SOME THEOREMS ON MINIMUM VARIANCE ESTIMATION

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

1. Introduction

Unbiased minimum variance estimation is applicable only in the case of distritions admitting uniformly minimum variance estimates of parameters. If such statistics do not exist then we need an additional criterion to choose among a number of alternative unbiased estimates having minimal values of the variance in different regions of the parameter space. Therefore it is of some importance to determine the class of distributions which admit statistics with uniformly minimum variance.

It was already shown in earlier papers that when distributions admit sufficient statistics, any parametric function possessing an unbiased estimate has a uniformly minimum variance estimate also. This result was proved in two stages. First it was shown that given any unbiased estimate it is possible to construct a function of the sufficient statistics which is also unbiased and has uniformly a smaller variance than the proposed one (Rao, 1945a). Secondly when the number of sufficient statistics equals that of the unknown paramters there is a unique function of the sufficient statistics which is unbiased for a particular parametric function (Rao, 1948).

It may now be asked whether this is an exclusive property of distributions admitting sufficient statistics. If so, the method of minimum variance estimation is applicable only to this class of distributions in which case the actual derivation of estimates is also simple.

In this paper some general theorems concerning unbiased estimates possessing minimum variance locally and uniformly have been deduced. The exclusive property of the distributions referred above has been demonstrated under some general conditions. Finally the usefulness of some expressions giving lower limits to the variance of estimates has been examined.

2. Some general theorems on minimum variance estimates

The following definitions are used throughout.

(i) The joint probability density of the observations is denoted by $P(x|\theta)$ where x stands for the observations $x_1, \dots, x_n$ and θ for the parameters $\theta_1, \dots, \theta_q$. If the observations are independent and arise from the same population then

$$P(x\,|\,\theta)=p(x_1\,|\,\theta)\dots p(x_n\,|\,\theta)$$

where $p(x\,|\,\theta)$ is the probability density in the parent distribution.

(ii) A function $\phi(x)$ of the observations is said to belong to L_2, $\phi(x)\epsilon L_2$ if

$$E[\phi(x)]^2 = \int [\phi(x)]^2 P(x|\theta)dx < \infty$$

where dx stands for the product of n differentials. It is not necessary that the variables should be continuous. In the case of discrete observations the integration can be replaced by the summation symbol without altering the steps of the proof.

(iii) A function $Z(x)$ or simply denoted by Z is said to belong to N, $Z \epsilon N$ if

$$E(Z) = \int ZP(x|\theta)dx = 0 \quad \text{for all } \theta.$$

(iv) A parametric function $\tau(\theta_1, ..., \theta_q)$ or $\tau(\theta)$ is said to belong to U, $\tau \epsilon U$ if it admits an unbiased estimate. An unbiased estimate of a parametric function $\tau(\theta)$ is represented by $T(x)$ or simply T.

(v) A minimum varince estimate will be referred to as *m.v.e.*

Theorem 1: *The necessary and sufficient condition that an unbiased estimate T of $\tau \epsilon U$ is the m.v.e. at a particular set of values $\theta^0 = (\theta^0_1, ..., \theta_q^0)$ is that*

$$E(ZT|\theta^0) = 0 \quad ... \quad (2.1)$$

whenever $Z \epsilon N$ and L_2

The necessity is easily proved by considering an alternative estimate $T+\lambda Z$ unbiased for τ, for arbitrary λ, and showing that its variance for a suitable choice of λ is smaller than that of T unless the condition (2.1) is satisfied. This is same as the proof given by Fisher (1921, 1925) and Stein (1949).

To prove sufficiency let us consider an alternative estimate T'. Then $(T-T')$ ϵN and also L_2 if $V(T)$ and $V(T')$ are $< \infty$. Hence

$$E\{T(T-T')|\theta^0\} = 0$$

or

$$\sigma_T^2 = \rho\sigma_T\sigma_{T'}, \text{ i.e., } \sigma_T \leqslant \sigma_{T'} \quad ... \quad (2.2)$$

This shows that T' is as good as T only when $\rho = 1$ in which case both the statistics are equivalent except for a set of observations of probability measure zero. It is also seen that $E(T-T')^2 = 0$ establishing equivalence of T and T'.

Also if T' is another unbiased estimate of τ satisfying (2.1) then $\sigma_{T'} \leqslant \sigma_T$ which together with (2.2) implies that $\sigma_T = \sigma_{T'}$ or $\rho = 1$.

A similar theorem involving the theory of functionals is given by Stein (1949) and another restricting to functions of a sufficient statistic by Lehmann and Scheffe (1950).

Corollary 1.1: *The necessary and sufficient condition that T is uniformly the best is that the condition (2.1) is true for all θ.*

Corollary 1.2: *If a statistic $T \epsilon L_2$ is such that $E(ZT|\theta^0) = 0$ for all $Z \epsilon N$ and L_2 then T is an estimate of some parametric function and has minimum variance at θ^0.*

This is proved if $E(T) \neq$ a constant C independently of θ. If so $(T-C) \epsilon N$ and L_2 which means that $V(T|\theta^0) = 0$.

