SUFFICIENT STATISTICS AND MINIMUM VARIANCE ESTIMATES

By C. RADHAKRISHNA RAO

Communicated by F. J. Anscombe

Received 12 March 1948

Definitions. Let the probability density of observations be denoted by $\phi(x \mid \theta)$, where* x stands for the variables and θ for the parameters. A function t of the observations is called an unbiased estimate of the function $\psi(\theta)$ of the parameters if

$$\int t\phi\, dx = \psi(\theta), \tag{1}$$

where dx stands for the product of differentials.

An unbiased estimate is said to be a *minimum variance estimate* if its variance is not less than that of any other unbiased estimate. This must be distinguished from the concept of efficiency as introduced by Fisher (4). Efficient statistics may be defined as those which, in large samples, contain almost the whole of the information concerning the parameters. When the class of consistent estimates which are normally distributed in large samples are considered, the information supplied by them can be measured by the reciprocal of their variances. The efficiency of a statistic, in the case of a single parameter, can be measured by the limit of I/v, where I is the information in the sample and v the variance of the statistic. The fully efficient statistics are those for which this ratio tends to unity. The object of the present article is to study some exact properties of minimum variance estimates in samples of a given size. These estimates will be, however, efficient in large samples provided their limiting distributions are normal.

A set of statistics $T_1, \ldots, T_m$ is called a *minimal sufficient set of statistics* if m is the smallest number for which

$$\phi(x \mid \theta) = \phi_1(T_1, \ldots, T_m \mid \theta)\, \phi_2(x \mid T),$$

where $\phi_1(T_1, \ldots, T_m \mid \theta)$ is the probability density of $T_1, \ldots, T_m$ and $\phi_2(x \mid T)$, the probability density of the sample given the statistics $T_1, \ldots, T_m$, is independent of the θ's. The integer m can be greater than, equal to or less than the number of parameters.

The author has shown elsewhere (8) that minimum variance estimates are explicit functions of the minimal sufficient set of statistics. For the derivation of such estimates, we need only consider the distribution of the sufficient statistics.

Mininum possible variance (*single parameter*). In the case of a single parameter it

* It is not necessary that the variables should be continuous. The integral sign can be replaced by the summation sign in the case of discrete variables.

https://doi.org/10.1017/S0305004100024737 Published online by Cambridge University Press

has been shown (3, 7) that the minimum variance attainable for an unbiased estimate of $\psi(\theta)$ is, under some conditions, not smaller than

$$\left(\frac{d\psi}{d\theta}\right)^2 \Big/ I, \quad I = E\left(\frac{1}{\phi}\frac{d\phi}{d\theta}\right)^2. \tag{2}$$

It may be noted that the quantity defined in (2) is not the *minimum attainable* for the estimation of any parametric function in any distribution. The estimate with the minimum attainable variance will be the minimum variance estimate. An estimate of $\psi(\theta)$ with variance $(d\psi/d\theta)^2/I$ will be a minimum variance estimate, but this is not a necessary condition.

Example 1. Consider n independent observations from the normal distribution

$$\frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\, dx. \tag{3}$$

Let μ^2 be the parameter to be estimated. Since $\bar{x}$, the mean, is a sufficient statistic for μ^2 (in fact for any function of μ), we need only consider functions of $\bar{x}$.

If σ is known, an unbiased estimate of μ^2 is

$$X = \bar{x}^2 - \sigma^2/n, \tag{4}$$

and

$$V(X) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^2}{n^2}.$$

This is greater than

$$\left(\frac{d\mu^2}{d\mu}\right)^2 \Big/ I = \frac{4\mu^2\sigma^2}{n}.$$

Let us consider an alternative statistic unbiased for μ^2 but which differs from X by $f(\bar{x})$. Then

$$E\{f(\bar{x})\} = \frac{\sqrt{n}}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{+\infty} f(\bar{x}) \exp\left\{-\frac{n}{2}\left(\frac{\bar{x}-\mu}{\sigma}\right)^2\right\} d\bar{x}.$$

