

SEQUENTIAL ESTIMATORS AND THE CRAMÉR-RAO LOWER BOUND

Gordon Simons\*

Summary

While all nonsequential unbiased estimators of the normal mean have variances which must obey the Cramér-Rao inequality, it is shown that some sequential unbiased estimators do not.

AMS 1970 Subject Classification: Primary 62F10, 62L12

Key Words and Phrases: Cramér-Rao inequality, Cramér-Rao lower bound, unbiased estimation, sequential estimation.

Short title: Sequential Estimation

---

\* The work of the author was supported by the National Science Foundation under Grant No. MCS-7801240.

## 1. Introduction

Beyond its limited practical importance, the main significance of Wolfowitz' (1947) discovery of a sequential version of the Cramér-Rao inequality was that it dampened the enthusiasm of early proponents of sequential sampling. The substantial savings in (expected) sample size from sampling sequentially, demonstrated by Wald and others in the area of hypothesis testing, would apparently not be realized in the area of estimation. The reasoning was as follows: Wolfowitz' result implies that no unbiased sequential estimator  $\hat{\theta}$  of the normal mean, satisfying certain regularity conditions, can have a smaller variance than  $\sigma^2/EN$ , where  $\sigma^2$  is the variance of the normal observations and  $EN$  is the expected sample size. Since the sample mean for a sample of size  $n$  has variance  $\sigma^2/n$ , the only way that  $\hat{\theta}$  can have a smaller variance is for  $EN$  to exceed  $n$ . The point of the present paper is to show that this reasoning is misleading; there do exist unbiased sequential estimators  $\hat{\theta}$  of the normal mean which have smaller variance than that permitted by the Cramér-Rao inequality. To be precise, the regularity conditions Wolfowitz assumes, while excessive, can not be eliminated altogether. Thus the Cramér-Rao bound does not need to hold. The conclusion to be drawn is that the limitations of sequential estimation, whatever they may be, must be assessed by other means.

One may reasonably object to an assessment which focuses completely on *unbiased* estimation. However, in defense of the current concern with unbiasedness, it should be said that:

(i) When one is estimating a location parameter such as the normal mean, a restriction to unbiased estimators seems fairly innocuous.

(ii) The current attempt to give a nonasymptotic theoretical assessment of sequential estimators is a difficult task made somewhat easier by restricting one's attention solely to unbiased estimators. Perhaps others will be encouraged to make a more comprehensive study which includes other estimators. The author believes that the present study sheds some light on the complexities one will need to consider.

(iii) The present study addresses itself to a long-standing, unsettled, theoretical question: It has been known for a long time that the regularity conditions one encounters in proofs of the Cramér-Rao inequality are usually unnecessary when one restricts one's attention to fixed sample size estimators. (This has been shown by Chapman and Robbins (1951), by Charles Stein in an unpublished note, and as an indirect consequence of the theory of complete and sufficient estimators.) It has not been known whether the same situation occurs with sequential estimators. This paper settles the issue; some regularity conditions *are* necessary. (A mild hint that regularity conditions may be needed is given by the fact that a widely encompassing theory of complete and sufficient statistics is impossible in a sequential setting; completeness too easily fails to occur. Cf., Lehmann and Stein (1950).)

As a historical note, it should be mentioned that sequential estimators have been shown to be *more* efficient than nonsequential estimators in a setting in which the Cramér-Rao bound has no applicability. This can occur when one is estimating a location parameter  $\theta$  for a family of densities  $\{f(x-\theta), -\infty < \theta < \infty\}$ , and  $f$  is a discontinuous density function. (Cf., Ibragimov and Khas'minskii (1974).) Of greater importance, many sequential estimators have been proposed which have attributes not obtainable by fixed sample size estimators.

Section 2 discusses three situations in which the Cramér-Rao bound holds for unbiased sequential estimators of the normal mean. Besides being of some interest in themselves, they shed light on the counter-examples discussed in Section 3.

