximum Likelihood is done a posteriori; it can be shown [16, pp. 13-22] that Maximum Likelihood estimators are consistent, tend to normality for large n, have minimum variance in the limit (at least) and provide sufficient estimators whenever they exist.

The postulates A and B given above suggest alternative approach to the problem of estimation of statistical parameters. Consider the following set C of postulates:

i) The estimator $t(x)$ is unbiased, i.e.,

$$E\Big((t(x)\Big) = \theta$$

ii) The variance of $t(x)$ is a minimum, i.e.,

$$E\left\{t(x) - \theta\right\}^2 = \text{minimum}$$

These postulates are indeed arbitrary, but so are those of any other proposed estimation procedure. The estimators amenable to postulates C are known as minimum variance unbiased estimators, which have been discussed rather extensively in the literature of the past twenty years, as will be seen in the next chapter.

It has been argued that the class of minimum variance unbiased estimators is much too restricted, for it excludes many convenient estimators that have a tolerable bias. A discussion of minimum variance estimators that are biased, however, without any consideration as to the bias, is a rather futile proposition, for the possibility of an unduly large bias is not eliminated. To give an extreme example, one could make the estimator equal to a constant, thus having a variance of zero. The bias, however, is uncontrolled and such an estimator is otherwise irrelevant for most practical and theoretical considerations.

A well known, often useful method of combining the bias and the variance is to form the 'Mean Square Error' (MSE) defined by

$$\text{MSE}\left(t(x)\right) = E\left\{t(x) - \theta\right\}^2$$

where $t(x)$ is an estimator of θ. Let

$$E\left(t(x)\right) = \mu(\theta).$$

Then

$$\text{MSE}\left(t(x)\right) = \text{Var } t(x) + \left\{\mu(\theta) - \theta\right\}^2$$

where $\mu(\theta) - \theta = \text{Bias in } t(x)$.

The minimization of the MSE contains the minimization of the variance as the special case when the bias is zero. However, even if the MSE is minimized, it would still be convenient to extricate the bias component from it, for there may be considerations of importance as regards the bias relative to the standard error. Furthermore, there could be no claim of uniqueness of the estimators obtained by minimizing the MSE, for all competitors with the same minimum MSE would have to be regarded as equally 'efficient' when in fact the bias in some of these would render them inappropriate for the purpose at hand.

A more flexible and useful approach is, therefore, to consider the study of minimum variance estimators with a fixed, prescribed bias. Here considerations of uniqueness are feasible, and it would be possible to select from these estimators the one with the most desirable properties.

In the approach followed in this study, the minimization was carried out by utilizing the Calculus of Variations. The general minimization of the MSE by certain estimators $t(x)$ would be formulated as follows:

To obtain the estimator $t(x)$ such that

$$\int_{R(x)} \left\{ t(x) - \theta \right\}^2 \phi(x;\, \theta)\, dx = \text{minimum}$$

The Calculus of Variations solution to this problem produces a trivial result from a statistical point of view, namely, $t(x)$ should be equal to θ. If a side-condition is imposed, however, that the class of competitor estimators should be subject to the same bias function, say $\mu(\theta) - \theta$, the problem now envisaged is one of the class of isoperimetric problems in the classical Calculus of Variations. The side-condition may be written as

$$\int_{R(x)} t(x)\, \phi(x;\, \theta)\, dx = \mu(\theta)$$

identically in θ.

The problem may be now formulated as follows: among all estimators $t(x)$, with finite variance and with expectation $\mu(\theta)$, to choose from them the one that minimizes the variance for <u>a</u> <u>particular</u> θ.

Now we would expect that the solution to this problem would still involve θ, for the fact is that the concept of $t(x)$ not involving θ is completely foreign to the Calculus of Variations. It is, indeed, a fortunate accident that there are density functions for which the solutions $t^*(x)$ are functions of x only, independent of θ, for it is these solutions only that make statistical sense. It is to this statistically sensible situation that the present study will be devoted.

II. REVIEW OF THE LITERATURE

The problem of minimum variance unbiased estimation, in the case of one population parameter θ, consists in determining the estimator $t(x)$ that makes

$$\int_{R(x)} \left\{ t(x) - \theta \right\}^2 \phi(x; \theta)\, dx \tag{1}$$

a minimum, where $t(x_1, \ldots, x_n) = t(x)$ is a function of a random sample $x_1, x_2, \ldots, x_n$, independent of θ, and with expectation θ; $\phi(x; \theta)$ is the joint density of the random sample $x_1, \ldots, x_n$.

The earliest formal approach to the theory of minimum variance unbiased estimation, without prior assumptions as to the model for the observational equations, was made by Aitken and Silverstone [2] in 1942. These authors considered the problem from the point of view of the Calculus of Variations. The equation they called the corresponding "Euler equation" for this problem was

$$\left\{ t(x) - \theta \right\} \cdot \phi(x; \theta) - \lambda(\theta) \frac{\partial}{\partial \theta} \phi(x; \theta) = 0 \tag{2}$$

or, also

$$t(x) - \theta = \lambda(\theta) \frac{\partial}{\partial \theta} \log \phi(x; \theta) \tag{3}$$

where $\lambda(\theta)$ is a constant, depending on θ but independent of x. The range of x was assumed independent of θ.

Aitken and Silverstone wrongly considered (2) the Euler equation, which is a necessary condition, for the minimization problems, with the unbiasedness property being an isoperimetric side-condition. The fact that (2) is only one of a class of limiting forms of the necessary condition is dealt with in the next chapter.

From Equation (2), Aitken and Silverstone showed that the estimator $t(x)$ that satisfies (2) has minimum variance given by $\lambda(\theta)$, because

$$\operatorname{Var} t(x) = \int_{R(x)} t^2(x)\,\phi(x;\theta)\,dx - \theta^2$$

$$= \int_{R(x)} t(x)\cdot t(x)\,\phi(x;\theta)\,dx - \theta^2$$

and by (2),

$$= \int_{R(x)} t(x)\left\{\lambda(\theta)\frac{\partial}{\partial\theta}(x;\theta) + \theta\,\phi\,(x;\theta)\right\} dx - \theta^2$$

$$= \lambda(\theta)\int_{R(x)} t(x)\frac{\partial}{\partial\theta}\phi\,(x;\theta)\,dx + \theta\int_{R(x)} t(x)\,\phi(x;\theta)\,dx - \theta^2$$

and, since the range of x is independent of θ,

$$= \lambda(\theta)\frac{\partial}{\partial\theta}\int_{R(x)} t(x)\,\phi\,(x;\theta)\,dx + \theta^2 - \theta^2$$

$$= \lambda(\theta), \text{ since } E\,t(x) = \theta.$$

By integrating (3), Aitken and Silverstone found that the form of the densities amenable to this estimation method must be of the form

$$\phi(x;\theta) = e^{P(\theta) + t(x)\,Q(\theta) + R(x)} \tag{4}$$

where $P(\theta)$, $Q(\theta)$ are functions of θ only, with $\frac{d}{d\theta}\,Q(\theta) = Q'(\theta) \neq 0$, and $t(x)$, $R(x)$ are functions of x only. The function of θ that would be unbiasedly estimated by $t(x)$ with minimum variance is determined from (4) to be

$$\mu(\theta) = -\frac{\partial P(\theta)}{\partial Q(\theta)} = -\frac{P'(\theta)}{Q'(\theta)} \tag{5}$$

and the minimum variance is given by

$$\sigma^2_{min} = -\frac{1}{n}\frac{P''(\theta)}{Q''(\theta)} \tag{6}$$

Earlier, Koopman [17] had found that the general form of a density function admitting the determination of a sufficient statistic for θ, in the one-parameter case, is precisely the form given by (4); the density $\phi(x;\ \theta)$ is assumed analytic and non-zero over some continuous range of θ, and $P(\theta)$, $Q(\theta)$ are single-valued, analytic functions of their arguments.

The link between minimum variance unbiased estimators and sufficient statistics was thus established. This link has been emphasized in the literature since 1945, especially in the writings of C. R. Rao [21, 22, 23].

In 1945, Rao [21] showed that the variance of any unbiased estimate of $\mu(\theta)$, a function of θ, has a lower bound given by the inverse of I, where

$$I = \text{Var}\left\{\frac{\partial}{\partial\theta}\log\phi(x;\ \theta)\right\} = E\left\{-\frac{\partial^2}{\partial\theta^2}\log\phi(x;\ \theta)\right\} \tag{7}$$

is the information on θ supplied by a sample of n observations, as defined by R. A. Fisher. The condition under which this result is applicable is that the range of x be independent of θ. The quantity I is defined independently of any method of estimation. Rao's approach is as follows:

Let $t(x)$ be unbiased for θ. Then

$$\int_{R(x)} t(x)\, \phi(x;\, \theta)\, dx = \theta \tag{8}$$

If the range of x is independent of θ, it is possible to differentiate under the integral sign with respect to θ,

$$\frac{\partial}{\partial\theta} \int_{R(x)} t(x)\, \phi(x;\, \theta)\, dx = \int_{R(x)} t(x) \frac{\partial}{\partial\theta} \phi(x;\, \theta)\, dx = 1 \tag{9}$$

which means that the covariance of $t(x)$ and $\frac{\partial}{\partial\theta}\{\log \phi(x;\, \theta)\}$ is unity. But

$$1 = \mathrm{Cov}^2 \left(t(x),\ \frac{\partial}{\partial\theta}\{\log \phi(x;\, \theta)\}\right) \leq \mathrm{Var}\ t(x)\ \mathrm{Var}\left(\frac{\partial}{\partial\theta}\{\log \phi(x;\, \theta)\}\right).$$

Hence,

$$\mathrm{Var}\ t(x) \geq \frac{1}{\mathrm{Var}\left(\frac{\partial}{\partial\theta}\{\log \phi(x;\, \theta)\}\right)} \tag{10}$$

If the expectation of $t(x)$ is $\mu(\theta)$, then (10) becomes

$$\mathrm{Var}\ t(x) \geq \frac{\mu'(\theta)}{\mathrm{Var}\left(\frac{\partial}{\partial\theta}\{\log \phi(x;\, \theta)\}\right)} \tag{11}$$

These results were also obtained independently by Cramér [11, p. 480]. The lower bound for the variance given by (10) or (11) is known as the information limit to the variance [20, p. 130] or the Cramér-Rao lower bound.

The above results have been refined further by several authors. Wolfowitz [18, p. 2-3] has given a set of regularity conditions that replace the rather "unpleasant" [18, p. 2-3] conditions given by Cramér in his original derivation. Wolfowitz's conditions of regularity are the following [12, p. 103]:

i) θ lies in an open interval D (infinite or semi-infinite) of the real line.

ii) $\frac{\partial}{\partial\theta}\phi(x;\theta)$ exists for all θ in D and almost all x. The exceptional set must not depend on θ.

iii) $\int\phi(x;\theta)\,dx$ may be differentiated under the integral sign.

iv) $E\left\{\frac{\partial}{\partial\theta}\log\phi(x;\theta)\right\}^2 > 0$ for every $\theta \in D$.

From the information limit to the variance, Cramér defined an unbiased estimate as "efficient" if its variance is equal to this (Cramér-Rao) lower bound.

The fact that the Cramér-Rao lower bound is not the greatest lower bound for the variance, in general, has been shown by Bhattacharyya [5] and by Chapman and Robbins [9]. Bhattacharyya has shown that there exist lower bounds that are higher or equal to the Cramér-Rao lower bound. He obtained these bounds as the ratio of two determinants, $\left|\lambda_{ij}^{(1)}\right|$ and $\left|\lambda_{ij}^{(2)}\right|$, where i, j = 1, 2, ..., k for $\lambda_{ij}^{(1)}$, i, j = 2, 3, ..., k for $\lambda_{ij}^{(2)}$, and the λ_{ij} are defined as the expectations of

$$\left(\frac{1}{\phi(x;\theta)}\right)^2\left(\frac{\partial^i}{\partial\theta^i}\phi(x;\theta)\right)\left(\frac{\partial^j}{\partial\theta^j}\phi(x;\theta)\right) \qquad (12)$$

Chapman and Robbins [9] have obtained a lower bound for the variance of estimators without utilizing regularity conditions. The bound they found is at least as sharp as the Cramér-Rao lower bound. These authors utilize differences instead of differentials in their derivation.

Bhattacharyya's results have been extended by Seth [26b] to sequential estimation.

Based upon Koopman's result on the form of a distribution admitting sufficient statistics, Rao [21] concludes that if a sufficient statistic and an unbiased estimator exist for θ, then the "best" (in the sense of minimum variance) unbiased estimate of θ is an explicit function of the sufficient statistic. However, in his book [20, p. 150] Rao states that he has shown in his 1945 paper [21] that minimum variance estimates must necessarily be functions of sufficient statistics. This implies the strong assertion that minimum variance estimators do not exist for distributions that do not admit sufficient statistics. That this is not the case will be shown by example in Chapter III. Rao has attempted to prove precisely the statement in his book; the main theorem in this context is the following [22] :

<u>Theorem</u>: If the parent distribution is such that in independent samples of any size from it, uniformly minimum variance estimators can be constructed for any function $T(\theta) \in U$ (where $T(\theta) \in U$ if it admits an unbiased estimator), then the distribution is of a special type known to admit sufficient statistics.

Rao himself agrees that the assumption that every function $T(\theta) \in U$ admits a uniformly minimum variance estimator appears to be very restrictive.

An important step forward was taken by Blackwell [6] in 1947, with the introduction of the concept of complete sufficient statistics.

