ON THE VARIANCE OF ESTIMATES

By G. R. Seth

Columbia University

Summary. In this paper recent results on the lower bound to the variance of unbiased estimates have been brought together. Some of them have been extended to sequential estimates and the others have been improved to some extent. In the last section a general method for generating a system of orthogonal polynomials with respect to a certain class of weight functions is obtained together with a result on the conditions under which the class of unbiased estimates formed by all functions of an unbiased estimate consists of just one element.

1. Introduction.

§1.1. Let $X_1, X_2 \cdots$ be a sequence of chance variables whose distribution depends upon an unknown parameter θ and possibly also a finite number of other parameters. It is assumed that either all the X's are absolutely continuous or that they are all discrete. Let $p_M(x_1, x_2, \cdots, x_M; \theta)$ denote the joint probability density function or the probability of $(X_1, \cdots, X_M)$ according as the X's are continuous or discrete. Let $\theta^*(x_1, x_2, \cdots, x_n)$ be an unbiased estimate of θ, where $x_1, x_2, \cdots, x_n$ is a sequence of observations on $X_1, X_2, \cdots, X_n$.

In this paper, we shall make use of the following short forms and abbreviations:

$E(X)$ will represent the expectation of X.

$\sigma^2(X)$ will represent the variance of X.

$E(y \mid x)$ will represent the conditional expectation of y, given x.

θ^* will represent an abbreviation of $\theta^*(x_1, x_2, \cdots, x_n)$.

f will represent an abbreviation of $f(x; \theta)$ or $f(x; \theta_1, \theta_2, \cdots, \theta_T)$.

p_n will represent an abbreviation of $p_n(x_1, x_2, \cdots, x_n; \theta)$ or $p_n(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_T)$.

p_N will represent p_n for a fixed size sample, i.e., $n = N$.

g will represent an abbreviation of $g(\theta^*; \theta)$ or $g(\theta_1^*, \theta_2^*, \cdots, \theta_T^*; \theta_1, \theta_2, \cdots, \theta_T)$.

h will represent $h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta)$ or $h(\xi_1, \xi_2, \cdots, \xi_{N-T} \mid \theta_1^*, \theta_2^*, \cdots, \theta_T^*; \theta_1, \theta_2, \cdots \theta_T)$.

$\phi_{i_1, i_2, \cdots, i_T}(n)$ will represent $\dfrac{1}{p_n} \cdot \dfrac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\,\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}}\, p_n$.

$g_{i_1, i_2, \cdots, i_T}$ will represent $\dfrac{1}{g}\, \dfrac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\,\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}}\, g$.

$h_{i_1, i_2, \cdots, i_T}$ will represent $\dfrac{1}{h}\, \dfrac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\,\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}}\, h$.

In case differentiations with respect to one parameter are involved, the last three abbreviations will be shortened to $\phi_i(n)$, g_i and h_i respectively.

In §1.1, n is assumed to be a constant equal to N, that is, the sequence of chance variables is finite and fixed, consisting of $X_1, X_2, X_3, \cdots, X_N$.

Cramer [1] and Rao [2] have shown that under certain conditions of regularity, the variance of $\theta^*(x_1, x_2, \cdots, x_N)$ satisfies the inequality:

$$\sigma^2 \theta^*(x_1, x_2, \cdots, x_N) \geq \frac{1}{E\left(\frac{1}{p_N}\frac{\partial p_N}{\partial \theta}\right)^2}. \tag{1.1.1}$$

Cramér [1] has shown that the lower bound for the variance of $\theta^*(x_1, x_2, \cdots, x_N)$ given by (1.1.1) is achieved if and only if:

(1.1.2). There exists a sufficient statistic for estimating θ.

(1.1.3). The probability distribution $g(\theta^*; \theta)$ of the sufficient statistic $\theta^*(x_1, x_2, \cdots, x_N)$ is of the form

$$\theta^*(x_1, x_2, \cdots, x_N) - \theta = \frac{K}{g(\theta^*; \theta)} \frac{\partial}{\partial \theta} g(\theta^*; \theta), \quad \text{whenever} \quad g(\theta^*; \theta) > 0,$$

where K depends only upon N and the parameters in the distribution.

Cramer calls the statistic $\theta^*(x_1, x_2, \cdots, x_N)$ satisfying (1.1.2) and (1.1.3) an "efficient" statistic estimating θ and we will use the word "efficient" in this sense alone. Bhattacharyya [3] has shown that there exists a lower bound to the variance of $\theta^*(x_1, x_2, \cdots, x_N)$ which is higher than or equal to the one given in (1.1.1). This lower bound is ${}_{(m)}\lambda^{11}$, that is,

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_N)) \geq {}_{(m)}\lambda^{11} \tag{1.1.4}$$

where

$$\| {}_{(m)}\lambda^{ij} \| = \| \lambda_{ij} \|^{-1},$$

and

$$\lambda_{ij} = E\left(\frac{1}{p_N^2} \frac{\partial^i p_N}{\partial \theta^i} \frac{\partial^j p_N}{\partial \theta^j}\right), \qquad i, j = 1, 2, \cdots, m, \tag{1.1.5}$$

where m is any positive integer.

Let θ consist of T components $\theta_1, \theta_2, \cdots, \theta_T$, and $p_N(x_1, x_2, \cdots, x_N; \theta_T)$ be the same as $\prod_{i=1}^{N} f(x_i; \theta_1, \theta_2, \cdots, \theta_T)$. Further let $\theta_1^*(x_1, x_2, \cdots, x_N)$, $\theta_2^*(x_1, x_2, \cdots, x_N), \cdots, \theta_T^*(x_1, x_2, \cdots, x_N)$ be unbiased estimates of $\theta_1, \theta_2, \cdots, \theta_T$ respectively, with the non-singular covariance matrix $\| V_{ij} \|$ $(i, j = 1, 2, \cdots, T)$. Cramér [4] has proved that under certain regularity conditions, the ellipsoid

$$\sum_{i,j=1}^{T} V^{ij} t_i t_j = T + 2 \tag{1.1.6}$$

contains within itself the ellipsoid

$$\sum_{i,j=1}^{T} I_{ij} t_i t_j = T + 2, \tag{1.1.7}$$

where

$$\| V^{ij} \|^{-1} = \| V_{ij} \|, \tag{1.1.8}$$

and

$$I_{ij} = E\left(\frac{N}{f^2}\frac{\partial f}{\partial \theta_i} \cdot \frac{\partial f}{\partial \theta_j}\right). \tag{1.1.9}$$

This result is also implicitly contained in Rao [2].

§1.2. Let us now take n as a chance variable determined by a sequential procedure. $X_1, X_2, X_3, \cdots$ is a sequence of chance variables having the same probability density or probability $f(x; \theta)$, according as X is absolutely continuous or discrete. The sequential process tells us, after each successive observation has been drawn, whether the next observation is to be taken or not. Thus n will denote the total number of observations taken by the time the sequential process has been completed. Under certain regularity conditions, Wolfowitz [5] has shown that if $\theta^*(x_1, x_2, \cdots, x_n)$ is an unbiased estimate of θ, then

$$\sigma^2 \theta^*(x_1, x_2, \cdots, x_n) \geq \frac{1}{En \cdot E\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2}. \tag{1.2.1}$$

Furthermore, if θ consists of T components, $\theta_1, \theta_2, \cdots, \theta_T$, and $\theta_1^*(x_1, x_2, \cdots, x_n)$, $\theta_2^*(x_1, x_2, \cdots, x_n)$, $\cdots$, $\theta_T^*(x_1, x_2, \cdots, x_n)$ are unbiased estimates of $\theta_1, \theta_2, \cdots, \theta_T$ respectively, Wolfowitz [5] has proved that

$$\sum_{i,j=1}^{T} I_{ij} t_i t_j = T + 2 \tag{1.2.2}$$

is contained within the ellipsoid

$$\sum_{i,j=1}^{T} V^{ij} t_i t_j = T + 2, \tag{1.2.3}$$

where

$$I_{ij} = En \cdot E\left(\frac{\partial \log f}{\partial \theta_i}\frac{\partial \log f}{\partial \theta_j}\right), \qquad i, j = 1, \cdots, T.$$

Blackwell and Girshick [6] have shown that the lower bound given by (1.2.1) for the variance of an unbiased estimate of θ is attained only for the sequential process for which $\Pr(n = N) = 1$, if the probability density function $f(x; \theta)$ of X is such that $E(X) = \theta$ and $x_1 + x_2 + x_3, \cdots + x_M$ is a sufficient statistic for all integral values of M, for estimating θ; $x_1, x_2, \cdots, x_M$ being M independent observations on the chance variable X.

In this paper the following results have been obtained. The specific conditions under which the results hold are stated at their proper places along with the results:

(1.3.1) The lower bound in (1.1.4) is valid when n is considered a chance variable determined by a sequential procedure instead of being a fixed number N.

(1.3.2) The concentration ellipsoid defined in (1.2.3) contains within itself another ellipsoid

$$\sum_{i,j=1}^{T} \mu_{ij} t_i t_j = T + 2$$

where μ_{ij} is given by (3.1.18), which in turn contains the ellipsoid given by (1.2.2).

(1.3.3). The Blackwell and Girshick result [6] for the achievement of the lower bound for the variance of unbiased estimates given by (1.2.1) has been extended to the case where the probability density (or probability) $\prod_{i=1}^{M} f(x_i ; \theta)$, for all fixed $M \geq N$, where N is the least value for which $\Pr(n = N) \neq 0$, has an unbiased "efficient" estimate for θ in the sense defined by Cramer. This is illustrated by two examples of Wald sequential procedures.

