MINIMUM VARIANCE ESTIMATION WITHOUT REGULARITY ASSUMPTIONS

By Douglas G. Chapman[1] and Herbert Robbins

University of Washington and University of North Carolina

1. Summary and Introduction. Following the essential steps of the proof of the Cramér-Rao inequality [1, 2] but avoiding the need to transform coordinates or to differentiate under integral signs, a lower bound for the variance of estimators is obtained which is (a) free from regularity assumptions and (b) at least equal to and in some cases greater than that given by the Cramér-Rao inequality. The inequality of this paper might also be obtained from Barankin's general result[2] [3]. Only the simplest case—that of unbiased estimation of a single real parameter—is considered here but the same idea can be applied to more general problems of estimation.

2. Lower bound. Let μ be a fixed measure on Euclidean n-space X and let the random vector $x = (x_1, \cdots, x_n)$ have a probability distribution which is absolutely continuous with respect to μ, with density function $f(x, \alpha)$, where α is a real parameter belonging to some parameter set A. Define $S(\alpha)$ as follows:

$$f(x, \alpha) > 0, \qquad \text{a.e. } x \text{ in } S(\alpha),$$
$$f(x, \alpha) = 0, \qquad \text{a.e. } x \text{ in } X - S(\alpha).$$

Let $t = t(x)$ be any unbiased estimator of α, so that for every α in A,

$$\int_X tf(x, \alpha)\, d\mu = \alpha. \tag{1}$$

If $\alpha, \alpha + h (h \neq 0)$ are any two distinct values in A such that

$$S(\alpha + h) \subset S(\alpha), \tag{2}$$

then, writing S for $S(\alpha)$,

$$\int_S f(x, \alpha)\, d\mu = 1, \qquad \int_{S(\alpha+h)} f(x, \alpha + h)\, d\mu = \int_S f(x, \alpha + h)\, d\mu = 1,$$

$$\int_S tf(x, \alpha)\, d\mu = \alpha, \qquad \int_S tf(x, \alpha + h)\, d\mu = \alpha + h,$$

so that

$$\int_S [t - \alpha] \sqrt{f(x, \alpha)}\, \frac{f(x, \alpha + h) - f(x, \alpha)}{hf(x, \alpha)} \sqrt{f(x, \alpha)}\, d\mu = 1.$$

---

[1] This research was supported in part by the Office of Naval Research.

[2] But again with some additional restrictions.

Applying Schwarz's inequality we obtain the relation

$$
\begin{aligned}
(3)\qquad 1 &\le \int_S [t-\alpha]^2 f(x,\alpha)\, d\mu \cdot \int_S \left[\frac{f(x,\alpha+h)-f(x,\alpha)}{hf(x,\alpha)}\right]^2 f(x,\alpha)\, d\mu \\
&= \operatorname{Var}(t \mid \alpha)\cdot \frac{1}{h^2}\left\{\int_S \left[\frac{f(x,\alpha+h)}{f(x,\alpha)}\right]^2 f(x,\alpha)\, d\mu - 1\right\}.
\end{aligned}
$$

Let

$$J = J(\alpha, h) = \frac{1}{h^2}\left\{\left[\frac{f(x,\alpha+h)}{f(x,\alpha)}\right]^2 - 1\right\};$$

then (3) can be written in the form

$$(4)\qquad \operatorname{Var}(t \mid \alpha) \ge \frac{1}{E(J \mid \alpha)}.$$

Since (4) holds whenever α, $\alpha + h$ are any two distinct elements of A satisfying (2) we obtain the fundamental inequality

$$(5)\qquad \operatorname{Var}(t \mid \alpha) \ge \frac{1}{\inf_h E(J \mid \alpha)},$$

where the infimum is taken over all $h \ne 0$ such that (2) is satisfied. It should be noted that (5) holds without any restriction on $f(x, \alpha)$ and without any restriction on t other than (1).

It is possible that $E(J \mid \alpha)$ does not exist (finitely) for any h. With the usual convention that $E(J \mid \alpha) = \infty$, in this case, (5) is still a valid, though trivial, inequality.

In applications μ will often be Lebesgue measure on X. It could equally well be a discrete measure on a countable set of points in X. Furthermore, if the set where $f(x, \alpha) > 0$ is independent of α then (2) is trivially satisfied for all $\alpha + h$ in A.

We shall have occasion to compare (5) with the Cramér-Rao inequality

$$(6)\qquad \operatorname{Var}(t \mid \alpha) \ge \frac{1}{E(\psi^2 \mid \alpha)}; \qquad \psi = \psi(\alpha) = \frac{\partial}{\partial\alpha}\ln f(x,\alpha).$$

This inequality is usually derived for distributions with range independent of the parameter and under certain regularity conditions on both $f(x, \alpha)$ and the unbiased estimator t.