Let the vector $x = (x_1, \ldots, x_n)$ have a probability density function $\phi(x; \theta)$ and suppose there exists an estimator $t(x)$ that is sufficient for θ. Let $\phi_1\left(t(x) \mid \theta\right)$ denote the density of $t(x)$. Now, if for every function $g(t)$,

$$\int_{R(t)} g(t)\, \phi_1\left(t(x) \mid \theta\right) dt \equiv 0 \qquad (13)$$

implies $g(t) = 0$ almost everywhere, the statistic $t(x)$ is called a complete sufficient statistic for θ [12, p. 134]. This definition of "completeness" of a system of sufficient statistics, which states that there are no non-trivial unbiased estimators of zero, is in fact the definition of "completeness" (Vollständigkeit) of a system of functions given in Courant and Hilbert [10, p. 94], which states that a system of functions is complete if there is no function orthogonal to all the functions of the system.

The following theorem, due to Blackwell, is of considerable importance [6; 18, p. 3-6]:

Theorem: Let x have density $\phi(x; \theta)$, and let $t(x)$ be a complete sufficient statistic for θ. Then every estimable function $g(\theta)$ (that is, such that there exists an unbiased estimator of $g(\theta)$) – possesses an unbiased estimator with uniformly smallest variance, and this estimator is the unique (apart from a set of measure zero) estimator of $g(\theta)$ which is a function of $t(x)$.

The proof of this theorem utilizes some results on conditional expectation [12, p. 133; 18, p. 3-7], particularly the fact that, if $t(x)$ is sufficient for θ and $y(x)$ is any unbiased estimate of θ, then

$$E\left(y(x) \mid t(x)\right) \qquad (14)$$

which is a function of $t(x)$, is also unbiased for θ, but with the added convenience of having its variance smaller than or equal to the variance of $y(x)$. The theorem not only tells whether a given unbiased estimator has minimum variance uniformly, but it further enables us to find such optimal estimators. For example, once it has been established [18, p. 3-9] that $T = \Sigma x_i$ is complete sufficient for λ in the Poisson distribution with parameter λ, it can be shown that

$$\frac{T}{n-1} \left(\frac{n-1}{n}\right)^T \tag{15}$$

is the minimum variance unbiased estimator of $\lambda e^{-\lambda}$. For suppose bTa^T is the form of the uniform minimum variance unbiased (UMVU) estimator of $\lambda e^{-\lambda}$. Then

$$E(bTa^T) = e^{-n\lambda} \sum_{t=o}^{\infty} \frac{bta^t (n\lambda)^t}{t!}$$

$$= e^{-n\lambda} \quad abn\lambda \sum_{t=o}^{\infty} \frac{a^{t-1} (n\lambda)^{t-1}}{(t-1)!}$$

$$= abn\lambda \quad e^{-\lambda(n-an)}$$

Hence, take $a = \frac{n-1}{n}$; $b = \frac{1}{n\,a} = \frac{1}{n-1}$, to obtain (15) as the UMVU for $\lambda e^{-\lambda}$.

An interesting application of complete sufficient statistics was given by Tate and Goen [29]. They utilize a result due to Tukey [31] that if a family of distributions admits a set of sufficient statistics, then the family obtained by truncation to a fixed set, or by fixed selection, also admits the same set of sufficient statistics.

Tate and Goen considered the truncated Poisson distribution, and obtained the characteristic function of the sum of a sample x_1, x_2, ..., x_n drawn therefrom. When the truncated portion is the zero class, the statistic

$$T_o = \sum_{i=1}^{n} x_i$$

is a complete sufficient statistic for the parameter of the truncated Poisson, by Tukey's result quoted above. Then, if an unbiased estimator of the Poisson parameter λ, based on T_o, exists, it will be the UMVU for λ. Tate and Goen obtained the density of T_o by applying the inversion formula for characteristic functions [19, p. 434]. With this result, from the condition of unbiasedness, they found that the UMVU for λ is given by

$$\sum_{i=1}^{n} x_i \left(\frac{\mathfrak{S}_{t-1}^{n}}{\mathfrak{S}_{t}^{n}} \right) \tag{16a}$$

or, also

$$\bar{x} \left(1 - \frac{\mathfrak{S}_{t-1}^{n-1}}{\mathfrak{S}_{t}^{n}} \right) \quad . \tag{16b}$$

where $\mathfrak{S}_t^n$ is a Sterling number of the second kind [15, p. 169], and $t = \sum^n x_i$. This, incidentally, is an instance where the UMVU estimator does not attain the Cramér-Rao lower bound.

Two important contributions to the theory of minimum variance estimation were made by Barakin [4] and by Stein [27].

Barakin has given an expression for the exact smallest variance which can be achieved by an unbiased estimator, at each point $\theta = \theta_o$ of the Θ-space. His paper actually deals with the unbiased estimators that minimize the s^{th} absolute central moment, from which $s = 2$ gives the case of minimum variance. Barankin produced estimators that achieve this minimum variance at Θ_o, and showed the uniqueness of these estimators for every Θ_o, if they exist. Barankin gives general theorems from which the Cramér-Rao lower bounds, the Bhattacharyya inequalities and the sequential analogues of these may be deduced. He has proved, further, [3b, 4] that Blackwell's theorem on complete sufficient statistics applies also in the more general case of the s^{th} absolute central moment.

Stein [27] has given a theoretical approach similar to Barankin's results for $s = 2$. Stein defines the probability ratios

$$\pi(x \mid \Theta) = \frac{p(x \mid \Theta)}{p(x \mid \Theta_o)} \tag{17}$$

and assumes them to be finite for almost all x and all Θ. He defines a quantity

$$A(\Theta_1, \Theta_2) = \int \pi(x \mid \Theta_1)\, \pi(x \mid \Theta_2)\, d\,\nu(x) \tag{18}$$

where $\nu(c)$ is the measure

$$\int_c p(x \mid \Theta_o)\, dx \tag{19}$$

and c is a subset of R, a space of points x. It is assumed that $\pi(x \mid \Theta) \in L^2$, that is, $\pi(x \mid \Theta)$ is quadratic summable in the Lebesgue sense. For if the inequality

$$A(\Theta, \Theta) < \infty$$

for all θ, does not hold, it may happen that there exists no unbiased estimate with minimum variance (except, of course, for the trivial case of the estimator equal to a constant) even though there exist unbiased estimates. Stein gives an example when this happens. Consider the case where θ may take only the values 0 or 1. Let the function to be estimated be $g(\theta) = \theta$, and

$$p(x;\ \theta = 0) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x;\ \theta = 1) = \begin{cases} \dfrac{1}{2\sqrt{x}} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

It is clear that there exist unbiased estimates of θ with arbitrarily small variance at $\theta = 0$ but there is no estimator with zero variance.

The principal theorem in Stein's paper is the following:

Let $h(x) \in L^2$, and let G be the class of functions ϕ expressible as

$$\phi(\theta) = \int h(x)\, \pi(x \mid \theta)\, d\nu(x);$$

then, if $\pi(x \mid \theta)$ is finite for all θ and almost all x, and $A(\theta, \theta) < \infty$, and there exists an unbiased estimate of $g(\theta)$, then there exists an unbiased estimate $f^*(x)$ of $g(\theta)$ which minimizes

$$\int \bigl(f(x)\bigr)^2\, p(x \mid \theta_o)\, d\mu(x).$$

If f^* has finite variance then any other unbiased estimate of $g(\theta)$ with minimum variance at θ_o is essentially equal to f^*, that is, $f = f^*$ almost everywhere. A function f is an unbiased estimate of $g(\theta)$ with minimum variance at θ_o if and only if there exists a real-valued functional T on G for which

$$TA(\theta, \theta_1) = g(\theta_1), \text{ for all } \theta_1 \text{ in the } \theta \text{ space} \tag{20}$$

and

$$T \int h(x)\, \pi(x \mid \theta)\, d\, \nu(x) = \int f(x)\, h(x)\, d\, \nu(x) \text{ for all } h(x) \in L^2 \tag{21}$$

The minimum variance is

$$Tg(\theta) - \left\{g(\theta_o)\right\}^2 \tag{22}$$

The proof utilizes some results on weak convergence based on the theorem of choice [24, p. 64]. Stein regards of considerable interest to obtain a characterization of all possible functionals T in terms of the usual operations such as integration and differentiation.

The main disadvantage of Barankin's and Stein's results is that they obtain an unbiased estimator that has minimum variance at a specific value θ_o of θ. This may be a very poor estimator, for the variance at the true θ of the estimator with variance minimum at θ_o may be unduly large. Thus, unless a reasonable knowledge of the location of the true θ is available, the construction of Barankin and Stein's general results are quite useless. Perhaps in sequential estimation or in instances where the θ_o is taken in a small neighborhood of the true θ, the above results may be useful. This last, however, presupposes a knowledge of θ that is not common in practical applications.

Bahadur [3a] has recently given most of the previous results by Rao and Stein in terms of a different mathematical framework. For example, he uses "subfields" instead of estimators or statistics.

Bahadur considers the class T of all estimators that are the UMV estimators of their respective expected values in a given statistical framework. Let T_b be the class of bounded estimators in T. Bahadur's conclusions are as follows:

i) There exists a statistic such that T_b is the class of all bounded functions of this statistic, which means that there exists a statistic such that every bounded function of this statistic is the UMV estimator for its expectation. Moreover, every real-valued function of this statistic is in T. It follows that if the statistic $t \in T_b$, and $u = f(t)$, then $u \in T$.

ii) T contains an unbiased estimate of every estimable parameter, that is to say, T contains the UMVU estimate of every parameter that can be estimated unbiasedly, if and only if the framework admits a complete sufficient statistic.

Conclusion ii) may be stated as follows: Suppose that in the given framework the maximum possible reduction of the sample space by means of a sufficient statistic has already been carried out. Then either each estimable parameter has a unique unbiased estimator (which therefore would be minimum variance) or there exist estimable parameters that do not admit unbiased estimates of uniformly minimum variance. In other words, if a sufficient statistic exists and an unbiased estimate exists for $g(\theta)$, then the UMVU estimator of $g(\theta)$ must be a function of the sufficient statistic, if it exists. (The UMVU estimator of $g(\theta)$ will always exist if we have a complete sufficient statistic.) This conclusion had been reached by Rao [21] in 1945. Furthermore, it does not say whether or not a UMVU exists for θ, when there are no sufficient statistics. Thus, Bahadur does not go beyond Rao's results. Bahadur [3a] states that the main object of his paper is to show that the techniques of complete sufficient statistics are available whenever every estimate $g(\theta)$ admits a uniformly efficient estimate. This is a conclusion that follows from Rao's principal theorem [22] of 1952.

An important result of Bahadur's paper is the conclusion, which he proves by an example, that a real valued function of a uniformly efficient estimate is not necessarily uniformly efficient, that is to say, a real function of a UMVU estimator is not necessarily the UMVU of its expectation. The importance of this result lies in the fact that, if it could be possible to show that every function of a UMVU is UMVU for its expectation, then by Rao's principal theorem [22] of 1952, it would follow that UMVU estimators are necessarily found among the set of sufficient statistics, which is the main thesis in Rao's writings on this topic since his 1945 paper. Bahadur, with his example, shows that this approach cannot prove Rao's conjecture that the class of UMVU estimators is in the class of sufficient statistics for any $g(\theta)$. In fact, one of the most important results of the present study is the direct proof, by constructing an example, that there are UMVU estimators of a function $g(\theta) = \mu(\theta)$ in a case where the density does not admit a sufficient statistic; this example is given in Section C of the next chapter.

Schmetterer [26a] has given quite recently a splendid generalization of Bahadur's, Barankin's and Stein's results to cover convex lossfunctions, whereby he takes care of most pathological cases not previously covered by the California school. Schmetterer's results have some relation to results previously obtained by Barankin [4].

III. MINIMUM VARIANCE ESTIMATORS OBTAINED BY CALCULUS OF VARIATIONS

A. Definitions

The following definitions shall prove useful in the course of this chapter.

Definition A.1

Let G be a set of finite or infinite Lebesgue measure. The intersection of G and $[a, b]$ is denoted by $G \cdot [a, b]$, where a and b are real numbers such that $a < b$.

Let f(x) be a measurable (not necessarily bounded) function defined on G (not necessarily of finite measure). If f(x) has a Lebesgue integral over $G \cdot [a, b]$ for every pair of real numbers a, b and if

$$\lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{G \cdot [a, b]} |f(x)| \, dx$$

exists, that is, it has a finite real value, f(x) is said to be summable over G. The value of this limit is denoted by

$$\int_G f(x) \, dx$$

and is called the Lebesgue integral of f(x) [30, pp. 173-174].

Definition A.2

If a function (bounded or unbounded) defined on G (of finite or infinite measure) has a Lebesgue integral, then f(x) is said to be summable over G [30, p. 174].

These two definitions apply if we replace the Lebesgue integral concept with that of a Stieltjes-Lebesgue integral, by defining a more general measure (a probability measure) instead of the Lebesgue measure.

Definition A.3

Let $\alpha(\theta, x)$ be a non-decreasing function in $[a, b]$ which is not constant. If $a = -\infty$ or $b = +\infty$, it is required that

$$\alpha(-\infty) = \lim_{x \to -\infty} \alpha(\theta, x)$$

and

$$\alpha(+\infty) = \lim_{x \to +\infty} \alpha(\theta, x)$$

should be finite.

The class of functions f(x) which are measurable with respect to $\alpha(\theta, x)$ and for which the Stieltjes-Lebesgue integral

$$\int_a^b |f(x)|^p \, d\alpha; \quad p \geq 1$$

exists, is called the class L_α^p [28, p. 1].