(1.3.4). Let N be fixed and $p_N(x_1, x_2, \cdots, x_N ; \theta) \cdot | J | = g(\theta^*; \theta)$ $h(\xi_1, \xi_2, \cdots, \xi_{N-1} | \theta^*, \theta)$, where J denotes the Jacobian of the transformation from $x_1, x_2, \cdots, x_N$ to $\theta^*, \xi_1, \xi_2, \cdots, \xi_{N-1}$. Here $g(\theta^*; \theta)$, and $h(\xi_1, \xi_2, \cdots, \xi_{N-1} | \theta^*; \theta)$ are respectively the probability density function (or probability) of θ^* and the conditional probability density function (or probability) of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for a given value of θ^*.

The necessary and sufficient conditions under which the lower bound for the variance of unbiased estimates given by Bhattacharyya [3] may be achieved are that there should exist a statistic $\theta^*(x_1, x_2, \cdots, x_N)$ such that:

(a) $h_1, h_2, \cdots, h_m$ are linearly dependent considered as functions of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for given values of θ and $\theta^*(x_1, x_2, \cdots, x_N)$ and

(b) the probability density $g(\theta^*; \theta)$ of $\theta^*(x_1, x_2, \cdots, x_N)$ satisfies the following equation:

$$\theta^*(x_1, x_2, \cdots, x_N) - \theta = \sum_{i=1}^{n} \frac{K_i}{g(\theta^*; \theta)} \frac{\partial^i}{\partial \theta^i} g(\theta^*; \theta),$$

where K_i are independent of the $x_1, x_2, x_3, \cdots, x_N$.

Equivalent conditions for the multiparameter case have also been given.

(1.3.5). The following properties of $\phi_1(n), \phi_2(n), \cdots$ are derived:

(a) Under certain conditions $\phi_1(N), \phi_2(N) \cdots$ form a system of orthogonal polynomials in $\phi_1(N)$, the weight function being $p_N(x_1, x_2, \cdots, x_N ; \theta)$.

(b) $\sum_{i=1}^{m} K_i \phi_i(n)$ cannot be a function of $x_1, x_2, \cdots, x_n$, independent of θ except for the constant zero.

(c) If $\theta^*(x_1, x_2, \cdots, x_n)$ is linearly dependent upon $\phi_1(n)$, then no other statistic except of the form $a\theta^*(x_1, x_2, \cdots, x_n) + b$ where a and b are constant independent of θ, can be linearly related with $\phi_1(n)$.

(1.3.6). If a) $\theta^*(x_1, x_2, \cdots, x_N)$ is an unbiased estimate of θ and b) if among all functions of $\theta^*(x_1, x_2, \cdots, x_N)$ which are unbiased estimates of θ with finite variance, θ^* is the one with the least variance and such that the set of polynomials with respect to the distribution function of θ^* is complete, then there is no function of θ^* having a finite variance which is an unbiased estimate of θ.

2. Estimation of a single parameter.

§2.1. Let $X_1, X_2, \cdots$ and $p_M(x_1, x_2, \cdots x_M; \theta)$ be as given in the first paragraph of (1.1). Let Ω be the space of all possible infinite sequences (ω) of observations $x_1, x_2, \cdots$. Let there be given an infinite sequence of Borel measurable functions $\Phi_1(x_1)$, $\Phi_2(x_1, x_2)$, $\cdots$, $\Phi_j(x_1, x_2, x_3, \cdots, x_j) \cdots$, defined for all observable sequences in Ω such that each takes only the values zero and one. We further assume that everywhere in Ω, except possibly on a set whose probability is zero for all θ under consideration at least one of the functions $\Phi_1(x_1)$, $\Phi_2(x_1, x_2)$, $\cdots$ takes the value of one. Let n be the smallest integer for which this occurs. Thus $n(\omega)$ is a chance variable. The sequential process is then defined as follows:

Take an observation and find $\Phi_1(x_1)$. If it is unity, the sampling process stops; otherwise continue sampling. If a second observation is taken and the value of $\Phi_2(x_1, x_2)$ is unity, the process stops; otherwise continue sampling, and so on. In general, if after taking j observations

$$\Phi_1(x_1, x_2, \cdots, x_i) = 0 \text{ for } i = 1, 2, \cdots j - 1,$$

and $\Phi_j(x_1, x_2, \cdots, x_j) = 1$, sampling stops; otherwise it is continued. We will denote by R_j, the set of all points $(x_1, x_2, \cdots, x_j)$ for which the process stops with the jth observation.

Let $\theta^*(x_1, x_2, \cdots, x_n)$ be a statistic whose expectation is a real valued function $\gamma(\theta)$ of θ. The development proceeds on the assumption that $p_M(x_1, x_2, \cdots, x_M; \theta)$ is a probability density function. The result is equally valid if $p_M(x_1, x_2, \cdots, x_M; \theta)$ is the probability of discrete variables X_1, $X_2, \cdots, X_M$ provided that integration is replaced by summation whenever this is required. Further the phrase "almost all points" in a Euclidean space of any finite dimensionality is understood to mean all points in the space with the following possible exceptions:

(a). A set of Lebesgue measure zero where $p_M(x_1, x_2, \cdots, x_M; \theta)$ is the probability density function;

(b). The points which belong to the set Z, where $p_M(x_1, x_2, \cdots, x_M; \theta)$ is the probability function of the discrete chance variables $X_1, X_2, \cdots, X_M$. The set Z consists of all points $(x_1, x_2, \cdots, x_M)$ such that $p_M(x_1, x_2, \cdots, x_M; \theta) = 0$ identically for all θ under consideration.

§2.2. *Conditions of regularity.* We will postulate the following conditions to be satisfied by $p_M(x_1, x_2, \cdots, x_M; \theta)$ and $\theta^*(x_1, x_2, \cdots, x_n)$.

(2.2.1). $\theta^*(x_1, x_2, \cdots x_n)$ has an expectation $\gamma(\theta)$ and a finite variance. All the derivations of $\gamma(\theta)$ are assumed to be finite. The parameter θ lies in an open interval D of the real line. D may consist of the entire line or an entire half line.

(2.2.2). The derivatives

$$\frac{\partial^i p_M}{\partial \theta^i}, \qquad (i = 1, 2, \cdots, m),$$

exist for all θ in D and almost all $x_1, x_2, \cdots, x_M$ in R_M and for all M. We define

$$\frac{1}{p_M} \frac{\partial^i p_M}{\partial \theta^i} = 0,$$

whenever $p_M(x_1, x_2, \cdots, x_M; \theta) = 0$; thus,

$$\frac{1}{p_M} \frac{\partial^i p_M}{\partial \theta^i} = \phi_i(M)$$

is defined for all θ in D and almost all $(x_1, x_2, \cdots, x_M)$ in R_M.

(2.2.3). For any integral j there exists non-negative L-measurable functions $T_i(x_1, x_2, \cdots, x_j)$, $(i = 1, 2, \cdots, m)$, such that

$$\text{(a)} \quad \left| \theta^*(x_1, x_2, \cdots, x_j) \frac{\partial^i}{\partial \theta^i} p_j(x_1, x_2, \cdots, x_j; \theta) \right| < T_i(x_1, x_2, \cdots, x_j),$$

for all θ in D and almost all $(x_1, x_2, \cdots, x_j)$ in R_j.

$$\text{(b)} \quad \int_{R_j} T_i(x_1, x_2, \cdots, x_j) \sum_{u=1}^{j} dx_u, \ (i = 1, 2, \cdots, m),$$

are finite.

(2.2.4). Let $t_j(\theta) = \int_{R_j} \theta^*(x_1, x_2, \cdots, x_j) p_j(x_1, x_2, \cdots, x_j; \theta) \prod_{u=1}^{j} dx_u$.

We postulate the uniform convergence of

$$\sum_{j=1}^{\infty} \frac{d^i}{d\theta^i} t_j(\theta), \qquad (i = 1, 2, \cdots, m)$$

(the existence of $\frac{d^i}{d\theta^i} (t_j(\theta))$ is assured by the assumption (2.2.3).)

(2.2.5). There exist functions $S_i(x_1, x_2, \cdots, x_j)$ for every j, $(i = 1, 2, \cdots, m)$, such that when $\theta^*(x_1, x_2, \cdots, x_j)$ and $T_i(x_1, x_2, \cdots, x_j)$ are replaced by unity and $S_i(x_1, x_2, \cdots, x_j)$ respectively, conditions (2.2.3) and (2.2.4) still hold good.

(2.2.6). The covariance matrix of $\phi_i(n)$ $(i = 1, \cdots, m)$ exists and is non-singular for almost all θ in D and almost all $(x_1, x_2, \cdots, x_n)$.

§2.3. Let us consider the sequential process mentioned in §2.1 and the functions $\theta^*(x_1, x_2, \cdots, x_n)$ and $p_M(x_1, x_2, \cdots, x_M; \theta)$ which satisfy the regularity conditions in §2.2. We will now find a lower bound for the variance of such estimates.