Corollary 1.3: *If $Z \epsilon N$ and L_2, then from the relation*

$$\int ZP(x|\theta)dx=0 \quad \dots \quad (2.3)$$

it is formally permissible to deduce the relation

$$\int ZP_{a\,b}\dots dx=0 \quad \dots \quad (2.4)$$

$$P_{ab}\dots_q=\frac{\partial^r P(x|\theta)}{\partial^a\theta_1\dots\partial^q\theta_q}$$

From this it follows that if there exist functions of parameters $\lambda_{a\,b}\dots$ such that

$$\Sigma\lambda_{a\,b}\dots\frac{P_{a\,b}\dots}{P(x|\theta)} \quad \dots \quad (2.5)$$

where the summation is over all values of a, b,... reduces to $T(x)$ a function of the observations only then T is a uniformly *m.v.e.*

We thus have a simple demonstration of the result obtained earlier by various authors. A. C. Aitken informed the author (in personal correspondence dated 28.8.1946 and 29.9.1946) that this result (2.5) was mentioned in correspondence with him in 1942 by R. C. Geary of Dublin. Later Leon Salmon considered this problem in a thesis submitted to the University of Edinburgh in 1944. An extensive discussion of this result was given by Bhattacharya (1946, 1947, 1948).

Corollary 1.4: *A formal integration of the equation* (2.3) *gives the result*

$$\int\int_{\theta^1}^{\theta} ZP(x|\theta)\, A_1(\theta)d\theta dx=0 \quad \dots \quad (2.6)$$

From this it would follow that the function

$$\lambda_0(\theta)+\Sigma\,\frac{\lambda_i(\theta)}{P(x|\theta)}\int_{\theta^1} P(x\,|\,\theta)\,A_i(\theta)d\theta \quad \dots \quad (2.7)$$

if it is independent of θ, provides a uniformly m.v.e.

Also the same is true of a combination of (2.5) and (2.7). This result due to Bhattacharya (1950) can thus be formally deduced from the equation (2.3).

It may now be asked under what conditions the expressions (2.5), (2.7) or a combination of both reduce to functions of the observations only. In a later section it is shown that this situation arises only in a special class of distributions and all the statistics derivable this way can be easily written down without going through the processes of differentiation and integration. It is, however, unfortunate that these expressions (2.5) and (2.7) having some symbolic beauty do not throw any additional light in discovering minimum variance statistics.

It should also be noted that Davis (1951) made an attempt to show that statistics cannot be constructed by using anything more than the first differential of the probability density. This does not seem to be correct. For, consider the case

$$P(x|\theta) = \text{const. } e^{-\frac{1}{2}\Sigma(x_i-\sqrt{\theta})^2}$$

$$\frac{P'(x|\theta)}{P(x|\theta)} = \frac{1}{2}\cdot\frac{\Sigma(x_i-\sqrt{\theta})}{\sqrt{\theta}}$$

$$\frac{P''(x|\theta)}{P(x|\theta)} - \frac{[P'(x|\theta)]^2}{[P(x|\theta)]^2} = -\frac{1}{4\theta}\left\{ n + \frac{\Sigma(x_i-\sqrt{\theta})}{\sqrt{\theta}} \right\}$$

Then it is possible to choose λ_1 and λ_2 such that

$$\lambda_1 \frac{P''(x|\theta)}{P(x|\theta)} + \lambda_2 \frac{P'(x|\theta)}{P(x|\theta)} + \theta = \frac{(\Sigma x)^2 - n}{n^2}$$

which shows that

$$\bar{x}^2 - 1/n$$

is the *m.v.e.* of θ. This example is covered by the distribution derived in section 4 of this paper.

Corollary 1.5: *If T is an unbiased estimate of τ and attains the minimum value at θ^0 and if a uniformly m.v.e. T' exists for τ then T and T' are equivalent provided that the range of observations giving positive values to the probability density is independent of θ.*

This is true because it follows from theorem 1 that T and T' have unit correlation at θ^0.

Consider the function obtained from the expressions (2.5), (2.7) or a combination of both by substituting a particular value of θ, say θ^0. This statistic is now a function of the observations only and may be considered as an estimate of its expected value say $\mu(\theta)$. Since the relations (2.4) and (2.6) are true it follows from theorem 1 that this statistic has the minimum variance at θ^0 as an unbiased estimate of $\mu(\theta)$. If $\mu(\theta)$ admits a uniformly *m.v.e.* then this must be equivalent to the above statistic.

We thus have a method of a formal derivation of statistics having minimum variance at specified points and which coincide with uniformly *m.v.* estimates when they exist.

Theorem 2: *If T is an unbiased estimate of some parametric function such that*

$$E(ZT|\theta) = 0 \qquad \dots \quad (2.8)$$

for any $Z \epsilon N$ then

(i) *T^r for any integral r is the m.v.e. of $E(T^r)$ if it exists,*

(ii) *any polynomial in T is m.v.e. of its expected value, and*

(iii) *in general, any function of T is the m.v.e. of its expected value.*

The *m.v.* estimates are uniformly so because the condition (2.8) is true for all values of θ.

Since $E(ZT)=0$, $ZT\epsilon N$ in which case $E(ZTT)=0$ and in general $E(ZT^r)=0$ showing thereby that T^r satisfies the condition of theorem 1. Hence (i) is proved. It may be possible that $V(T^r)$ does not exist in which case it has to be concluded that no estimate of $E(T^r)$ exists with a finite variance.

The proof of (ii) follows automatically from the proof of (i).