Definition A.4

A system of functions is called closed when no function exists that is orthonormal to all the functions in the system [10, p. 94].

B. The Calculus of Variations Approach

Consider the one-parameter distribution function whose probability density function is given by

$$\phi(x_i;\ \theta)$$

The joint density of a random sample $x_1, x_2, \ldots, x_n$ will be represented by

$$\phi(x_1, x_2, \ldots, x_n;\ \theta) = \phi(x;\ \theta) \tag{23}$$

Let Ω be a proper non-empty subclass of the class L^2_α as defined in Definition A.3, where a, b may be $-\infty$, $+\infty$ respectively. The measurable function $\alpha(\theta, x)$ considered will be a general cummulative distribution function such that

$$d\,\alpha(x;\ \theta) = \phi(x;\ \theta)\ dx$$

where $\phi(x;\ \theta)$ is a probability density function as given in (23). The elements of the class Ω are the estimators of θ

$$t(x_1, x_2, \ldots, x_n) = t(x) \tag{24}$$

independent of θ, such that the expectation of every $t(x) \in \Omega$ is $\mu(\theta)$, identically in θ. That is to say,

$$\int_{R(x)} t(x)\ \phi(x;\ \theta)\ dx = \mu(\theta) \tag{25}$$

identically in θ, where $R(x)$ indicates the range of x. The class Ω, being a proper subclass of L^2_α, contains only estimators with finite variance.

Consider the following problem:

It is required to find that estimator $t^*(x) \in \Omega$, if it exists, such that

$$\int_{R(x)} (t(x) - \theta)^2 \phi(x; \theta) \, dx \qquad (26)$$

is a minimum for $t(x) = t^*(x)$. The expression (26) is called the Mean Square Error (MSE) of $t(x)$; in the particular case when $\mu(\theta) = \theta$, $t^*(x)$ becomes the minimum variance unbiased estimator of θ. The set of estimators $t(x)$ in the class Ω is called the set of 'competitor' estimators of θ, with prescribed bias $\mu(\theta) - \theta$.

The problem as stated may be treated as an isoperimetric problem in the classical Calculus of Variations. This problem, however, requires a more careful specification since various possibilities arise:

We may require to minimize (26) for

1) one particular value of θ, say θ_o

2) a 'dense' finite set of θ-values: $\theta = \theta_i$ $(i = 1, 2, \ldots)$

3) all values of θ in the parameter space.

For each of these cases we may specify several isoperimetric side-conditions. We may consider that the 'competitor' functions $t(x)$ have a prescribed expectation given by

a) $\mu(\theta) = \mu(\theta_o)$, for a particular value $\theta = \theta_o$

b) $\mu(\theta) = \mu(\theta_i)$, for a 'dense' finite set of θ-values θ_i

c) $\mu(\theta)$ for all θ values in the parameter space.

We thus envisage nine possibilities, each one consisting of the combination of one of the cases of MSE minimum with one of the iso-perimetric conditions, as set out in Table 1 below.

Table 1. Estimators $t(x)$ of θ with minimum Mean Square Error (MSE) and prescribed expectation function

	Minimize MSE = $\int (t(x) - \theta)^2 \phi(x; \theta)\, dx$ for:		
	1	2	3
Competitor functions $t(x)$ have prescribed expectation $\mu(\tau)$ for:	One particular value $\theta = \theta_o$	A 'dense' or finite set of θ-values $\theta = \theta_i$	All values of θ in the parameter space
a. One particular value $\tau = \tau_o = \theta_o$	Problem (a, 1)	Problem (a, 2)	Problem (a, 3)
b. A 'dense' or finite set of τ-values $\tau = \tau_i = \theta_i$	Problem (b, 1)	Problem (b, 2)	Problem (b, 3)
c. All values of τ in parameter space	Problem (c, 1)	Problem (c, 2)	Problem (c, 3)

It is readily found that the Solution to Problems (a, 1), (a, 2), (a, 3) and (b, 3), obtained by the Calculus of Variations, depends on the parameter θ. The Solution to Problem (a, 1), for example, is

$$t^*(x) = \theta_o \tag{27}$$

which is quite useless from the standpoint of Statistics. The above four problems where the Solution $t^*(x)$ always depends on θ shall not be considered any further. The remaining five problems in Table 1 require further study. Let us consider a finite range of θ values, if necessary by means of a reparametrization of the original range of θ. Consider a Cartesian k-dimensional grid, for the case of k parameters, of equidistant θ values, the common distance being denoted by Δ. Let σ be the variable that covers the range of θ. The set of isoperimetric conditions is now given by

$$\int_{R(x)} t(x)\ \phi(x;\ \sigma_j)\ dx = \mu(\sigma);\ j = 1,\ 2,\ \ldots,\ T. \tag{28}$$

From the classical Calculus of Variations it is known [7, p. 206; 10, p. 204] that, if there exists a solution to the proposed problem, it should come from solving for $t_i(x)$ the equation

$$\lambda(\theta_i)\ t_i(x)\ \phi(x;\ \theta_i) = \sum_j \lambda(\sigma_j,\ \theta_i)\ \phi(x;\ \sigma_j) \tag{29}$$

identically in x. If $\phi(x;\ \theta_i) = 0$ for a set of x's of non-zero measure on which not all $\phi(x;\ \sigma_j)$ are zero, it is implied that the set of $\phi(x;\ \sigma_j)$ must be linearly dependent for the solution to exist. The solutions $t_i(x)$ are called 'extremals', as suggested by Kneser [7, p. 27]. Equation (29) is a necessary and sufficient condition for $t_i(x)$ to solve Problem (b, 2) in Table 1. It is necessary since (29) is the Euler equation arising from the isoperimetric problem we are now considering. To prove sufficiency, let $f_i(x)$ be any other competitor, that is, an estimator of θ with the same bias function as $t_i(x)$. We may write

$$f_i(x) = t_i(x) + e_i(x) \tag{30}$$

where, clearly, $E(e_i(x)\ ) = 0$. Consider now

$$\lambda(\theta_i) \int_{R(x)} f_i^2(x)\, \phi(x;\, \theta_i)\, dx = \lambda(\theta_i) \int_{R(x)} t_i^2(x)\, \phi(x;\, \theta_i)\, dx$$

$$+ \lambda(\theta_i) \int_{R(x)} e_i^2(x)\, \phi(x;\, \theta_i)\, dx$$

$$+ 2\, \lambda(\theta_i) \int_{R(x)} e_i(x)\, t_i(x)\, \phi(x;\, \theta_i)\, dx \qquad (31)$$

In the last integral of (31) we may write

$$\int_{R(x)} e_i(x)\, \lambda(\theta_i)\, t_i(x)\, \phi(x;\, \theta_i)\, dx$$

$$= \int_{R(x)} e_i(x) \sum_j \lambda(\theta_i,\, \sigma_j)\, \phi(x;\, \sigma_j)\, dx$$

since $\lambda(\theta_i)$ and $\lambda(\theta_i, \sigma_j)$ are independent of x. Hence

$$= \sum_j \lambda(\theta_i,\, \sigma_j) \int_{R(x)} e_i(x)\, \phi(x;\, \sigma_j)\, dx = 0 \qquad (32)$$

because $E(e_i(x)) = 0$. We may write, therefore, from (32) and (31),

$$\int_{R(x)} f_i^2(x)\, \phi(x;\, \theta_i)\, dx \geq \int_{R(x)} t_i^2(x)\, \phi(x;\, \theta_i)\, dx$$

Q.E.D.

Let us now assume that the grid distance Δ is defined by

$$\Delta_j = 2^{-j};\; j = 1, 2, \ldots, T \qquad (33)$$

Let $t_j(x) \epsilon \Omega \subset L_\alpha^2$ be the extremal from (29) that corresponds, for a particular θ-value, to the grid with distance $\Delta_j = 2^{-j}$.

Now, since $t_j(x) \epsilon \bar{\Omega}$, the sequence $\{t_j(x)\}$ is bounded. For otherwise there would be some $t_j(x)$ not in L_α^2. By the theorem of choice

[24, p. 64], there exists a subsequence of the $\{t_j(x)\}$, say $\{t_{mj}(x)\}$, that is weakly convergent, that is to say, there exists a $t^*(x)$ such that

$$\lim_{j\to\infty} \int_{R(x)} t_{mj}(x)\, g(x)\, d\,\alpha = \int_{R(x)} t^*(x)\, g\,(x)\, d\alpha \qquad (34)$$

for every $g(x) \in L^2_\alpha$. But since $g(x) = 1 \in L^2_\alpha$, it follows that $t^*(x) \in \Omega$, that is,

$$E(t^*(x)\,) = \mu(\theta)$$

and

$$\int_{R(x)} t^{*2}(x)\, \phi(x;\, \theta)\, dx$$

exists.

Further, in the case of weak convergence [24, p. 60],

$$\int_{R(x)} t^{*2}(x)\, \phi(x;\, \theta)\, dx \leq \liminf \int_{R(x)} t^2_{mj}(x)\, \phi(x;\, \theta)\, dx \qquad (35)$$

which is equivalent to stating that $t^*(x)$ has minimum variance.

Now $t^*(x)$ is unique (almost everywhere), for let $f\,(x)$ be another estimator such that its expectation is $\mu(\theta)$ and

$$\text{Var } f(x) \;=\; \text{Var } t^*(x) \qquad (36)$$

It will suffice to show that the correlation between $f(x)$ and $t^*(x)$ is unity. Consider

$$\text{Var}\left\{ \tfrac{1}{2}\left(t^*(x) + f(x)\right)\right\}$$
$$= \tfrac{1}{4}\left\{ \text{Var } t^*(x) + \text{Var } f(x) + 2\,\text{Cov}\left(t^*(x),\, f(x)\right)\right\} \qquad (37)$$

From (36) it follows that

$$\text{Var}\left\{\frac{t^*(x) + f(x)}{2}\right\} = \frac{1}{2}\left\{\text{Var } t^*(x) + \text{Cov } \left(t^*(x), f(x)\right)\right\} \quad (38)$$

Consider now

$$\text{Cov } \left\{\left(\frac{t^*(x) + f(x)}{2}\right), t^*(x)\right\}$$

$$= E\left\{\left(\frac{t^*(x) + f(x)}{2}\right) \cdot t^*(x)\right\} - E\left\{\frac{t^*(x) + f(x)}{2}\right\} \cdot E\, t^*(x)$$

$$= \frac{1}{2}\left\{ E\left(t^{*2}(x)\right) + E\left(f(x)\, t^*(x)\right)\right\} - \mu^2(\theta)$$

$$= \frac{1}{2}\left\{\left[E\left(t^{*2}(x)\right) - \mu^2(\theta)\right] + \left[E\left(f(x)\, t^*(x)\right) - \mu^2(\theta)\right]\right\}$$

$$= \frac{1}{2}\left\{\text{Var } t^*(x) + \text{Cov } \left(t^*(x), f(x)\right)\right\}$$

and by (38),

$$= \text{Var } \left(\frac{t^*(x) + f(x)}{2}\right)$$

$$\therefore \text{Cov } \left\{\left(\frac{t^*(x) + f(x)}{2}\right), t^*(x)\right\} = \text{Var } \left(\frac{t^*(x) + f(x)}{2}\right) \quad (39)$$

Let ρ^* be the correlation between $t^*(x)$ and $\left(\frac{f(x) + t^*(x)}{2}\right)$. By definition, then,

$$(\rho^*)^2 = \frac{\text{Cov}^2 \left\{\left(\frac{t^* + f(x)}{2}\right), t^*(x)\right\}}{\text{Var } \left(\frac{t^*(x) + f(x)}{2}\right) \cdot \text{Var } t^*(x)}$$

which, from (39), reduces to

$$(\rho^*)^2 = \frac{\text{Var } \left(\frac{t^*(x) + f(x)}{2}\right)}{\text{Var } t^*(x)}$$

But $0 < (\rho^*)^2 \leq 1$. Hence,

$$\text{Var}\left(\frac{t^*(x) + f(x)}{2}\right) \leq \text{Var } t^*(x) = \sigma^2_{\min} \tag{40}$$

This is a contradiction, if $(\rho^*)^2 < 1$, for then the estimator

$$\frac{1}{2}\left(t^*(x) + f(x)\right)$$

unbiased for $\mu(\theta)$, would have a variance less than $\sigma^2_{\min}$. Hence, it must be that $(\rho^*) = 1$ and hence

$$\text{Var}\left(\frac{t^*(x) + f(x)}{2}\right) = \text{Var } t^*(x) \tag{41}$$

Now, from (38) we may write (41) as

$$\frac{1}{2}\left\{\text{Var } t^*(x) + \text{Cov}\left(t^*(x), f(x)\right)\right\} = \text{Var } t^*(x)$$

$$\therefore \text{Cov}\left(t^*(x), f(x)\right) = \text{Var } t^*(x) \tag{42}$$

which implies that the correlation between $t^*(x)$ and $f(x)$ is unity. Thus we conclude that $t^*(x)$ is unique (almost everywhere).

From these results we may write (29), after taking the limit, as

$$\lambda_o(\theta)\ t\ (x)\ \phi(x;\ \theta) = \lim_{j \to \infty} \sum_j \lambda(\tau_j,\ \theta)\ \phi(x;\ \tau_j) \tag{43}$$

The limit in the right-hand side of (43) exists, since it is equal to the limit represented by the left-hand side, which was shown to exist in (34). The limit expression in (43) may be considered as defining a linear operator on the $\phi(x;\ \tau)$; $\phi(x;\ \tau) > 0$ almost everywhere. We write, therefore, as a general limiting form of (29),

$$\lambda_o(\theta)\ t^*(x)\ \phi(x;\ \theta) = O_\tau(\theta)\ \phi(x;\ \tau) \tag{44}$$

where $O_\tau(\theta)$ stands for the linear operator on the $\phi(x;\ \tau)$.

