AN EXTENSION OF THE CRAMÉR-RAO INEQUALITY[1]

By John J. Gart[2]

Virginia Polytechnic Institute

1. Review of the literature and summary. Cramér ([6], p. 474 ff.), Darmois [8], Fréchet [10], and Rao [14] derived independently a lower bound for the mean square error of an estimate t of a parameter which appears in a frequency function of a specified form. This epxression, alternately termed the Cramér-Rao inequality or the information limit, is

$$E(t - \alpha)^2 \geqq [E(t) - \alpha]^2 + \frac{\left[\frac{\partial E(t)}{\partial \alpha}\right]^2}{E\left(\frac{\partial \ln \phi}{\partial \alpha}\right)^2} \tag{1.1}$$

where ϕ is the likelihood of the sample. The expression $E(\partial \ln \phi/\partial\alpha)^2$ is called the information on α and is sometimes denoted by $I(\alpha)$. Under rather general conditions it can be shown equal to $E(-\partial^2 \ln \phi/\partial\alpha^2)$.

The equality in (1.1) is reached if and only if,

$$\phi = \phi_1 e^{tV(\alpha)+W(\alpha)} \tag{1.2}$$

where t and ϕ_1 are functions of the observations alone and $V(\alpha)$ and $W(\alpha)$ are functions of α alone. By the results of Pitman [13] and Koopman [12], the form of (1.2) implies that t must be a sufficient statistic. The fact that this form of the likelihood yields a minimum variance estimate was first pointed out by Aitken and Silverstone [1]. If we have n observations which are independently and identically distributed, the frequency function of the underlying population must be of the so-called Pitman-Koopman form,

$$f(x; \alpha) = u(\alpha)h(x)e^{P(\alpha)g(x)} \tag{1.3}$$

and t must be a function of $\sum_{i=1}^n g(x_i)$ for the equality in (1.2) to hold.

Several extensions of the basic inequality have been derived. Bhattacharyya [4] and Chapman and Robbins [5] have derived results which yield more stringent inequalities in certain instances. Wolfowitz [21] has extended the result to sequential sampling situations. Cramér [7], Darmois [8], and Barankin [2] have considered joint bounds on sets of estimates of parameters and Hammersley [11] has derived a lower bound of the mean square error of an estimate for the situation in which the parameter to be estimated can only assume discrete values. Barankin [3] has also considered lower bounds on the general absolute central moments of the estimate.

All these results assume that the parameters involved are constants. Here we

Received May 21, 1957; revised May 25, 1958.

[1] Research sponsored by the National Science Foundation under grant NSF G-1858.

[2] Present address: Department of Biostatistics, The Johns Hopkins University.

shall consider the case where the parameters are random variables. Thus the lower bound of the mean square error of an estimate will take into account the variability due to both the observations and the parameters involved. Necessary and sufficient conditions for equality of the extended inequality are derived. Most unfavorable distributions, i.e., distributions which maximize the lower bound, are defined, and several examples are given. Extensions analogous to those of Bhattacharyya [4] and Wolfowitz [21] are also considered. Finally, bounds on the variance of linear estimates of the mean of the parameter are derived.

2. Notation. Consider a frequency function $f(x \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2), \cdots, \theta_s)$, the function being specified when $\boldsymbol{\theta}$ is specified. Further, $\boldsymbol{\theta}$ is a random variable having the distribution $G(\boldsymbol{\theta})$ defined over a non-degenerate range A_s. Let $\mathbf{X} = (x_1, x_2, \cdots, x_n)$ be a random sample from a randomly chosen population having the specified frequency function. Let $t_k = t_k(\mathbf{X})$ be an estimate of θ_k, $1 \leqq k \leqq s$, functionally independent of θ_k. Denote $E(t_k \mid \boldsymbol{\theta})$ by $\psi_k(\boldsymbol{\theta})$ and the conditional likelihood of the sample by $\phi(\mathbf{X} \mid \boldsymbol{\theta})$, which in general will be $\prod_{i=1}^n f(x_i \mid \boldsymbol{\theta})$.

3. The continuous case.[3] If $f(x \mid \boldsymbol{\theta})$ is a density, assume $\partial\phi/\partial\theta_k$ exists for all $\boldsymbol{\theta}$ in A_s and $\mid \partial\phi/\partial\theta_k \mid < H(\mathbf{X})$ where H and t_kH are integrable over R_n, the range of $\mathbf{X}$, which is independent of θ_k. We have

$$1 = \int_{R_n} \phi \, d\mathbf{X} \tag{3.1}$$

and

$$\psi_k(\boldsymbol{\theta}) = \int_{R_n} t_k \phi \, d\mathbf{X}. \tag{3.2}$$

By the assumptions just made (see Cramér [6], p. 66 and p. 475), we may differentiate under the integral signs in (3.1) and (3.2) and obtain

$$0 = \int_{R_n} \frac{\partial\phi}{\partial\theta_k} d\mathbf{X} = \int_{R_n} \frac{\partial \ln \phi}{\partial\theta_k} \phi \, d\mathbf{X} \tag{3.3}$$

and

$$\frac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k} = \int_{R_n} t_k \frac{\partial\phi}{\partial\theta_k} d\mathbf{X} = \int_{R_n} t_k \frac{\partial \ln \phi}{\partial\theta_k} \phi \, d\mathbf{X}. \tag{3.4}$$

Finding expectations of (3.3) and (3.4) with respect to $\boldsymbol{\theta}$, we have

$$0 = \int_{A_s} \int_{R_n} \frac{\partial \ln \phi}{\partial\theta_k} \phi \, d\mathbf{X} \, dG(\boldsymbol{\theta}) \tag{3.5}$$

and

$$E\left(\frac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k}\right) = \int_{A_s} \frac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k} dG(\boldsymbol{\theta}) = \int_{A_s} \int_{R_n} t_k \frac{\partial \ln \phi}{\partial\theta_k} \phi \, d\mathbf{X} \, dG(\boldsymbol{\theta}). \tag{3.6}$$

[3] Results analogous to those of this and the two succeeding paragraphs have been obtained by Schützenberger [15] for the *a posteriori* distribution of θ_k.