## 2. Cramér-Rao bounds.

The Cramér-Rao inequality for an unbiased estimator  $\hat{r}$  of a real parameter  $r(\theta)$ , when  $\theta$  is real and  $r$  is differentiable, is a statement that its variance is bounded below by the ratio of the square of the derivative of  $r(\theta)$  and the Fisher information in the sample. In the case of a random sample of size  $n$ , this takes the form

$$(1) \quad \text{Var}_{\theta} \hat{r} \geq \frac{(r'(\theta))^2}{nI(\theta)},$$

where  $I(\theta)$  is the Fisher information in a single observation. Wolfowitz' (1947) version for a sequential random sample with stopping time  $N$  takes the form

$$(2) \quad \text{Var}_{\theta} \hat{r} \geq \frac{(r'(\theta))^2}{E_{\theta} N \cdot I(\theta)}.$$

When particularized to an unbiased sequential estimator  $\hat{\theta}$  of the normal mean  $\theta$ , (2) reduces to

$$(3) \quad \text{Var}_{\theta} \hat{\theta} \geq \frac{\sigma^2}{E_{\theta} N}, \quad -\infty < \theta < \infty,$$

where  $\sigma^2$  is the variance of the normal observations. Before we proceed to show that (3) can be violated, we shall discuss three situations in which (3) is valid:

(a) Suppose  $\hat{\theta} = \bar{X}_N$ , the sample mean for the observations  $X_1, \dots, X_N$  where the random sample size  $N \geq 2$  is chosen so that the occurrence or non-occurrence of the event  $[N=n]$  depends on the first  $n$  observations

$X_1, \dots, X_n$  only through the values of the differences  $X_2 - X_1, X_3 - X_1, \dots, X_n - X_1$  for each  $n \geq 2$ . (For instance, if one were uncertain of the value of  $\sigma^2$ , one might use the sequence of sample variances to define  $N$ .) Then it is easily checked that  $\hat{\theta}$  is an unbiased estimator of  $\theta$  (assuming  $N$  is almost surely finite for all  $\theta$ ), the distribution of  $N$  is independent of  $\theta$ , and  $\text{Var}_\theta \hat{\theta} = \sigma^2 E_\theta N^{-1}$ . In such a case, (3) is equivalent to  $E_\theta N^{-1} E_\theta N \geq 1$ ,  $-\infty < \theta < \infty$ , which, of course, holds. Moreover, it is easily seen that an equality is obtained in (3) for  $\hat{\theta} = \bar{X}_N$  only when  $N$  is almost surely a constant, i.e., when  $\hat{\theta}$  is a nonsequential estimator. Even when  $N$  is not constant, if  $E_\theta N < \infty$ , there exists for each fixed  $\theta_0$  an unbiased sequential estimator which attains the lower bound given in (3) at  $\theta = \theta_0$ . For the special case  $\theta_0 = 0$ , this estimator takes the form  $\hat{\theta} = S_N / E_\theta N$ , where  $S_N = X_1 + \dots + X_N$ .

(b) Blackwell and Girshick (1947) show, for a general stopping time  $N$ , that the unbiased sequential estimator  $\hat{\theta} = E_\theta(X_1 | S_N, N)$  satisfies (3). (Some assumptions appearing in their paper are not needed.) The main aspects of their argument are: (i)  $\hat{\theta}$  is a statistic  $((S_N, N)$  is a sufficient statistic for  $\theta$ ) and an unbiased estimator for  $\theta$ ; (ii) when  $E_\theta N < \infty$ ,  $\hat{\theta}(S_N - N\theta)$  and  $X_1(S_N - N\theta)$  have finite expectations, and

$$E_\theta \{\hat{\theta}(S_N - N\theta)\} = E_\theta \{X_1(S_N - N\theta)\} = \sigma^2$$

(the first equality being a trivial consequence of the definition of  $\hat{\theta}$ , and the second being established with an interesting martingale argument); and (iii) using Schwartz' inequality,

$$\text{Var}_\theta \hat{\theta} \geq \sigma^4 / E_\theta (S_N - N\theta)^2 = \sigma^4 / \{\sigma^2 E_\theta N\} = \sigma^2 / E_\theta N .$$

This latter step requires a version of a Wald identity not available in 1947

(cf., Chow, Robbins and Teicher (1965)). Inequality (3) is a triviality for any  $\theta$  for which  $E_\theta N = \infty$ .