The operator $O_\tau(\theta)$ is, therefore, defined in the set of $\phi(x;\ \tau)$ functions depending on θ, identically in θ. This means that $O_\tau(\theta)$ commutes with $\int_{R(x)} dx$.

We prove next that (44) is a sufficient condition for $t^*(x)$ to be a solution to Problem (c, 3).

Theorem. (Sufficient condition.)

If $t^*(x) \in \Omega$ is an estimator that satisfies the Euler-type equation

$$\lambda_o(\theta)\ t^*(x)\ \phi(x;\ \theta) = O_\tau(\theta)\ \phi(x;\ \tau)$$

identically in x, then $t^*(x)$ is the unique minimum variance estimator of θ with bias $\mu(\theta) - \theta$.

Proof: Let $t(x)$ and $t^*(x)$ both be elements of the class , and let

$$t(x) = t^*(x) + e(x)$$

where $e(x) \in L^2_\alpha$. Since $t(x) \in \Omega$,

$$\int_{R(x)} t(x)\ \phi(x;\ \theta)\ dx = \mu(\theta)$$

Hence,

$$\int_{R(x)} e(x)\ \phi(x;\ \theta)\ dx = 0 \quad \text{identically in } \theta.$$

Also,

$$\lambda_o(\theta) \int_{R(x)} t^2(x)\, \phi(x;\, \theta)\, dx = \lambda_o(\theta) \int_{R(x)} t^{*2}(x)\, \phi(x;\, \theta)\, dx$$

$$+ \lambda_o(\theta) \int_{R(x)} e^2(x)\, \phi(x;\, \theta)\, dx$$

$$+ 2\, \lambda_o(\theta) \int_{R(x)} e(x)\, t^*(x)\, \phi(x;\, \theta)\, dx \qquad (45)$$

The last integral in the right-hand side of (45) may be written, because of (44), as

$$2 \int_{R(x)} e(x)\, O_{\tilde{\theta}}(\theta)\, \phi(x;\, \tilde{\theta})\, dx = 2\, O_{\tilde{\theta}}\, (\theta) \int_{R(x)} e(x)\, \phi(x;\, \tilde{\theta})\, dx = 0$$

Hence, we conclude that

$$\int_{R(x)} t^2(x)\, \phi(x;\, \theta)\, dx \geq \int_{R(x)} t^{*2}(x)\, \phi(x;\, \theta)\, dx \qquad (46)$$

and the equality holds only when $t(x)$ is essentially (i.e., except in a set of measure zero) equal to $t^*(x)$.

C. Special Cases of the Operator $O_{\tilde{\theta}}(\theta)$

Certain limiting forms of (29) provide results which have been previously derived in the literature.

1. The linear differential operator

Let the range of α be independent of θ. Consider the linear differential equation

$$t^*(x)\ \phi(x;\ \theta) = \lambda(\theta)\ \phi_\theta(x;\ \theta) + \omega(\theta)\ \phi(x;\ \theta) \tag{47}$$

where ϕ_θ denotes the partial derivative of $\phi(x;\ \theta)$ with respect to θ. This equation is a special case of (44). Aitken and Silverstone [2] claimed incorrectly that (47) was a necessary condition for the solution $t^*(x)$ of Problem (c, 3), whereas in fact it is only a special case of (44).

By integrating (47) with respect to θ it is easily found that the solution $t^*(x)$ in (47) is amenable to a density function of the form

$$\phi(x;\ \theta) = e^{P(\theta)+t^*(x)\ Q(\theta)+R(x)} \tag{48}$$

which is the Koopman [17] form of $\phi(x;\ \theta)$ and represents the special case of a distribution admitting sufficient statistics. It is readily shown that $t^*(x)$ is a minimum variance unbiased estimator of

$$\mu(\theta) = -\frac{\dfrac{\partial P(\theta)}{\partial \theta}}{\dfrac{\partial Q(\theta)}{\partial \theta}} = -\frac{P'(\theta)}{Q'(\theta)}\ ;\ Q'(\theta) \neq 0 \tag{49}$$

It can be shown [2], as seen in Chapter II, that

$$\text{Var } t^*(x) = \sigma^2_{\min} = \lambda(\theta) \tag{50}$$

As an example of the case in which (47) arises we may consider the density

$$\phi(x_i\ \theta) = \theta^{-n}\ e^{-n\bar{x}/\theta}\ ;\ \theta > 0 \tag{51}$$

which is the joint density of a gamma distribution with parameters $1/\theta$

and 1. By comparison with (48) we find that

$$t^*(x) = \bar{x}; \quad P(\theta) = -\,n \log \theta; \quad Q = -\frac{n}{\theta}$$

and $\mu(\theta) = E(t^*(x)\,) = \omega(\theta) = \theta$, which is obtained by integrating over x in (47),

$$E\ \ t^*(x) \quad = \lambda(\theta) \int_{R(x)} \phi_\theta(x;\,\theta)\,dx + \omega(\theta) \int_{R(x)} \phi(x;\,\theta)\,dx = \omega(\theta)$$

because

$$\int_{R(x)} \phi(x;\,\theta)\,dx = 1$$

and hence $\dfrac{\partial}{\partial\theta} \displaystyle\int_{R(x)} \phi(x;\,\theta)\,dx = \int_{R(x)} \phi_\theta(x;\,\theta)\,dx = 0$

and, by (49)

$$\mu(\theta) \quad = \quad -\frac{P'(\theta)}{Q'(\theta)} \quad = \quad \theta\,.$$

Also $\qquad \operatorname{Var} t^*(x) \;=\; \lambda(\theta) \;=\; \dfrac{\theta^2}{n}\,.$

2. <u>The integral equation for $\phi(x;\,\theta)$</u>

Consider the special case of (44) given by the integral equation

$$t^*(x)\ \phi(x;\,\theta) \quad = \int_{R(\tau)} \lambda(\theta,\,\tau)\ \phi(x;\,\tau)\ d\,\tau \tag{52}$$

with kernel $\lambda(\theta,\,\tau)$. If we assume that $\lambda(\theta,\,\tau)$ is of the type

$$\lambda(\theta,\,\tau) \;=\; \lambda_1(\theta)\ \ \lambda_2(\tau) \tag{53}$$

then (52) expresses a functional onto $\phi(x;\, \tau)$, which, according to a theorem by F. Riesz [24, p. 61; 28, p. 12] represents a linear operation on $\phi(x;\, \tau)$.

Consider the range of x independent of θ. The expected value of $t^*(x)$ is given by

$$\begin{aligned} E\ t^*(x) &= \int_{R(x)} \int_{R(\tau)} \lambda(\theta,\, \tau)\, \phi(x;\, \tau)\, d\tau\, dx \\ &= \int_{R(\tau)} \lambda(\theta,\, \tau)\, d\tau \int_{R(x)} \phi(x;\, \tau)\, dx \\ &= \int_{R(\tau)} \lambda(\theta,\, \tau)\, d\tau = \mu(\theta) \end{aligned} \tag{54}$$

The above integration is justified by Fubini's theorem [24, p. 85]. Let us examine some examples where (52) applies.

Consider the density of a gamma distribution with parameters θ and p, given by

$$\phi(x_i;\, \theta,\, p) = \frac{\theta^p}{\Gamma(p)}\, e^{-\theta x_i}\, x_i^{\,p-1} \tag{55}$$

where $0 \leq x_i < \infty$; $p \geq 1$.

Assume p is a known integer. The joint density of a random sample $x_1, \ldots, x_n$ from this population is given by

$$\phi(x;\, \theta) = \phi_1(x,\, p)\, \theta^{np}\, e^{-\theta n\bar{x}} \tag{56}$$

where $\phi_1(x,\, p)$ does not involve θ.

Let now

$$\lambda(\theta, \tau) = \theta^{np} / \tau^{np} ; \quad \tau \geq \theta \qquad (57)$$
$$= 0 \qquad \tau < \theta$$

By (52) we have

$$t^*(x)\ \phi(x;\ \theta) = \phi_1(x;\ p)\ \theta^{np} \int_{\theta}^{\infty} e^{-\tau n\bar{x}}\, d\tau$$

$$= \phi_1(x;\ p)\ \frac{\theta^{np}}{n\bar{x}} \left[- e^{-\tau n\bar{x}} \right]_{\theta}^{\infty}$$

$$= \frac{1}{n\bar{x}}\, \phi(x;\ \theta)$$

$$\therefore\ t^*(x) = \frac{1}{n\bar{x}} \qquad (58)$$

and the expectation of $t^*(x)$ is given, by (54), as

$$\mu(\theta) = \int_{R(\tau)} \lambda(\theta, \tau)\, d\tau = \theta^{np} \int_{\theta}^{\infty} d\tau / \tau^{np}$$

$$= \frac{\theta}{np - 1} \qquad (59)$$

Hence, we have that

$$\frac{np - 1}{n\bar{x}} \qquad (60)$$

is the minimum variance unbiased estimator of θ for the gamma defined in (55).

Now it is known [20, p. 136] that a monotone function of a sufficient statistic is also sufficient for its expectation. Here we have clearly $\bar{x}$ sufficient, since $\phi(x;\ \theta)$, from (56), is of Koopman form and $\bar{x}$ is the coefficient of $Q(\theta) = n\theta$, in the notation of (48). Also,

(60) defines a monotone function of $\bar{x}$, and hence the estimator (60) is sufficient for θ. Therefore, (60) is an estimator of θ that has the minimum variance.

This example shows that some operators, other than the one (differential) found by Aitken and Silverstone, exist that will yield minimum variance unbiased estimators. It is clear, therefore, that (47) does not constitute a necessary condition for $t^*(x)$ to be the minimum variance estimator of its expectation.

As a further illustration consider the normal density

$$\phi(x;\ \sigma^2) = (\sqrt{2\pi})^{-n} \quad \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \qquad (61)$$

where μ is known and $\sigma > 0$.

Let the kernel $\lambda(\theta, \tau) = \lambda(\sigma, \tau)$ be defined by

$$\begin{aligned} \lambda(\sigma, \tau) &= (n-2)\,\tau^{n-3}/\sigma^n; \quad 0 \le \tau < \sigma \\ &= 0 \qquad\qquad\qquad\quad\ \tau > \sigma \end{aligned} \qquad (62)$$

In this case (52) becomes

$$t^*(x)\ \phi(x;\sigma) = (\sqrt{2\pi})^{-n}(n-2)\,\sigma^{-n}\int_0^{\sigma}\exp\left\{-\sum^{n}(x_i-\mu)^2/2\,\tau^2\right\}\tau^{-3}\,d\tau$$

$$= (\sqrt{2\pi})^{-n}\,\frac{(n-2)\,\sigma^{-n}}{\sum^{n}(x_i-\mu)^2}\left[\exp\left\{-\sum^{n}(x_i-\mu)^2/2\,\tau^2\right\}\right]_0^{\sigma}$$

$$= \frac{n-2}{\sum^{n}(x_i-\mu)^2}\cdot\phi(x;\sigma)$$

$$\therefore\ t^*(x) = \frac{n-2}{\sum^{n}(x_i-\mu)^2} \qquad (63)$$

and the expectation of $t^*(x)$ is given, from (54), by

$$\mu(\sigma) = \frac{n-2}{\sigma^n} \int_0^{\sigma} \tau^{n-3} \, d\tau = \frac{1}{\sigma^2} \tag{64}$$

Hence, the estimator given in (63) is the minimum variance unbiased estimator of $1/\sigma^2$ in the normal (μ, σ^2) with μ known. This result implies, for example, that

$$\frac{\nu - 2}{\chi^2_{\nu}} \tag{65}$$

is the minimum variance unbiased estimator of $1/\sigma^2$ if χ^2_{ν} is based on ν degrees of freedom.

To obtain the variance of $t^*(x)$ in this case, we write

$$\text{Var } t^*(x) = \int_{R(x)} t^{*2}(x) \, \phi(x; \theta) \, dx - \mu^2(\theta)$$

Now,

$$\int_{R(x)} t^{*2}(x) \, \phi(x; \theta) \, dx = \int_{R(x)} t^*(x) \cdot t^*(x) \, \phi(x; \theta) \, dx$$

and by the integral equation (52)

$$= \int_{R(x)} t^*(x) \int_{R(\tau)} \lambda(\theta, \tau) \, \phi(x; \tau) \, d\tau \, dx$$

$$= \int_{R(\tau)} \lambda(\theta, \tau) \left(\int_{R(x)} t^*(x) \, \phi(x; \tau) \, dx \right) d\tau$$

$$= \int_{R(\tau)} \lambda(\theta, \tau) \, \mu(\tau) \, d\tau$$

Hence,

$$\text{Var } t^*(x) = \int_{R(\bar{\tau})} \lambda(\theta, \tau)\, \mu(\tau)\, d\,\tau - \mu^2(\theta) \qquad (66)$$

Thus, for the example of the gamma density (55), with the kernel $\lambda(\theta, \tau)$ defined by (57), we find

$$\text{Var } \frac{1}{n\,\bar{x}} = \theta^{np} \int_{\theta}^{\infty} \frac{d\,\tau}{\tau^{np-1}} - \frac{\theta^2}{(np - 1)^2} = \frac{\theta^2}{(np - 1)^2\,(np - 2)} \qquad (67)$$

Hence,

$$\text{Var}\left(\frac{np - 1}{n\,\bar{x}}\right) = \frac{\theta^2}{np - 2} \qquad (68)$$

For the example of the normal density given by (61) with the kernel $\lambda(\theta, \tau)$ defined by (62) we have

$$\text{Var } \left(\frac{n - 2}{\overset{n}{\Sigma}(x_i - \mu)^2}\right) = \frac{n - 2}{\sigma^n} \int_0^{\sigma} \tau^{n-3} \cdot \frac{1}{\tau^2}\, d\,\tau - \frac{1}{\sigma^4}$$

$$= \frac{n - 2}{n - 4} \cdot \frac{1}{\sigma^4} - \frac{1}{\sigma^4}$$

$$= \left(\frac{2}{n - 4}\right) \cdot \frac{1}{\sigma^4} \qquad (69)$$

3. <u>A second order linear differential operator</u>

So far we have considered examples where the density function was of Koopman form, that is to say, a density that admits a sufficient statistic. The minimum variance estimators obtained were, as could be expected, functions of the corresponding sufficient statistic in each case.