3. Examples.

Example 1. *Unbiased estimation of the mean of a normal distribution based on a random sample of size n.* Here

$$f(x,\alpha) = (2\pi)^{-(n/2)}\sigma^{-n} e^{-(1/2\sigma^2)\sum_{i=1}^n (x_i-\alpha)^2},$$

where σ is a positive constant, and

$$J = \frac{1}{h^2}\{e^{-(1/\sigma^2)\sum_{i=1}^n[(x_i-\alpha-h)^2-(x_i-\alpha)^2]} - 1\} = \frac{n}{\sigma^2 k^2}\{e^{-k^2}e^{2ku} - 1\},$$

where we have set $u = \sum_{i=1}^{n}(x_i - \alpha)/(\sigma\sqrt{n})$, $\quad k = h\sqrt{n}/\sigma \neq 0$.

When the mean is α, u is normally distributed with mean 0 and variance 1, and we find after a simple computation that

$$E(J \mid \alpha) = n(e^{k^2} - 1)/(\sigma^2 k^2),$$
$$\inf_h E(J \mid \alpha) = \lim_{k \to 0} [n(e^{k^2} - 1)/(\sigma^2 k^2)] = n/\sigma^2 = [E(\psi^2 \mid \alpha)]. \tag{7}$$

Hence if t is any unbiased estimator of α it follows from (5) that

$$\operatorname{Var}(t \mid \alpha) \geq \sigma^2/n. \tag{8}$$

Since the sample mean $\bar{x}$ is an unbiased estimator of α with $\operatorname{Var}(\bar{x} \mid \alpha) = \sigma^2/n$, it follows that $\bar{x}$ has minimum variance in the class of *all* unbiased estimators of α.

In this example the Cramér-Rao inequality (6) yields precisely the same bound (8).

Corresponding results hold for the unbiased estimation of the variance when the mean is known. Both (5) and (6) yield the inequality

$$\operatorname{Var}(t \mid \alpha) \geq 2\alpha^2/n,$$

where α is the unknown variance. The equality sign holds for

$$t = n^{-1} \sum_{i=1}^{n} (x_i - m)^2,$$

where m is the mean of the normal population.

Example 2. *Unbiased estimation of the standard deviation of a normal population with known mean.* Here

$$f(x, \alpha) = (2\pi)^{-(n/2)} \alpha^{-n} e^{-(1/2\alpha^2)\Sigma_{i=1}^{n}(x_i - m)^2}.$$

Setting $k = h/\alpha$ we find that for $-1 < k < \sqrt{2} - 1$, $k \neq 0$,

$$E(J \mid \alpha) = \{(1 + k)^{-n}[1 - k(2 + k)]^{-(n/2)} - 1\}/(\alpha^2 k^2). \tag{9}$$

In this case also, $\lim_{k \to 0} E(J \mid \alpha) = 2n/\alpha^2 = E(\psi^2 \mid \alpha)$. But the minimum value of $E(J \mid \alpha)$ is not approached in the neighborhood of $h = k = 0$, and the inequality (5) is sharper than (6). We shall consider only the case $n = 2$. Equation (9) then becomes

$$E(J \mid \alpha) = (p + 1)^2/[\alpha^2 p^2(2 - p^2)],$$

where we have set $p = 1 + k$ and $0 < p < \sqrt{2}$. We have for $p = .8393$, $1/E(J \mid \alpha) = .2698\, \alpha^2$, so that by (5)

$$\operatorname{Var}\ (t \mid \alpha) \geq .2698\, \alpha^2 > .25\alpha^2 = \frac{1}{E(\psi^2 \mid \alpha)}.$$

It is interesting to note that the unbiased estimator

$$t = \sqrt{\tfrac{1}{2}n}\,\frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(n+1)]}\sqrt{\sum_{i=1}^{n}(x_i - m)^2 \Big/ n}$$

has variance

$$\alpha^2\left[\tfrac{1}{2}n\,\frac{\Gamma^2\frac{1}{2}n}{\Gamma^2[\frac{1}{2}(n+1)]} - 1\right],$$

which for $n = 2$ becomes

$$\alpha^2\left[\frac{4}{\pi} - 1\right] = .2732\alpha^2.$$

But it can be shown using results of Lehmann and Scheffé [4], or of Hoel [5], which were derived from Blackwell's theorem on conditional expectation [6], that no other unbiased estimator can have smaller variance than t. Thus (5) does not give the *greatest* lower bound in this case.