$\hat{\theta}$ , in general, is not  $\bar{X}_N$  even when  $N$  is a stopping variable of the type described in (a) above. Of course, when  $N$  is almost surely a constant,  $\hat{\theta}$  reduces to a sample mean. Blackwell and Girshick show, in their paper, that an equality holds in (3) only when  $N$  is a constant. This pertains to their estimator  $\hat{\theta} = E_\theta(X_1 | S_N, N)$  and does not contradict what is said about  $\hat{\theta} = S_N/E_\theta N$  in (a) above.

(c) Suppose  $N$  has a finite moment generating function  $m_\theta(t) = E_\theta e^{tN}$  for all real  $t$  and  $\theta$ . (This occurs, for instance, when  $N$  is a bounded stopping time.) Further, let

$$p_\theta = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^N (X_i - \theta)^2\right\}, \quad -\infty < \theta < \infty,$$

denote the likelihood function. Since  $m_\theta(h^2/\sigma^2) = E_{\theta+2h} \left(\frac{p_{\theta+h}}{p_{\theta+2h}}\right)^2$ , it is easily seen that the finiteness of  $m_\theta(t)$  for all  $\theta$  and  $t$  is equivalent to

$$(4) \quad E_\theta \left(\frac{p_{\theta'}}{p_\theta}\right)^2 < \infty, \quad -\infty < \theta, \theta' < \infty.$$

With (4) as an assumption, it follows from Theorem 1 of Stein (1950) that, for each fixed  $\theta_0$ , there exists an unbiased sequential estimator  $\hat{\theta}$ , based upon the sample  $X_1, \dots, X_N$ , which has minimum variance at  $\theta = \theta_0$  among all unbiased sequential estimators using the same sample. The form taken by  $\hat{\theta}$  is not specified, and whether or not (3) holds for this estimator is not addressed. However, using the same argument given by Chapman and Robbins (1951), one can show for every unbiased sequential estimator  $\hat{\theta}$  that  $\text{Var}_\theta \hat{\theta} \geq 1/I_N(\theta)$ ,  $-\infty < \theta < \infty$ , providing one defines the Fisher information for the sample in a nonstandard way by

$$I_N(\theta) = \liminf_{h \rightarrow 0} E_\theta h^{-2} \left( \frac{P_{\theta+h}}{P_\theta} - 1 \right)^2, \quad -\infty < \theta < \infty.$$

Finally, by using the dominated convergence theorem, together with the fact that

$$E_\theta h^{-2} \left( \frac{P_{\theta+h}}{P_\theta} - 1 \right)^2 = E_\theta \left\{ \frac{P_{\theta+2h}}{P_\theta} \cdot \frac{e^{h^2 N / \sigma^2} - 1}{h^2} \right\},$$

one easily obtains  $I_N(\theta) = E_\theta N / \sigma^2$ , from which (3) follows.

Two counter-examples. For both counter-examples given below, one is concerned with estimating the normal mean  $\theta$  unbiasedly, and in such a way that the variance of the estimator  $\hat{\theta}$  violates the Cramér-Rao bound, given in (3), when  $\theta = 0$ . The first counter-example is defective in the sense that the stopping time assumes an infinite value with positive probability when  $\theta \neq 0$ . Nevertheless, it is quite simple and it motivates the second counter-example, which is more difficult to describe.