The question remains of whether it is possible to find a uniformly minimum variance estimator of a parameter Θ in a distribution that does not admit a sufficient statistic. Rao [20, p. 150] has stated the conjecture that minimum variance estimators must be found among the set of sufficient statistics. The statement seems to imply that whenever sufficient statistics do not exist, there are no minimum variance estimators. The following example, utilizing a special case of the operator $O_{\sigma}(\Theta)$ in (44) disproves this conjecture.

Let

$$\phi(x;\, \Theta) = \rho(\Theta)\, h(x)\, e^{-\Theta f(x)} + \omega(\Theta)\, g(x)\, e^{\Theta f(x)} \tag{70}$$

be a density function, where the range of the vector $x = (x_1, \ldots, x_n)$ is independent of Θ, and $h(x) \neq k\, g(x)$, for any value of the proportionality constant $k \cdot (k \neq 0)$.

Since a necessary and sufficient condition for a density function to admit sufficient statistics is that it factors as

$$\phi(x;\, \Theta) = \phi_1\left(t(x),\, \Theta\right) \cdot \phi_2(x)$$

where ϕ_1 is the density of $t(x)$ and $\phi_2(x)$ is independent of Θ [20, p. 135], it follows that $\phi(x;\, \Theta)$ as given in (70) does not admit a sufficient statistic.

Consider the second order linear differential operator $O_{\sigma}(\Theta)$ such that

$$\begin{aligned} t^*(x)\, \phi(x;\, \Theta) = \psi(\Theta)\, \phi(x;\, \Theta) &+ \lambda(\Theta)\, \phi_{\Theta}(x;\, \Theta) \\ &+ \eta(\Theta)\, \phi_{\Theta\Theta}(x;\, \Theta) \end{aligned} \tag{71}$$

where

$$\phi_\theta = \frac{\partial \phi}{\partial \theta} \quad \text{and} \quad \phi_{\theta\theta} = \frac{\partial^2 \phi}{\partial \theta^2} .$$

We have

$$\phi_\theta(x;\ \theta) = \rho'(\theta)\ h(x)\ e^{-\theta f(x)} - \rho(\theta)\ h(x)\ f(x)\ e^{-\theta f(x)}$$

$$+ \ \omega'(\theta)\ g(x)\ e^{\theta f(x)} + \ \omega(\theta)\ g(x)\ f(x)\ e^{\theta f(x)}$$

$$= \rho'(\theta)\ h(x)\ e^{-\theta f(x)} + \ \omega'(\theta)\ g(x)\ e^{\theta f(x)}$$

$$+ f(x)\ \left[- \rho(\theta)\ h(x)\ e^{-\theta f(x)} + \omega(\theta)\ g(x)\ e^{\theta f(x)}\right] \quad (72)$$

Also

$$\phi_{\theta\theta}(x;\ \theta) = \rho''(\theta)\ h(x)\ e^{-\theta f(x)} - 2\ \rho'(\theta)\ h(x)\ f(x)\ e^{-\theta f(x)}$$

$$+ \rho(\theta)\ h(x)\ f^2(x)\ e^{-\theta f(x)} + \ \omega''(\theta)\ g(x)\ e^{\theta f(x)}$$

$$+ 2\ \omega'(\theta)\ g(x)\ f(x)\ e^{\theta f(x)} + \ \omega(\theta)\ g(x)\ f^2(x)\ e^{\theta f(x)}$$

$$\therefore\ \phi_{\theta\theta}(x;\ \theta) = \rho''(\theta)\ h(x)\ e^{-\theta f(x)} + \ \omega''(\theta)\ g(x)\ e^{\theta f(x)}$$

$$+ 2\ f(x)\ \left[- \rho'(\theta)\ h(x)\ e^{-\theta f(x)} + \omega'(\theta)\ g(x)\ e^{\theta f(x)}\right]$$

$$+ f^2(x)\left[\rho(\theta)\ h(x)\ e^{-\theta f(x)} + \ \omega(\theta)\ g(x)\ e^{\theta f(x)}\right] \quad (73)$$

where, in the last brackets we recognize $\phi(x;\ \theta)$ given in (70).

Let $\lambda(\theta)$ and $\eta(\theta)$ be so chosen as to make

$$\begin{aligned} -\ \lambda(\theta)\,\rho(\theta) - 2\ \eta(\theta)\,\rho'(\theta) &\equiv 0 \\ \lambda(\theta)\,\omega(\theta) + 2\ \eta(\theta)\,\omega'(\theta) &\equiv 0 \end{aligned} \tag{74}$$

From (74) we have, clearly,

$$\omega(\theta) = \gamma\rho(\theta) \tag{75}$$

where $\gamma > 0$ is a constant.

Also,

$$-\frac{1}{2}\,\frac{\lambda(\theta)}{\eta(\theta)} = \frac{\rho'(\theta)}{\rho(\theta)} \tag{76}$$

Let $\not\!\beta(\theta)$ be chosen so as to have

$$\begin{aligned} \not\!\beta(\theta)\,\rho(\theta) + \lambda(\theta)\,\rho'(\theta) + \eta(\theta)\,\rho''(\theta) &\equiv 0 \\ \not\!\beta(\theta)\,\omega(\theta) + \lambda(\theta)\,\omega'(\theta) + \eta(\theta)\,\omega''(\theta) &\equiv 0 \end{aligned} \tag{77}$$

which is also satisfied by (75). Equations (77) may be taken as the definition of $\not\!\beta(\theta)$.

From (74) and (77) we may write the density $\phi(x;\,\theta)$ under the linear operator, as defined in (71), after cancellations, as

$$\begin{aligned} f^2(x)\,\phi(x;\,\theta) = \not\!\beta(\theta)\,\phi(x;\,\theta) &+ \lambda(\theta)\,\phi_\theta(x;\,\theta) \\ &+ \eta(\theta)\,\phi_{\theta\theta}(x;\,\theta) \end{aligned} \tag{78}$$

which indicates that

$$t^*(x) = f^2(x)\ . \tag{79}$$

The expectation of $t^*(x)$ will be denoted by $\mu(\theta)$,

$$\mu(\theta) = \int_{R(x)} f^2(x)\ \phi(x;\ \theta)\ dx \tag{80}$$

But

$$\int_{R(x)} \phi(x;\ \theta)\ dx = 1 \quad \text{identically in } \theta.$$

Since the range of x is independent of θ, we have

$$\frac{\partial}{\partial\theta} \int_{R(x)} \phi(x;\ \theta) = \int_{R(x)} \phi_\theta(x;\ \theta)\ dx = 0$$

and hence

$$\frac{\partial}{\partial\theta} \int_{R(x)} \phi_\theta(x;\ \theta)\ dx = \int_{R(x)} \phi_{\theta\theta}(x;\ \theta)\ dx = 0$$

Therefore, substituting $f^2(x)\ \phi(x;\ \theta)$ in (80) from the expression in (78), we have

$$\mu(\theta) = \psi(\theta) \int_{R(x)} \phi(x;\ \theta)\ dx = \psi(\theta) \tag{81}$$

We may write the density given in (70), substituting (75), as

$$\phi(x;\ \theta) = \left[h(x)\ e^{-\theta f(x)} + \gamma\, g(x)\ e^{\theta f(x)} \right] \rho(\theta) \tag{82}$$

Integrating with respect to x we obtain

$$1 = \left[\int_{R(x)} h(x)\ e^{-\theta f(x)} dx + \gamma \int_{R(x)} g(x)\ e^{\theta f(x)} dx \right] \rho(\theta) \tag{83}$$

Let

$$\int_{R(x)} h(x)\ e^{-\theta f(x)} dx = E_1(\theta) > 0 \tag{84}$$

and

$$\int_{R(x)} g(x)\ e^{\theta f(x)} dx = E_2(\theta) > 0 \tag{85}$$

We may write now

$$\rho(\theta) = \frac{1}{E_1(\theta) + \gamma E_2(\theta)} > 0 \tag{86}$$

From (77) we have now

$$\phi(\theta) = -\ \lambda(\theta)\ \frac{\rho'(\theta)}{\rho(\theta)} - \ \eta(\theta)\ \frac{\rho''(\theta)}{\rho(\theta)} \tag{87}$$

and from (76),

$$-\ \lambda(\theta) = 2\ \eta(\theta)\ \frac{\rho'(\theta)}{\rho(\theta)}\ .$$

Hence,

$$\phi(\theta) = 2\ \eta(\theta)\ \left(\frac{\rho'(\theta)}{\rho(\theta)}\right)^2 - \ \eta(\theta)\ \frac{\rho''(\theta)}{\rho(\theta)}$$

$$= \frac{\eta(\theta)}{\rho^2(\theta)} \left\{2\ \left(\rho'(\theta)\right)^2 - \rho(\theta)\ \rho''(\theta)\right\} \tag{88}$$

But now, from the definition of $\rho(\theta)$ given in (86), we have

$$\rho'(\theta) = -\ \frac{E_1'(\theta) + \gamma E_2'(\theta)}{\left\{E_1(\theta) + \gamma E_2(\theta)\right\}^2} \tag{89}$$

and

$$\rho''(\theta) = \frac{2\left\{E_1'(\theta) + \gamma E_2'(\theta)\right\}^2 - \left\{E_1(\theta) + \gamma E_2(\theta)\right\}\ \left\{E_1''(\theta) + \gamma E_2''(\theta)\right\}}{\left\{E_1(\theta) + \gamma E_2(\theta)\right\}^3} \tag{90}$$

Substituting the last two results in (88) we have, after simplification,

$$\phi(\theta) = \frac{\eta(\theta)}{\rho^2(\theta)} \left(\frac{E_1''(\theta) + \gamma E_2''(\theta)}{\{E_1(\theta) + E_2(\theta)\}^3} \right) = \eta(\theta) \left(\frac{E_1''(\theta) + \gamma E_2''(\theta)}{E_1(\theta) \pm \gamma E_2(\theta)} \right) \quad (91)$$

But from (84) and (85) we have now

$$E_1'(\theta) + \gamma E_2'(\theta) = - \int_{R(x)} f(x)\, h(x)\, e^{-\theta f(x)} dx + \gamma \int_{R(x)} f(x)\, g(x)\, e^{\theta f(x)} dx \quad (92)$$

and hence

$$E_1''(\theta) + \gamma E_2''(\theta) = \int_{R(x)} f^2(x) \left[h(x)\, e^{-\theta f(x)} + \gamma g(x)\, e^{\theta f(x)} \right] dx \quad (93)$$

Therefore, in (91) we have

$$\phi(\theta) = \eta(\theta) \left\{ \rho(\theta) \int_{R(x)} f^2(x) \left[h(x)\, e^{-\theta f(x)} + \gamma g(x)\, e^{\theta f(x)} \right] dx \right\}$$

$$= \eta(\theta) \int_{R(x)} f^2(x)\, \phi(x;\, \theta)\, dx$$

which, by definition of $\mu(\theta)$, given in (80), is simply

$$= \eta(\theta)\, \mu(\theta) \quad (94)$$

However, from (81) we have that $\phi(\theta) = \mu(\theta)$. Therefore, from (94) we conclude that

$$\eta(\theta) \equiv 1 \quad (95)$$

identically in θ. Therefore, from (76) we have

$$\lambda(\theta) = -2\,\frac{\rho'(\theta)}{\rho(\theta)} \tag{96}$$

and from (88), since $\eta(\theta)$ is identically unity,

$$\mu(\theta) = \phi(\theta) = 2\left(\frac{\rho'(\theta)}{\rho(\theta)}\right)^2 - \frac{\rho''(\theta)}{\rho(\theta)} \tag{97}$$

The linear operator equation (78) may be spelled out now as

$$\begin{aligned} f^2(x)\,\phi(x;\,\theta) = {} & \left[2\left(\frac{\rho'(\theta)}{\rho(\theta)}\right)^2 - \frac{\rho''(\theta)}{\rho(\theta)}\right]\phi(x;\,\theta) \\ & - 2\,\frac{\rho'(\theta)}{\rho(\theta)}\,\phi_\theta(x;\,\theta) + \phi_{\theta\theta}(x;\,\theta) \end{aligned} \tag{98}$$

which is easily verified to hold, for let $\omega(\theta)$ be substituted by $\gamma\rho(\theta)$ in (72) and (73). We can write now in (98)

$$\begin{aligned} f^2(x)\,\phi(x;\,\theta) = {} & 2\left(\frac{\rho'(\theta)}{\rho(\theta)}\right)^2\phi(x;\,\theta) && \text{(A)} \\ & - \frac{\rho''(\theta)}{\rho(\theta)}\,\phi(x;\,\theta) && \text{(B)} \\ & - 2\,\frac{\rho'(\theta)}{\rho(\theta)}\cdot\rho'(\theta)\,f(x)\left[-h(x)\,e^{-\theta f(x)} + \gamma g(x) e^{\theta f(x)}\right] && \text{(C)} \\ & - 2\,\frac{\rho'(\theta)}{\rho(\theta)}\cdot\rho'(\theta)\left[h(x)\,e^{-\theta f(x)} + \gamma g(x)\,e^{\theta f(x)}\right] && \text{(D)} \\ & + f^2(x)\,\rho(\theta)\left[h(x)\,e^{-\theta f(x)} + \gamma g(x)\,e^{\theta f(x)}\right] && \text{(E)} \\ & + 2\,\rho'(\theta)\,f(x)\left[-h(x)\,e^{-\theta f(x)} + \gamma g(x)\,e^{\theta f(x)}\right] && \text{(F)} \\ & + \rho''(\theta)\left[h(x)\,e^{-\theta f(x)} + \gamma g(x)\,e^{\theta f(x)}\right] && \text{(G)} \end{aligned} \tag{99}$$

Now, term (C) cancels with term (F). Term (D) may be written as

$$- 2 \left(\frac{\rho'(\theta)}{\rho(\theta)}\right)^2 \phi(x;\ \theta).$$

Hence, term (D) cancels with term (A). Term (G) may be written as

$$+ \frac{\rho''(\theta)}{\rho(\theta)} \phi(x;\ \theta).$$

Hence, term (G) cancels with term (B). The remaining term, (E), is clearly $f^2(x)\ \phi(x;\ \theta)$. Thus (98) is seen to be an identity. It is worth while noticing that $\mu(\theta)$, being the expectation of $f^2(x)$, is essentially positive.