By the Schwarz inequality we may write

$$
\begin{aligned}
&\left\{\int_{A_s}\int_{R_n}[t_k-\theta_k-E[\psi_k(\boldsymbol{\theta})]+E(\theta_k)]^2\phi\,d\mathbf{X}\,dG(\boldsymbol{\theta})\right\}\\
(3.7)\qquad &\cdot\left\{\int_{A_s}\int_{R_n}\left(\frac{\partial\ln\phi}{\partial\theta_k}\right)^2\phi\,d\mathbf{X}\,dG(\boldsymbol{\theta})\right\}\\
&\geqq\left\{\int_{A_s}\int_{R_n}[t_k-\theta_k-E[\psi_k(\boldsymbol{\theta})]+E(\theta_k)]\left(\frac{\partial\ln\phi}{\partial\theta_k}\right)\phi\,d\mathbf{X}\,dG(\boldsymbol{\theta})\right\}^2.
\end{aligned}
$$

In view of (3.3), (3.5), and (3.6); (3.7) may be written,[4]

$$
\operatorname{Var}(t_k-\theta_k)EE\left[\left(\frac{\partial\ln\phi}{\partial\theta_k}\right)^2\Big|\,\boldsymbol{\theta}\right]\geqq E^2\left(\frac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k}\right),
$$

and if $EE[(\partial\ln\phi/\partial\theta_k)^2\mid\boldsymbol{\theta}]\neq 0$, then

$$
\operatorname{Var}(t_k-\theta_k)\geqq\frac{E^2\left(\dfrac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k}\right)}{E[I_k(\boldsymbol{\theta})]},
$$

where $I_k(\boldsymbol{\theta})=E[(\partial\ln\phi/\partial\theta_k)^2\mid\boldsymbol{\theta}]$. Since $\operatorname{Var}(t_k-\theta_k)=EE[(t_k-\theta_k)^2\mid\boldsymbol{\theta}]-E^2[\psi_k(\boldsymbol{\theta})-\theta_k]$, we may write

$$
(3.8)\qquad EE[(t_k-\theta_k)^2\mid\boldsymbol{\theta}]\geqq E^2[\psi_k(\boldsymbol{\theta})-\theta_k]+\frac{E^2\left(\dfrac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k}\right)}{E[I_k(\boldsymbol{\theta})]}.
$$

If $\psi_k(\boldsymbol{\theta})=\theta_k$, (3.8) may be written

$$
(3.9)\qquad EE[(t_k-\theta_k)^2\mid\boldsymbol{\theta}]\geqq\frac{1}{E[I_k(\boldsymbol{\theta})]}.
$$

When $[\partial^2\phi/\partial\theta_k^2]<K(\mathbf{X})$, where $K(\mathbf{X})$ is integrable over R_n, it is well known that

$$
E\left[\left(\frac{\partial\ln\phi}{\partial\theta_k}\right)^2\Big|\,\boldsymbol{\theta}\right]=E\left[-\frac{\partial^2\ln\phi}{\partial\theta_k^2}\Big|\,\boldsymbol{\theta}\right].
$$

Then we can write $I_k(\boldsymbol{\theta})=E[-\partial^2\ln\phi/\partial\theta_k^2\mid\boldsymbol{\theta}]$ in (3.8) and (3.9). Since $I_k(\boldsymbol{\theta})$ is called the amount of information, $E[I_k(\boldsymbol{\theta})]$ may logically be termed the mean amount of information.

It should be noted that the derivation of these inequalities is equally applicable to samples from multivariate populations.

4. The discrete case. Suppose that $f(x\mid\boldsymbol{\theta})$ is a discrete frequency function whose range, R_1, may be finite or denumerably infinite but independent of θ_k. Assume ϕ is a continuous function of θ_k for all $\mathbf{X}$ in R_n and $\boldsymbol{\theta}$ in A_s, and that $\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\partial\phi/\partial\theta_k$ and $\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}t_k(\partial\phi/\partial\theta_k)$ converge uni-

[4] EE symbolizes taking the expectation with respect to $\mathbf{X}$ for fixed $\boldsymbol{\theta}$ and then with respect to $\boldsymbol{\theta}$.

formly in A_s. By operations similar to those employed in paragraph 3, we find

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\frac{\partial \ln \phi}{\partial \theta_k}\phi = 0 \tag{4.1}$$

and

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} t_k\frac{\partial \ln \phi}{\partial \theta_k}\phi = \frac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k} \tag{4.2}$$

since the assumptions just made allow differentiation under the summation signs. By the Schwarz inequality we may write

$$\left\{\int_{A_s}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}[t_k - \theta_k - E[\psi_k(\boldsymbol{\theta})] + E(\theta_k)]^2\phi\, dG(\boldsymbol{\theta})\right\}$$
$$\cdot\left\{\int_{A_s}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(\frac{\partial \ln \phi}{\partial \theta_k}\right)^2\phi\, dG(\boldsymbol{\theta})\right\}$$
$$\geqq \left\{\int_{A_s}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}[t_k - \theta_k - E[\psi_k(\boldsymbol{\theta})] + E(\theta_k)]\frac{\partial \ln \phi}{\partial \theta_k}\phi\, dG(\boldsymbol{\theta})\right\}^2.$$

Following steps analogous to those in paragraph 3, we arrive at (3.8) and (3.9), for the discrete case.

5. Conditions for Equality. The condition under which the equalities in (3.8) and (3.9) hold are set forth in the following three theorems.

THEOREM 1. *If*

(i) $\Pr.\ \{E[(t_k - \theta_k)^2 \mid \boldsymbol{\theta}] = c_1\} = 1,$

(ii) $\Pr.\ [\psi_k(\boldsymbol{\theta}) = c_2] = 1,$

(iii) $\Pr.\ [\theta_k = c_3] = 1,$

(iv) $\Pr.\left[\dfrac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k} = c_4\right] = 1,$ *and*

(v) $\Pr.\ [I_k(\boldsymbol{\theta}) = c_5] = 1,$

where c_i, $i = 1, 2, \cdots, 5$, are constants, then the equality in (3.8) *holds if and only if t_k is a sufficient estimate of θ_k.*

PROOF. Under the conditions of the theorem (3.8) reduces to (1.1), for which it has been shown by Rao [14] that the equality holds if and only if t_k is a sufficient estimate of θ_k.