First counter-example: Let

$$(5) \quad N = \inf\{n \geq 1: |S_n| < n^\gamma\} \quad (= \infty \text{ if } |S_n| \geq n^\gamma, n \geq 1),$$

where  $S_n = X_1 + \dots + X_n$ , and  $\gamma \in (\frac{1}{2}, 1)$  is a fixed constant. Then  $P_0(N < \infty) = 1$ , in fact  $E_0 N < \infty$ , while  $P_\theta(N = \infty) > 0$ ,  $\theta \neq 0$ . Estimate  $\theta$  by

$$\begin{aligned} \hat{\theta} &= 0 && \text{when } N < \infty \\ &= \bar{X}_\infty / P_{\bar{X}_\infty}(N = \infty) && \text{when } N = \infty, \end{aligned}$$

where  $\bar{X}_\infty = \lim_{n \rightarrow \infty} \bar{X}_n$ . In view of the strong law of large numbers, this estimator is well-defined, and is clearly unbiased. Since  $\text{Var}_0 \hat{\theta} = 0$  and  $E_0 N < \infty$ ,

(3) is violated when  $\theta = 0$ . I.e.,

$$\text{Var}_0 \hat{\theta} = 0 < \frac{\sigma^2}{E_0 N}.$$

Observe that  $\text{Var}_\theta \hat{\theta} = \theta^2 P_\theta(N < \infty) / P_\theta(N = \infty)$  is finite when  $\theta \neq 0$ . But, of course, there is no way of contradicting (3) for such  $\theta$  since  $E_\theta N = \infty$ .

Summarizing,  $\hat{\theta}$  is an unbiased estimator of  $\theta$  with a finite variance for all  $\theta$  and with variance zero for  $\theta = 0$ ; it does not satisfy the Cramér-Rao bound when  $\theta = 0$ .

Second counter-example: Relabel the observable normal random variables as  $X, Y, Z, X_1, X_2, \dots$  and continue to define  $S_n$  as  $X_1 + \dots + X_n$ ,  $n \geq 1$ . We shall exploit the facts that  $X$  is an unbiased estimator of  $\theta$ , and  $Z - Y$  is a random variable whose distribution is independent of  $\theta$ . A positive integer-valued function of  $Z - Y$  is a random variable  $M$  which is independent of  $X, X_1, X_2, \dots$ , and which has the same distribution for all  $\theta$ . Moreover, one can choose any distribution on  $\{1, 2, \dots\}$  for  $M$  as need may dictate. The estimator  $\hat{\theta}$  will be defined for a sample of size

$$(6) \quad L = 3 + \min(M, N) \quad (\text{a stopping time}),$$

where  $N$  is defined by (5). Thus  $L$  is almost surely finite for all  $\theta$  and, in fact, the expected sample size  $E_\theta L$  is uniformly bounded for all  $\theta$  providing  $M$  has a finite expectation. Since  $E_0 L \leq 3 + E_0 N < \infty$ , one can contradict the Cramér-Rao bound, i.e., show that

$$\text{Var}_0 \hat{\theta} \cdot E_0 L < \sigma^2,$$

by showing that, with one's freedom to choose a distribution for  $M$ , the value of  $E_0 \hat{\theta}^2$  can be made arbitrarily small.

a) Definition of  $\hat{\theta}$ : Let

$$\begin{aligned} \hat{\theta} &= 0 && \text{when } N < M \\ &= \frac{P_0(N_M = n | S_n)}{P_0(N \geq n, M = n | S_n)} && \text{when } N \geq M = n, \end{aligned}$$



$n \geq 1$ , where  $\gamma \in (\frac{1}{2}, 1)$  is as in (5), and

$$(7) \quad N_m = \inf\{n \geq m: |S_k| \geq k^\gamma, k_n < k \leq n\}, \quad m \geq 1.$$

