EXAMPLE 1. We have n observations from a rectangular distribution from 0 to θ ($\Omega = \{\theta \mid \theta > 0\}$). It suffices to consider the maximum Y of the observations, whose density is ny^{n-1}/θ^n for $0 \leqq y \leqq \theta$, and 0 elsewhere. For $n = 1$, the denominator of the right side of (4) becomes $\inf_{-\theta<h<0}\{-1/[h(\theta + h])\}$, so that (4) gives the bound $\theta^2/4$. It would be too tedious to carry this calculation out for each n, but it can be shown that, as $n \to \infty$, (4) asymptotically gives the bound $.648\theta^2/n^2$. On the other hand, if we put $d\lambda_1(h) = [(n + 1)/\theta]\ (h/\theta + 1)^n\, dh$ for $-\theta < h < 0$, the term in braces on the right side of (3) becomes $\theta^2/[n(n + 2)]$, which is in fact attained as the variance of the unbiased estimator $[(n + 1)/n]Y$.

EXAMPLE 2. We have m observations from the distribution with density $e^{-(x-\theta)}$ for $x \geqq \theta$ and 0 elsewhere (Ω is the real line). Here the minimum Z of the observations is sufficient and has density $me^{-m(z-\theta)}$, $z \geqq \theta$. The denominator of (4) is $\inf_{h>0}\ ([e^{mh} - 1]/h^2)$. The infimum is attained for $mh = 1.5936$, and yields $.648/m^2$ as the bound given by (4). On the other hand, putting $d\lambda_1(h) = me^{-mh}\, dh$ for $0 < h < \infty$ and 0 otherwise, the expression in braces of (3) becomes $1/m^2$, which is actually attained as the variance of the unbiased estimator $Z - 1/m$.

REFERENCES

[1] D. C. CHAPMAN AND H. ROBBINS, "Minimum variance estimation without regularity assumptions," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 581–586.

[2] E. BARANKIN, "Locally best unbiased estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 477–501.

BHATTACHARYYA BOUNDS WITHOUT REGULARITY ASSUMPTIONS

BY D. A. S. FRASER AND IRWIN GUTTMAN

University of Toronto

1. Summary. In [1] a method for removing the regularity conditions from the Cramér-Rao Inequality was given and applied to the estimation of a single real parameter. It was noted there that the method would extend to problems more general than estimating a single real parameter. However, the method extends also for the estimation of a single real parameter and produces analogues of the Bhattacharyya bounds with and without nuisance parameters.

2. Introduction. Let $\mu(x)$ be a σ-finite measure defined over an additive class $\mathcal{A}$ of subsets of a space $\mathfrak{X}$, and let X be a random variable with density

$$f(x; \theta_1, \cdots, \theta_k)$$

with respect to $\mu(x)$. $\theta_1, \cdots, \theta_k$ are real with $(\theta_1, \cdots, \theta_k) = \Theta\ \varepsilon\ A \subset R^k$. The carrier $S(\theta_1, \cdots, \theta_k)$ of the distribution is defined by

$$S(\theta_1, \cdots, \theta_k) = \{x \mid f(x; \theta_1, \cdots, \theta_k) > 0\}.$$

We restrict our consideration of $S(\theta_1, \cdots, \theta_k)$ to the positive sample space of the measure μ.

The following lemma given in [2] will be needed:

LEMMA. *If the real valued random variables $T, S_1, \cdots, S_r$ satisfy*

$$\begin{aligned} E(S_i) &= 0, & \\ E(TS_i) &= 1, & i &= 1, \\ &= 0, & i &= 2, \cdots, r, \end{aligned}$$

then $\sigma_T^2 \geqq 1/\sigma^2_{S_1 \cdot S_2 \cdots S_r}$, where $\sigma^2_{S_1 \cdot S_2 \cdots S_r}$ is the variance of the residual of the regression fit of $S_2, \cdots, S_r$ to S_1.

PROOF. Since the covariance of T and $S_1 - \sum_2^r l_i S_i$ is 1, then the product of the variances is greater than or equal to 1; $\sigma_T^2 \geqq 1/\sigma^2_{S_1 - \Sigma l_i S_i}$. The sharpest inequality is obtained by a regression fit.