The variance of $f^2(x)$ may be written as follows:

$$\begin{aligned}
\operatorname{Var} f^2(x) &= \int_{R(x)} f^2(x) \cdot f^2(x)\ \phi(x;\ \theta)\ dx - \mu^2(\theta) \\
&= \mu^2(\theta) - 2\, \frac{\rho'(\theta)}{\rho(\theta)} \frac{\partial}{\partial\theta} \int_{R(x)} f^2(x)\ \phi(x;\ \theta)\ dx \\
&\qquad + \frac{\partial^2}{\partial\theta^2} \int_{R(x)} f^2(x)\ \phi(x;\ \theta)\ dx - \mu^2(\theta) \\
&= -\ 2\, \frac{\rho'(\theta)}{\rho(\theta)}\, \mu_\theta(\theta) + \mu_{\theta\theta}(\theta).
\end{aligned} \tag{100}$$

If the linear operator utilized for this example is denoted by O_θ, we see that the variance is given by

$$\operatorname{Var} f^2(x) = O_\theta(\mu(\theta)\ ) - \mu^2(\theta) \tag{101}$$

which is of the form given in (66).

IV. ESTIMATORS WITH MINIMUM WEIGHTED MEAN SQUARE ERROR AND PRESCRIBED BIAS FUNCTION

The theory developed so far has been concerned with the problem of obtaining estimators with minimum mean square error (MMSE) and with prescribed bias function. There is the possibility, however, of enlarging the scope of the theory with the introduction of the concept of a minimum weighted mean square error (MWMSE). We may postulate the problem, now, as follows: it is proposed to obtain the estimator $t^*(x)$ of the parameter θ that would minimize a weighted expression for the mean square error, with arbitrary known weights, and subject to the condition that the optimal estimator should come from the class Ω of competitor estimators $t(x)$ that are quadratic summable (i.e., have finite variance) and that have a prescribed expectation $\mu(\theta)$, identically in θ.

Let us consider a finite set of θ-values, if necessary, by reparameterization of the original range of θ, and let us consider a grid of T values of θ, equally spaced and with common distance Δ in the k-dimensional grid for k parameters. Let $\tilde{\theta}$ denote a variable that ranges over all the T values of θ. Let α_θ, the weight at θ, be known for all θ-values, where α_θ is such that

$$\alpha_\theta \geq 0 \text{ and } \sum_{\theta}^{T} \alpha_\theta \neq 0$$

It is proposed to find the estimator $t^*(x)$ from among the class of estimators $t(x)$, such that

$$\sum_{\theta}^{T} \alpha_\theta \int_{R(x)} (t(x) - \theta)^2 \, \phi(x;\, \theta) \, dx \tag{102}$$

becomes a minimum for $t(x) = t^*(x)$. We have, since $t(x) \in \Omega$, that

$$\int_{R(x)} t(x) \, \phi(x;\, \tau) \, dx = \mu(\tau) \tag{103}$$

for every value of τ.

The problem is again an isoperimetric problem in the Calculus of Variations; we obtain, as before, that a necessary condition for $t^*(x)$ to minimize (102) subject to (103) is that $t^*(x)$ satisfies the following Euler-type equation

$$\sum_{\theta} \alpha_\theta \left(t^*(x) - \theta \right) \phi(x;\, \theta) - \sum_{\tau} \eta_\tau \, \phi(x;\, \tau) = 0 \tag{104}$$

identically in x. We may write

$$t^*(x) \sum_{\theta} \alpha_\theta \, \phi(x;\, \theta) - \sum_{\theta} \alpha_\theta \cdot \theta \, \phi(x;\, \theta) = \sum_{\tau} \eta_\tau \, \phi(x;\, \tau)$$

which can be written as

$$t^*(x) \sum_{\theta} \alpha_\theta \, \phi(x;\, \theta) = \sum_{\tau} \lambda_\tau \, \phi(x;\, \tau) \tag{105}$$

where

$$\tau = \begin{cases} \eta_\theta + \alpha_\theta \cdot \theta & \text{for } \tau = \theta \\ \\ \eta_\tau & \text{for } \tau \neq \theta \end{cases}$$

$$\therefore \quad t^*(x) = \frac{\sum_{\zeta} \lambda_\zeta \, \phi(x;\, \zeta)}{\sum_{\theta} \alpha_\theta \, \phi(x;\, \theta)} \tag{106}$$

Now, from the side condition (103), we have

$$\mu(\theta) = \int_{R(x)} t^*(x)\, \phi(x;\, \theta)\, dx = \sum_{\zeta} \lambda_\zeta \int_{R(x)} \frac{\phi(x;\, \theta)\cdot\phi(x;\, \zeta)}{\sum_{\omega} \alpha_\omega \phi(x;\omega)}\, dx \tag{107}$$

Let us denote the inner product in (107) as

$$\int_{R(x)} \frac{\phi(x;\, \theta)\cdot\, \phi(x;\, \zeta)}{\sum_{\omega} \alpha_\omega \phi(x;\omega)}\, dx = \rho_\theta \cdot \rho_\zeta \tag{108}$$

Then

$$\mu(\theta) = \sum_{\zeta} \lambda_\zeta \, (\rho_\theta \cdot \rho_\zeta) \tag{109}$$

and this equation determines the λ_ζ's for arbitrary $\mu(\theta)$, since $(\rho_\theta \cdot \rho_\zeta)$ is known.

We shall prove now that Equation (105) is a sufficient condition for $t^*(x)$ to minimize (102). Let

$$t(x) = t^*(x) + e(x) \tag{110}$$

be any competitor estimator in the class Ω. Since the expectation of $t(x)$ is $\mu(\theta)$ it follows that

$$E(e(x)\,) = \int_{R(x)} e(x)\, \phi(x;\, \theta)\, dx = 0 \tag{111}$$

identically in θ. Now

$$\text{weighted MSE}(t(x)\ ) = \Sigma\, \alpha_o \int_{R(x)} \{(t^*(x) - \theta) + e(x)\}^2\, \phi(x;\, \theta)dx$$

$$= \sum_\theta \alpha_\theta \int_{R(x)} \left\{t^*(x) - \theta\right\}^2 \phi(x;\, \theta)\ dx + \sum_\theta \alpha_\theta \int_{R(x)} e^2(x)\ \phi(x;\, \theta)dx$$

$$+ 2 \sum_\theta \alpha_\theta \int_{R(x)} e(x) \left\{t^*(x) - \theta\right\} \phi(x;\, \theta)\ dx \tag{112}$$

The last integral in the above expression may be written as

$$2 \sum_\theta \alpha_\theta \int_{R(x)} e(x)\ t^*(x)\ \phi(x;\, \theta)\ dx - 2 \sum_\theta \theta\, \alpha_\theta \int_{R(x)} e(x)\ \phi(x;\, \theta)\ dx$$

$$= 2 \sum_\theta \alpha_\theta \int_{R(x)} e(x)\ t^*(x)\ \phi(x;\, \theta)\ dx \tag{113}$$

since $E\ e(x) = 0$. We can write (113), further, as

$$2 \int_{R(x)} e(x) \sum_\theta \alpha_\theta\ t^*(x)\ \phi(x;\, \theta)\ dx$$

$$= \quad 2 \int_{R(x)} e(x) \sum_{\tilde{6}} \lambda_{\tilde{6}}\ \phi(x;\, \tilde{6})\ dx, \text{ from (105)},$$

$$= \quad 2 \sum_{\tilde{6}} \lambda_{\tilde{6}} \int_{R(x)} e(x)\ \phi(x;\, \tilde{6})\ dx = 0.$$

Hence, from this result and (112), we have

$$\sum_\theta \alpha_\theta \int_{R(x)} \left\{t(x) - \theta)\right\}^2 \phi(x;\, \theta)\ dx$$

$$\geq \sum_\theta \alpha_\theta \int_{R(x)} \left\{t^*(x) - \theta\right\}^2 \phi(x;\, \theta)\ dx \tag{114}$$

with equality holding only if $t(x)$ is essentially equal to $t^*(x)$. Q.E.D.

Suppose now that a UMV estimator exists for $\mu(\theta)$. Then, from the preceding chapter, this estimator $t^*(x)$ must satisfy the equation

$$t^*(x)\ \phi(x;\ \theta) = \sum_{\tau} \lambda(\theta,\ \tau)\ \phi(x;\ \tau).$$

Multiplying both sides of this expression by any arbitrary weights α_θ and summing over θ we obtain

$$t^*(x) \sum_{\theta} \alpha_\theta\ \phi(x;\ \theta) = \sum_{\theta} \alpha_\theta \sum_{\tau} \lambda(\theta,\ \tau)\ \phi(x;\ \tau)$$

$$= \sum_{\tau} \left(\sum_{\theta} \alpha_\theta\ \lambda(\theta,\ \tau) \right) \phi(x;\ \tau)$$

Let

$$\sum_{\theta} \alpha_\theta\ \lambda(\theta,\ \tau) = \lambda'_\tau\ .$$

We may write, therefore,

$$t^*(x) \sum_{\theta} \alpha_\theta\ \phi(x;\ \theta) = \sum_{\tau} \lambda'_\tau\ \phi(x;\ \tau)$$

which is Equation (105). Thus, the following important result has been established, since the UMV estimator, if it exists, is unique.

If a UMV estimator exists for $\mu(\theta)$, the method of minimum weighted variance will reproduce it for <u>any</u> set of weights α_θ.

Let us consider now the rank of the quadratic form

$$\sum_{\tau} \lambda_\tau \int_{R(x)} \frac{\phi(x;\ \theta)\ \phi(x;\ \tau)}{\sum_{\omega} \alpha_\omega\ \phi(x;\ \omega)}\ dx = \sum_{\tau}\ {}_\tau(\rho_\theta \cdot \rho_\tau) \tag{115}$$

that appears in (107). Without explicitly determining the rank of the corresponding matrix, we may decide the question of the linear dependence of the ρ_θ's [10, p. 29]. Consider the quadratic form

$$G = \left(\sum_{\theta}^{T} u_\theta \rho_\theta\right)^2 = \sum_{\theta}^{T} \sum_{\zeta}^{T} (\rho_\theta \cdot \rho_\zeta) u_\theta u_\zeta \tag{116}$$

Clearly, $G \geq 0$. Assume that

$$\Sigma u_\theta^2 = 1 \tag{117}$$

Then, the ρ_θ's are linearly dependent if there exists a set of u_θ's satisfying (117) for which $G = 0$. Thus if the ρ_θ's are linearly dependent the minimum of G subject to (117) must be equal to zero. But this minimum is just the smallest eigenvalue of G, i.e., the least root of the equation

$$\begin{vmatrix} \rho_{(1)}^2 - k & \rho_{(1)} \cdot \rho_{(2)} & \cdots & \rho_{(1)} \cdot \rho_{(T)} \\ \rho_{(2)} \cdot \rho_{(1)} & \rho_{(2)}^2 - k & \cdots & \rho_{(2)} \cdot \rho_{(T)} \\ \vdots & \vdots & & \\ \rho_{(T)} \cdot \rho_{(1)} & \rho_{(T)} \cdot \rho_{(2)} & \cdots & \rho_{(T)}^2 - k \end{vmatrix} = 0$$

Hence [10, p. 29], a necessary and sufficient condition for the linear dependence of the ρ_θ's is the vanishing of the Gram determinant

$$\Gamma = \begin{vmatrix} \rho_{(1)}^2 & \rho_{(1)} \cdot \rho_{(2)} & \cdots & \rho_{(1)} \cdot \rho_{(T)} \\ \rho_{(2)} \cdot \rho_{(1)} & \rho_{(2)}^2 & \cdots & \rho_{(2)} \cdot \rho_{(T)} \\ \vdots & \vdots & & \vdots \\ \rho_{(T)} \cdot \rho_{(1)} & \rho_{(T)} \cdot \rho_{(2)} & \cdots & \rho_{(T)}^2 \end{vmatrix} \tag{118}$$

Now, the Gram determinant of an arbitrary system of vectors is never negative. The relation [10, p. 30]

$$\Gamma = \left| (\rho_\theta \cdot \rho_\tau) \right| \geq 0 \tag{119}$$

in which the equality holds only for linearly dependent ρ_θ's, is a generalization of the Schwarz inequality

$$\left(\int_{R(x)} \frac{\phi^2(x;\ \theta)}{\sum_\omega \alpha_\omega \phi(x;\omega)}\, dx \right) \left(\int_{R(x)} \frac{\phi^2(x;\ \tau)}{\sum_\omega \alpha_\omega \phi(x;\omega)}\, dx \right)$$

$$- \left(\int_{R(x)} \frac{\phi(x;\ \theta)\ \phi(x;\ \tau)}{\sum_\omega \alpha_\omega \phi(x;\omega)}\, dx \right)^2 = \begin{vmatrix} \rho_\theta^2 & \rho_\theta \cdot \rho_\tau \\ \rho_\theta \cdot \rho_\tau & \rho_\tau^2 \end{vmatrix} \geq 0 \tag{120}$$

The equality is attained when

$$\phi(x;\ \theta) = k\ \phi(x;\ \tau); \quad k \neq 0 \tag{121}$$

for $\theta \neq \tau$. Therefore, the expression (115) is of full rank unless (121) holds, in which case the Gram determinant $\Gamma = 0$. The majority of the important density functions in statistics do not accept (121), and hence, the corresponding (115) quadratic form will be positive definite. There are some instances where (121) holds, however, like in the case of a rectangular distribution.