Let us examine

$$F = E\left(\theta^*(x_1, x_2, \cdots x_n) - \gamma(\theta) - \sum_{i=1}^{m} K_i \phi_i(n)\right)^2, \tag{2.3.1}$$

where K_i $(i = 1, 2, \cdots, m)$ are independent of $(x_1, x_2, \cdots x_n)$. Now (2.3.1) can be written as

$$F = \sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) - 2\sum_{i=1}^{m} K_i E\theta^*(x_1, x_2, \cdots, x_n)\phi_i(n) \tag{2.3.2}$$
$$+ 2\gamma(\theta)\sum_{i=1}^{m} K_i E\phi_i(n) + \sum_{i,j=1}^{m} K_i K_j \lambda_{ij},$$

where

$$\lambda_{ij} = E(\phi_i(n)\phi_j(n)) \qquad (i, j = 1, 2, \cdots, m).$$

Now

$$E(\theta^*(x_1, x_2, \cdots, x_n)\phi_i(n)) = \sum_{j=1}^{\infty}\int_{R_j}\theta^*(x_1, x_2, \cdots, x_j)\frac{\partial^i p_j}{\partial\theta^i}\prod_{u=1}^{j} dx_u. \tag{2.3.4}$$

We also know that

$$\sum_{j=1}^{\infty}\int_{R_j}\theta^*(x_1, x_2, \cdots, x_j)p_j\prod_{u=1}^{j} dx_u = \gamma(\theta). \tag{2.3.5}$$

Differentiating both sides of (2.3.5) i times $(i = 1, 2, \cdots, m)$ we have, because of conditions (2.2.3) and (2.2.4):

$$\sum_{j=1}^{\infty}\int_{R_j}\theta^*(x_1, x_2, \cdots, x_j)\frac{\partial^i p_j}{\partial\theta^i}\prod_{u=1}^{i} dx_u = \frac{d^i\gamma(\theta)}{d\theta^i}, \qquad (i = 1, 2, \cdots, m). \tag{2.3.6}$$

From (2.3.4) and (2.3.6), we obtain

$$E(\theta^*(x_1, x_2, \cdots, x_n)\phi_i(n)) = \frac{d^i}{d\theta^i}\gamma(\theta). \tag{2.3.7}$$

Differentiating

$$1 = \sum_{j=1}^{\infty}\int_{R_j} p_j\prod_{u=1}^{j} dx_u \tag{2.3.8}$$

i times $(i = 1, 2, \cdots, m)$ with respect to θ, we obtain because of conditions (2.2.5)

$$0 = \sum_{j=1}^{\infty}\int_{R_j}\frac{\partial^i p_j}{\partial\theta^i}\prod_{u=1}^{j} dx_u, \qquad (i = 1, \cdots, m). \tag{2.3.9}$$

(2.3.8) is valid on account of the type of sequential process (2.1). Now

$$E(\phi_i(n)) = \sum_{j=1}^{\infty}\int_{R_j}\frac{\partial^i p_j}{\partial\theta^i}\prod_{u=1}^{j} dx_u, \qquad (i = 1, \cdots, m). \tag{2.3.10}$$

By (2.3.7) and (2.3.10), (2.3.2) reduces to

$$F = \sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) - 2\sum_{i=1}^{m} K_i\frac{d^i\gamma(\theta)}{d\theta^i} + \sum_{i,j=1}^{m} K_i K_j \lambda_{ij}. \tag{2.3.11}$$

Now $\|\lambda_{ij}\|$ being non-singular on account of condition (2.2.6), we get just one set of values of K's which minimize F. These values are given by

$$K_j = \sum_{i=1}^{m} {}_{(m)}\lambda^{ij}\frac{d^i\gamma(\theta)}{d\theta^i}, \tag{2.3.12}$$

where

$$\| {}_{(m)}\lambda^{ij} \|^{-1} = \| \lambda_{ij} \|, \qquad (i, j = 1, 2, \cdots, m). \tag{2.3.13}$$

Putting the above values of $K_j (j = 1, 2, \cdots, m)$ in (2.3.11), we obtain

$$F = \sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) - \sum_{i,j=1}^{m} {}_{(m)}\lambda^{ij} \cdot \frac{d^i\gamma(\theta)}{d\theta^i} \cdot \frac{d^j\gamma(\theta)}{d\theta^j}. \tag{2.3.14}$$

Hence, F being non-negative by (2.3.1), we have

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq \sum_{i,j=1}^{m} {}_{(m)}\lambda^{ij} \cdot \frac{d^i\gamma(\theta)}{d\theta^i} \cdot \frac{d^j\gamma(\theta)}{d\theta^j}. \tag{2.3.15}$$

Thus R.H.S. of the above inequality gives the lower bound to the variance of unbiased estimates of $\gamma(\theta)$.[1] When $\gamma(\theta) = \theta$, the above reduces to

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq {}_{(m)}\lambda^{11}. \tag{2.3.16}$$

When $m = 1$ and $p_n(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$, (2.3.16) reduces to

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq \frac{1}{En \cdot E\left(\left(\frac{\partial}{\partial\theta} \log f(x; \theta)\right)^2\right)}, \tag{2.3.17}$$

which is the result given by Wolfowitz [5].

When n, the chance variable, is constant and equal to N, then (2.3.15) and (2.3.16) correspond to those given by Bhattacharyya [3]. Although the conditions of regularity under which Bhattacharyya proves his results are not clear from his paper, they are likely to be slightly different from those in §2.3, as the results in [3] are obtained only for a fixed size sample.

§2.4. We will now investigate the necessary and sufficient conditions under which the lower bound given in (2.3.16) is actually higher than that given in (2.3.17).

We can easily see that

$${}_{(m)}\lambda^{11} = \frac{1}{\lambda_{11}(1 - R^2_{1\cdot 23\cdots m})}, \tag{2.4.1}$$

where $R_{1\cdot 23\cdots m}$ is the multiple correlation coefficient between $\phi_1(n)$ and $\phi_2(n)$, $\phi_3(n), \cdots, \phi_m(n)$.

The excess of the lower bound given by (2.3.16) over that when we use $m = 1$ is given by

$$\frac{1}{\lambda_{11}(1 - R^2_{1\cdot 23\cdots m})} - \frac{1}{\lambda_{11}} \tag{2.4.2}$$

---

[1] Under certain weak restrictions, an optimum lower bound to the variance of unbiased estimates has been obtained by me along the lines of a similar result for fixed size samples in an unpublished paper by A. Wald. Independently C. Stein has obtained the same result in a paper not yet published.

which is further equal to

$$\frac{R^2_{1\cdot 23\cdots m}}{\lambda_{11}(1 - R^2_{1\cdot 23\cdots m})}. \tag{2.4.3}$$

Thus the lower bound for the variance of unbiased estimates of θ is obtained by using $m > 1$ is higher than that obtained by employing $m = 1$ if and only if $R_{1\cdot 23\cdots m}$ is not zero for some $m \geq 2$. This is equivalent to the condition that for at least one $i \geq 2$, λ_{1i}, the correlation coefficient between $\phi_1(n)$ and $\phi_i(n)$ $(i > 1)$, is different from zero. Suppose further that we have used $m = \alpha$ and that we wish to find the increase in the lower bound if α were replaced by $\alpha + 1$. The increase in this case is given by

$$\frac{\rho^2_{1(\alpha+1)\cdot 23\cdots\alpha}}{\lambda_{11}(1 - R^2_{1\cdot 23\cdots(\alpha+1)})} \tag{2.4.4}$$

where $\rho_{1(\alpha+1)\cdot 23\cdots\alpha}$ is the partial correlation coefficient between $\phi_1(n)$ and $\phi_{\alpha+1}(n)$ keeping $\phi_2(n), \cdots, \phi_\alpha(n)$ fixed. It is greater than zero if and only if $\rho_{1(\alpha+1)\cdot 23\cdots\alpha}$ is not equal to zero.

§2.5. If $p_n(x_1, x_2, \cdots, x_n; \theta_1)$ also depends upon a finite number of other parameters $\theta_2, \theta_3, \cdots, \theta_T$, then a lower bound higher than or equal to that given in (2.3.16) can be obtained by using

$$\sum_{i_1+i_2+\cdots+i_T\leq m} K_{i_1,i_2\cdots,i_T}\cdot\phi_{i_1,i_2,\cdots,i_T}(n) \quad \text{instead of} \quad \sum_{i_1=1}^{m} K_{i_1}\cdot\phi_{i_1}(n) \quad \text{in } (2\cdot 3\cdot 1).$$

The lower bound in this case is given by (3.1.14) (see section 3) by taking $s = 1$, that is,

$$\sigma^2(\theta^*(x_1, x_2, \cdots, x_n)) \geq C(1, 1) \tag{2.5.1}$$

where $C(1, 1)$ is the element in the first row and first column of the inverse of W defined in (3.1.9).

The result for $n = N$, N fixed, is obtained by Bhattacharyya [3, 1947]. Let us illustrate it by an example. Take samples of fixed size N. Suppose we are required to find the lower bound to the variance of unbiased estimates of θ_1 in the normal population

$$f(x; \theta_1\, \theta_2) = \frac{1}{\sqrt{2\pi\theta_1}}\cdot e^{-(x-\theta_2)^2/2\theta_1} \tag{2.5.2}$$

on the basis of N independent observations $x_1, x_2, \cdots, x_N$. The lower bound for the variance of the unbiased estimates of θ_1, when we use

$$\sum_{i_1=1}^{m} K_{i_1}\cdot\phi_{i_1}(N) \text{ in (2.3.1)} \qquad \text{is given by } \frac{2\theta_1^2}{N}.$$

However, if $\sum_{i_1+i_2\leq 2} K_{i_1,i_1}\cdot\phi_{i_1,i_2}(N)$ is used, the lower bound, by the help of (2.5.1), is found to be equal to $2\theta_1^2/(N - 1)$. In fact there exists the statistic

$$\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N - 1}$$

whose variance is equal to $2\theta_1^2/(N - 1)$ where $\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}$. Thus the use of $\sum_{i_1+i_2 \leq 2} K_{i_1,i_2} \cdot \phi_{i_1,i_2}(N)$ brings into relief the unbiased estimate with the least variance.

