NOTE ON UNIFORMLY BEST UNBIASED ESTIMATES[1]

By R. C. Davis

Pasadena, California

1. Summary and introduction. Bhattacharyya [1] has considered recently the following problem in statistical estimation. Let $X_1, X_2, \cdots, X_n$ be n stochastic variables distributed according to the probability law $f(x_1, x_2, \cdots, x_n; \theta) dx_1 dx_2 \cdots dx_n$, where θ is the unknown parameter. Consider the class of all functions $T(x_1, x_2, \cdots, x_n)$ of the stochastic variables such that the expectation of each function in this class is equal to a preassigned function $\tau(\theta)$. Usually $\tau(\theta)$ admits of more than one unbiased estimate, and the problem posed by various authors is to obtain a lower bound of the variances of all such estimates, this lower bound to be independent of the estimates themselves but depending on $\tau(\theta)$ and the distribution function of the n stochastic variables. Under certain regularity conditions Bhattacharyya obtained a lower bound of the above type which is never lower than the one obtained earlier independently by Cramér [2] and Rao [3], although the conditions assumed by Bhattacharyya are more restrictive than those assumed by the latter authors. Recently E. W. Barankin in a remarkable paper [4] has developed a procedure which yields the class of lower bounds of unbiased estimates having minimum sth absolute central moment ($s > 1$) at a preassigned parameter value θ_0. In this note we are concerned with the attainment of a lower bound obtained first by Bhattacharyya. Bhattacharyya discusses the case in which his lower bound is attained and derives some interesting properties of the distribution of such a statistic (which might be called a generalized efficient statistic).

The purpose of this note is to prove that in the case in which the variables $X_1, X_2, \cdots, X_n$ are independently and identically distributed with a common distribution function $F(x; \theta)$ depending upon a single unknown parameter, one obtains the following result: under the regularity conditions assumed by Bhattacharyya in which the parameter θ may assume values in an interval of the real axis, and with an additional slight restriction on the cumulative distribution function $F(x; \theta)$, no generalized efficient statistic exists which is constructed by use of both the first and second derivatives of the likelihood function with respect to the parameter. It follows that if an efficient estimate (in the sense originally defined by Fisher [5]) for the single unknown parameter does not exist, then no distribution $F(x; \theta)$ exists possessing a uniformly minimum variance unbiased estimate of $\tau(\theta)$ which is constructed by using a linear combination of the first and second partial logarithmic derivatives of the likelihood function. This result for the case involving a single unknown parameter is peculiarly of interest in view of the fact that Seth [6] has given an example

[1] Presented at the meeting of the Institute of Mathematical Statistics on 18 March 1950 at the University of North Carolina.

in which the above construction is possible if the distribution involves two unknown parameters.

2. A theorem. Let X be a chance variable possessing an absolutely continuous probability distribution $F(x; \theta)$ in which θ is the single unknown parameter. Denote by $f(x; \theta)$ the probability density function of X, this function existing almost everywhere. Consider a finite sequence $\{X_i\}$, $i = 1, 2, \cdots, n$, of independently distributed chance variables possessing the common distribution function $F(x; \theta)$. We restrict ourselves in this note to unbiased estimates of $\tau(\theta)$ which are functions $T_n(x_1, x_2, \cdots, x_n)$, where x_i is a random observation of X_i. Denote also by $L(x_1, x_2, \cdots, x_n; \theta)$ the likelihood function of $X_1, X_2, \cdots, X_n$, so that in the case considered in this note

$$L(x_1, x_2, \cdots, x_n; \theta) = f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta).$$

We denote by E_n n-dimensional Euclidean space.

Following Bhattacharyya we make the following assumptions concerning $F(x; \theta)$:

ASSUMPTION A. *X assumes values x in E_1 and the true parameter value θ lies in an interval $I \subset E_1$.*

ASSUMPTION B. *$F(x; \theta)$ is absolutely continuous in x.*

ASSUMPTION C. $\frac{\partial L}{\partial \theta}, \frac{\partial^2 L}{\partial \theta^2}, \frac{\partial^3 L}{\partial \theta^3}$ *exist almost everywhere in E_n and for every $\theta \varepsilon I$.*

ASSUMPTION D. $\frac{\partial L}{\partial \theta}$ *and* $\frac{\partial^2 L}{\partial \theta^2}$ *are linearly independent for almost all points $x_1, x_2, \cdots, x_n$ in E_n.*

ASSUMPTION E. $\left| \frac{\partial^i L}{\partial \theta^i} \right| < G_i(x_1, x_2, \cdots, x_n)$, $i = 1, 2$, *for all $\theta \varepsilon I$, where $G_i(x_1, x_2, \cdots, x_n)$ is integrable with respect to F over $(-\infty, \infty)$.*

ASSUMPTION F. $\frac{d\tau}{d\theta}$ *and* $\frac{d^2\tau}{d\theta^2}$ *exist for all $\theta \varepsilon I$.*

ASSUMPTION G. $J_{ij} = E\left[\frac{1}{L}\frac{\partial^i L}{\partial \theta^i} \cdot \frac{1}{L}\frac{\partial^j L}{\partial \theta^j}\right]$ *exists for each i, j ($i = 1, 2; j = 1, 2$).*

In this note we make an additional assumption concerning the density function $F(x; \theta)$.