To prove (*iii*) we observe that

$$\int\int ZT^r P(T)P(Z|T)dZdT=0$$

for all r which implies

$$\int T^r f(T)\, P(T)\, dT=0 \qquad \text{for all } r$$

where $f(T)=E(Z\,|\,T)$. We can obviously restrict the class of functions $f(T)$ to the condition $V\{f(T)\}<\infty$ which means $Z\,\epsilon\, N$ and L_2 or $|f(T)|\,P(T)$ is integrable. Suppose that

$$e^{k\sqrt{T}}\,|f(T)|\,P(T) \text{ when } T \text{ is in } (0, \infty)$$

or

$$e^{k|T|}\,|f(T)|\,P(T) \quad \text{when } T \text{ is in } (-\infty, \infty)$$

is integrable for some positive k then by Hardy's (1917) result it follows that the moment problem is uniquely soluble and

$$f(T)P(T)=0 \quad \text{or} \quad f(T)=0 \qquad \dots \quad (2.9)$$

except for a set of points of probability measure zero.

The result (iii) follows from (2.9) because this condition implies that $E[Z\,|\,\psi(T)]=0$ for any arbitrary function $\psi(T)$. Hence $E(Z\psi)=0$ implying that ψ is *m.v.e.*

In the statement of theorem 2 it is assumed that the statistic T has zero covariance with all functions $Z\epsilon N$. The condition for the minimum variance statistic is less stringent viz., that $Z\epsilon N$, L_2. Suppose it is given that T is the minimum variance statistic then under what conditions is theorem 2 true?

Let us assume that all moments of T exist and $E(T^r)$ admits a minimum variance statistic with a finite fourth moment then theorem 2 is ture. The author however feels that all these assumptions may not be necessary.

From the assumptions made above it follows that $E(ZT)=0$ for $Z\epsilon N$, L_2. To deduce that $E(ZT^2)=0$, it is necessary to show that T^2 is a minimum variance estimate. If not let T' be so. Then

$$E\{T(T^2-T')\}=0$$

and

$$V\{T(T^2-T')\}<\infty$$

because $E(T')^4<\infty$ and $E(T^r)<\infty$. Therefore $T(T^2-T')\epsilon N$, L_2 in which case

$$E\{T.T(T^2-T')\}=0$$

which shows that T' and T^2 are equivalent and hence the result.

Corollary 2.1: *Let T_1 and T_2 be two statistics such that*

$$E(ZT_1|\theta)=0 \text{ and } E(TZ_2|\theta)=0$$

for all $Z \epsilon N$. Following the proof of theorem 2 it can be shown that any function of T_1 and T_2 is an m. v. e. under some general conditions.

When the condition is restricted to $Z \epsilon N$ and L_2 the additional assumptions needed are that the moments of T_1 and T_2 exist and the fourth moment of the *m.v.e.* of $E(T_1{}^r T_2{}^s)$ exists. This result can be extended to a number of statistics having zero covariance with functions $Z \epsilon N$ and $Z \epsilon N$ and L_2. The statistics T_1 and T_2 considered in corollary 2.1 have necessarily different expectations.

Illustration 1. *The complete class theorem of sufficient statistics*

Consider a distribution admitting sufficient statistics equal in number to the unknown parameters. The probability density can be written in the form (Koopman, 1936)

$$P(x|\theta)=e^{t_1\Theta_1+t_2\Theta_2+\ldots+t_q\Theta_q+\Theta+t}$$

where t_i and t are functions of the observations only and Θ_i and Θ are functions of the parameters only. Let $Z \epsilon N$, then for all θ

$$\int ZP(x|\theta)dx=0$$

In such a case it is *strictly permissible* to differentiate under the integral sign with respect to Θ_i and obtain the result

$$\int Zt_iP(x|\theta)dx=0$$

which is true except perhaps at the extremeties of the range of Θ_i. To make the argument more rigorous we integrate the expression

$$\int ZP(x|\theta)dx=0$$

over constant values of t_1, assuming that the ranges of $t_2, \ldots, t_q$ are independent of Θ_1, to obtain the identity

$$\int f(t_1)e^{t_1\Theta_1}\, dt_1=0$$

in the entire range of Θ_1. The function $f(t_1)$ does not invlve Θ_1 if we assume that $\Theta_1, \ldots, \Theta_q$ are functionally independent. Using the powerful results due to Widder we can differentiate under very general conditions to obtain the above result.

Thus each t_i satisfies the condition of theorem 2 and therefore any function of $t_1, \ldots, t_q$ is an *m.v.e.* This result together with a previous finding (Rao, 1945a) that any parametric function $\tau(\theta) \epsilon U$ admits a function of the sufficient statistics as an unbiased estimate shows that the complete class of unbiased *m.v* estimates is supplied by functions of sufficient statistics. That is, every member of this class is the *m.v.e.* of some parametric function and also every *m.v.e.* is a member of this class. It also follows that there cannot to be two functions of sufficient statistics having the same expectation. In may be recalled that this unicity property was earlier used by the author (Rao, 1948) and Lehman and Scheffe (1950) to prove the complete class theorem.