THEOREM 2. *If t_k is an unbiased sufficient estimate of θ_k, then the equality in* (3.9) *holds if and only if* Pr. $[I_k(\boldsymbol{\theta}) = c_5] = 1$, *where c_5 is a constant.*

PROOF. Since t_k is an unbiased sufficient estimate of θ_k, we have from Rao [14], $E[(t_k - \theta_k)^2 \mid \boldsymbol{\theta}] = I_k(\boldsymbol{\theta})^{-1}$. Taking expectations with respect to $\boldsymbol{\theta}$, we have

$$EE[(t_k - \theta_k)^2 \mid \boldsymbol{\theta}] = E[I_k(\boldsymbol{\theta})^{-1}]. \tag{5.1}$$

Now equality of (3.9) requires that,

$$EE[(t_k - \theta_k)^2 \mid \boldsymbol{\theta}] = \frac{1}{E[I_k(\boldsymbol{\theta})]}. \tag{5.2}$$

Combining (5.1) and (5.2), we have $E[I_k(\boldsymbol\theta)^{-1}] = 1/E[I_k(\boldsymbol\theta)]$, or

$$E[I_k(\boldsymbol\theta)^{-1}]E[I_k(\boldsymbol\theta)] = 1. \tag{5.3}$$

This can be written,

$$\left\{\int_{A_s} I_k(\boldsymbol\theta)^{-1}\, dG(\boldsymbol\theta)\right\}\left\{\int_{A_s} I_k(\boldsymbol\theta)\, dG(\boldsymbol\theta)\right\} = 1. \tag{5.4}$$

Now by the Schwarz inequality we have

$$\left\{\int_{A_s} I_k(\boldsymbol\theta)^{-1}\, dG(\boldsymbol\theta)\right\}\left\{\int_{A_s} I_k(\boldsymbol\theta)\, dG(\boldsymbol\theta)\right\} \geqq \left\{\int_{A_s} dG(\boldsymbol\theta)\right\}^2 = 1. \tag{5.5}$$

Obviously when the equality holds in (5.5) it is equivalent to (5.4). But the equality in (5.5) is achieved if and only if for a constant c_5 independent of θ, $c_5[I_k(\boldsymbol\theta)]^{-\frac{1}{2}} = [I_k(\boldsymbol\theta)]^{\frac{1}{2}}$ with probability one; that is, if and only if

$$\Pr\,[I_k(\boldsymbol\theta) = c_5] = 1,$$

which proves the theorem.

Before proceeding to theorem 3, we cite the following definition.

DEFINITION. Any pair of $\phi(\mathbf{X} \mid \boldsymbol\theta)$ and $G(\boldsymbol\theta)$ wherein any one of the assumptions (i)–(iv) inclusive of Theorem 1 does not hold for the θ_k under consideration is termed the non-trivial estimation case.

THEOREM 3. For the non-trivial estimation case the equality in (3.8) is achieved if and only if t_k is an unbiased sufficient estimate of θ_k which is normally distributed with constant variance equal to $I_k(\boldsymbol\theta)^{-1}$. Consequently the equality in (3.9) is achieved under the same conditions.

PROOF. For the non-trivial estimation case the equality in (3.7) and consequently in (3.8) and (3.9) is achieved if and only if there exists a λ independent of $\mathbf{X}$ and $\boldsymbol\theta$ such that,

$$\lambda \frac{\partial \ln \phi}{\partial \theta_k} = t_k - \theta_k - E[\psi_k(\boldsymbol\theta)] + E(\theta_k),$$

for almost all $\mathbf{X}$ in R_n and $\boldsymbol\theta$ in A_s. Integrating, we have

$$\lambda \ln \phi = \theta_k t_k - \theta_k^2/2 - \theta_k E[\psi_k(\boldsymbol\theta)] + \theta_k E(\theta_k) + C_1(\mathbf{X}, \boldsymbol\theta^*),$$

where $\boldsymbol\theta^* = (\theta_1, \theta_2, \cdots, \theta_{k-1}, \theta_{k+1}, \cdots, \theta_s)$. We thus have

$$\phi = C_2(\mathbf{X}, \boldsymbol\theta^*) \exp\,(1/\lambda)\{\theta_k t_k - (\theta_k^2/2) - \theta_k E[\psi_k(\boldsymbol\theta)] + \theta_k E(\theta_k)\}.$$

This is a special case of the form, found by Pitman [13] and Koopman [12], wherein t_k is a sufficient statistic for θ_k. Integrating both sides of the above equation over R_n, we have

$$\exp\,\{(\theta_k/\lambda)[E(\theta_k) - E(\psi_k(\boldsymbol\theta))]\} \int_{R_n} C_2(\mathbf{X}, \boldsymbol\theta^*) \exp\,(\theta_k t_k/\lambda)\, d\mathbf{X} = \exp\,(\theta_k^2/2\lambda).$$

Make the change of variables in the integral, $z_i = z_i(\mathbf{X})$, $i = 1, 2, \cdots, n-1$, $t_k = t_k(\mathbf{X})$, where $t_k(\mathbf{X})$ and the $z_i(\mathbf{X})$ are unique, continuous, and possess continuous partial derivatives. Further the transformation is one-to-one. Then we have,

$$\exp\{(\theta_k/\lambda)[E(\theta_k) - E[\psi_k(\mathbf{\theta})]]\} \int_{B_n} C_3(\mathbf{Z}, t_k, \mathbf{\theta}^*) \exp(\theta_k t_k/\lambda)\, d\mathbf{Z}\, dt_k = \exp(\theta_k^2/2\lambda),$$

where B_n is the range of $(\mathbf{Z}, t_k) = (z_1, z_2, \cdots, z_{n-1}, t_k)$. If we integrate out $\mathbf{Z}$, we have

$$\int_{b_k} C_4(t_k, \mathbf{\theta}) e^{\theta_k t_k/\lambda}\, dt_k = e^{\theta_k^2/2\lambda}, \tag{5.6}$$