The sequence  $k_n$  is required to go to infinity as  $n \rightarrow \infty$  and to satisfy  $0 \leq k_n < n$ ,  $n \geq 1$ . The properties of  $\hat{\theta}$  depend upon the values of  $k_n$  and  $p_n = \Pr(M=n)$ ,  $n \geq 1$ , which remain to be specified. It is easily checked that  $\hat{\theta}$  is a function of the observed sample  $X, Y, Z, X_1, X_2, \dots, X_{M \wedge N}$ , where  $M \wedge N$  denotes the minimum of  $M$  and  $N$ , and, hence, can be used to estimate  $\theta$ .

b) Moments of  $\hat{\theta}$ : Since  $S_n$  is a sufficient statistic for  $\theta$  relative to the observations  $X_1, \dots, X_n$ , the probabilities

$$P_\theta(N_M=n|S_n), \quad P_\theta(N \geq n, M=n|S_n)$$

are actually independent of  $\theta$ . Let  $I = \{n \geq 1: p_n > 0\}$ . For integer  $r \geq 1$ ,

$$\begin{aligned} E_\theta |\hat{\theta}|^r &= \sum_{n \in I} \int_{[M=n \leq N]} |X|^r \left| \frac{P_\theta(N_M=n|S_n)}{P_\theta(N \geq n, M=n|S_n)} \right|^r dP_\theta \\ (8) \quad &= E_\theta |X|^r \sum_{n \in I} E_\theta \left\{ \left| \frac{P_\theta(N_M=n|S_n)}{P_\theta(N \geq n, M=n|S_n)} \right|^r P_\theta(M=n \leq N|S_n) \right\} \\ &= E_\theta |X|^r \sum_{n \in I} E_\theta \{ P_\theta^r(N_M=n|S_n) / P_\theta^{(r-1)}(N \geq n, M=n|S_n) \} \end{aligned}$$

and, consequently,

$$E_\theta |\hat{\theta}| = E_\theta |X| P_\theta(N_M \in I) < \infty, \quad -\infty < \theta < \infty.$$

Thus  $E_\theta \hat{\theta}$  is defined, and it follows, as above, that

$$(9) \quad E_\theta \hat{\theta} = E_\theta X P_\theta(N_M \in I) = \theta P_\theta(N_M \in I).$$

(c) The unbiasedness of  $\hat{\theta}$ : In view of (9),  $\hat{\theta}$  is unbiased whenever  $P_{\theta}(N_M \in I) = 1$ ,  $\theta \neq 0$ . First, observe that this can not hold if  $I$  has a largest index  $n_0$ , for there is positive probability (for every  $\theta$ ) that  $|S_k| < k^\gamma$ ,  $1 \leq k \leq n_0$ , forcing  $N_M$  to exceed  $n_0$  (cf., (7)). On the other hand, if  $I$  is an unbounded continuous sequence of integers with minimal element  $n_0$ , then  $N_M \geq M \geq n_0$  (cf., (7)) and, consequently,  $N_M \in I$  whenever  $N_M < \infty$ . Now, when  $\theta \neq 0$ ,  $P_{\theta}(|S_k| \leq k^\gamma \text{ i.o.}) = 0$ , since  $\gamma < 1$ . And since  $k_n \rightarrow \infty$ ,  $P_{\theta}(N_M < \infty) = 1$  (cf., (7)),  $\theta \neq 0$ . Thus  $\hat{\theta}$  is unbiased whenever the support of  $M$  is an unbounded interval of integers. We note, in passing, that when  $I$  is bounded, so is the stopping time  $L$ , and, from the discussion in part (c) of Section 2, it is apparent that, in such a situation, every unbiased estimator satisfies the Cramér-Rao bound.

(d) A needed lemma: For some positive constant  $c$  depending only on  $\gamma$ ,

$$(10) \quad P_0(|S_k| \geq k^\gamma, m+1 \leq k \leq n | S_n) \leq c P_0(|S_k| \geq k^\gamma, m \leq k \leq n | S_n)$$

almost surely,  $1 \leq m \leq n < \infty$ .