We could avoid differentiation under the integral sign by first proving that $t_1' = e^{t_1\theta}$ has zero covariance with all $Z \epsilon X$ for a sufficiently big range of Θ_1. This is secured by splitting Θ_1 into δ and $\Theta_1 - \delta$ in the relation expressing the condition $E(Z) = 0$ and choosing δ small. The argument is repeated with $t_1', t_2', \ldots$

Illustration 2. *Linear estimation*

If the class of estimates is restricted to linear functions of the observations only then it follows from theorem 1 that the complete class of estimates is defined by the linear functions

$$c_1 x_1 + \ldots + c_n x_n \qquad \ldots \quad (2.10)$$

which have zero covariance with all linear functions

$$d_1 x_1 + \ldots + d_n x_n \qquad \ldots \quad (2.11)$$

having zero expectation. Suppose that

$$E(x_i) = a_{i1}\tau_1 + \ldots + a_{im}\tau_m$$

where $\tau_1, \ldots, \tau_m$ are unknown parameters. The condition that (2.11) has zero expectation means

$$\Sigma a_{ij} d_i = 0, \quad j = 1, \ldots, m \qquad \ldots \quad (2.12)$$

and the condition of zero covariance between (2.10) and (2.11) reduces to

$$\Sigma\Sigma d_i c_j \rho_{ij} = 0 \qquad \ldots \quad (2.13)$$

where (ρ_{ij}) is the covarince matrix of $x_1, \ldots, x_n$. The condition (2.13) is true for all $d_1, \ldots, d_n$ satisfing (2.12). Hence there exist suitable multipliers $\lambda_1, \ldots, \lambda_m$ such that

$$\Sigma c_j \rho_{ij} = \Sigma \lambda_j a_{ij}, \quad i = 1, \ldots, n$$

or

$$c_j = \sum_i \rho^{ji} \sum_r \lambda_r a_{ir}$$

where (ρ^{ij}) is the matrix inverse to (ρ_{ij}). Consequently

$$\Sigma c_i x_i = \Sigma \lambda_j Q_j \qquad \ldots \quad (2.14)$$

where

$$Q_j = \sum_s (\sum_i \rho^{si} a_{ij}) x_s$$

Suppose that $\Sigma \lambda'_j Q_j$ is an alternative expression having the same expectation then

$$\Sigma(\lambda_j - \lambda'_j) Q_j \qquad \ldots \quad (2.15)$$

belongs to the class (2.14), but has zero covariance with all members of this class and hence with itself. This is impossible unless (2.15) is identically zero. Thus in the problem of linear estimation we are thrown on the problem of adjusting the coefficients

λ_j in (2.14) to have a given expectation irrespective of any consideration about its variance.

A simple way of doing this is given by the author in Rao (1945b). First the normal equations are written

$$E(Q_i)=l_{i1}\tau_1+\ldots+l_{im}\tau_m$$
$$i=1,\ldots,m$$

where
$$l_{ij}=\Sigma\ \Sigma\ \rho^{st}a_{si}a_{tj}$$

If $t_1,\ldots,t_m$ is a solution of the equations

$$Q_i=l_{i1}t_1+\ldots+l_{im}t_m$$
$$i=1,\ldots,m$$

then the best estimate of an estimable parametric function

$$p_1\tau_1+\ldots+p_m\tau_m$$

is simply $p_1t_1+\ldots+p_mt_m$. This gives a linear function of Q_i which has $p_1\tau_1+\ldots+p_m\tau_m$ as its expectation.

Illustration 3. *Quadratic estimate of variance*

Hsu (1938) presented a solution to the problem of determining the best unbiased quadratic estimate of variance when it is known that x_i are independent variables such that

$$E(x_i)=a_{i1}\tau_1+\ldots+a_{im}\tau_m$$

and $V(x_i)=\sigma^2$ independently of i. Hsu considered the class of quadratic forms whose variance is independent of the unknown parameters τ. The alternative assumption made here is that the quadratic form is definite.

If the matrix (a_{ij}) is denoted by A and has rank r then there exists an orthogonal transformation

$$(\underline{z}\vdots\ \underline{y})=\underline{x}(B\vdots\ C)$$

where
$$B'=(b_{ij}),\ i=1,\ldots,n-r,\ j=1,\ldots,n$$

and
$$C'=(c_{ij}),\ i=1,\ldots,r,\quad j=1,\ldots,n$$

such that

$$E(z_i)=0 \quad \text{and} \quad E(y_i)=\xi_i\neq 0$$

A general quadratic form in x can now be written

$$\underline{z}D\underline{z}'+\underline{z}P\underline{y}'+\underline{y}Q\underline{y}'$$

which has the expected vlue

$$\sigma^2\,(\text{trace of } D+\text{trace of } Q)\ +\ \underline{\xi}\ Q\ \underline{\xi}'$$

Since the quadratic form has to be unbiased $\underline{\xi}Q\underline{\xi}'=0$ for all ξ which means that $Q=0$. The quadratic form, by assumption, is definite so that $P=0$. The admissible class is now $\underline{z}D\underline{z}'$ with the condition that the trace of D is unity. Let R be a matrix with trace zero in which case $E(\underline{z}R\underline{z}')=0$. If $BDB'=(\lambda_{ij})$ and $BRB'=(\mu_{ij})$ then the covariance between $\underline{z}D\underline{z}$ and $\underline{z}R\underline{z}'$ is the same as that between $\Sigma\Sigma\lambda_{ij}\, x_i x_j$ and $\Sigma\Sigma\mu_{ij}\, x_i x_j$ or that between the same expressions with x_i considered as deviations from the mean values. This value is easily seen to be

$$\sigma^4\{(\Sigma\lambda_{ii}\mu_{ii})(\beta_2-3)+2\Sigma\Sigma\lambda_{ij}\mu_{ij}\}$$

where β_2 is the value of μ_4/σ^4 assuming that all x have the same fourth moment.

The necessary and sufficient conditions that $\underline{z}D\underline{z}'$ is uniformly the best are

(i) $$\Sigma\Sigma\lambda_{ij}\,\mu_{ij}=0, \quad \text{when} \quad \beta_2=3$$

and (ii) $\Sigma\Sigma\lambda_{ij}\mu_{ij}=0$, $\Sigma\lambda_{ii}\,\mu_{ii}=0$ when the value of β_2 is unknown.