Differentiating twice with respect to μ we find

$$\int_{-\infty}^{+\infty} \bar{x}^2 f(\bar{x}) \exp\left\{-\frac{n}{2}\left(\frac{\bar{x}-\mu}{\sigma}\right)^2\right\} d\bar{x} = 0,$$

which shows that the covariance between $\bar{x}^2$ and $f(\bar{x})$ is zero. The variance of the alternative statistic is

$$V\{X+f(\bar{x})\} = V(X) + V\{f(\bar{x})\} + 2\operatorname{cov}\{X f(\bar{x})\} = V(X) + V\{f(\bar{x})\} \geqslant V(X),$$

which shows that X as defined in (4) is the minimum variance estimate, although its variance is greater than the quantity defined in (2).

Example 2. Consider n independent observations from the population

$$\frac{\alpha^p}{\Gamma(p)} e^{-\alpha x} x^{p-1}\, dx.$$

If p is known, $\bar{x}$ is sufficient for α. Since

$$E(\bar{x}) = p/\alpha, \quad V(\bar{x}) = p/(n\alpha^2) = \left\{\frac{d(p/\alpha)}{d\alpha}\right\}^2 \Big/ I \quad \left(I = \frac{pn}{\alpha^2}\right),$$

it follows that $\bar{x}$ is the minimum variance estimate of p/α. Let us consider α itself as

https://doi.org/10.1017/S0305004100024737 Published online by Cambridge University Press

the parameter to be estimated. The statistic $X = (np-1)/\bar{x}$ is unbiased for α and its variance is

$$V\left(\frac{np-1}{\bar{x}}\right) = \frac{\alpha^2}{np-2},$$

which is greater than $1/I = \alpha^2/np$. Let an alternative statistic unbiased for α differ from X by $f(\bar{x})$. Then

$$\int_0^\infty f(\bar{x})\, e^{-\alpha n\bar{x}} \bar{x}^{np-1}\, d\bar{x} = 0 \quad \text{for all } \alpha > 0.$$

Successive differentiations with respect to α give

$$\int_0^\infty \bar{x}^k f(\bar{x})\, e^{-\alpha n\bar{x}} \bar{x}^{np-1}\, d\bar{x} = 0 \quad (k = 1, 2, \ldots),$$

which means that

$$\int_0^\infty e^{it\bar{x}} f(\bar{x})\, e^{-\alpha n\bar{x}} \bar{x}^{np-1}\, d\bar{x} = 0,$$

where $i = \sqrt{(-1)}$ and t is real. Then from the well-known inversion theorem it follows that $f(\bar{x})$ is zero everywhere if it is continuous. This shows that there exists no other function of $\bar{x}$ which has the expectation α. Therefore $(np-1)/\bar{x}$ is the minimum variance estimate of α. It is incorrect to say that this statistic is not efficient since it has not attained some lower bound derived for any general distribution. In fact Bhattacharya (2) has shown that more stringent lower bounds than the one considered here could be derived for any general distribution. It is not proper, in small samples, to judge the efficiency of an estimate by comparing its variance with some lower limit as the one considered here.

Properties of distributions admitting sufficient statistics (single parameter). In the case of a single parameter it has been shown by Fisher (5) and Koopman (6) that the frequency function $\phi(x \mid \theta)$ admitting a sufficient statistic can be expressed in the form

$$\phi(x \mid \theta) = \exp(\Theta_1 X_1 + \Theta_2 + X_2),$$

where Θ_1 and Θ_2 are functions of θ only and X_1 and X_2 are functions of the observations only.