3. Bhattacharyya bounds. Let T be an unbiased estimate of the parameter θ_1, and define $\{S_{i_1 \cdots i_k}\}$ as follows:

$$\begin{aligned} S_1 &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(1)}}{\Delta} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \\ &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \cdot \frac{f(x; \theta_1^{(1)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)}) - f(x; \theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)})}{\theta_1^{(1)} - \theta_1^{(0)}}, \\ S_2 &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(1)}, \theta_1^{(2)}}{\Delta^2} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \\ &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(2)}}{\Delta} \underset{\theta_1^{(1)}}{\Delta} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \qquad (3.1) \\ &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \Bigg[\frac{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})}{(\theta_1^{(0)} - \theta_1^{(1)})(\theta_1^{(0)} - \theta_1^{(2)})} \\ &\qquad + \frac{f(x; \theta_1^{(1)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)})}{(\theta_1^{(1)} - \theta_1^{(0)})(\theta_1^{(1)} - \theta_1^{(2)})} + \frac{f(x; \theta_1^{(2)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)})}{(\theta_1^{(2)} - \theta_1^{(0)})(\theta_1^{(2)} - \theta_1^{(1)})} \Bigg], \\ S_{i_1, \cdots, i_k} &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(1)}, \cdots, \theta_1^{(i_1)}}{\Delta^{i_1}} \cdots \underset{\theta_k^{(1)}, \cdots, \theta_k^{(i_k)}}{\Delta^{i_k}} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}), \end{aligned}$$

where $\Delta^i_{\theta^{(1)} \cdots \theta^{(k)}} g(\theta^{(0)})$ is the ith divided difference

$$\underset{\theta^{(1)}, \cdots, \theta^{(i)}}{\Delta^i} g(\theta^{(0)}) = \underset{\theta^{(i)}}{\Delta} \cdots \underset{\theta^{(1)}}{\Delta} g(\theta^{(0)}),$$

$$\underset{\theta^{(1)}}{\Delta} g(\theta^{(0)}) = \frac{g(\theta^{(1)}) - g(\theta^{(0)})}{\theta^{(1)} - \theta^{(0)}},$$

where the expressions are to be considered as functions of $\theta^{(0)}$ for further differencing. Also we introduce the following assumption concerning the carrier of the distribution.

ASSUMPTION A. *$S(\theta_1^{(0)}, \cdots, \theta_k^{(0)}) \supset S(\theta_1^{(i_1)}, \cdots, \theta_k^{(i_k)})$ for all $(i_1, \cdots, i_k)$ for which S's have been defined.*

On the basis of Assumption A, we have

$$E_{\Theta_0}(S_{i_1 \cdots i_k}) = \int \Delta^{i_1}_{\theta_1^{(1)}, \cdots, \theta_1^{(i_1)}} \cdots \Delta^{i_k}_{\theta_k^{(1)}, \cdots, \theta_k^{(i_k)}} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})\, d\mu(x)$$
$$= 0,$$

$$E_{\Theta_0}(TS_{i_1 \cdots i_k}) = \int \Delta^{i_1}_{\theta_1^{(1)}, \cdots, \theta_1^{(i_1)}} \cdots \Delta^{i_k}_{\theta_k^{(1)}, \cdots, \theta_k^{(i_k)}} f(x; \Theta_0) T\, d\mu(x)$$
$$= 1, \qquad \text{if } i_1 = 1, i_2 = \cdots = i_k = 0,$$
$$= 0, \qquad \text{otherwise.}$$

Letting S_β stand for any one of the above defined S's except S_1 and applying the lemma of Section 2, the following inequality is obtained (subject to Assumption A) for the variance of the unbiased estimate T:

$$\text{var}_{\Theta_0} T \geqq \inf_{\substack{\theta_1^{(1)}, \theta_1^{(2)}, \cdots \\ \cdots\cdots\cdots \\ \theta_k^{(1)}, \theta_k^{(2)}, \cdots}} \frac{1}{\sigma^2_{S_1 \cdot S_{\beta'} S_{\beta''} \cdots S_{\beta(t)}}}. \tag{3.2}$$

If the usual regularity conditions are assumed it is easily seen that this bound is at least as large as the ordinary Bhattacharyya bound.