ASSUMPTION H. *There exists a closed inteval Δ ($\Delta \subset E_1$) such that, for $\theta \varepsilon I$ and $x \varepsilon \Delta$, $f(x, \theta) > 0$ and is continuous in x. Moreover, $\frac{\partial f}{\partial \theta}(x; \theta) \not\equiv 0$.*

If we denote by (J^{ij}) the matrix inverse to (J_{ij}), Bhattacharyya deduces the following inequality for the variance, $V(T_n)$, of any unbiased estimate $T_n(x_1, x_2, \cdots, x_n)$ of $\tau(\theta)$:

$$V(T_n) \geq \sum_{i=1}^{2} \sum_{j=1}^{2} \tau^i \tau^j J^{ij}, \tag{1}$$

where

$$\tau^i = \frac{d^i \tau(\theta)}{d\theta^i}.$$

Equation (1) can become an equality if and only if the following equation holds for almost all $x_1, x_2, \cdots, x_n$ in E_n:

$$T_n - \tau = \sum_{i=1}^{2} \lambda_i^0 \frac{1}{L} \frac{\partial^i L}{\partial \theta^i}, \tag{2}$$

where

$$\lambda_i^0 = \sum_{j=1}^{2} \tau^j J^{ij}.$$

When equation (2) holds, it is clear that the statistic T_n becomes a generalization of Fisher's efficient statistic. In view of the desirable property of minimizing the variance among the class of unbiased estimates of $\tau(\theta)$, it is clearly worth while to attempt to characterize the distributions for which equation (2) is valid. We prove the following theorem:

THEOREM 1. *In the absence of an efficient statistic for $\tau(\theta)$ there exists no cumulative distribution function $F(x; \theta)$ satisfying Assumptions* (A)–(H) *which yields for any sample size, n, a statistic $T_n(x_1, x_2, \cdots, x_n)$ satisfying equation* (2) *for all $\theta \in I$ and almost all $(x_1, x_2, \cdots, x_n)$ in E_n.*

PROOF. We prove the theorem by showing that equation (2) leads to an impossibility if $F(x; \theta)$ satisfies Assumptions A–H. First we transform equation (2) into a partial differential equation in $\ln L(x_1, x_2, \cdots, x_n)$. To simplify the presentation we give the proof for the case $\tau(\theta) = \theta$, but we lose no generality in doing so. Equation (2) then assumes the form

$$T_n - \theta = \lambda_1^0 \frac{\partial}{\partial \theta} \ln L + \lambda_2^0 \left[\frac{\partial^2}{\partial \theta^2} \ln L + \left(\frac{\partial}{\partial \theta} \ln L \right)^2 \right], \tag{3}$$

in which

$$\lambda_1^0 = j^{11} = \frac{j_{22} + 2(n-1) j_{11}^2}{n[j_{11} j_{22} - j_{12}^2 + 2(n-1) j_{11}^3]},$$

$$\lambda_2^0 = j^{21} = \frac{j_{12}}{n[j_{11} j_{22} - j_{12}^2 + 2(n-1) j_{11}^3]},$$

and

$$j_{ik} = E\left[\frac{1}{f} \frac{\partial^i f}{\partial \theta^i} \cdot \frac{1}{f} \frac{\partial^k f}{\partial \theta^k} \right].$$

As stated already, equation (3) need not be valid for a set of points $(x_1, x_2, \cdots, x_n)$ having measure zero. Denote by Δ^n the closed cube in E_n defined by $x_i \in \Delta$, where Δ is the closed interval defined in Assumption H. It is clear from Assumption H that there exist points in Δ^n for which equation (3) is valid. We choose one such point $(x_1, x_2, \cdots, x_n)$ and consider this as a fixed point in Δ^n for the subsequent analysis. For each θ in the interval I, we denote by $x(\theta)$ the value of x given by the transformation

$$\ln f(x;\theta) = \frac{1}{n}\ln L(x_1, x_2, \cdots, x_n; \theta). \tag{4}$$

Since $\frac{\partial}{\partial\theta}\ln L$ and $\frac{\partial^2}{\partial\theta^2}\ln L$ exist almost everywhere in E_n by Assumption C, it is clear that $\frac{d}{d\theta}\ln f(x(\theta);\theta)$ and $\frac{d^2}{d\theta^2}\ln f(x(\theta);\theta)$ exist. We choose the fixed point $(x_1, x_2, \cdots, x_n)$ so that the above derivatives exist and also so that equation (3) is valid. For each $\theta \varepsilon I$, the following equation is valid:

$$T_n - \theta = \lambda_1^0 n \frac{d}{d\theta}\ln f(x;\theta) + \lambda_2^0\left[n\frac{d^2}{d\theta^2}\ln f(x;\theta) + n^2\left(\frac{d}{d\theta}\ln f(x;\theta)\right)^2\right]. \tag{5}$$