Various examples of the application of (5) can be given where $S(\alpha)$ is not a constant and where the Cramér-Rao formula is invalid (see for example Cramér [1], p. 485). It should be noted, however, that in many of the standard problems of this type stronger results can be obtained by other methods.

Another class of estimation problems where (5) may be applied occurs if the parameter space is discrete. Again in this case the Cramér-Rao formula does not hold. An example of this type has been given by Chapman ([7], pp. 149–150). Other applications of this type and some results related to this paper were obtained recently by Hammersley [8].

4. General comparison with the Cramér-Rao inequality. Let

$$\bar{J} = \bar{J}(\alpha, h) = \left[\frac{f(x, \alpha + h) - f(x, \alpha)}{hf(x, \alpha)}\right]^2; \tag{10}$$

then

$$E(\bar{J} \mid \alpha) = E(J \mid \alpha).$$

Hence in the fundamental inequality (5) we can replace J by $\bar{J}$. But from (10) it is clear that

$$\lim_{h\to 0} \bar{J}(\alpha, h) = \left[\frac{\partial}{\partial\alpha}\ln f(x, \alpha)\right]^2 = \psi^2(\alpha)$$

whenever the latter exists.

Assuming now the usual regularity conditions under which the Cramér-Rao lower bound is derived, that $S(\alpha)$ is independent of α and that $f(x, \alpha)$ is sufficiently regular that we may pass to the limit inside the integral sign,

$$E(\psi^2 \mid \alpha) = E\left[\lim_{h\to 0}(\bar{J} \mid \alpha)\right] = \lim_{h\to 0} E(\bar{J} \mid \alpha) \geq \inf_h E(\bar{J} \mid \alpha) = \inf_h E(J \mid \alpha), \tag{11}$$

the infimum being taken over admissible values of h. It follows that *the inequality* (5) *is at least as sharp as that given by the Cramér-Rao formula* (6).

On the other hand, when $x = (x_1, \cdots, x_n)$ is a random sample from a regular distribution, and when $E(\psi^2 \mid \alpha) < \infty$, then for any *fixed* $h \neq 0$, there exists an n_0 such that for $n > n_0$

$$E(\psi^2 \mid \alpha) \leq E(J \mid \alpha). \tag{12}$$

Without loss of generality assume $E(J \mid \alpha) < \infty$. Letting $g(t, \alpha)$ denote the density function of a single x_i and ν the one-dimensional measure which generates μ, it is easily verified that

$$E(J \mid \alpha) = \frac{1}{h^2}\left(\left[\int_X \frac{g^2(t, \alpha + h)}{g(t, \alpha)}\, d\nu\right]^n - 1\right).$$

By hypothesis, except on a set of measure 0,

$$g(t, \alpha + h) = g(t, \alpha) + h \left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)}; \qquad \alpha \leq \alpha(h) \leq \alpha + h.$$

Hence

$$\int_X \frac{g^2(t, \alpha + h)}{g(t, \alpha)}\, d\nu = 1 + 2h \int_X \left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)} d\nu + h^2 \int_X g^{-1}\left(\left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)}\right)^2 d\nu. \tag{13}$$

Denoting the last integral of the right hand side of (13) by $R(\alpha, h)$ and noting that the relation

$$\int_X g(t, \alpha)\, d\nu = 1$$

may be differentiated under the integral sign so that the middle term vanishes, it follows that

$$E(J \mid \alpha) = \frac{[1 + h^2 R(\alpha, h)]^n - 1}{h^2} \geq nR(\alpha, h) + \tfrac{1}{2}n(n-1)h^2 R^2(\alpha, h). \tag{14}$$

On the other hand, from (11) and (14),

$$E(\psi^2 \mid \alpha) = nR(\alpha, 0). \tag{15}$$

In order that different parameters may be distinguishable we must have

$$\left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)} \neq 0$$

for a set of positive measure on the t-axis, and hence $R(\alpha, h) > 0$. From this and the fact that $R(\alpha, 0)$ is independent of n, (12) follows at once, for sufficiently large n, from (14) and (15).

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[2] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.

[3] E. W. BARANKIN, "Locally best unbiased estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 477–501. (More complete references to the general problem are given in this paper.)

[4] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation, Part 1," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[5] P. G. HOEL, "Conditional expectation and the efficiency of estimates," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 299–301.

[6] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[7] D. G. CHAPMAN, "Some properties of the hypergeometric distribution with applications to zoological sample censuses," *Univ. of California Publ. Statist.*, Vol. 1, No. 7 (1951), pp. 131–160.

[8] J. M. HAMMERSLEY, "On estimating restricted parameters," *Jour. Roy. Stat. Soc., Ser. B*, Vol. 12 (1950), pp. 192–229.