The condition $\Sigma\Sigma\lambda_{ij}\,\mu_{ij}=0$ implies that the trace of $(\lambda_{ij})(\mu_{ij})=0$ or trace of $BDB'\ BRD'=0$, and $B'B=1$. This must be true whenever the trace of $BRB'=0$. The solution $D=I$ (multiplied by a scalar) satisfies the requirements and from uniqueness of minimum variance estimates it follows that this is the only solution. Hence when $\beta_2=3$ the unbiased minimum variance estimate is supplied by

$$(z_1^2+\ldots+z^2_{n-r})/(n-r)$$

which is same as the least square estimate obtained by dividing the minimum sum of squares

$$\Sigma(x_i-a_{i1}\,\tau_1-\ldots)^2$$

by $(n-r)$. When $\beta_2\neq3$ and unknown the further condition to be satisfied is

$$\Sigma\lambda_{ii}\mu_{ii}=0$$

a sufficient condition for which is obviously $\lambda_{ii}=\lambda$ independent of i since $\Sigma\mu_{ii}=0$. This result has been noted by Hsu. To obtain a necessary and sufficient condition we observe that

$$\lambda_{ii}=b^2_{1i}+\ldots+b^2_{n-r\cdot i}$$

where the b coefficients are specified by the orthogonal transformation. Consider now, the special values of

$$\mu_{ii}=b^2_{ri}-b^2_{si}$$

and

$$\mu_{ii}=b_{ri}b_{si}$$

arising out of the quadratic forms $z^2_r-z^2_s$ and $z_r z_s$ having zero expectation. The necessary and sufficient condition may be stated in the form

$$\Sigma(b^2_{ri}-b^2_{si})(b^2_{1i}+b^2_{2i}+\ldots)=0$$
$$\Sigma b_{ri}b_{si}(b^2_{1i}+b^2_{2i}+\ldots)=0$$

for all r and s. If λ_{ii} is independent of i, then the above conditions are automatically satisfied.

Thus the least square estimate is uniformly the best only when the necessary and sufficient conditions stated above are true in the case β_i is unknown. To show that there can exist minimum varince estimates without λ_{ii} being equal we give the following example where the z functions turn out to be

$$z_1 = \frac{x_1}{\sqrt{2}} + \frac{x_2}{\sqrt{4}} + \frac{x_4}{\sqrt{4}}$$

$$z_2 = \frac{x_1}{\sqrt{4}} + \frac{x_3}{\sqrt{4}} - \frac{x_4}{\sqrt{2}}$$

The above conditions are satisfied. The best estimate is $(z_1^2+z_2^2)/2$ whose expansion gives different values to the coefficients of the square terms.

Illustration 4. *Markoff's set up with linear restrictions on parameters*

The independent variables $x_1,\ldots,x_n$ have the expectations

$$E(x_i)=a_{i1}\tau_1+\ldots+a_{im}\tau_m, \quad i=1,\ldots,n$$

where the unknown parameters $\tau_1,\ldots,\tau_m$ are subject to s restrictions

$$g_i=r_{i1}\tau_1+\ldots+r_{im}\tau_m, \qquad i=1,\ldots,s$$

The class of minimum variance linear estimates is defined by

$$\lambda_1 Q_1+\ldots+\lambda_m Q_m+C$$

where $\quad Q_i=a_{1i}x_1+\ldots+a_{ni}x_n$ and $\lambda_1,\ldots,\lambda_m$ are such that

$$\lambda_1 r_{i1}+\ldots+\lambda_m r_{im}=0$$

and C is a constant independent of the variables. This is obtained by expressing the condition that the estimate has zero covariance with all linear functions whose expectations are independent of τ's subject to the above restrictions. It may be recalled when there are no restrictions on τ's the complete class is defined by

$$\lambda_1 Q_1+\ldots+\lambda_m Q_m$$

for all arbitrary $\lambda_1,\ldots,\lambda_m$. To estimate any desired parametric function we need only adjust the λ's suitably to satisfy the condition of unbiasedness.

To obtain the best quadratic estimate of variance we observe that there exists an orthogonal transformation of $x_1,\ldots,x_n$ to

$$y_1,\ldots,y_k,z_1,\ldots,z_{n-k}$$

such that $E(y_i)$ is dependent on τ and $E(z_i)=b_i$ independent of τ. If q is the rank of the matrix

$$\begin{matrix} a_{11} & \ldots & a_{1m} \\ \cdot & \ldots & \cdot \\ a_{n1} & \ldots & a_{nm} \\ r_{11} & \ldots & r_{1m} \\ \cdot & \ldots & \cdot \\ r_{s1} & \ldots & r_{sm} \end{matrix}$$

and t that of

$$\begin{matrix} r_{11} & \cdots & r_{1m} \\ \cdot & \cdots & \cdot \\ r_{s1} & \cdots & r_{sm} \end{matrix}$$

then the value of k is determined from the formula

$$k=n-q+t$$

The quadratic estimate (including linear terms) assumed to be non-negative can now be written in the form

$$\Sigma\Sigma c_{ij}(z_i-b_i)(z_j-b_j)$$

The conditions under which

$$\Sigma(z_i-b_i)^2$$

is the best estimate are same as those in the illustration 3 in terms of the coefficients of the variables $x_1,\ldots,x_n$ in the transformation from x to z. It is easily seen that $\Sigma(z_i-b_i)^2$ is the least value of

$$\Sigma(x_i-a_{i1}\tau_1\ldots-a_{im}\tau_m)^2$$

subject to the restrictive conditions on τ's mentioned above.