Substituting the values of λ_1^0 and λ_2^0 in (5) and simplifying, we obtain the expression

$$T_n - \theta = \frac{(j_{22} - 2j_{11}^2)\frac{d}{d\theta}\ln f(x;\theta) + j_{12}\frac{d^2}{d\theta^2}\ln f(x;\theta) + n\left[2j_{11}^2\frac{d}{d\theta}\ln f(x;\theta) + j_{12}\left(\frac{d}{d\theta}\ln f(x;\theta)\right)^2\right]}{j_{11}j_{22} - j_{12}^2 - 2j_{11}^3 + 2j_{11}^3 n}. \tag{6}$$

To simplify matters we write this simply as

$$T_n - \theta = \frac{a(x(\theta),\theta) + nb(x(\theta),\theta)}{c(\theta) + ne(\theta)}. \tag{7}$$

Since this equation is valid for every $\theta \varepsilon I$, we can differentiate both sides of the equation with respect to θ. We obtain a quadratic polynomial in n, which we write as

$$\alpha n^2 + \beta n + \gamma = 0, \tag{8}$$

in which

$$\alpha = e^2\frac{d}{d\theta}\left(\frac{b}{e} + \theta\right),$$

$$\beta = c^2e^2\left[\frac{d}{d\theta}\left(\frac{a}{e} + \frac{b}{c}\right) + \frac{2}{ce}\right],$$

$$\gamma = c^2\frac{d}{d\theta}\left[\frac{a}{c} + \theta\right].$$

The two roots of (8) are given by

$$n = \frac{-\beta \pm\sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}.$$

Since by assumption at least one of these roots is a positive integer, we obtain the relations

$$\beta = -2\alpha N_1,$$
$$\sqrt{\beta^2 - 4\alpha\gamma} = 2\alpha N_2,$$

where N_1 and N_2 are integers. From these two relationships, we can deduce that $\gamma = (N_1^2 - N_2^2)\alpha$.

Referring to the definition of α and γ and performing two quadratures, we obtain the equations

$$\frac{1}{\dot{j}_{11}} \frac{d}{d\theta} \ln f(x; \theta) + \frac{\dot{j}_{12}}{2\dot{j}_{11}^3} \left[\frac{d}{d\theta} \ln f(x; \theta)\right]^2 = T_n - \theta, \tag{9}$$

$$\left[\frac{\dot{j}_{22} - 2\dot{j}_{11}^2}{\dot{j}_{11}\dot{j}_{22} - \dot{j}_{12}^2 - 2\dot{j}_{11}^3}\right] \frac{d}{d\theta} \ln f(x; \theta) + \left[\frac{\dot{j}_{12}}{\dot{j}_{11}\dot{j}_{22} - \dot{j}_{12}^2 - 2\dot{j}_{11}^3}\right] \frac{d^2}{d\theta^2} \ln f(x; \theta) = T_n - \theta. \tag{10}$$

The solution of the quadratic equation in $\frac{d}{d\theta} \ln f(x; \theta)$ in (9) yields

$$\frac{d}{d\theta} \ln f(x; \theta) = M(\theta) + \sqrt{N(\theta) + Q(\theta)[T_n - \theta]}, \tag{11}$$

and the integration of the first order differential equation in $\frac{d}{d\theta} \ln f(x; \theta)$ given in (10) yields

$$\frac{d}{d\theta} \ln f(x; \theta) = G(\theta)T_n + H(\theta) + R(\theta). \tag{12}$$

It is clear from inspection of equations (9) through (12) that the solutions to (9) and (10) are identical if and only if $\dot{j}_{12} \equiv 0$. Since $\dot{j}_{12}$ is proportional to λ_2^0 in (3), the vanishing of $\dot{j}_{12}$ implies that the statistic $T_n(x_1, x_2, \cdots, x_n)$ is formed only from the first partial derivative of the likelihood function and hence is an efficient statistic. This is contrary to the assumption of the theorem. The possibility that each side of equations (9) and (10) vanish identically is ruled out by the part of Assumption H in which it is stated that $\frac{\partial}{\partial\theta} \ln f(x; \theta) \not\equiv 0$ for $x \in \Delta$. Hence our assumption that equation (2) holds leads to a contradiction of the assumptions of the theorem. We conclude that there exists no cumulative probability distribution function (satisfying the assumptions of the theorem) which yields a generalized efficient statistic for any sample size n.

REFERENCES

[1] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946), pp. 1–14.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946, chap. 32.
[3] C. R. RAO, "Information and accuracy obtainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), p. 81.
[4] E. W. BARANKIN, "Locally best unbiased estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 477–501.
[5] R. A. FISHER, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, Vol. 22 (1925), pp. 700–725.
[6] G. R. SETH, "On the variance of estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 1–27.