3. Multi-parameter case. In this section we will prove the result mentioned in (1.3.2) of §1.3.

§3.1. *Let* θ *consist of* T *components* $(\theta_1, \theta_2, \cdots, \theta_T)$ *and* $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$ *be unbiased estimates of* $\theta_1, \theta_2, \cdots, \theta_T$ *respectively. Also, let a sequential process of the type described in* §2.1 *be given. We postulate the following regularity conditions:*

(3.1.1). *The covariance matrix* $\| V_{ij} \|$ *of the estimates* $\theta_i^* (i = 1, 2, \cdots, T)$ *is non-singular in* D, *where* D *is an open interval of the* T*-dimensional parameter space.*

(3.1.2). *The conditions of section* (2.2) *are satisfied for each one of* $\theta_i^*(i = 1, 2, \cdots, T)$ *and* $\phi_{i_1,i_2,\ldots,i_T}(n)$, $(i_1 + i_2 + \cdots + i_T \leq m)$.

(3.1.3). *The covariance matrix of* $\phi_{i_1,i_2,\ldots,i_T}(n)$, $i_1 + i_2 + \cdots + i_T \leq m$ *exists and is non-singular. Under the assumptions* (3.1.1)–(3.1.3), *we prove the result* (1.3.2) *in section* 1.3.

PROOF: Using the same arguments of §2.3, we obtain

$$E(\theta_j^*(x_1, x_2, \cdots, x_n) \cdot \phi_{i_1,i_2,\cdots,i_T}(n)) = 1, \quad \begin{cases} i_\beta = \delta_{\beta j}(\beta = 1, 2, \cdots, T), \\ j = 1, 2, \cdots, T \end{cases} \tag{3.1.4}$$

$$= 0 \text{ otherwise.} \tag{3.1.5}$$

Let the covariance matrix of $\theta_j^*(j = 1, 2, \cdots, s; s \leq T)$ and $\phi_{i_1,i_2,\ldots,i_T}(n)$, $(i_1 + i_2 + \cdots + i_T \leq m)$ be given by

$$U = \begin{Vmatrix} A & B \\ B' & W \end{Vmatrix} \tag{3.1.6}$$

where

$$A = \| V_{ij} \|, \; i, j = 1, 2, \cdots, s; \qquad s \leq T; \tag{3.1.7}$$

$$B = \| I, 0 \|; \tag{3.1.8}$$

(3.1.9) and W = covariance matrix of the set

$$[\phi_{i_1,i_2,\ldots,i_T}(n); i_1 + i_2 + \cdots + i_T \leq m],$$

arranged such that the jth term in the leading diagonal is given by

$$E(\phi^2_{i_1,i_2,\ldots,i_T}(n)), \qquad \text{where } i_j = 1, i_\beta = 0, \beta \neq j, \quad (j = 1, 2, \cdots, T), \tag{3.1.10}$$

and B' is the transpose of B.

As U is positive semi-definite, we have

$$| U | \geq 0. \tag{3.1.11}$$

The above can further be reduced to

(3.1.12) $$|W| \cdot |A - BW^{-1}B'| \geq 0,$$

which leads to

(3.1.13) $|A - B \cdot W^{-1} \cdot B'| \geq 0$, as W is positive definite.

By the use of (3.1.8) we obtain from above

(3.1.14) $$|A - C| > 0$$

where C is the top left part of W^{-1}, consisting of s rows and s columns.

Let us now consider the matrix

(3.1.15) $$\| V_{ij} - v_{ij} \|, \qquad (i, j = 1, 2, \cdots, T),$$

where $\| v_{ij} \|$ is the topleft part of W^{-1} consisting of T rows and T columns, and is equal to

(3.1.16) $$\| W_{11} - W_{12}W_{22}^{-1}W_{21} \|^{-1},$$

when W is written as

(3.1.17) $$W = \begin{Vmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{Vmatrix},$$

where W_{11} has T rows and T columns.

By the repeated application of (3.1.14), we are led to the conclusion that all the leading minors of the matrix in (3.1.15) are either positive or zero. Hence the matrix in (3.1.15) is semi-positive definite.

If now we put

(3.1.18) $$\| \mu_{ij} \| = \| v_{ij} \|^{-1}$$

we obtain

(3.1.19) $$\| \mu_{ij} - V^{ij} \|$$

to be semi-positive definite. Thus the ellipsoid

(3.1.20) $$\sum_{i,j=1}^{T} V^{ij} \cdot t_i \cdot t_j = T + 2$$

contains within itself the ellipsoid

(3.1.21) $$\sum_{i,j=1}^{T} \mu_{ij} \cdot t_i \cdot t_j = T + 2.$$

Cramer calls the ellipsoid in (3.1.20) a "concentration" ellipsoid.

We will now show that the ellipsoid given by (3.1.21) contains within itself the ellipsoid

(3.1.22) $$\sum_{i,j=1}^{T} I_{ij} \cdot t_i \cdot t_j = T + 2$$

where $|| I_{ij} ||$ is the information matrix given by W_{11} in (3.1.17). We will prove the above by showing

$$|| I_{ij} - \mu_{ij} ||, \qquad (i, j = 1, \cdots, T), \tag{3.1.23}$$

to be semi-positive definite.

We obtain, from (3.1.16) and (3.1.18),

$$|| \mu_{ij} || = W_{11} - W_{12}W_{22}^{-1}W_{21}, \qquad (i, j = 1, 2, \cdots, T). \tag{3.1.24}$$

From the above it follows that

$$|| I_{ij} - \mu_{ij} || = W_{12}W_{22}^{-1}W_{21}. \tag{3.1.25}$$

Thus the matrix on the right hand side is semi-positive definite since W_{22} is positive definite, we see that the ellipsoid (3.1.21) contains within itself the ellipsoid given by (3.1.22). This proves the assertion made in (1.3.2) of §1.3. It may be seen that (3.1.22) is strictly contained in (3.1.21) if and only if $W_{12} \neq 0$. It may be mentioned that in this section as well as elsewhere, $T + 2$, appearing on the right hand side of the equation of an ellipsoid, can be replaced by any positive constant. Also the ellipsoid in (3.1.21) depends upon the choice of m and it can be shown that for any two positive integers m_1, m_2 ($m_1 > m_2$) the ellipsoid for $m = m_1$ contains within itself the one for $m = m_2$.

§3.2. In general, let $\theta_i^*(x_1, x_2, \cdots, x_n)$ be statistics whose expectations are $\gamma_i(\theta_1, \theta_2, \cdots, \theta_T)$, $(i = 1, 2, \cdots, T)$, the latter being assumed to admit partial derivatives of all possible orders. Under the postulates enumerated in §3.1, we see that the ellipsoid in (3.1.20) contains within itself the ellipsoid

$$\sum_{i,j=1}^{T} S_{ij} \cdot t_i \cdot t_j = T + 2 \tag{3.2.1}$$

where

$$|| S_{ij} || = || RW^{-1}R' ||^{-1}, \qquad i, j = 1, 2, \cdots, T, \tag{3.2.2}$$

and

$$R = \left\| \frac{\partial^{i_1 + i_2 + \cdots + i_T}}{\partial\theta_1^{i_1}\, \partial\theta_2^{i_2} \cdots \partial\theta_T^{i_T}} \gamma_j(\theta_1, \theta_2, \cdots, \theta_T) \right\|, \tag{3.2.3}$$

$$(j = 1, 2, \cdots, T; i_1 + i_2 + \cdots + i_T \leq m),$$

where j and $i_1 + i_2 + \cdots + i_T$ indicate the number of the row and the column respectively and is arranged to correspond to the arrangement of W, where W is the same as given in (3.1.9).

4. Achievement of the different lower bounds. In §4.1 we will demonstrate the desirability of finding a higher lower bound to the variance of sequential estimates than that given by Wolfowitz, by giving two examples in which the latter is not achieved. From §2.4 it is clear that this will be so if $E(\phi_1(n) \cdot \phi_i(n))$ is not zero for at least one value of $i \geq 2$. We will demonstrate that this is true for $i = 2$. In §4.2 we show that if "efficient" statistic exists for all $M \geq N$, the bound is achieved only in the case when the sample size is fixed. In §4.3 we obtain necessary and sufficient conditions for the attainment of the bound given in (1.1.4). In §4.4 we discuss the conditions under which there exists a "concentration ellipsoid" which coincides with the ellipsoid given in (3.1.21) for samples of fixed size N.

§4.1. Ex. 1. The Wald sequential procedure for testing $\theta = \theta_1$, against $\theta = \theta_2$ in a normal population

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} \tag{4.1.1}$$

is given as follows: *If*

$$B < \sum_{i=1}^{s} \left(x_i - \frac{\theta_1 + \theta_2}{2}\right) < A, \ (s = 1, 2, \cdots, j-1), \tag{4.1.2}$$

and

$$\sum_{i=1}^{j} \left(x_i - \frac{\theta_1 + \theta_2}{2}\right) \textit{ is either } \geq A \textit{ or } \leq B, \tag{4.1.3}$$

we cease sampling and make a decision. Here A and B are constants fixed by the probability levels of making a correct decision.