where b_k, the range of t_k, may be taken from $-\infty$ to ∞. Then the left hand side of (5.6) is a bilateral Laplace transform of $C_4(t_k, \mathbf{\theta})$ with argument θ_k/λ. Recall that θ_k has a non-degenerate range say $\gamma_1 < \theta_k < \gamma_2$. Obviously $e^{\theta_k/2\lambda}$ exists at $\theta_k = \gamma_1 + \epsilon_1$ and $\theta_k = \gamma_2 - \epsilon_2$, where $\epsilon_1, \epsilon_2 > 0$, such that $\epsilon_2 + \epsilon_1 < \gamma_2 - \gamma_1$. Thus we can apply the theorem of Widder ([19], p. 238), and conclude that the integral in (5.6) converges for θ_k in the vertical strip of the complex plane, $\gamma_1 + \epsilon_1 < \theta_k < \gamma_2 - \epsilon_2$. Thus we can apply the uniqueness theorem of the bilateral Laplace transform (see Widder [19], p. 243) and conclude that $C_4(t_k, \mathbf{\theta}) = (1/\sqrt{2\pi\lambda}) \exp(-t_k^2/2\lambda)$. Therefore, for equality, the frequency function of t_k must be

$$h(t_k \mid \theta_k) = \frac{1}{\sqrt{2\pi\lambda}} e^{-(t_k-\theta_k)^2/2\lambda}, \qquad -\infty < t_k < \infty,$$

where obviously $\lambda = I_k(\mathbf{\theta})^{-1}$. Further the equality holds regardless of the form of the marginal distribution of θ_k.

It should be noted that though theorems 2 and 3 require that $I_k(\mathbf{\theta})$ be a constant, it is not necessary that all the components of $\mathbf{\theta}$ occurring in $I_k(\mathbf{\theta})$ be constants. It is possible, for instance, that the components of $\mathbf{\theta}$ occurring in $I_k(\mathbf{\theta})$ have a singular multivariate distribution such that all the probability is located on the hyperplane $I_k(\mathbf{\theta}) = $ constant.

Obviously the sample mean from a normal population with constant variance satisfies theorem 3. However, it is by no means the only such estimate. Let

$$f(x \mid \theta) = \frac{1}{x\sqrt{2\pi c}} e^{-(\ln x - \theta)^2/2c}, \qquad 0 < x < \infty,$$

where c is a constant, which is the so-called logarithmico-normal distribution (see Cramér [6], p. 220). Then, $t = \sum_{i=1}^{n} \ln x_i/n$ is normally distributed with mean θ and variance c/n, which is the minimum variance attainable under the extended inequality.

A situation in which a parameter is assumed to be a random variable is the analysis of variance model II of Eisenhart [9]. The simplest case is the one way classification. Here the model is $x_{ij} = \alpha_i + \epsilon_{ij}$, where x_{ij} is the jth observation in the ith class. We assume there are k classes where the ith class has n_i observations. Suppose α_i and ϵ_{ij} are random samples of size m and N, $N =$

$\sum_{i=1}^{k} n_i$, from two normally distributed populations having means μ and zero respectively, and variances σ_α^2 and σ_ϵ^2 respectively. Then,

$$\phi = \prod_{i=1}^{m} \prod_{j=1}^{n_i} f(x_{ij} \mid \alpha_i) = \frac{1}{(\sqrt{2\pi}\sigma_\epsilon)^N} \exp\left\{-\sum_i \sum_j \frac{(x_{ij} - \alpha_i)^2}{2\sigma_\epsilon^2}\right\},$$

and

$$EE\left(-\frac{\partial^2 \ln \phi}{\partial \alpha_i^2}\middle| \alpha_i\right) = \frac{n_i}{\sigma_\epsilon^2}, \qquad i = 1, 2, \cdots, m. \tag{5.7}$$

The ML estimate for α_i is $\hat{\alpha}_i = (\sum_{j=1}^{n_i} x_{ij})/n_i$, $i = 1, 2, \cdots, m$. Here $E(\hat{\alpha}_i \mid \alpha_i) = \alpha_i$, and $EE[(\hat{\alpha}_i - \alpha_i)^2 \mid \alpha_i] = \sigma_\epsilon^2/n_i$, which by (5.7) is the minimum mean square error. Notice the assumption of normality of α_i was not required for equality.

6. Most unfavorable distributions.[5] In most cases the $G(\boldsymbol{\theta})$ is not known, so the lower bound on the mean square error cannot be found. If $\psi_k(\boldsymbol{\theta}) = \theta_k$, it is of interest to know the greatest value the lower bound can attain, as well as the set of $G(\boldsymbol{\theta})$ which produces it. To this end define $G_k^*(\boldsymbol{\theta})$ to be a most unfavorable distribution with respect to θ_k if $\int_{A_s} I_k(\boldsymbol{\theta})\, dG_k^*(\boldsymbol{\theta}) \leqq \int_{A_s} I_k(\boldsymbol{\theta})\, dG(\boldsymbol{\theta})$, for all $G(\boldsymbol{\theta})$ defined over A_s.

If $I_k(\boldsymbol{\theta})$ has a unique minimum with respect to that subset of the parameters appearing therein, then a most unfavorable distribution is one for which the marginal distribution of these parameters is trivial. It may be that $I_k(\boldsymbol{\theta})$ is independent of all parameters so that all $G(\boldsymbol{\theta})$ are most unfavorable distributions. A case in point is the Cauchy distribution,

$$f(x \mid \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \qquad -\infty < x < \infty,$$

where

$$I(\theta) = \frac{4n}{\pi} \int_{-\infty}^{\infty} \frac{(x - \theta)^2}{[1 + (x - \theta)^2]^3}\, dx = \frac{n}{2}.$$

Here $EE[(t - \theta)^2 \mid \theta] \geqq 2/n$ regardless of the form of $G(\theta)$. There are also cases in which no most unfavorable distribution exists except possibly when from some prior information A_s is restricted.