Consider the limiting form of (102), where the problem would be to obtain the estimator $t^*(x)$ that minimizes the WMSE given by

$$\int_{R(\theta)} \alpha(\theta)\ d\theta \quad \int_{R(x)} \left\{ t(x) - \theta \right\}^2 \phi(x;\ \theta)\ dx \tag{122}$$

From the Calculus of Variations we find that the Euler equation is, in this instance,

$$\int_{R(\theta)} \alpha(\theta)\, d\theta \,\{ t^*(x) - \theta\} \phi(x;\, \theta) = 0 \tag{123}$$

Therefore,

$$\int_{R(\theta)} t^*(x)\, \phi(x;\, \theta)\, \alpha(\theta)\, d\theta - \int_{R(\theta)} \theta\, \phi(x;\, \theta)\, \alpha(\theta)\, d\theta = 0 \tag{124}$$

which gives

$$t^*(x) = \frac{\int_{R(\theta)} \theta\, \phi(x;\, \theta)\, \alpha(\theta)\, d\theta}{\int_{R(\theta)} \phi(x;\, \theta)\, \alpha(\theta)\, d\theta} \tag{125}$$

that is, $t^*(x)$ is equal to the 'fiducial' expectation of θ.

To show that (125) is a sufficient condition for $t^*(x)$ to minimize the WMSE given in (122), let

$$t(x) = t^*(x) + e(x) \tag{126}$$

where $t(x)$ is any other estimator of θ, and hence $e(x)$ is completely arbitrary, except that it should be quadratic summable. Consider the WMSE for $t(x)$. We have

$$\int_{R(\theta)} \alpha(\theta)\, d\theta \int_{R(x)} \{t^*(x) + e(x) - \theta\}^2 \phi(x;\, \theta)\, dx$$

$$\begin{aligned} &= \int_{R(\theta)} \alpha(\theta)\, d\theta \int_{R(x)} \{t^*(x) - \theta\}^2 \phi(x;\, \theta)\, dx \\ &\quad + \int_{R(\theta)} \alpha(\theta)\, d\theta \int_{R(x)} e^2(x)\, \phi(x;\, \theta)\, dx \\ &\quad - 2 \int_{R(\theta)} \alpha(\theta)\, d\theta \int_{R(x)} \{t^*(x) - \theta\}\, e(x)\, \phi(x;\, \theta)\, dx \end{aligned} \tag{127}$$

The last integral in (127) may be written as

$$- 2 \int_{R(x)} e(x) \left\{ \int_{R(\theta)} t^*(x)\, \phi(x;\, \theta)\, \alpha(\theta)\, d\theta \right.$$

$$\left. - \int_{R(\theta)} \theta\, \phi(x;\, \theta)\, \alpha(\theta)\, d\theta \right\} dx \;=\; 0$$

since the expression in brackets is zero, by (124). Hence, we have that

$$\text{WMSE } (t(x)\ ) \geq \text{WMSE } (t^*(x)\ ) \tag{128}$$

Suppose now that for some expectation function $\mu(\theta)$ there exists a UMMSE. This estimator would then be unique, as was shown in the preceding chapter. From (44) we have that the UMMSE estimator for this $\mu(\theta)$ would satisfy the Euler equation

$$t^*(x)\, \phi(x;\, \theta) = 0_\tau\, (\theta)\, \phi(x;\, \tau) \tag{129}$$

which, integrating with respect to x, yields

$$\mu(\theta) = 0_\tau\, (\theta) \tag{130}$$

From (129) we obtain

$$\int_{R(\theta)} \alpha(\theta)\, t^*(x)\, \phi(x;\, \theta)\, d\theta = \int_{R(\theta)} \alpha(\theta)\, 0_\tau\, (\theta)\, \phi(x;\, \tau)\, d\theta$$

Let

$$0_\tau\, (\theta)\, \phi(x;\, \tau) = \int_{R(\tau)} \lambda(\theta,\, \tau)\, \phi(x;\, \tau)\, d\, \tau \tag{131}$$

Then

$$\int_{R(\theta)} \alpha(\theta)\, t^*(x)\, \phi(x;\,\theta)\, d\theta = \int_{R(\theta)} \alpha(\theta) \int_{R(\tau)} \lambda(\theta,\,\tau)\, \phi(x;\,\tau)\, d\tau d\theta$$

$$= \int_{R(\tau)} \left(\int_{R(\theta)} \alpha(\theta)\, \lambda(\theta,\,\tau)\, d\theta \right) d\tau \qquad (132)$$

Let

$$\int_{R(\theta)} \alpha(\theta)\, \lambda(\theta,\,\tau)\, d\theta = \tau\, \alpha(\tau)\ . \qquad (133)$$

We then have the Euler equation given by (124), for the estimator $t^*(x)$ that minimizes the WMSE given in (122). The following result has, therefore, been established.

If the UMMSE estimator $t^*(x)$ exists for <u>some</u> expectation function $\mu(\theta)$, then $t^*(x)$ will be the **UMWMSE** estimator for any $\alpha(\theta)$ and competitors having that same expectation function.

The universal MMSE estimator $t^*(x)$ for some weight function $\alpha(\theta)$ is given by (125), i.e., $t^*(x)$ is the fiducial expectation of θ with <u>a priori</u> probability distribution $\alpha(\theta)$. For $\alpha(\theta) \equiv 1$,

$$t^*(x) = \frac{\int_{R(\theta)} \theta\, \phi(x;\,\theta)\, d\theta}{\int_{R(\theta)} \phi(x;\,\theta)\, d\theta} \qquad (134)$$

It is interesting to point out that the result obtained for $t^*(x)$ gives the fiducial expectation of θ, whereas the maximum likelihood estimator is the fiducial mode. There are situations where the two fiducial results coincide. For example, the fiducial expectation in the normal density is the fiducial mode.

Suppose we have a gamma-type exponential density,

$$\phi(x;\ \theta) = \theta^{-n}\ e^{-n\bar{x}/\theta} \tag{135}$$

Let the weight function be

$$\alpha(\theta) = \theta^{-k} \tag{136}$$

where, for the present time, k is an integer without further qualification. The estimator $t^*(x)$ of θ is, from (125),

$$t^*(x) = \frac{\int_0^\infty \theta^{-n-k+1}\ e^{-n\bar{x}/\theta}\ d\theta}{\int_0^\infty \theta^{-n-k}\ e^{-\bar{x}/\theta}\ d\theta} \tag{137}$$

Let

$$q = \theta^{-1};\ \text{hence } d\theta = -q^{-2}\ dq.$$

The numerator in (137) now becomes

$$\int_0^\infty q^{n+k-3}\ e^{-n\bar{x}q}\ dq = (n\bar{x})^{-n-k+2}\,\Gamma(n+k-2) \tag{138}$$

and the denominator in (137) becomes

$$\int_0^\infty q^{n+k-2}\ e^{-n\bar{x}q}\ dq = (n\bar{x})^{-n-k+1}\ \Gamma(n+k-1) \tag{139}$$

Hence, from the last two expressions we find

$$t^*(x) = \frac{\Gamma(n+k-2)}{\Gamma(n+k-1)} \cdot (n\bar{x})$$

and since $\Gamma(p+1) = p\,\Gamma(p)$ for any real p, we may write finally

$$t^*(x) = \left(\frac{n}{n+k-2}\right) \bar{x}$$

For simplicity, let

$$m = k-2 \tag{140}$$

$$\therefore \quad t^*(x) = \left(\frac{n}{n+m}\right) \bar{x} \tag{141}$$

Now, the UMVU estimator of θ is $\bar{x}$, as we have seen before [see p. 37]. Clearly, for $m = 0$, that is, for $k = 2$, $t^*(x)$ becomes $\bar{x}$.

Since $\bar{x}$ is unbiased for θ, it is seen immediately that

$$\begin{aligned} \text{Bias } t^*(x) &= \mu(\theta) - \theta = E(t^*(x)) - \theta \\ &= -\theta\left(\frac{m}{n+m}\right) \end{aligned} \tag{142}$$

which shows that, for fixed m, $t^*(x)$ is consistent.

The MSE of $t^*(x)$ is, since $\text{Var } \bar{x} = \theta^2/n$ [see p. 37],

$$\begin{aligned} \text{MSE}(t^*(x)) &= \text{Var } t^*(x) + \text{Bias}^2\, t^*(x) \\ &= \frac{n^2}{(n+m)^2} \cdot \frac{\theta^2}{n} + \frac{m^2}{(n+m)^2}\theta^2 \\ &= \theta^2 \left[\frac{n+m^2}{(n+m)^2}\right] \end{aligned} \tag{143}$$

Let us compare the MSE of $t^*(x)$ with the variance of $\bar{x}$, the UMVU estimator of θ; we shall examine the inequality

$$\text{MSE}(t^*(x)) \leq \text{Var } \bar{x}$$

$$\therefore\ \theta^2 \left[\frac{n + m^2}{(n + m)^2} \right] \leq \frac{\theta^2}{n}$$

which leads to

$$m^2\, n \leq m^2 + 2\, nm \tag{144}$$

The equality holds for m=0 and $m = \frac{2n}{n-1}$. When m=0, $t^*(x)$ becomes $\bar{x}$, the unique UMVU estimator of θ. Hence k=2 reproduces the UMVU estimator in this case.

The value $m = \frac{2n}{n-1}$ produces the estimator

$$\left(\frac{n - 1}{n + 1}\right)\ \bar{x} \tag{145}$$

whose MSE is equal to Var $\bar{x} = \theta^2/n$. This estimator, however, has a bias of

$$- \theta \left(\frac{2}{n + 1}\right)$$

which is 'larger' than the bias of $t^*(x)$ for either m=1 or m=2. Furthermore, the MSE of $t^*(x)$ for either m=1 or m=2 is smaller than the MSE of the estimator given in (145). We conclude, therefore, that the value for the equality in (144) given by

$$m = \frac{2n}{n - 1}$$

is not a solution to our problem of finding estimators with minimum MSE. Therefore, m=0 is the only solution for the equality sign to hold in (144). Hence, for $m \neq 0$, we may write (144) as a strict inequality,

$$m^2\, n < m^2 + 2nm \tag{146}$$

Suppose that $m > 0$. We have, then, from (146) after dividing through by m,

$$mn < m + 2n$$

or

$$m < \frac{2n}{n - 1} \tag{147}$$

For n=1, a trivial case, this inequality is always satisfied. For any sample size $n > 2$, however, (147) is satisfied only for m=1 and m=2. If m=3, (147) is satisfied only for n=2, which is again a useless result. If $m > 3$, the inequality is not satisfied for any sample size $n > 1$.

Suppose now that m is negative, say

$$m = -p \tag{148}$$

where p is an integer greater than zero. We may write now (146) as

$$p^2 n < p^2 - 2np$$

which leads to

$$p < -\frac{2n}{n - 1} \tag{149}$$

an expression that obviously cannot be satisfied for any sample size.

We have reached the following conclusions:

i) When m=0, or k=2, $t^*(x) = \bar{x}$ is the UMVU estimator of θ.

ii) When m=1, or k=3,

$$t^*(x) = \left(\frac{n}{n + 1}\right) \bar{x}$$

and its MSE is given by

$$\text{MSE} \quad \left(\frac{n}{n + 1}\right) \bar{x} \quad = \frac{\theta^2}{n + 1}$$

iii) When m=2, or k=4,

$$t^*(x) = \left(\frac{n}{n+2}\right) \bar{x}$$

with

$$\text{MSE} \quad \left(\frac{n}{n+2}\right) x \quad = \theta^2 \left(\frac{n+4}{(n+2)^2}\right)$$

iv) When m=1 or m=2, the estimators found have a MSE smaller than Var $\bar{x} = \theta^2/n$. In both cases the estimator obtained is consistent.

v) When m=3, or k=5, the estimator

$$t^*(x) = \left(\frac{n}{n+3}\right) \bar{x}$$

has a MSE given by

$$\text{MSE} \quad \left(\frac{n}{n+3}\right) \bar{x} \quad = \theta^2 \left(\frac{n+9}{(n+3)^2}\right)$$

which is smaller than θ^2/n only for $n \leq 2$.