Let us denote the set of points satisfying (4.1.2) and (4.1.3) by R_j. In this case

$$\phi_1(n) = \sum_{i=1}^{n} (x_i - \theta) = Z_n - n\theta, \quad \text{where} \quad Z_n = \sum_{i=1}^{n} x_i. \tag{4.1.4}$$

The above is differentiable with respect to θ. On differentiating we have

$$\phi_2(n) = (Z_n - n\theta)^2 - \text{n}. \tag{4.1.5}$$

Now

$$E(\phi_1(n) \cdot \phi_2(n)) = E(Z_n - n\theta)^3 - E(n(Z_n - n\theta)). \tag{4.1.6}$$

By theorem 7.3, Wolfowitz [5],

$$E(Z_n - n\theta)^3 = En \cdot E(X - \theta)^3 + 3E(n(Z_n - n\theta)), \tag{4.1.7}$$

where X has the distribution given in (4.1.1). As $E(X - \theta)^3$ is equal to zero, (4.1.6) reduces to

$$E(\phi_1(n) \cdot \phi_2(n)) = 2E(n(Z_n - n\theta)). \tag{4.1.8}$$

We will now show that right hand side of (4.1.8) is not identically zero in θ. Let us consider

$$E(n) = \sum_{j=1}^{\infty} \int_{R_j} \frac{j}{(2\pi)^{j/2}} \cdot \left[\exp\left(-\tfrac{1}{2} \sum_{i=1}^{j} (x_i - \theta)^2\right)\right] \prod_{u=1}^{j} dx_u. \tag{4.1.9}$$

Differentiating with respect to θ, we get

$$(4.1.10)\qquad \frac{d}{d\theta}(E(n)) = \sum_{j=1}^{\infty}\int_{R_j}\frac{j(z_j - j\theta)}{(2\pi)^{j/2}}\cdot\left[\exp\left(-\tfrac{1}{2}\sum_{i=1}^{j}(x_i-\theta)^2\right)\right]\prod_{u=1}^{j} dx_u .$$

The righthand side of the above equation being equal to $E(n(Z_n - n\theta))$, the latter does not vanish identically in θ, because the lefthand side is not identically zero. The step from (4.1.9) to (4.1.10) can be easily seen to be valid.

Ex. 2. The Wald sequential procedure for testing $p = p_1$ against $p = p_2$ in a binomial distribution, where p is the probability of the event occurring, is given as follows: *If*

$$(4.1.11)\qquad B < \sum_{i=1}^{s}(x_i - d) < A, \qquad s = 1, 2, \cdots, j-1,$$

and

$$(4.1.12)\qquad \sum_{i=1}^{j}(x_i - d) \text{ is either } \geq A \text{ or } \leq B,$$

where d is given by $[\log(1-p_1)/(1-p_2)]/\log[(p_2(1-p_1)/p_1(1-p_2)]$, *the process stops with the* jth *observation and a decision is taken.* Here, x_i is the characteristic function of the event at the ith trial, that is:

$x_i = 1$, when the event occurs at the ith trial;

$= 0$, otherwise.

Let us denote the set of points satisfying (4.1.11) and (4.1.12) by R_j. In this case we find

$$(4.1.13)\qquad E[\phi_1(n)\cdot\phi_2(n)] = \frac{2}{p^2\cdot(1-p)^2}\cdot E(n(Z_n - np)),$$

where $Z_n = \sum_{i=1}^{n} x_i$. We have now to show that the righthand side is not identically zero. Differentiating

$$(4.1.14)\qquad E(n) = \sum_{j=1}^{\infty}\sum_{R_j} jp^{z_j}(1-p)^{j-z_j}$$

with regard to p, we obtain

$$(4.1.15)\qquad \frac{d}{dp}(E(n)) = \sum_{j=1}^{\infty}\sum_{R_j}\frac{j(Z_j - jp)}{p(1-p)}\cdot p^{z_j}\cdot(1-p)^{j-z_j}.$$

The righthand side of the above is the same as

$$(4.1.16)\qquad \frac{1}{p(1-p)}E(n(Z_n - np)).$$

Thus, the lefthand side of (4.1.15) being not identically zero, the same is true for (4.1.16), and consequently the bound given by Wolfowitz is not achieved in this case.

The step from (4.1.14) to (4.1.15) is valid as

$$\sum_{j=1}^{\infty}\left(\sum_{R_j} j\cdot p^{z_j}(1-p)^{j-z_j}\cdot\frac{z_j-j\cdot p}{p(1-p)}\right) \tag{4.1.17}$$

is absolutely and uniformly convergent.

§4.2. Let θ^* be some unbiased estimate of θ, where x_i's are successive independent observations on the chance variable X having the probability density function or probability function $f(x;\theta)$. We adopt a sequential procedure mentioned in §2.1 satisfying the regularity conditions in §2.2 and also postulate the following:

(i) For all positive integral values of $M \geq N$

$$p_M(x_1, x_2, \cdots, x_M; \theta) = \prod_{i=1}^{M} f(x_i; \theta)$$

possesses an 'efficient' estimate for θ, where N is the least value of n for which $\Pr(n = N) \neq 0$.

(ii) $E(n)$ exists and admits derivatives up to the second order with respect to θ. Furthermore, $\frac{d}{d\theta}(E(n))$ is either zero for all θ under consideration or is never zero.

Under the above conditions the Wolfowitz lower bound for the variance of unbiased estimates is achieved only when $\Pr(n = N) = 1$.

PROOF: This bound will be attained if and only if there exists an unbiased estimate θ^* of θ such that

$$E(\theta^* - \theta - K\phi_1(n))^2 = 0, \tag{4.2.1}$$

that is,

$$\theta^* - \theta = K\phi_1(n) \tag{4.2.2}$$

with probability one, where K is independent of all x_i's and n. As there exists an 'efficient' estimate, say $\psi(M)$ for all $M \geq N$, we have

$$\psi(M) - \theta = \frac{1}{M\cdot E\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]} \tag{4.2.3}$$

for all $M \geq N$. From (4.2.2) and (4.2.3), it follows that

$$\theta^* - \theta = K\cdot n\cdot(\psi(n) - \theta)\cdot E\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]. \tag{4.2.4}$$

Now as

$$K = \frac{1}{En\cdot E\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]}, \tag{4.2.5}$$

we have

$$\theta^* - \theta = \frac{n \cdot (\psi(n) - \theta)}{En}. \tag{4.2.6}$$

If $E(n)$ is independent of θ, then from (4.2.6), we obtain

$$n/E(n) = 1, \tag{4.2.7}$$

that is, n is constant with probability one and the sequential procedure reduces to a fixed size sample case. If $E(n)$ is not independent of θ, then differentiating (4.2.6) with regard to θ, we obtain

$$1 = \frac{n \cdot (\psi(n) - \theta) \cdot \frac{d}{d\theta}(En)}{(En)^2} + \frac{n}{En}. \tag{4.2.8}$$

As $\frac{d}{d\theta}(En)$ is not equal to zero for any θ under consideration, substituting the value of $\psi(n)$ from (4.2.8) in (4.2.6), the latter takes the form:

$$\theta^* - \theta = \frac{En - n}{\frac{d}{d\theta}(En)}. \tag{4.2.9}$$

Differentiating the above with respect to θ, the result is:

$$-1 = -\frac{En - n}{\left[\frac{d}{d\theta}(En)\right]^2} \cdot \frac{d^2}{d\theta^2}(En) + 1. \tag{4.2.10}$$

Now if $\frac{d^2}{d\theta^2}(En) = 0$, then (4.2.10) is not valid, thereby contradicting (4.2.2). If $\frac{d^2}{d\theta^2}(En) \neq 0$, then rearranging (4.2.10), we obtain

$$n = -\frac{2\left(\frac{d}{d\theta}En\right)^2}{\frac{d^2}{d\theta^2}(En)} + En, \tag{4.2.11}$$

that is, n is a constant with probability one. This proves that Wolfowitz bound is achieved only in the case when $n = N$ with probability one. This generalizes the result of Blackwell and Girshick [6][2] to the extent that in [6] the existence of an efficient estimate is assumed for all integral values of M instead of $M \geq N$, as assumed here. Moreover the proof given here, with slight modifications, is also valid when the successive observations are not independent.

[2] In [6] the assumption that "$x_1 + x_2 + \cdots + x_M$ be a sufficient statistic for all M" really amounts to the postulate that "$x_1 + x_2 + \cdots + x_M$ be an "efficient" statistic for all M," when we restrict ourselves to probability density functions satisfying the conditions given by Koopman in [7].

§4.3. Let us consider a sample of fixed size N. Let θ^* together with the probability density function p_N satisfy the following regularity conditions:

(i). There exists a transformation T from $(x_1, x_2, \cdots, x_N)$ to the variables

$$\xi_i = \xi_i(x_1, x_2, \cdots, x_N), \qquad \theta^* = \theta^*(x_1, x_2, \cdots, x_N), \qquad i = 1, 2, \cdots, N-1, \tag{4.3.1}$$

such that

(a). The functions ξ_i are everywhere unique and continuous, and have continuous partial derivatives

$$\frac{\partial \xi_i}{\partial x_u}, \frac{\partial \theta^*}{\partial x_u} \ (i = 1, 2, \cdots, N-1, u = 1, 2, \cdots, N)$$

in all points $(x_1, x_2, \cdots, x_N)$ except possibly in certain points belonging to a finite number of hyper-surfaces.

(b). The relation (4.3.1) define a one-to-one correspondence between the points $x = (x_1, x_2, \cdots, x_N)$ and $y = (\xi_1, \xi_2, \cdots, \xi_{N-1}, \theta^*)$ so that conversely $x_i = \eta_i(\xi_1, \xi_2, \cdots, \xi_{N-1}, \theta^*)$ where η_i are unique.