7. Most unfavorable distributions for some Laplacian distributions. M. C. K. Tweedie [16], [17] has called a distribution Laplacian if it belongs to the general class of distributions for which the sample mean is a sufficient statistic for one of its parameters. The general form of such a distribution's frequency function is

$$f(x \mid \theta_1, \theta_2) = e^{-xq(\theta_1) - \theta_2 F(\theta_1)} h(x, \theta_2).$$

This is, of course, a special case of the Pitman-Koopman form (1.3). Here we have,

$$E[I(\theta_1, \theta_2)] = \int_{A_2} [m(\theta_1, \theta_2) q''(\theta_1) + \theta_2 F''(\theta_1)] g(\theta_1, \theta_2)\, d\theta_1\, d\theta_2,$$

[5] For an analogous concept, least favorable distributions, see Wald ([18], p. 18).

where $E(x \mid \theta_1, \theta_2) = m(\theta_1, \theta_2)$. If $I(\theta_1, \theta_2)$ has an absolute minimum for some subset A_2' of A_2, and (θ_1^0, θ_2^0) is an element of A_2', then $E[I(\theta_1, \theta_2)] \geqq m(\theta_1^0, \theta_2^0)q''(\theta_1^0) + \theta_2^0 F''(\theta_1^0)$, for all (θ_1, θ_2) in A_2. Further, this is the absolute minimum attainable by $E[I(\theta_1, \theta_2)]$. It is reached when $dG(\theta_1, \theta_2) = 0$, for (θ_1, θ_2) not an element of A_2'. Thus, we have found a set of most unfavorable distributions. This result will now be applied to several specific Laplacian distributions.

Type a. $\theta_2 = 1$. This includes the binomial, Pascal and Poisson distributions.

(1) Binomial distribution.

$$f(x \mid \theta_1) = e^{x \ln(\theta/1-\theta_1) - \ln(1-\theta_1)}, \qquad x = 0, 1, \qquad 0 < \theta_1 < 1,$$

$$q(\theta_1) = -\ln\left(\frac{\theta_1}{1-\theta_1}\right), \qquad F(\theta_1) = \ln(1-\theta_1), \qquad m(\theta_1) = \theta_1,$$

and

$$I(\theta_1) = \frac{1}{\theta_1(1-\theta_1)}.$$

$I(\theta_1)$ has a unique absolute minimum at $\theta_1 = \frac{1}{2}$, so that[6] $G^*(\theta_1) = \epsilon(\theta_1 - \frac{1}{2})$ is the only most unfavorable distribution. Further $E[I(\theta_1)] \nless 4N$.

(2) Pascal distribution.

$$f(x \mid \theta_1) = e^{x \ln(1-\theta_1) - r \ln(1-\theta_1/\theta_1)} \binom{x-1}{r-1}, \qquad x = r, r+1, \cdots; \qquad 0 \leqq \theta_1 \leqq 1,$$

$$q(\theta_1) = -\ln(1-\theta_1), \qquad F(\theta_1) = \ln\left(\frac{1-\theta_1}{\theta_1}\right), \qquad m(\theta_1) = r/\theta_1,$$

where r is a known fixed positive integer. $I(\theta_1) = r/\theta_1^2(1-\theta_1)$, which has a unique absolute minimum at $\theta_1 = \frac{2}{3}$. Therefore $G^*(\theta_1) = \epsilon(\theta_1 - \frac{2}{3})$ is the only most unfavorable distribution and $E[I(\theta_1)] \nless 17N/4$. It should be noted x/r is not an unbiased estimate of θ_1. If we consider $\alpha = 1/\theta_1$ as the parameter to be estimated, then

$$f(x \mid \alpha) = e^{-x \ln(\alpha/\alpha-1) - \ln(\alpha-1)} \binom{x-1}{r-1}, \qquad x = r, r+1, \cdots; 1 < \alpha < \infty,$$

$$q(\alpha) = \ln\left(\frac{\alpha}{\alpha-1}\right), \qquad F(\alpha) = \ln(\alpha-1), \qquad m(\alpha) = r\alpha,$$

and

$$I(\alpha) = \frac{r}{\alpha(\alpha-1)}.$$

Here x/r is an unbiased estimate of α. However the expression $I(\alpha)$ does not have an absolute minimum in A_1, i.e., $1 \leqq \alpha < \infty$, but rather it has a limit

[6] Following Cramér ([6], p. 192), the distribution function $\epsilon(x - a) = \begin{cases} 0 \text{ for } x < a \\ 1 \text{ for } x \geqq a. \end{cases}$

of zero as $\alpha \to \infty$. Thus letting $\alpha \to \infty$ produces a most unfavorable situation, which is equivalent to letting $\theta_1 = 0$. It is interesting to note that though $\theta_1 = \frac{2}{3}$ was the most unfavorable situation when estimating θ_1, when estimating $\alpha = 1/\theta_1$ we have, in effect, that $\theta_1 = 0$ is the most unfavorable situation. Thus we have established that "most unfavorableness" is not an invariant property.

(3) Poisson distribution.

$$f(x \mid \theta_1) = e^{x\ln\theta_1 - \theta_1} \frac{1}{x!}, \qquad x = 0, 1, 2, \cdots; \qquad 0 < \theta_1 < \infty$$

$$q(\theta_1) = -\ln\theta_1, \qquad F(\theta_1) = \theta_1, \qquad m(\theta_1) = \theta_1,$$

$$I(\theta_1) = 1/\theta_1.$$

$I(\theta_1)$ has no absolute minimum in A_1, but rather has a limit of zero as $\theta_1 \to \infty$. However, if from some prior consideration we can restrict $\theta_1 < a$, then a most unfavorable distribution is $G^*(\theta_1) = \epsilon(\theta_1 - a)$.

Type b. $\theta_2 \neq 1$, $q(\theta_1) = \theta_1$. Immediately we have $q''(\theta_1) = 0$, and $I(\theta_1, \theta_2) = \theta_2 F''(\theta_1)$.

(1) Gamma distribution.

$$f(x \mid \theta_1, \theta_2) = e^{-x\theta_1 + \theta_2 \ln\theta_1} [x^{\theta_1 - 1}/\Gamma(\theta_2)], \qquad x > 0, \qquad \theta_2 > 0, \qquad \theta_1 > 0.$$

$$F(\theta_1) = -\ln\theta_1, \qquad I(\theta_1, \theta_2) = \theta_2/\theta_1^2.$$

Here $I(\theta_1, \theta_2) \to 0$, as $\theta_1 \to \infty$ and/or $\theta_2 \to 0$, but no most unfavorable distribution can be cited unless we assume $\theta_2/\theta_1^2 \leqq a$.