For $m > 3$, $t^*(x)$ is not more efficient than $\bar{x}$ except for n=1, a trivial result.

vi) For $m < 0$, no estimator is more efficient than $\bar{x}$, for any sample size.

Since we are interested in results that hold for general sample size n, we find that the only acceptable values of k are k=2, k=3, k=4. Therefore, when the weight function

$$\alpha(\theta) = \theta^{-k}$$

given in (136), is proposed, we should qualify the values of k by imposing the condition that

$$2 \leq k \leq 4 \tag{150}$$

We may examine now the relative efficiencies and the bias for the cases when k=3 and k=4, since k=2 corresponds to $t^*(x) = \bar{x}$, the UMVU estimator of θ.

Take k=3, or m=1. Then the bias of $t^*(x) = \left(\frac{n}{n+1}\right)\bar{x}$ is

$$-\theta\left(\frac{1}{n+1}\right).$$

The relative efficiency of $\left(\frac{n}{n+1}\right)\bar{x}$ as compared with $\bar{x}$ is

$$\text{Rel. Eff.}\left\{\left(t^*(x) \mid m = 1\right) \text{ vs. } \bar{x}\right\} = \frac{\frac{\theta^2}{n}}{\frac{\theta^2}{n+1}} = 1 + \frac{1}{n} \tag{151}$$

The gain in efficiency attained with $(t^*(x) \mid m = 1)$ is $(100/n)\%$.

Take now k=4 or m=2. The bias of $t^*(x) = \left(\frac{n}{n+2}\right)\bar{x}$ is

$$-\theta\left(\frac{2}{n+2}\right).$$

$$\text{Rel. Eff.}\left\{\left(t^*(x) \mid m = 2\right) \text{ vs. } \bar{x}\right\} = \frac{\frac{\theta^2}{n}}{\theta^2\left(\frac{n+4}{(n+2)^2}\right)} = 1 + \frac{4}{n^2 + 4n}$$

$$= 1 + \frac{1}{n(\frac{n}{4} + 1)} \tag{152}$$

We thus find that the estimator obtained when m=1 or k=3 is most efficient, and has smallest bias among the estimators considered. Therefore,

$$t^*(x) = \left(\frac{n}{n+1}\right)\bar{x}$$

is UMWMSE for θ in the gamma type considered in (135).

As a further illustration, let us investigate the estimators with

MWMSE of the parameter θ in the gamma density given by

$$\phi(x;\ \theta) = \theta^{n}\ e^{-\theta n\bar{x}} \tag{153}$$

Let the weight function be

$$\alpha(\theta) = \theta^{k} \tag{154}$$

where k is an integer. The possible values of k will be examined later on, when we assess the relative efficiency of whatever estimators are obtained by applying Equation (125) to this problem.

The estimator proposed will be

$$t^*(x) = \frac{\int_{R(\theta)} \alpha(\theta) \cdot \theta\ \phi(x;\ \theta)\ d\theta}{\int_{R(\theta)} \alpha(\theta)\ \phi(x;\ \theta)\ d\theta} = \frac{\int_0^{\infty} \theta^{k+n+1}\ e^{-\theta n\bar{x}}\ dx}{\int_0^{\infty} \theta^{k+n}\ e^{-\theta n\bar{x}}\ dx}$$

$$= \frac{(n\bar{x})^{-k-n-2}\ \Gamma(k+n+2)}{(n\bar{x})^{-k-n-1}\ \Gamma(k+n+1)} = \left(\frac{n+k+1}{n}\right)\frac{1}{\bar{x}} \tag{155}$$

Let k+l=p, an integer. We may write now

$$t^*(x) = \left(\frac{n+p}{n}\right)\frac{1}{\bar{x}} \tag{156}$$

If p=-1, or k=-2, the estimator becomes the UMVU estimator of θ, as seen in (60), with

$$\text{Var}\ \left(\frac{n-1}{n}\right)\ \frac{1}{\bar{x}} = \frac{\theta^2}{n-2}\ ;\ n > 2,$$

as has been given in (68). The variance of (156) is, therefore,

$$\mathrm{Var}\left(\frac{n+p}{n}\right)\frac{1}{\bar{x}} = \left(\frac{n+p}{n-1}\right)^2 \mathrm{Var}\left(\frac{n-1}{n}\right)\frac{1}{\bar{x}}$$

$$= \left(\frac{n+p}{n-1}\right)^2 \cdot \frac{\theta^2}{n-2} \qquad (157)$$

The expectation of $t^*(x)$ is found to be

$$E\ t^*(x) = \left(\frac{n+p}{n-1}\right) E\left(\frac{n-1}{n\bar{x}}\right) = \left(\frac{n+p}{n-1}\right)\theta \qquad (158)$$

From this result we find that the bias of $t^*(x)$ is

$$\mathrm{Bias}\ t^*(x) = \left(\frac{n+p}{n-1}\right)\theta - \theta = \theta\left(\frac{p+1}{n-1}\right) \qquad (159)$$

The MSE of $t^*(x)$ is, therefore,

$$\mathrm{MSE}\ t^*(x) = \left(\frac{n+p}{n-1}\right)^2 \cdot \frac{\theta^2}{n-2} + \theta^2 \frac{(p+1)^2}{(n-1)^2}$$

$$= \frac{\theta^2}{(n-1)^2}\left\{\frac{(n+p)^2}{n-2} + (p+1)^2\right\}$$

$$= \frac{\theta^2}{(n-1)^2(n-2)}\left\{(n+p)^2 + (n-2)(p+1)^2\right\} \qquad (160)$$

Let us consider now the values of p that will make $t^*(x)$ to be more efficient than the UMVU estimator of θ. We want to satisfy the inequality

$$\frac{\theta^2}{(n-1)^2(n-2)}\left\{(n+p)^2 + (n-2)(p+1)^2\right\} \leq \frac{\theta^2}{n-2} \qquad (161)$$

which leads to

$$(n-2)(p+1)^2 \leq (n-1)^2 - (n+p)^2$$

$$\therefore\ (p+1)(2n+p-1) \leq -(n-2)(p+1)^2 \qquad (162)$$

The equality in (162) is satisfied for $p = -1$, which is the value of p that makes $t^*(x)$ the UMVU estimator of θ, in the gamma density under consideration.

Let p be different from -1, and let $p+1$ be positive, that is, $p=0, 1, 2, \ldots$. Dividing through in (162) by $(p+1)$ we have

$$2n + p - 1 < - (n - 2)(p + 1) = - np - n + 2p + 2$$

which gives

$$3n - 3 < p - np$$

or, also

$$3(n - 1) < - p(n - 1)$$

$$\therefore \quad 3 < - p \tag{163}$$

which is impossible. We conclude that p cannot be zero or any positive integer, if (161) is to be satisfied; we must look at negative values of p, other than -1, to see if there are any estimators $t^*(x)$ with MSE smaller than $\theta^2/(n-2)$.

Consider $p < - 1$, an integer. Then $(p+1) < 0$. Dividing now through (162) by $(p+1)$ we have

$$2n + p - 1 > - (n - 2)(p + 1)$$

which gives

$$3n - 3 > - np + p$$

$$3(n - 1) > - p(n - 1)$$

$$\therefore \quad 3 > - p \tag{164}$$

Therefore, from (164) and (163) we conclude that p may take the values -1, -2, only, in order to satisfy the inequality (161). Now, when $p = -1$ we have seen that $t^*(x)$ is the unique UMVU estimator of θ. For $p = -2$, we have

$$t^*(x) = \left(\frac{n-2}{n}\right) \frac{1}{\bar{x}} \tag{165}$$

with MSE given by

$$\text{MSE}\left(t^*(x) \mid p = -2\right) = \frac{\theta^2}{(n-1)^2 (n-2)} \left\{ (n-2)^2 + (n-2) \right\}$$

$$= \frac{\theta^2}{n-1} \tag{166}$$

The bias of $\left(\frac{n-2}{n}\right) \frac{1}{\bar{x}}$ is

$$-\frac{\theta}{n-1} \tag{167}$$

which shows that $t^*(x)$, for $p = -2$, is consistent. The relative efficiency of $t^*(x)$, with respect to the UMVU estimator of θ, is

$$\text{Rel. Eff. } \left\{ \left(t^*(x) \mid p = -2\right) \text{ vs. UMVU estimator} \right\}$$

$$= \frac{\theta^2/(n-2)}{\theta^2/(n-1)} = \frac{n-1}{n-2} = 1 + \frac{1}{n-2} \tag{168}$$

The gain in efficiency attained by utilizing $t^*(x)$ over the UMVU estimator is $100/(n-2)\%$.

We see now that the only values of k in the weight function $\alpha(\theta)$ given in (154) that make the estimator $t^*(x)$ at least as efficient as the UMVU estimator of θ, are $k = -2$ and $k = -3$.

V. LITERATURE CITED

1. Aitken, A. C. On the estimation of many statistical parameters. Proc. Roy. Soc. Edin. Sec. A, 62: 369-377. 1947.

2. ________ and Silverstone, H. On the estimation of statistical parameters. Proc. Roy. Soc. Edin. Sec. A, 61: 186-194. 1942.

3a. Bahadur, R. R. On unbiased estimates of uniformly minimum variance. Sankhyā 18: 211-224. 1957.

3b. Barankin, E. W. Extension of a theorem of Blackwell. Annals Math. Stat. 21: 280-284. 1950.

4. ________. Locally best unbiased estimates. Annals Math. Stat. 20: 477-501. 1949.

5. Bhattacharyya, A. On some analogues of the amount of information and their use in statistical estimation. Sankhyā 8: 1-13. 1947.

6. Blackwell, D. Conditional expectation and unbiased sequential estimation. Annals Math. Stat. 18: 105-110. 1947.

7. Bolza, O. Lectures on the calculus of variations. Chicago, The University of Chicago Press. 1904.

8. Brand, L. Advanced calculus. New York, John Wiley and Sons, Inc. 1955.

9. Chapman, D. G. and Robbins, H. Minimum variance estimation without regularity assumptions. Annals Math. Stat. 22: 581-586. 1951.

10. Courant, R. and Hilbert, D. Methoden der mathematischen physik. Vol. 1, 2nd ed. Berlin, Julius Springer. 1931.

11. Cramér, H. Mathematical methods of statistics. Princeton, N. J., Princeton University Press. 1946.

12. Gurland, J. Lectures on tests of hypotheses and on estimation. (Mimeographed) Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology. 1957.

13. Halmos, P. R. and Savage, L. J. Applications of the Radon-Nikodym theorem to the theory of sufficient statistics. Annals Math. Stat. 20: 225-241. 1949.

14. Jackson, D. Fourier series and orthogonal polynomials. Mathematics Association of America. Carus Mathematical Monographs No. 6. 1941.

15. Jordan, C. Calculus of finite differences. 2nd ed. New York, Chelsea Publishing Co. 1960.

16. Kendall, M. G. The advanced theory of statistics. Vol. 2, 2nd ed. London, Charles Griffin and Co., Ltd. 1951.

17. Koopman, B. O. On distributions admitting a sufficient statistic. Trans. Amer. Math. Soc. 29: 399-409. 1936.

18. Lehman, E. L. Notes on the theory of estimation. (Mimeographed) Berkeley, University of California. 1950.

19. Lèvy, P. Problèmes concrets d'analyse fonctionelle. 2nd ed. Paris, Gauthier-Villars. 1951.

20. Rao, C. Radhakrishna. Advanced statistical methods in biometric research. New York, John Wiley and Sons, Inc. 1952.

21. _________. Information and the accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37: 81-91. 1945.

22. _________. Some theorems on minimum variance estimation. Sankhyā 12: 27-42. 1952.

23. _________. Sufficient statistics and minimum variance estimates. Proc. Cam. Phil. Soc. 45: 213-218. 1949.

24. Riesz, F. and Sz. Nagy, B. Leçons d'analyse fonctionelle. 3rd ed. Paris, Gauthier-Villars. 1955.

25. Robbins, H. Mixture of distributions. Annals Math. Stat. 19: 360-365. 1948.

26a. Schmetterer, L. On unbiased estimation. Annals Math. Stat. 31: 1154-1163. 1960.

26b. Seth, G. R. On the variance of estimates. Annals Math. Stat. 20: 1-27. 1949.

27. Stein, C. Unbiased estimates with minimum variance. Annals Math. Stat. 21: 406-415. 1950.

28. Szegö, G. Orthogonal polynomials. Rev. ed. Am. Math. Soc. Coll. Pub. Vol. 23. New York, American Mathematical Society. 1959.

29. Tate, R. F. and Goen, R. L. Minimum variance unbiased estimation for the truncated Poisson distribution. Annals Math. Stat. 29: 755-765. 1958.

30. Thielman, H. P. Theory of functions of real variables. New York, Prentice-Hall, Inc. 1953.

31. Tukey, J W. Sufficiency, truncation and selection. Annals Math. Stat. 20: 309-311. 1949.

VI. ACKNOWLEDGMENTS

The author considers it a privilege to have worked under Professor H. O. Hartley, to whom he expresses his gratitude. Professor Hartley suggested the topic, contributed many ideas and had the patience to guide the writer throughout the various stages in the preparation of this thesis.

The author is especially grateful to Professor T. A. Bancroft who, through his personal kindness, enabled him to pursue his studies at the Graduate College of Iowa State University.

Acknowledgment is made to the National Science Foundation, which supported this research with Grant Number NSF-G 14236.