When the variables are normally distributed this is the minimum variance estimate of the variance in the whole class of estimates and is distributed as $\sigma^2\chi^2$ on $(n-q+t)$ degrees of freedom being the sum of squares of $(n-q+t)$ independent norml variates (Rao, 1951).

It is of interest to note that when the observations are normally distributed the complete class of minimum variance estimates of linear parametric functions is defined by

$$C_1Q_1+\ldots+C_mQ_m$$

in the entire class of estimating functions. For this it is enough to show that each Q_i is orthogonal to zero functions. The condition for a zero function is

$$\int ZPdv=0$$

Differentiating this with respect to τ_i which occurs in the expression

$$P=\text{const } e^{-\frac{1}{2}\Sigma(x_j-a_{j1}\tau_1-\ldots)^2}$$

we find

$$Q_i=\Sigma x_j a_{ji}$$

is a minimum variance estimate. To complete the proof it is necessary to show that linear parametric functions which cannot be estimated by linear functions of $Q_1,\ldots,Q_m$ do not admit unbiased estimates and should be regarded as non-estimable. If $b_1\tau_1+\ldots+b_m\tau_m$ is a linear parametric function admitting an unbiased estimate T then differentiation of the equation of unbiasedness leads to the results

$$\begin{gathered} C_1a_{11}+C_2a_{21}+\ldots=b_1 \\ \ldots \quad \ldots \\ C_1a_{1m}+C_2a_{2m}+\ldots=b_m \end{gathered}$$

where C_i is the covariance between T and x_i. The condition of estimability is then that there exist $C_1, C_2 \ldots$ satisfying the above linear equations. When such C_i exist it is easy to show that there exists a linear function of the observations whose expectation is $l_1\tau_1+\ldots+l_m\tau_m$ and hence its minimum variance estimate belongs to the complete class determined above.

Illustration 5. *Estimation of moments*

Let $x_1,\ldots,x_n$ be independent observations from the same parent distribution. Suppose a quadratic estimate of the second moment μ_2 of x is sought. Using the result mentioned by Basu (1952) we have to search for the *m.v.* estimates from the class of symmetric functions only. Consider a quadratic expression of the form

$$a\Sigma x_i^2+b\Sigma x_i x_j \qquad \ldots \quad (2.16)$$

which has the expectation

$$na(\mu_2+\mu_1^2)+n(n-1)b\mu_1^2$$

vanishing identically (when μ_1 and μ_2 are both unknown) only when $a=0, b=0$. Therefore (2.16) is *m.v.e.* for all a and b in the class of quadratic estimates. In particular if $a=1/n$ and $b=-a/(n-1)$ then this expression reduces to

$$\Sigma(x_i-\bar{x})^2/(n-1)$$

as the best quadratic estimte of the variance.

Similarly, if nothing is known about the first k moments or their inter relations any symmetric function

$$\Sigma x_i^r x_j^s \ldots \qquad \ldots \quad (2.17)$$

where the summation is over $i, j,\ldots$ keeping the indices fixed is an *m.v.e.* in the class of $(r+s+\ldots)$th degree polynomials in n variables. Then the problem of estimation reduces to that of finding coefficients for the functions of the type (2.17) to make the final expression unbiased for a moment of order $(r+s+\ldots)$. From this it would follow that the expressions for sample semi-invariants as defined by Fisher are *m.v.e.*'s when nothing is known about the population values, it being assumed that the class of estimates is restricted to polynomials in the variables.

These results can also be deduced by expressing the conditions that the estimate of the $(r+s+\ldots)$th moment has zero covariance with functions of the type

$$x_1^{r+s+\ldots}-x_2^{r+s+\ldots}$$
$$x_1^r x_2^s \ldots - x_1^s x_2^r \ldots$$

all of which have zero expectation. Thus in the case of a quadratic estimate of variance

$$\Sigma\Sigma a_{ij} x_i x_j$$

the condition of zero covarince with expressions of the type $x_1^2-x_2^2,\ldots,\ x_1x_2-x_1x_3\ldots$ leads to an estimate of the form

$$a\Sigma x_i^2+b\Sigma x_i x_j$$

which is same as that considered in (2.16).

In illustrations 2 and 3 it is shown that uniformly *m.v.e.* can be obtained if the class of estimates is restricted to functions of certain types. This seems inevitable when the criterion of minimum variance fails due to the non-existence of uniformly *m.v.e.* The usefulness of this technique is also, however, limited to a special class of distribution laws which will be separately discussed in a forthcoming publication. An example is given here because of its long standing interest.