(ii). There exists partial derivatives of $g(\theta^*;\theta)$, $h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta)$ with regard to θ of all orders up to and including m, where m is some finite integer. The variances of θ^*, h_i and $g_i \cdot g_j$, $i, j = 1, 2, \cdots, m$, are finite, where h_i and g_i are defined in section 1.

(iii). There exist functions

$$T_{ij}\begin{pmatrix} i = 1, 2, \cdots, m \\ j = 1, 2, 3 \end{pmatrix}$$

such that

$$\left| \frac{\partial^i p_N}{\partial \theta^i} \right| < T_{i1}(x_1, x_2, \cdots, x_N);$$

$$\left| \frac{\partial^i g}{\partial \theta^i} \right| < T_{i2}(\theta^*);$$

$$\left| \frac{\partial^i h}{\partial \theta^i} \right| < T_{i3}(\xi_1, \xi_2, \cdots, \xi_{N-1}; \theta^*),$$

for all θ in D and for almost all $(x_1, x_2, \cdots, x_N)$ where D is an open interval. Further

$$\int T_{i1}(x_1, x_2, \cdots, x_N) \prod_{u=1}^{N} dx_u,$$

$$\int T_{i2}(\theta^*)\, d\theta^* \quad \text{and} \quad \int T_{i3}(\xi_1, \xi_2, \cdots, \xi_{N-1}; \theta^*) \prod_{i=1}^{N-1} d\xi_u$$

are all finite, the range of integration, in each case, being the whole range for the arguments indicated. Then the necessary and sufficient conditions that the variance of θ^* equals the lower bound given in (1.1.4) are (iv). $h_1, h_2, \cdots, h_m$ are linearly dependent considered as functions of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for any given θ^* and θ, and

(v). The probability density function g of θ^* is of the form

$$\theta^* - \theta = \sum_{i=1}^{m} K_i g_i$$

where K_i may depend upon θ and N only.

The proof here is given when p_N is a probability density function. It is also valid with slight modification when p_N is the probability of discrete variables.

PROOF: Let J be the Jacobian of the transformation T in (4.3.1). Then because of conditions (i) and (ii) above, we have,

$$p_N(x_1, x_2, \cdots, x_N; \theta) \cdot | J | = g(\theta^*; \theta) \cdot h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta) \tag{4.3.2}$$

Further

$$\int h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta^*; \theta) \prod_{u=1}^{N-1} d\xi_u = 1, \tag{4.3.3}$$

the range of integration being the space of $\xi_1, \xi_2, \cdots, \xi_{N-1}$. Differentiating the above i times under the integral sign, it follows that

$$E(h_i \mid \theta^*; \theta) = 0. \tag{4.3.4}$$

Similarly we have

$$E(g_i \cdot h_i) = 0 \tag{4.3.5}$$

as the expectation of the quantity on the L.H.S. is finite by virtue of (ii). More generally, we have

$$E(F(\theta^*) \cdot h_i) = E[F(\theta^*) \cdot E(h_i \mid \theta^*)] = 0 \tag{4.3.6}$$

if $E(F(\theta^*) \cdot h_i)$ is finite. Let us now examine

$$E\left(\theta^* - \theta - \sum_{i=1}^{m} K_i \phi_i(N)\right)^2, \tag{4.3.7}$$

where $K_i\phi_i(N)$ can also be written as

$$K_i\left(g_i + \binom{i}{1} h_i g_{i-1} + \cdots + h_i\right). \tag{4.3.8}$$

Now (4.3.7) can be put in the form

$$E\left(\theta^* - \theta - \sum_{i=1}^{m} K_i g_i - \sum_{i=1}^{m} L_i \cdot h_i\right)^2, \tag{4.3.9}$$

where

$$L_i = \sum_{i=1}^{m} K_j \cdot \binom{j}{i} \cdot g_{j-i}, \qquad (i = 1, 2, \cdots, m), \tag{4.3.10}$$

clearly depend on θ and θ^* only.

By virtue of (4.3.4–4.3.6) and $F(\theta^*)$ involved in (4.3.9) being such that $E[F(\theta^*) \cdot h_i](i = 1, 2, \cdots, m)$ is finite because of (ii), we can further reduce (4.3.9) to

$$(4.3.11) \qquad E\left(\theta^* - \theta - \sum_{i=1}^{m} K_i g_i\right)^2 + E\left[E\left(\left(\sum_{i=1}^{m} L_i h_i\right)^2 \middle| \theta^*\right)\right].$$

The lower bound will be achieved if and only if the above expression is zero, the necessary and sufficient conditions for which are:

$$(4.3.12) \qquad \theta^* - \theta = \sum_{i=1}^{m} K_i \cdot g_i ,$$

and

$$(4.3.13) \qquad \sum_{i=1}^{m} L_i h_i \equiv 0 \quad \text{in} \quad \xi_1, \xi_2, \cdots, \xi_{N-1}$$

for any given values of θ^* and θ.

(4.3.13) is equivalent to the condition that h_i, $(i = 1, 2, \cdots, m)$ are linearly dependent considered as functions of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for any given values of θ and θ^*.

When m takes the value one, the above reduces to the Cramer conditions for the existence of an "efficient" estimate.

§4.4. *Multiparameter case.* Let $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$ be the unbiased estimates of $\theta_1, \theta_2, \cdots, \theta_T$ in the probability density function

$$p_N(x_1, x_2, \cdots, x_N; \theta_1, \theta_2, \cdots, \theta_T)$$

and the regularity conditions of §4.3 are satisfied when θ^* and $\frac{\partial^i}{\partial\theta^i}$ $(i = 1, 2, \cdots, m)$ are replaced by θ_j^* $(j = 1, 2, \cdots, T)$ and

$$\frac{\partial^{i_1+i_2+\cdots+i_T}}{\partial\theta_1^{i_1}\,\partial\theta_2^{i_2}\cdots\partial\theta_T^{i_T}} \qquad i_1 + i_2 + \cdots + i_T \leq m)$$

respectively. Further let

$$\begin{aligned} & p_N(x_1, x_2, \cdots, x_N; \theta_1, \theta_2, \cdots, \theta_T) \cdot |J| \\ (4.4.1) \qquad & = g(\theta_1^*, \theta_2^*, \cdots, \theta_T^*; \theta_1, \theta_2, \cdots, \theta_T) \\ & \cdot h(\xi_1, \xi_2, \cdots, \xi_{N-1} \mid \theta_1^*, \theta_2^*, \cdots, \theta_T^*) \end{aligned}$$

where g and h are respectively the joint probability distribution functions of $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$ and the conditional probability distribution of $\xi_1, \xi_2, \cdots, \xi_{N-1}$ for a given set of values of $\theta_1^*, \theta_2^*, \cdots, \theta_T^*$. In order that the ellipsoid (3.1.20) coincides with the one given by (3.1.21), it is necessary and sufficient that the following be satisfied for each $t(t = 1, 2, \cdots, T)$

$$(4.4.2) \qquad E\left(\theta_t^* - \theta_t - \sum_{i_1+i_2+\cdots+i_T\leq m} {}^{(t)}K_{i_1,i_2,\cdots,i_T} \cdot \phi_{i_1,i_2,\cdots,i_T}^{(N)}\right)^2 = 0.$$

Now reasoning similar to that in §4.3, we conclude from the above that the necessary and sufficient conditions are:

There exist T independent linear combinations of

$$h_{i_1,i_2,\dots,i_T}; \qquad i_1 + i_2 + \cdots + i_T \leq m \tag{4.4.3}$$

which vanish with probability one for any given values of the sets

$$(\theta_1^*, \theta_2^*, \cdots, \theta_T^*) \quad \text{and} \quad (\theta_1, \theta_2, \cdots, \theta_T),$$

and

$$\theta_t^* - \theta_t = \sum_{i_1+i_2+\cdots+i_T\leq m} {}^{(t)}K_{i_1,i_2,\dots,i_T}\, g_{i_1,i_2,\dots,i_T}, \qquad t = 1, 2, \cdots, T, \tag{4.4.4}$$

where the K's do not depend upon θ_i^* and ξ_i's. For $T = 1$, the above reduce to the conditions in §4.3. We will now give an example in which (4.4.3) and (4.4.4) are satisfied. Let

$$p_N(x_1, x_2, \cdots, x_N; \theta_1, \theta_2) = \frac{1}{(2\pi\theta_1)^{N/2}}\left[\exp - \frac{1}{2\theta_1} \cdot \sum_{i=1}^{N} (x_i - \theta_2)^2\right] \tag{4.4.5}$$

We have

$$\theta_1^* = \sum_{i=1}^{N} (x_i - \bar{x})^2/(N - 1), \tag{4.4.6}$$

$$\theta_2^* = \sum_{i=1}^{N} x_i/N = \bar{x}, \tag{4.4.7}$$

unbiased estimates of θ_1 and θ_2 in (4.4.5). The joint distribution of θ_1^* and θ_2^* is given by

$$\begin{aligned} g(\theta_1^*, \theta_2^*; \theta_1, \theta_2) &= C \\ &\cdot \exp\left[\frac{-N(\theta_2^* - \theta_2)^2\,(N - 1)\theta_1^*}{2\theta_1} \cdot (\theta_1^*)^{N-1/2} \cdot (\theta_1)^{-N/2}.\right] \end{aligned} \tag{4.4.8}$$