(2) Normal distribution (parameters adjusted).

$$f(x \mid \theta_1, \theta_2) = e^{-\theta_1 x - \theta_2(\theta_1^2/2)} \frac{e^{-(x^2/2\theta_2)}}{\sqrt{2\pi\theta_2}},$$

$$-\infty < x < \infty, \qquad \theta_2 > 0, \qquad -\infty < \theta_1 < \infty,$$

where $\theta_2 = \sigma^2$, $\theta_1 = -\mu/\sigma^2$, in the usual notation.

$$F(\theta_1) = \theta_1^2/2 \quad \text{and} \quad I(\theta_1, \theta_2) = \theta_2.$$

Here $\theta_2 \to 0$ establishes a minimum, so that if we can restrict $\theta_2 \geqq \alpha$, then $\epsilon(\theta_2 - a)$ is a most unfavorable distribution.

8. More stringent inequalities. Bhattacharyya [4] has found greater lower bounds for the mean square errors of estimates in the case of constant parameters. This admits of direct extension to the present case. We can write (3.3) and (3.4) respectively as

$$E\left(\frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k} \,\middle|\, \boldsymbol{\theta}\right) = 0 \tag{8.1}$$

and

$$\frac{\partial\psi_k(\boldsymbol{\theta})}{\partial\theta_k} = \operatorname{Cov}\left(t_k - \theta_k, \frac{1}{\phi}\frac{\partial\phi}{\partial\theta_k} \,\middle|\, \boldsymbol{\theta}\right). \tag{8.2}$$

By the result of the appendix,

$$\operatorname{Cov}\left(t_k - \theta_k, \frac{1}{\phi}\frac{\partial \phi}{\partial \theta_k}\right) = E\left[\operatorname{Cov}\left(t_k - \theta_k, \frac{1}{\phi}\frac{\partial \phi}{\partial \theta_k}\right)\bigg| \boldsymbol{\theta}\right] + \operatorname{Cov}\left[E\,(t_k - \theta_k \mid \boldsymbol{\theta}), E\left(\frac{1}{\phi}\frac{\partial \phi}{\partial \theta_k}\bigg| \boldsymbol{\theta}\right)\right].$$

From which by using (8.1) and (8.2), we obtain

$$E\left(\frac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}\right) = \operatorname{Cov}\left(t_k - \theta_k, \frac{1}{\phi}\frac{\partial \phi}{\partial \theta_k}\right). \tag{8.3}$$

With suitable regularity conditions on ϕ and its derivatives similar to those cited in paragraph 3, we can differentiate (3.2) p times and obtain as in (8.3) that

$$E\left(\frac{\partial^\beta \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\beta}\right) = \operatorname{Cov}\left(t_k - \theta_k, \frac{1}{\phi}\frac{\partial^\beta \phi}{\partial \theta_k^\beta}\right), \qquad \beta = 1, 2, \cdots, p. \tag{8.4}$$

Define,

$$J_{\alpha\beta} = \operatorname{Cov}\left(\frac{1}{\phi}\frac{\partial^\alpha \phi}{\partial \theta_k^\alpha}, \frac{1}{\phi}\frac{\partial^\beta \phi}{\partial \theta_k^\beta}\right), \qquad \alpha, \beta = 1, 2, \cdots, p.$$

Let $J = [J_{\alpha\beta}]$ and $J^{-1} = [J^{\alpha\beta}]$. Denote by $R_{0\cdot 12\cdots p}$ the multiple correlation coefficient between $t_k - \theta_k$ and $(1/\phi)(\partial\phi/\partial\theta_k)$, $(1/\phi)(\partial^2\phi/\partial\theta_k^2)$, $\cdots$, $(1/\phi)(\partial^p\phi/\partial\theta_k^p)$. Then by a result cited by Wilks ([20], p. 42 ff.),

$$R^2_{0\cdot 123\cdots p} = \frac{\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} E\left(\frac{\partial^\alpha \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\alpha}\right) E\left(\frac{\partial^\beta \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\beta}\right) J^{\alpha\beta}}{\operatorname{Var}\,(t_k - \theta_k)}$$

Since $R^2_{0\cdot 123\cdots p} \leqq 1$, we may write

$$\operatorname{Var}\,(t_k - \theta_k) \geqq \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} E\left(\frac{\partial^\alpha \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\alpha}\right) E\left(\frac{\partial^\beta \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\beta}\right) J^{\alpha\beta},$$

from which we have

$$EE[(t_k - \theta_k)^2 \mid \boldsymbol{\theta})] \geqq E^2(\psi_k(\boldsymbol{\theta}) - \theta_k) + \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} E\left(\frac{\partial^\alpha \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\alpha}\right) E\left(\frac{\partial^\beta \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\beta}\right) J^{\alpha\beta}. \tag{8.5}$$

This is a greater lower bound than that of (3.8) since the multiple correlation between $t_k - \theta_k$ and the above series of variates will be larger than the simple correlation between $t_k - \theta_k$ and $(1/\phi)(\partial\phi/\partial\theta_k)$. This latter correlation is essentially what was used in deriving (3.8). It should be noted that this method of obtaining a higher lower bound applies only if $\psi_k(\boldsymbol{\theta})$ is non-linear function of θ_k and consequently is not applicable in the unbiased case. As noted in Wilks ([20], p. 46), the equality holds if and only if all the probability in the $p + 1$

dimensional space of the random variables lies on the surface,

$$t_k - \theta_k - E[\psi_k(\boldsymbol{\theta})] + E(\theta_k) = \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \frac{1}{\phi} \frac{\partial^\alpha \phi}{\partial \theta_k^\alpha} \cdot E\left(\frac{\partial^\beta \psi_k(\boldsymbol{\theta})}{\partial \theta_k^\beta}\right) J^{\alpha\beta}$$

9. The Sequential Case. Wolfowitz [21] has extended the Cramér-Rao Inequality to situations where the sample size is a random variable depending on the sequence of observations. In our notation this result is

$$E(t_k - \theta_k)^2 \geq [\psi_k(\boldsymbol{\theta}) - \theta_k]^2 + \frac{\left(\dfrac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}\right)^2}{E(n \mid \boldsymbol{\theta}) E\left[\left(\dfrac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_k}\right)^2 \middle| \boldsymbol{\theta}\right]}.$$

We shall proceed to extend this result to the case where $\boldsymbol{\theta}$ is a random variable.