Consider the class of estimating functions $A(x_1, \ldots, x_n)$ which possess the property

$$A(x_1+\lambda, \ldots, x_n+\lambda) = \lambda + A(x_1, \ldots, x_n) \qquad \ldots \quad (2.18)$$

and the joint density of the form ($x_1, x_2, \ldots$ not necessarily independent)

$$P(x_1-\theta, \ldots, x_n-\theta)$$

where θ is the location parameter to be estimated. Defining $\xi_1 = x_1 - x_2, \ldots, \xi_n = x_{n-1} - x_n$, the conditional expectation of A can be written

$$E(A \mid \xi, \theta) = \theta + E(A \mid \xi, 0)$$

where $E(A \mid \xi, 0) = f_A(\xi)$ is a function of ξ only. Let B be any other statistic of the type (2.18) then the difference

$$B - A_c, \quad A_c = A - f_A(\xi)$$

is a function of ξ only. Consequently

$$E(B-\theta)^2 = E(A_c-\theta)^2 + E(B-A_c)^2$$

the product term being zero. Therefore the statistic A_c obtained by correcting any arbitrary function A of the type (2.18) has smaller expected square of the difference and is therefore a uniformly *m.v.e.* in the class of estimates assumed. Since A is arbitrary we can choose any single observation say x_1 or $\bar{x}$ in which case the estimate is

$$x_1 - E(x_1 \mid \xi, 0) \quad \text{or} \quad \bar{x} - E(\bar{x} \mid \xi, 0)$$

both leading to the same expression. This result is stated in a extremely interesting form by Pitman (1938) who gave the estimate as

$$\int_{-\infty}^{+\infty} \theta P(x_1-\theta, \ldots, x_n-\theta)\, d\theta \Big/ \int_{-\infty}^{+\infty} P(x_1-\theta, \ldots, x_n-\theta)\, d\theta.$$

3. The main theorem

Theorem 3: *If the parent distribution $p(x \mid \theta)$ is such that in independent samples of any size from it, uniformly m.v.e. can be constructed for any $\tau(\theta) \epsilon U$ then the distribution is of a special type known to admit sufficient statistics.*

This is the converse proposition of what has been proved before (Rao, 1945a, 1948) and also in illustration 1 of the last section. The assumption that every function tion $\tau(\theta) \epsilon U$ admits a uniformly *m.v.e.* appears to be very restricted and it may be possible to deduce the result from the existence of minimum variance estimates of a particular type. However, this proposition is of great interest and is proved under the assumptions made above. There are three stages in the proof.

Firstly if $p(x_1 \mid \theta) \ldots p(x_n \mid \theta)$ is the probability density of the observations then the statistic

$$\Sigma \frac{p'(x_i \mid \theta^0)}{p(x_i \mid \theta^0)} = \Sigma y_i \qquad \ldots \quad (3.1)$$

has minimum variance at θ^0 as an unbiased estimate of $E(\Sigma y_i)$ and following the arguments of corollary 1.5, we find under the assumptions made above that Σy_i is the uniformly *m.v.e.* of $E(\Sigma y_i)$. Transforming x_i to y_i we observe that the distribution of y is such that in independent samples of any size n, $\bar{y}$ is the uniformly *m.v.e.* of ϕ the expected value of y. If the variance of y is σ^2 then the variance of the *m.v.e.* is σ^2/n.

The second stage of the proof is just a survey of the properties of maximum likelihood estimates. If M_n is the maximum likelihood estimate of ϕ and is such that

$$E(nM_n^2)<\infty$$

then M_n is a consistent estimate of ϕ (Wald, 1949) and

$$E(M_n)=\phi+b_n(\phi),\ b_n(\phi)\to 0 \text{ as } n\to\infty,$$

and $nV(M_n)\to 1/i$ where i is the information on ϕ per single observation (Fisher, 1925). Also if C_n is any other consistent estimate of ϕ then

$$\frac{V(M_n)}{V(C_n)}\to \text{a limit} \leqslant 1 \qquad \ldots \quad (3.2)$$

In the third stage we use the assumption of the existence of *m.v.e.* estimates of all $\tau(\phi)\epsilon U$. Let B_n be the *m.v.e.* of $b_n(\phi)$ and $\bar{y}_n$ that of ϕ. Then $\bar{y}_n+B_n$ is the *m.v.e.* of $\phi+b_n(\phi)$. Therefore

$$V(M_n)\geqslant V(\bar{y}_n+B_n)\geqslant V(\bar{y}_n)+V(B_n)+2C(y_nB_n) \qquad \ldots \quad (3.3)$$

Let us suppose that

$$\frac{V(B_n)}{V(\bar{y}_n)}\to 0 \quad \text{or} \quad \frac{C(y_nB_n)}{V(\bar{y}_n)}\to 0 \qquad \ldots \quad (3.4)$$

in which case

$$\frac{V(M_n)}{V(\bar{y}_n)}\to \text{a limit} \geqslant 1 \qquad \ldots \quad (3.5)$$

which together with (3.2) implies that the limit is unity or $i\sigma^2=1$, a situation which is realised only when the parent distribution satisfies the condition

$$\frac{p'(y|\phi)}{p(y,\phi)}=\lambda(y-\phi)$$

and this proves the proposition. It may happen that none of the conditions (3.4) is true in which case let

$$\frac{V(B_n)}{V(\bar{y}_n)}\to\alpha^2 \text{ and } \frac{C(B_n\bar{y}_n)}{V(\bar{y}_n)}\to\rho\alpha,\ \rho\leqslant 1$$

Consider the statistic

$$\bar{y}_n-\frac{\rho}{\alpha}B_n$$

which is consistent and by using the result (3.2) we observe that

$$\frac{V(M_n)}{V\left(\bar{y}_n-\frac{\rho}{\alpha}B_n\right)}=\frac{V(M_n)/V(\bar{y}_n)}{V\left(\bar{y}_n-\frac{\rho}{\alpha}B_n\right)/V(\bar{y}_n)}\to \text{a limit} \leqslant 1 \qquad \ldots \quad (3.6)$$

$\{V(M_n)/V(\bar{y}_n)\}\to$ a limit $\geqslant 1+\alpha^2+2\rho\alpha$ because of (3.3). Also

$$\{V(\bar{y}_n-\frac{\rho}{\alpha}B_n)/V(\bar{y}_n)\}\to 1-\rho^2$$

Therefore the ratio in (3.6) tends to a limit not less than

$$\frac{1+\alpha^2+2\rho\alpha}{1-\rho^2}$$

which is always greater than unity. This is impossible unless $\rho=0$ and also $\alpha=0$. Hence the theorem is proved. This proof can be generalized to the case of many parameters.