It can be easily seen that the condition (4.4.3) is satisfied, and the estimates themselves can be put in the form

$$\theta_1^* = \theta_1 + \frac{2\theta_1^2}{N - 1} \cdot \frac{1}{g}\frac{\partial g}{\partial\theta_1} - \frac{\theta_1^2}{N(N - 1)} \cdot \frac{1}{g}\frac{\partial^2 g}{\partial\theta_1^2}, \tag{4.4.9}$$

$$\theta_2^* = \theta_2 + \frac{\theta_1}{N}\frac{1}{g}\frac{\partial g}{\partial\theta_2}. \tag{4.4.10}$$

It is thus seen that the 'concentration' ellipsoid for θ_1^*, θ_2^* coincides with the ellipsoid (3.1.21) for $m = 2$. On the other hand if we use $m = 1$, the condition (4.4.3) is satisfied but not the one in (4.4.4), as can be seen from (4.4.9), and thus the concentration ellipsoid strictly contains within itself the one given by the information matrix. It may be noted that for $m = 1$, the condition (4.4.3) merely requires that a system of sufficient statistics exists for estimating θ_1, $\theta_2, \cdots, \theta_T$. The reason is that the condition (4.4.3) takes the equivalent form

$$\frac{\partial h}{\partial \theta_i} = 0 \tag{4.4.11}$$

for $i = 1, 2, \cdots, T$ identically in $\xi_1, \xi_2, \cdots, \xi_{N-T}$ that is, that h is free of $\theta_1, \theta_2, \cdots, \theta_T$.

5. Miscellaneous. In §5.1–§5.3 we discuss certain properties of $\phi_i(n)$. In §5.4 we obtain conditions under which there exists no unbiased estimate of θ, having a finite variance, which is functionally dependent upon a given unbiased estimate θ^* of θ.

§5.1. Assume that there exists an "efficient" statistic $\theta^*(x_1, x_2, \cdots, x_N)$ for estimating θ, in probability density function (or probability)

$$p_N(x_1, x_2, \cdots, x_N; \theta).$$

That is,

$$\theta^*(x_1, x_2, \cdots, x_N) - \theta = K \cdot \phi_1(N) \tag{5.1.1}$$

where K as usual may only depend on θ. We postulate as usual the existence of all partial derivatives of p_N of all orders and also of K up to the third order with

$$\frac{d^3 K}{d\theta^3} = 0. \tag{5.1.2}$$

Further we assume that

$$\left| \frac{\partial^i p_N}{\partial \theta^i} \right| < T_i(x_1, x_2, \cdots, x_N)$$

where

$$\int T_i(x_1, x_2, \cdots, x_N) \prod_{u=1}^{N} dx_u \qquad \text{is finite for all } i$$

Under the above assumptions we will show that

$$\phi_0(N) = 1, \phi_1(N), \phi_2(N), \cdots, \phi_i(N), \cdots$$

form a set of orthogonal polynomials in $\phi_1(N)$ with respect to the weight function

$$p_N(x_1, x_2, \cdots, x_N; \theta).$$

PROOF: We can easily see that

$$\frac{\partial \phi_i}{\partial \theta} = \phi_{i+1} - \phi_i \cdot \phi_i \tag{5.1.3}$$

where $\phi_i(N)$ is shortened to ϕ_i for convenience. Differentiating (5.1.1) with respect to θ,

$$\frac{\partial \phi_1}{\partial \theta} = -\frac{1}{K}\frac{dK}{d\theta}\phi_1 - \frac{1}{K}. \tag{5.1.4}$$

Let us designate

$$z_i = \frac{1}{K}\frac{d^i K}{d\theta^i} \tag{5.1.5}$$

for all integral values of i. From (5.1.3) and (5.1.4), it follows that

$$\phi_2 - \phi_1^2 = -z_1\phi_1 - \frac{1}{K}. \tag{5.1.6}$$

Differentiating (5.1.6) further with regard to θ and using (5.1.3) and (5.1.6) we obtain

$$\phi_3 - \phi_1\phi_2 = -2z_1\phi_2 - \left(\frac{2}{K} + z_2\right)\phi_1. \tag{5.1.7}$$

Differentiating (5.1.7) with regard to θ, and using (5.1.2) we get

$$\phi_4 - \phi_1\phi_3 = -3z_1\phi_3 - \left(3z_2 + \frac{3}{K}\right)\phi_2. \tag{5.1.8}$$

We assume generally that

$$\phi_{i+1} - \phi_1\phi_i = -iz_1\phi_i - \left(\frac{i(i-1)}{2}z_2 + \frac{i}{K}\right)\phi_{i-1}. \tag{5.1.9}$$

Differentiating (5.1.9), and employing (5.1.3), (5.1.3) and (5.1.9) we obtain

$$\phi_{i+2} - \phi_1\phi_{i+1} = -(i+1)z_1\phi_{i+1} - \left(\frac{i(i+1)}{2}z_2 + \frac{i+1}{K}\right)\phi_i. \tag{5.1.10}$$

We know that (5.1.9) holds for $i = 1, 2, 3$; ϕ_0 being taken equal to one, and we have proved that if (5.1.9) is true for $i = j$, it is true for $i = j + 1$. Thus by mathematical induction (5.1.9) holds good for all integral values of i.

It is also clear from (5.1.6) and (5.1.9) that ϕ_i can be expressed as a polynomial in ϕ_1 of the ith degree, the coefficient of ϕ_1^i being equal to unity.

To complete the proof of our assertion we will now prove that

$$E(\phi_i \cdot \phi_j) = 0, \qquad i \neq j. \tag{5.1.11}$$

From (5.1.9)

$$\phi_1 \cdot \phi_i = \phi_{i+1} + iz_1\phi_i + \left(\frac{i(i-1)}{2}z_2 + \frac{i}{K}\right)\phi_{i-1}, \tag{5.1.12}$$

where i is any positive integer. We multiply both sides of (5.1.12) by ϕ_1 and reduce every product $\phi_i\phi_j$ to a linear combination of ϕ_{j+1}, ϕ_j and ϕ_{j-1} with the help of (5.1.12). Repeating this process $j - 1$ times ($j < i$) it follows that:

$$\phi_1^j \cdot \phi_i = \phi_{i+j} + \sum_{u=1}^{2j-1} d_u^i \cdot \phi_{i+j-u} + d_{2j}^i \cdot \phi_{i-j} \tag{5.1.13}$$

where d_u^i are functions of K, z_1 and z_2. From (5.1.13), by taking expectations of both sides,

$$E(\phi_i^j \cdot \phi_i) = 0, \qquad (j < i). \tag{5.1.14}$$

Now, since ϕ_j is a polynomial of the jth degree in ϕ_1 we conclude that (5.1.11) is true for all integral (positive) values of i.

Thus we obtain

$$\phi_0(N) = 1, \quad \phi_1(N), \quad \phi_2(N), \cdots, \quad \phi_i(N), \cdots, \tag{5.1.15}$$

as a set of orthogonal polynomials in $\phi_1(N)$, the weight function being

$$p_N(x_1, x_2, \cdots, x_N; \theta).$$

Furthermore

$$\phi_1^{i-1}\cdot\phi_i = \phi_{2i-1} + \sum_{u=1}^{2i-3} d_u^{i-1}\cdot\phi_{2i+1-u} + d_{2i-2}^{i-1}\cdot\phi_1 \tag{5.1.16}$$

where

$$d_{2i-2}^{i-1} = \prod_{j=2}^{i} B_j \tag{5.1.17}$$

and

$$B_j = \frac{j(j-1)}{2}\cdot z_2 + \frac{j}{K}. \tag{5.1.18}$$

Hence

$$E(\phi_1^i\cdot\phi_i) = \prod_{j=1}^{i} B_j. \tag{5.1.19}$$

Thus if we divide ϕ_i by $\sqrt{\prod_{j=1}^{i} B_j}$, (5.1.15) becomes the orthonormal set.

Some cases, where we obtain ϕ_i as orthogonal polynomials, are given below,

1. $p_N = \dfrac{1}{(\sqrt{2\pi})^r}\cdot e^{-\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2}, \qquad \phi_1 = \sum_{i=1}^{N}(x_i - \theta).$

2. $p_N = \dfrac{1}{(2\pi\theta)^{N/2}}\cdot e^{-\frac{1}{2\theta}\sum_{i=1}^{N}x_i^2}, \qquad \phi_1 = \dfrac{\sum_{i=1}^{N} x_i^2}{2\theta^2} - \dfrac{N}{2\theta}.$

3. $p_N = \theta^{\sum_{i=1}^{N}x_i}\cdot(1-\theta)^{N-\sum_{i=1}^{N}x_i} \quad \begin{pmatrix} x_i = 1 \text{ with prob. } \theta \\ \quad = 0 \text{ with prob. } 1-\theta \end{pmatrix},$

$$\phi_1 = \frac{\sum_{i=1}^{N} x_i - N\theta}{\theta(1-\theta)}.$$

4. $p_N = \dfrac{e^{-N\theta}\cdot\theta^{\sum_{i=1}^{N}x_i}}{\prod_{i=1}^{N} x_i!}, \qquad \phi_1 = \dfrac{\sum_{i=1}^{N} x_i - N\theta}{\theta}.$

A_i and B_i, the coefficients of ϕ_i and ϕ_{i-1} respectively in (5.1.12) for the above four cases are given as below:

	A_i	B_i
1.	0	$i \cdot N$
2.	$2i/\theta$	$\frac{i(i-1)}{\theta^2} + \frac{iN}{2\theta^2}$
3.	$\frac{i(1-2\theta)}{\theta(1-\theta)}$	$\frac{-i(i-1)}{\theta(1-\theta)} + \frac{iN}{\theta(1-\theta)}$
4.	i/θ	iN/θ

It may be mentioned that in all these cases $\{\phi_i\}$ are also a complete set of polynomials.