Under suitable regularity conditions Wolfowitz has shown,

$$E\left(\frac{\partial \ln \phi}{\partial \theta_k} \middle| \boldsymbol{\theta}\right) = 0,$$

and

$$E\left[\left(\frac{\partial \ln \phi}{\partial \theta_k}\right)^2 \middle| \boldsymbol{\theta}\right] = E(n \mid \boldsymbol{\theta}) E\left[\left(\frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_k}\right)^2 \middle| \boldsymbol{\theta}\right]. \tag{9.1}$$

By definition,

$$\sum_{j=1}^{\infty} \int_{R_j} t_k(x_1, x_2, \cdots, x_j) \prod_{i=1}^{j} f(x_i \mid \boldsymbol{\theta})\, dx_i = \psi_k(\boldsymbol{\theta}).$$

Under the regularity conditions cited by Wolfowitz we may differentiate under the integral sign and obtain

$$E\left(t_k \cdot \frac{\partial \ln \phi}{\partial \theta_k} \middle| \boldsymbol{\theta}\right) = \operatorname{Cov}\left(t_k - \theta_k, \frac{\partial \ln \phi}{\partial \theta_k} \middle| \boldsymbol{\theta}\right) = \frac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}.$$

The result of the appendix yields,

$$E\left[\frac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}\right] = \operatorname{Cov}\left(t_k - \theta_k, \frac{\partial \ln \phi}{\partial \theta_k}\right).$$

Since the square of the correlation coefficient of any two variates cannot exceed unity, we have,

$$E^2\left(\frac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}\right) \leqq \operatorname{Var}\left(\frac{\partial \ln \phi}{\partial \theta_k}\right) \cdot \operatorname{Var}(t_k - \theta_k),$$

and

$$\operatorname{Var}(t_k - \theta_k) \geqq \frac{E^2\left(\dfrac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}\right)}{E\left\{E(n \mid \boldsymbol{\theta}) E\left[\left(\dfrac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_k}\right)^2 \middle| \boldsymbol{\theta}\right]\right\}}.$$

This may be written

$$EE[(t_k - \theta_k)^2 \mid \boldsymbol{\theta}] \geqq E^2[\psi_k(\boldsymbol{\theta}) - \theta_k] \tag{9.2}$$
$$+ \frac{E^2\left(\dfrac{\partial \psi_k(\boldsymbol{\theta})}{\partial \theta_k}\right)}{E\left\{E(n \mid \boldsymbol{\theta}) \cdot E\left[\left(\dfrac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_k}\right)^2 \,\middle|\, \boldsymbol{\theta}\right]\right\}}$$

When $\psi_k(\boldsymbol{\theta}) = \theta_k$, (9.2) becomes,

$$EE[(t_k - \theta_k)^2 \mid \boldsymbol{\theta}] \geqq \frac{1}{E\left\{E(n \mid \boldsymbol{\theta}) E\left[\left(\dfrac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_k}\right)^2 \,\middle|\, \boldsymbol{\theta}\right]\right\}} \tag{9.3}$$

These results are valid for discrete as well as continuous distributions.

A simple example of sequential estimation involves sampling from a binomial population until a specified number of successes, say r, occur. Here $f(x \mid \theta) = \theta^x(1 - \theta)^{1-x}$, $x = 0, 1$, and $E(n \mid \theta) = r/\theta$. Therefore, for $\psi(\theta) = \theta$,

$$EE[(t - \theta)^2 \mid \theta] \geqq \frac{1}{rE\left(\dfrac{1}{\theta^2(1 - \theta)}\right)}.$$

This result corresponds exactly to that obtained in paragraph 7 for the Pascal distribution.

10. Linear Estimation of $E(\theta_k)$. Consider m samples $\mathbf{X}_i = (x_{i1}, x_{i2}, \cdots, x_{in_i})$, $i = 1, 2, \cdots m$, chosen from a population $f(x \mid {}^i\boldsymbol{\theta})$, which are randomly and independently chosen from a super-population of populations with frequency functions of the form $f(x \mid \boldsymbol{\theta})$. Thus for each sample $\mathbf{X}_i$, there is associated an unobserved random variable ${}^i\boldsymbol{\theta}$, $i = 1, 2, \cdots, m$, with distribution $G({}^i\boldsymbol{\theta})$. We seek to find an estimate of $E(\theta_k)$, say T_k, where $1 \leqq k \leqq s$. It is supposed that for each sample there exists an unbiased estimate of ${}^i\theta_k$, namely ${}^it_k$, and we restrict our discussion to the set of T_k which are linear functions of the ${}^it_k$, that is,

$$T_k = \sum_{i=1}^{m} c_i \, {}^it_k$$

where $(c_1, c_2, \cdots, c_m)$ is a vector of real numbers. If we further restrict ourselves to unbiased estimates of $E(\theta_k)$, it follows that $\sum_{i=1}^m c_i = 1$.

The minimum variance unbiased estimate of $E(\theta_k)$, $\hat{T}_k$, is found by minimizing the expression, $\text{Var}(T_k) = \sum_{i=1}^m c_i^2 \text{Var}({}^it_k)$, with respect to the c's, subject to the restriction $\sum_{i=1}^m c_i = 1$. This yields the normal equations:

$$\hat{c}_i \text{Var}({}^it_k) + \lambda = 0, \qquad i = 1, 2, \cdots, m, \sum_{i=1}^{m} \hat{c}_i = 1, \tag{10.1}$$

where λ is a Lagrangian multiplier.

Consider now the variance of the minimum estimate $\hat{T}_k$, found by solving (10.1). We have

$$\text{Var}(\hat{T}_k) = \sum_{i=1}^{m} \hat{c}_i^2 \text{Var}({}^it_k). \tag{10.2}$$

By the result of the appendix,

$$\operatorname{Var}({}^i t_k) = E[\operatorname{Var}({}^i t_k \mid \boldsymbol{\theta})] + \operatorname{Var}[E({}^i t_k \mid \boldsymbol{\theta})],$$

from which it follows that,

$$\operatorname{Var}({}^i t_k) = EE[({}^i t_k - {}^i\theta_k)^2 \mid \boldsymbol{\theta}] + \operatorname{Var}({}^i\theta_k). \tag{10.3}$$

So that (10.2) may be written

$$\operatorname{Var}(\hat{T}_k) = \sum_{i=1}^{m} \hat{c}_i^2 EE[({}^i t_k - {}^i\theta_k)^2 \mid \boldsymbol{\theta}] + \operatorname{Var}(\theta_k) \sum_{i=1}^{m} \hat{c}_i^2 . \tag{10.4}$$

Applying (3.9) to (10.4), we have

$$\operatorname{Var}(\hat{T}_k) \geqq \sum_{i=1}^{m} \frac{\hat{c}_i^2}{EE\left[\left(\dfrac{\partial \ln \phi(X_i \mid \boldsymbol{\theta})}{\partial \theta_k}\right)^2 \middle| \boldsymbol{\theta}\right]} + \operatorname{Var}(\theta_k) \sum_{i=1}^{m} \hat{c}_i^2 , \tag{10.5}$$

the equality being achieved under the conditions cited in paragraph 5.