For a biased estimate T having $E_\Theta(T) = g(\Theta)$, the following inequality is obtained (subject to Assumption A):

$$\text{var}_{\Theta_0} T \geqq \inf_{\substack{\theta_1^{(1)}, \theta_1^{(2)}, \cdots \\ \cdots\cdots\cdots \\ \theta_k^{(1)}, \theta_k^{(2)}, \cdots \\ l_{\beta'}, l_{\beta''}, \cdots}} \frac{\left[\Delta_{\theta_1^{(1)}} g(\Theta_0) - \sum l_{i_1 \cdots i_k} \Delta^{i_1}_{\theta_1^{(1)} \cdots \theta_1^{(i_1)}} \cdots \Delta^{i_k}_{\theta_k^{(1)} \cdots \theta_k^{(i_k)}} g(\Theta_0)\right]^2}{\text{var}_{\Theta_0}(S_1 - \sum l_\beta S_\beta)}. \tag{3.3}$$

4. Multistatistic case. For more than one statistic, say $(T_1, \cdots, T_m) = \mathbf{T}$, there is an immediate generalization of the inequality (3.3). It is obtained from the covariance relation that $[\sum_{yy} - \sum_{yx} \sum_{xx}^{-1} \sum_{xy}]$ is positive semi-definite, where $\sum_{yy}$, $\sum_{xx}$ and $\sum_{xy}$ are respectively the covariance matrices for a vector $\mathbf{y}$, for a vector $\mathbf{x}$, and between the vectors $\mathbf{x}$ and $\mathbf{y}$. Letting $\mathbf{y}$ be the statistic and $\mathbf{x}$ be a set of the S's defined by (3.1), then $\sum_{xy}$ becomes a matrix of differences of $E_\Theta(\mathbf{T})$ (as in the numerator on the right hand side of (3.3)).

5. Binomial distribution. For the unbiased estimation of the parameter p of the binomial distribution, the following lower bound for the variance at p is obtained using S and an interval h for differencing:

$$\sigma_T^2 \geqq \frac{h^2}{\left[1 + h^2 \frac{1}{p_0 q_0}\right]^n - 1}.$$

The greatest lower bound is obtained by letting $h \to 0$.

This greatest lower bound can also be obtained by using the Bhattacharyya bound (3.2) without applying a limiting operation:

$$\frac{p_0 q_0}{n} = \frac{1}{\sigma^2_{S_1 - S_2/2 + \cdots - (-1)^n S_n/n}},$$

where the S's are defined (3.1) as the divided differences corresponding to ordinary differences with interval h (h being chosen sufficiently small that all p's fall between 0 and 1).

REFERENCES

[1] D. G. CHAPMAN AND H. ROBBINS, "Minimum variance estimation without regularity assumptions," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 581–586.

[2] E. L. LEHMANN, "Notes on the Theory of Estimation," mimeographed notes.

ON THE ANALYSIS OF SAMPLES FROM k LISTS[1]

BY LEO A. GOODMAN

The University of Chicago

1. Introduction and summary. Suppose we have k lists of names, no name appearing more than once in each list. We are interested in estimating the following parameters: (a) the number of names occurring in common in pairs, triples, $\cdots$, of lists; (b) the number of names occurring in 1, 2, $\cdots$, k lists. This note presents unbiased estimators for these parameters when a random sample is drawn from each list. It is also observed that the estimators presented are the only real-valued statistics which are unbiased estimators of the parameters, and hence must be the minimum variance unbiased estimators. This yields another example in which "insufficient" statistics have been used to obtain minimum variance unbiased estimators.

These unbiased estimators may at times give unreasonable estimates. In such cases, it is suggested that the statistics be modified so that the nearest reasonable estimate is used. Although this procedure introduces some bias, it usually reduces the mean square error.

This problem arises when we are interested in tracing the interrelations of agencies through the individual members. The problem also arises in the work of H. H. Fussler and J. M. Dawson of the University Library, University of Chicago, who are interested in comparing the acquisitions of various libraries. For special problems other sampling schemes may be more economical or more efficient than taking a sample from each list. Professor F. F. Stephan of Princeton University pointed out to the author that, in the special case of the "library problem," the Book Catalog and author cards used by many libraries provide a convenient means of drawing matched samples. (There is a brief discussion

[1] This work was prepared in connection with research supported by the Office of Naval Research.