§5.2. Let $\sum_{i=1}^{m} K_i\phi_i(n)$, where $K_i(i = 1, 2, \cdots, m)$ depends upon θ be such that $\sum_{i=1}^{m} K_i\phi_i(n)$ and $\phi_i(n)$ satisfy the regularity conditions mentioned in §2.2. Then we will show that $\sum_{i=1}^{m} K_i\phi_i(n)$ cannot be a function of $x_1, x_2, \cdots, x_n$ alone except for constant zero.

Proof: Let us assume that $\sum_{i=1}^{m} K_i \cdot \phi_i(n)$ is independent of θ, that is, it is some statistic, say,

$$\theta^*(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{m} K_i \cdot \phi_i(n). \tag{5.2.1}$$

Taking expectations of both the sides, we obtain:

$$E(\theta^*(x_1, x_2, \cdots, x_n)) = \sum_{i=1}^{m} K_i \cdot E\phi_i(n) = 0. \tag{5.2.2}$$

Differentiating (5.2.2) i times with regard to θ, we have, because of the regularity conditions on $\phi_i(n)$ and $\theta^*(x_1, x_2, \cdots, x_n)$,

$$E(\theta^*(x_1, x_2, \cdots, x_n) \cdot \phi_i(n)] = 0, \qquad i = 1, 2, \cdots, m. \tag{5.2.3}$$

It may be noted this is similar to the result in (2.3). From (5.2.3) and (5.2.1) it follows that

$$E[\theta^*(x_1, x_2, \cdots, x_n)]^2 = 0. \tag{5.2.4}$$

Thus $\theta^*(x_1, x_2, \cdots, x_n)$ is zero with probability one, that is,

$$\sum_{i=1}^{m} K_i \cdot \phi_i(n),$$

if independent of θ, is zero with probability one. This proves our assertion that this cannot be a function of $x_1, x_2, \cdots, x_n$ alone except for constant zero.

From the foregoing we deduce the following conclusions:

I. $\phi_i(n)$ or any power of it cannot be a function of the observations free of θ.

II. If a statistic $\theta^*(x_1, x_2, \cdots, x_n)$, which is not a constant with probability one, can be put in the form

$$\theta^*(x_1, x_2, \cdots, x_n) = K_0 + \sum_{i=1}^{m} K_i \cdot \phi_i(n), \tag{5.2.5}$$

where m is some finite positive integer, then

(i) K_0 must depend upon θ,

(ii) The expression (5.2.5) for $\theta^*(x_1, x_2, \cdots, x_n)$ in $\phi_i(n)$ is unique.

(iii) No other unbiased estimate of K_0 satisfying the regularity conditions can be put in the form (5.2.5).

(iv) When $m = 1$, there is no other statistic except $a\theta^* + b$, where a and b are constants independent of θ, which can be put in the above form $K_0 + K_1 \cdot \phi_1(n)$, K_0 and K_1 are differentiable functions of θ and K_1 does not vanish for any θ under consideration.

(v) Let ξ be any function of $x_1, x_2, \cdots, x_n$ free of θ, satisfying the regularity conditions of §2.2 with $E(\xi) = 0$. Since the covariance between ξ and $\theta^*(x_1, x_2, \cdots, x_n)$ in (5.2.5) is equal to zero, the statistic of the form (5.2.5) has the least variance of all unbiased estimates of K_0 that satisfy the regularity conditions of §2.2.

Also, if the probability density or the probability function depends on more than one parameter, then all the above results except (iv) hold good if

$$\sum_{i=1}^{m} K_i \cdot \phi_i \cdot (n)$$

is replaced by

$$\sum_{i_1+i_2+\cdots+i_T \leq m} K_{i_1, i_2, \cdots, i_T} \cdot \phi_{i_1, i_2, \cdots, i_T}(n).$$

§5.3. Let us now prove the assertion made in (iv) of §5.2, when m is equal to one.

Suppose the contrary that there is a statistic $\theta_1^*(x_1 \cdot x_2, \cdots, x_n)$ which is of the form

$$\theta_1^*(x_1, x_2, \cdots, x_n) = L_0 + L_1 \cdot \phi_1(n). \tag{5.3.1}$$

$\theta^*(x_1, x_2, \cdots, x_n)$, of course, has the form

$$\theta^*(x_1, x_2, \cdots, x_n) = K_0 + K_1 \cdot \phi_1(n). \tag{5.3.2}$$

We will assume K_0, K_1, L_0, L_1 to be differentiable functions of θ and that K_1, L_1 do not vanish for values of θ under consideration.

Differentiating, with respect to θ, the expressions in (5.3.1) and (5.3.2 , we have

$$\frac{dL_0}{d\theta} + \frac{dL_1}{d\theta} \cdot \phi_1 + L_1(\phi_2 - \phi_1^2) = 0; \tag{5.3.3}$$

$$\frac{dK_0}{d\theta} + \frac{dK_1}{d\theta} \cdot \phi_1 + K_1(\phi_2 - \phi_1^2) = 0, \tag{5.3.4}$$

where ϕ_i is short for $\phi_i(n)$. Taking the expectations of the above and rearranging, it follows that

$$E(\phi_1^2) = \frac{1}{L_1}\frac{dL_0}{d\theta} = \frac{1}{K_1}\frac{dK_0}{d\theta}. \tag{5.3.5}$$

From (5.3.3) to (5.3.5), we deduce that

$$\frac{1}{L_1}\frac{dL_1}{d\theta} = \frac{1}{K_1}\frac{dK_1}{d\theta}. \tag{5.3.6}$$

Now solving the above differential equation, we get

$$L_1 = aK_1, \tag{5.3.7}$$

where a is a constant independent of θ. From (5.3.5) and (5.3.7) it follows that

$$L_0 = aK_0 + b, \tag{5.3.8}$$

where b is a constant independent of θ. From (5.3.7) and (5.3.8) we conclude that the statistic in (5.3.1) must be of the form a $\theta^* + b$, which proves our assertion. An immediate consequence is that if there exists an efficient statistic for estimating $\gamma(\theta)$, then no other function of θ except a $\gamma(\theta) + b$ can have an efficient estimate.[3]

§5.4. If $\theta^*(x_1, x_2, \cdots, x_n)$ is an unbiased estimate of θ satisfying the following conditions:

(i) Among all unbiased estimates of θ having finite variances, which are also functions of θ^*, θ^* is one with the least variance,

(ii) For all θ there exists a complete set of polynomials with respect to the distribution function of θ^*, then there exists no unbiased estimate of θ with a variance, which is functionally dependent upon θ^*, except θ^* itself.

PROOF: Let θ^* be the unbiased estimate of θ which has the least variance among all unbiased estimates of θ which are functions of θ^*. Further let $S(\theta^*)$ be any function of θ^*, free of θ, whose expectation exists and is equal to zero. Let the variance of $S(\theta^*)$ be finite. It is well known that for any such $S(\theta^*)$

$$E(\theta^* S(\theta^*)) = 0. \tag{5.4.1}$$

Now $\theta^* S(\theta^*)$ in turn having expectation equal to zero, we obtain

$$E(\theta^{*2} S(\theta^*)) = 0. \tag{5.4.2}$$

Repeating the above i times we obtain, in general, that

$$E(\theta^{*i} S(\theta^*)) = 0 \tag{5.4.3}$$

[3] We assume the existence of $\frac{d^i\gamma}{d\theta^i}$ $(i = 1, 2)$ and $\frac{d}{d\theta}(E\phi_1^2)$ for all θ, and also postulate that $\frac{d\gamma(\theta)}{d\theta}$ and $E(\phi_1^2)$ do not vanish for any θ under consideration.

for all positive integers i. From the above, with the help of condition (ii), we conclude that $S(\theta^*)$ must be equal to zero. Thus if $H(\theta^*)$ is an unbiased estimate of θ with finite variance, then from above, $H(\theta^*) - \theta^*$, having the expectation zero and a finite variance, must be zero with probability one. Thus $H(\theta^*)$ is the same as θ^*, which proves the result.

EXAMPLE. If θ^* is of the form (5.2.7) and condition (ii) is satisfied, then there is no function of θ^*, free of θ and having a finite variance, whose expectation is K_0.

Conditions (i) and (ii) above are satisfied for estimating θ in the examples quoted at the end of the section 5.1, and thus in these cases the result holds good when θ^* is the efficient estimate.

I am highly thankful to Professor J. Wolfowitz for his guidance and help in this research.

REFERENCES

[1] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946, p. 480.

[2] C. R. RAO, "Information and the accuracy attainable in the estimation of statistical parameters," *Calcutta Math. Soc. Bull.*, September, 1945.

[3] A. BHATTACHARYYA, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946); also "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1947).

[4] H. CRAMÉR, "Contributions to the theory of statistical estimation," *Skandinavisk Aktuar. tids.*, Vol. 29 (1946), pp. 5–94.

[5] J. WOLFOWITZ, "Efficiency of sequential estimates," *Annals of Math. Stat.*, Vol. 18 (1947).

[6] BLACKWELL AND GIRSHICK, "A lower bound for the variance of some unbiased sequential estimates," *Annals of Math. Stat.*, Vol. 18 (1947).

[7] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Am. Math. Soc. Trans.*, Vol. 39 (1936), p. 399.