Since

$$\int \exp(\Theta_1 X_1 + \Theta_2 + X_2)\, dx = 1,$$

we get, on differentiating with respect to Θ_1,

$$E(X_1) = -\frac{d\Theta_2}{d\Theta_1}, \quad V(X_1) = -\frac{d^2\Theta_2}{d\Theta_1^2}.$$

If we choose $-d\Theta_2/d\Theta_1$ as the parameter to be estimated, then the lower bound of the variance as defined in (2) is

$$-\left\{-\frac{d}{d\theta}\frac{d\Theta_2}{d\Theta_1}\right\}^2 \Big/ \left\{\left(\frac{d\Theta_1}{d\theta}\right)^2 \frac{d^2\Theta_2}{d\Theta_1^2}\right\} = -\frac{d^2\Theta_2}{d\Theta_1^2} = V(X_1).$$

Hence X_1 as an estimate of $-d\Theta_2/d\Theta_1$ has the minimum variance defined in (2). The following theorem is established.

THEOREM 1. *If (a) a sufficient statistic exists so that the probability density $\phi(x \mid \theta)$ can*

https://doi.org/10.1017/S0305004100024737 Published online by Cambridge University Press

be expressed in the form $\exp(\Theta_1 X_1 + \Theta_2 + X_2)$, *and* (*b*) *the differentiation under the integral sign is possible yielding the results*

$$E(X_1) = -d\Theta_2/d\Theta_1 \quad \text{and} \quad V(X_1) = -d^2\Theta_2/d\Theta_1^2,^*$$

then the function $-d\Theta_2/d\Theta_1$ *admits an estimate whose variance is exactly equal to the lower bound defined in* (2).

Some results similar to that in Theorem 1 are obtained by Aitken and Silverstone (1). If the $\phi(x \mid \theta)$ is of the form

$$\exp(\Theta_1 X_1 + \Theta_2 + X_2),$$

then X_1 is a sufficient statistic for θ. On integrating out over constant values of X_1, the distribution of X_1 may be written as

$$\exp(\Theta_1 X_1 + \Theta + X)\, dX_1,$$

where X is a function of X_1 only. In this case we need only consider functions of X_1 to search for minimum variance estimates.

Let $F(X_1)$ be any function such that $E\{F(X_1)\} = \chi(\theta)$. If an alternative function, differing from $F(X_1)$ by a continuous function $f(X_1)$ of X_1, has the same expectation $\chi(\theta)$, then

$$\int f(X_1) \exp(\Theta_1 X_1 + \Theta + X)\, dX_1 \equiv 0. \tag{5}$$

Successive differentiations with respect to Θ_1 give

$$\int X_1^k f(X_1) \exp(\Theta_1 X_1 + \Theta + X)\, dX_1 = 0 \quad (k = 1, 2, \ldots), \tag{6}$$

which means that

$$\int e^{uX_1} f(X_1) \exp(\Theta_1 X_1 + \Theta + X)\, dX_1 = 0 \tag{7}$$

or $f(X_1)$ is identically zero. The result (7) depends on the validity of successive differentiations involved in the derivation of the equations (6). Some sufficient conditions under which this holds are given below.

THEOREM 2. *If the range of integration is finite and* $\exp(\Theta_1 X_1 + \Theta + X)$ *is a continuous function of both the variables* X_1 *in the range of integration and* Θ_1 *in some continuous interval, then successive differentiation under the integral sign is valid.*

The theorem follows from the fact that under the conditions made above

$$X_1^k f(X_1) \exp(\Theta_1 X_1 + \Theta + X)$$

is a continuous function of both the variables.

THEOREM 3. *If* $\exp(\Theta_1 X_1 + \Theta + X)$ *is a continuous function of both the variables* X_1 *in* (a, ∞) *where* a *is finite and* Θ_1 *in some closed interval* (Θ_1', Θ_1'') *and*

$$\int_a^\infty f(X_1) \exp(\Theta_1 X_1 + \Theta + X)\, dX_1$$

converges uniformly, then successive differentiation under the integral sign is valid for some range of Θ_1.