We may apply these results to the analysis of variance model II cited in paragraph 5. To simplify the normal equations above, let $n_i = n$, $i = 1, 2, \cdots, m$. In the notation of this paragraph, ${}^i\theta_k = \alpha_i$, ${}^i t_k = \hat{\alpha}_i = (\sum_{j=1}^{n} x_{ij})/n$, $i = 1, 2, \cdots, m$, and (10.1) becomes $\hat{c}_i((\sigma_\epsilon^2/n) + \sigma_\alpha^2) + \lambda = 0$, $\sum_{i=1}^{n} \hat{c}_i = 1$. Solving, we have $\hat{c}_i = 1/m$, $i = 1, 2, \cdots, m$. Thus

$$\hat{T}_k = \left(\sum_{i=1}^{m} \hat{\alpha}_i\right) \bigg/ m = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}\right) \bigg/ mn ,$$

and

$$\operatorname{Var}(\hat{T}_k) = \frac{1}{m}\left(\frac{\sigma_\epsilon^2}{n} + \sigma_\alpha^2\right),$$

which equals the lower bound given by (10.5).

11. Acknowledgements. The author wishes to express his appreciation to Professors C. W. Clunies-Ross and J. E. Freund for their suggestions.

APPENDIX

The Covariance in Terms of Conditional Expectations.

Let $\mathbf{U} = (u_1, u_2, \cdots, u_i)$ and $\mathbf{V} = (v_1, v_2, \cdots, v_j)$ be random variables. Assume $p = p(\mathbf{U}, \mathbf{V})$ and $q = q(\mathbf{U}, \mathbf{V})$ have finite means and variances. Then we have

(i) $\operatorname{Cov}(p, q) = EE(p \cdot q \mid \mathbf{V}) - EE(p \mid \mathbf{V})EE(q \mid \mathbf{V})$.

But,

$$\operatorname{Cov}(p, q \mid \mathbf{V}) = E(p \cdot q \mid \mathbf{V}) - E(p \mid \mathbf{V})E(q \mid \mathbf{V}),$$

from which, taking expectations with respect to $\mathbf{V}$, we have,

$$E[\operatorname{Cov}(p, q \mid \mathbf{V})] = EE(p \cdot q \mid \mathbf{V}) - E[E(p \mid \mathbf{V}) \cdot E(q \mid \mathbf{V})].$$

Substituting this result in (i) gives

(ii) $\text{Cov}(p, q)$

$$= \text{E}[\text{Cov}(p, q \mid \mathbf{V})] + E[E(p \mid \mathbf{V}) \cdot E(q \mid \mathbf{V})] - EE(p \mid \mathbf{V}) EE(q \mid \mathbf{V}),$$

(iii) $\text{Cov}(p, q)$

$$= E[\text{Cov}(p, q \mid \mathbf{V})] + \text{Cov}[E(p \mid \mathbf{V}), E(q \mid \mathbf{V})]. \qquad \text{Q.E.D.}$$

REFERENCES

[1] A. C. AITKEN AND H. SILVERSTONE, "On the estimation of statistical parameters," *Proc. Roy. Soc. Edinburgh, Sect. A*, Vol. 61 (1942), pp. 186–194.
[2] EDWARD W. BARANKIN, "Concerning some inequalities in the theory of statistical estimation," *Skand. Aktuarietids.*, Vol. 34 (1951), pp. 35–40.
[3] E. W. BARANKIN, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477–501.
[4] A. BHATTACHARYYA, "On some analogues to the amount of information and their uses in statistical estimation," *Sankhyā*, Vol. 8 (1946), pp. 1–14.
[5] DOUGLAS G. CHAPMAN AND HERBERT ROBBINS, "Minimum variance estimation without regularity conditions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 581–586.
[6] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[7] HARALD CRAMÉR, "Contributions to the theory of statistical estimation," *Skand. Aktuarietids.*, Vol. 29 (1946), pp. 85–94.
[8] G. DARMOIS, "Sur les limites de la dispersion de certaines estimations," *Rev. Inst. Int. Statist.*, Vol. 13 (1945), pp. 9–15.
[9] CHURCHILL EISENHARDT, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1–26.
[10] M. FRÉCHET, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons," *Rev. Inst. Int. Statist.*, Vol. 11 (1943), pp. 182–205.
[11] J. M. HAMMERSLEY, "On estimating restricted parameters," *Jour. Roy. Stat. Soc. B.*, Vol. 12 (1950), pp. 192–240.
[12] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399–409.
[13] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy," *Proc. Cambridge Philos. Soc.*, Vol. 32 (1936), pp. 567–579.
[14] C. RADHAKRISHNA RAO, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.
[15] M. P. SCHÜTZENBERGER, "A generalization of the Fréchet-Cramér inequality to the case of Bayes' estimation" (Abstract), *Bull. Amer. Math. Soc.*, Vol. 63 (1957), p. 142.
[16] M. C. K. TWEEDIE, "Functions of a statistical variate with special reference to Laplacian distributions," *Proc. Cambridge Philos. Soc.*, Vol. 43 (1947), pp. 41–49.
[17] M. C. K. TWEEDIE, *Notes on Generating Functions and Distribution Theory*, Virginia Polytechnic Institute, 1953.
[18] ABRAHAM WALD, *Statistical Decision Functions*, John Wiley and Sons, Inc., New York, 1950.
[19] DAVID VERNON WIDDER, *The Laplace Transform*, Princeton University Press, 1946.
[20] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1950.
[21] J. WOLFOWITZ, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.