4. Lower limits to variances

It is of some importance to examine, in view of the results of the previous sections, how far the expressions to the lower limits to the variances obtained by various authors are useful in discovering minimum variance estimates. The first in the series, established for any sample size, is the information limit to the variance (Cramer, 1946; Rao, 1945a). This limit is attained when the probability density of the observations is such that

$$\frac{P'(x|\theta)}{P(x|\theta)}=\lambda(\theta)\{T-\psi(\theta)\}$$

in which case T is the *m.v.e.* This result was also deduced by the method of calculus of variation by Aitken and Silverstone (1942). Obviously T is a sufficient statistic and consequently all powers of T and in general any function of T is *m.v.* estimate under very general conditions. The realization of the information limit and the existence of a sufficient statistic are synonimous in this sense.

Bhattacharya (1946) observed that there may exist estimates expressible in the form

$$T=\psi(\theta)+\lambda_1\frac{P'(x|\theta)}{P(x|\theta)}+\lambda_2\frac{P''(x|\theta)}{P(x|\theta)} \qquad \dots \quad (4.1)$$

Let us suppose that $x_1,\dots,x_n$ are independent, in which case defining

$$y_i=p'(x_i|\theta)/p(x_i|\theta) \text{ and } z_i=dy_i/d\theta$$

the expression (4.1) can be written in the form

$$T=\psi(\theta)+\lambda_1(\Sigma y_i)+\lambda_2\{(\Sigma y_i)^2+\Sigma z_i\} \qquad \dots \quad (4.2)$$

Differentiating continuously with respect to x_1 and x_2 we find

$$\frac{\partial^2 T}{\partial x_1 \partial x_2}=2\lambda_2\frac{dy_1}{dx_1}\frac{dy_2}{dx_2}$$

which is true for all x_1 and x_2. Choosing a particular value of x_2

$$\frac{dy_1}{dx_1}=A(\theta)B(x_1)$$

where A is a function of θ only and B of x only. This shows that the parent distribution $p(x|\theta)$ is of the form admitting a sufficient statistic. It would then follow that the expression (4.1) can utmost yield polynomials of the sufficient statistic and for this purpose it is unnecessary to go through the second and higher derivatives of the probability density because of the complete class theorem.

Bhattacharya (1950) also considered the possibility of statistics being of the form

$$T = \psi(\theta) + \lambda_1 \frac{P'(x,\theta)}{P(x,\theta)} + \lambda_2 \frac{\int_{\theta^*}^{\theta} P(x,\theta) C(\theta) d\theta}{P(x,\theta)} \qquad \dots \quad (4.3)$$

Multiplying both sides by $P(x,\theta)$, differentiating with respect to θ and then with respect to x_1, x_2 and also in the reverse order x_2, x_1 and subtracting one from the other the following result is obtained

$$\frac{\partial T}{\partial x_1} \frac{dy_2}{dx_2} = \frac{\partial T}{\partial x_2} \frac{dy_1}{dx_1}$$

This is true for all x_1 and x_2. Hence as before

$$\frac{dy_1}{dx_1} = A(\theta) B(x_1)$$

It would then follow that the expression (4.3) can only be a special function of the sufficient statistic.

REFERENCES

AITKEN, A. C. and H. SILVERSTONE (1942): On the estimation of statistical parameters. *Proc. Roy. Soc. Edin.*, 61, 186.

BHATTACHARYA, A. (1946, 47, 48): On some analogues of the amount of information and their use in statistical estimation. *Sankhyā*, 8, 1, 201, 315.

——— (1950): Unbiased estimates with minimum variance. *Proc. Roy. Soc. Edin.*, 53, 69.

CRAMÉR, H. (1946): *Mathematical methods of statistics*, Princeton University Press.

DAVIS, R. C. (1951): Note on uniformly best unbiased estimates. *Ann. Math. Stat.*, 22, 440.

FISHER, R. A. (1921): On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. A*, 222, 309.

HARDY, G. H. (1917): On Stieltje's 'probleme des moments. *Mess. Math.*, XLVI, 175.

HSU, P. L. (1938): On the best unbiased quadratic estimate of the variance. *Stat. Res. Mem.*, 2, 91

KOOPMAN, B. O. (1936): On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.*, 39, 399.

LEHMAN, E. L. and H. SCHEFFE (1950): Completeness, similar regions and unbaised estimation. *Sankhyā*, 10, 305.

PITMAN, E. J. G. (1938): The estimation of location and scale parameters of a continuous population of any given form. *Biometrika*, 30, 391.

RAO, C. R. (1945b): Generalisation of Markoff's theorem and tests of linear hypotheses. *Sankhyā*, 7, 9.

——— (1945a): Information and accuracy attainable in the estimation of statistical parameters *Bull. Cal. Math. Soc.*, 37, 81.

——— (1948): Sufficient statistics and minimum variance estimates. *Proc. Camb. Phil. Soc.*, 45, 213.

STEIN CHARLES (1950): Unbiased estimates with minimum variance. *Ann. Math. Stat.*, 21, 406.

WALD, A. (1949): Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, 20, 595.