* This is ensured if the ranges of integration do not involve the parameter θ, and the analytical conditions required for differentiation under the integral sign are satisfied by the integrand.

https://doi.org/10.1017/S0305004100024737 Published online by Cambridge University Press

Consider

$$\int_a^\infty X_1^k f(X_1) \exp(\Theta_1 X_1 + \Theta + X)\, dX_1 = \int_a^\infty X_1^k e^{-\delta X_1} f(X_1) \exp[(\Theta_1 + \delta) X_1 + \Theta + X]\, dX_1, \tag{8}$$

where δ is a positive quantity less than $|\Theta_1' - \Theta_1''|$. Since

$$\int_a^\infty f(X_1) \exp[(\Theta_1 + \delta) X_1 + \Theta + X]\, dX_1$$

is uniformly convergent in the closed interval $(\Theta_1', \Theta_1'' - \delta)$ and $X_1^k e^{-\delta X_1}$ is bounded and continuous in (a, ∞), it follows that the integral in (8) is uniformly convergent in $(\Theta_1', \Theta_1'' - \delta)$. Since δ can be chosen as small as possible the integral in (8) is uniformly convergent in the closed interval (Θ_1', θ_1'') where $\theta_1'' < \Theta_1''$, in which case the differentiation is valid for the range (Θ_1', θ_1'').

Similar results can be established even when the range of integration is from $-\infty$ to $+\infty$.

Under the conditions* assumed above it follows that every function of X_1 is unique as an unbiased estimate of its expected value, so that every function of X_1 is a minimum variance estimate of its expected value. The theorems given above supply a general proof of the properties observed in Examples 1 and 2.

The case of several parameters. Koopman (6) has shown that under some general conditions the distribution function $\phi(x_1, \ldots, x_n \mid \theta_1, \ldots, \theta_q)$ admitting a set of sufficient statistics $T_1, \ldots, T_q$ can be expressed in the form

$$\phi = \exp(\Theta_1 X_1 + \ldots + \Theta_q X_q + \Theta + X), \tag{9}$$

where the Θ's are functions of θ only and $X_1, \ldots, X_q$ of the T's only and X of the observations only. The author has derived elsewhere (7) lower bounds to the variances similar to that defined in (2) in the case of several parameters and shown that for distributions of the form (9) there exist functions of parameters for which estimates with the minimum variances so defined could be constructed. By an extension of the analysis developed in the case of a single variable it can be shown that every function of $T_1, \ldots, T_q$ is unique as the estimate of its expected value and hence the best possible. The discussion is valid subject to some analytical conditions required for differentiation under the integral sign, etc.

The method of maximum likelihood. The method of maximum likelihood yields estimates which are functions of the minimal set of sufficient statistics. This satisfies the necessary condition for an estimate to have the minimum attainable variance. When the number of sufficient statistics is equal to the number of parameters to be estimated the maximum likelihood estimates will be, as shown above under some conditions, the minimum variance estimates of their expected values even in small samples. In the general case, it has been formally demonstrated by the author (8) that the maximum likelihood estimates are fully efficient in large samples.

I wish to thank Dr H. E. Daniels for drawing my attention to this problem.

* The result is true under more general conditions discussed by Widder (9).

https://doi.org/10.1017/S0305004100024737 Published online by Cambridge University Press

REFERENCES

(1) AITKEN, A. C. and SILVERSTONE, H. *Proc. Roy. Soc. Edinburgh*, A, 61 (1942), 186.
(2) BHATTACHARYA, A. *Sankhya*, 8 (1947), 1.
(3) CRAMÉR, H. *Mathematical methods of statistics* (Uppsala, 1945) (reprinted in English, Princeton Univ. Press).
(4) FISHER, R. A. *Philos. Trans.* A, 222 (1921), 309.
(5) FISHER, R. A. *Proc. Roy. Soc.* A, 144 (1934), 285.
(6) KOOPMAN, B. O. *Trans. American Math. Soc.* 39 (1936), 399.
(7) RAO, C. R. *Bull. Calcutta Math. Soc.* 37 (1945), 81.
(8) RAO, C. R. *Proc. Cambridge Phil. Soc.* 43 (1947), 280.
(9) WIDDER, D. V. *The Laplace transform.* (Princeton, 1941).

KING'S COLLEGE
CAMBRIDGE

https://doi.org/10.1017/S0305004100024737 Published online by Cambridge University Press