Proof. For definiteness, let  $\sigma^2 = 1$ . Then, conditional on  $S_{m+1}$  being  $v$ ,  $S_m$  is normally distributed with mean  $\frac{m}{m+1}v$  and variance  $\frac{m}{m+1}$ . Consequently,

$$\begin{aligned} P_0(|S_m| \geq m^\gamma | S_{m+1} = v) &= 1 - \Phi\left(\frac{m^\gamma - \frac{m}{m+1}v}{\sqrt{m/(m+1)}}\right) + \Phi\left(\frac{-m^\gamma - \frac{m}{m+1}v}{\sqrt{m/(m+1)}}\right) \\ &\geq \Phi\left(\frac{\frac{m}{m+1}|v| - m^\gamma}{\sqrt{m/(m+1)}}\right). \end{aligned}$$

When  $|v| \geq (m+1)^\gamma$ ,

$$(11) \quad P_0(|S_m| \geq m^\gamma |S_{m+1} = v) \geq \phi \left( \frac{\frac{m}{m+1}(m+1)^\gamma - m^\gamma}{\sqrt{m/(m+1)}} \right) \geq c^{-1}$$

for some sufficiently large constant  $c$  depending on  $\gamma$ , but not on  $m$  (or  $n$ ). (The argument of  $\phi$  in (11) approximately equals  $-(1-\gamma)m^{-(1-\gamma)}$ , which is bounded below for  $\gamma < 1$ .) Inequality (10) easily follows from (11).  $\square$

Since  $N > n$  says that  $|S_k| \geq k^\gamma$  for  $1 \leq k \leq n$ , and  $N_M = n$  implies that  $|S_k| \geq k^\gamma$  for  $k_n+1 \leq k \leq n$ , it follows from the lemma that

$$(12) \quad P_0(N_M = n | S_n) \leq c^{k_n} P_0(N > n | S_n) \quad \text{almost surely, } n \geq 1.$$

(e) The variance of  $\hat{\theta}$  at  $\theta = 0$ : According to (8),

$$(13) \quad E_0 \hat{\theta}^2 = E_0 X^2 \sum_{n \in I} E_0 \{ P_0^2(N_M = n | S_n) / P_0(N \geq n, M = n | S_n) \}.$$

Since

$$P_0(N \geq n, M = n | S_n) \geq p_n P_0(N > n | S_n),$$

it follows from (12) and (13), and then (7), that

$$\begin{aligned} E_0 \hat{\theta}^2 &= \sigma^2 \sum_{n \in I} c^{k_n} p_n^{-1} P_0(N_M = n) \\ &\leq \sigma^2 \sum_{n \in I} c^{k_n} p_n^{-1} P_0(|S_n| \geq n^\gamma). \end{aligned}$$

The latter expression can be made arbitrarily small by letting  $k_n \rightarrow \infty$  very slowly, by letting  $p_n$  be zero until  $n$  is quite large, and then letting  $p_n \rightarrow 0$  at a reasonably slow rate. This is because  $P_0(|S_n| \geq n^\gamma) = o(e^{-(n^{2\gamma-1}/2\sigma^2)})$  as  $n \rightarrow \infty$  and, hence, it goes to zero at a suitably fast rate. In fact, one may let  $p_n \rightarrow 0$  sufficiently fast that  $M$  has all of its moments finite. Thus, not only can one find an unbiased estimator  $\hat{\theta}$  which violates the Cramér-Rao inequality at  $\theta = 0$ , but this can be done with a stopping time  $L$  whose  $r$ -th

moment is uniformly bounded in  $\theta$  (cf., (6)) for every  $r \geq 1$ .

We have not been able to prove that  $\hat{\theta}$  has a finite variance for all  $\theta$  or found a way to modify the approach used in this second counter-example to obtain a finite variance for all  $\theta$ . While this is not a necessary attribute in a counter-example, it would have been nice to have been able to do so. Recall that the variance of  $\hat{\theta}$  is finite for all  $\theta$  in the first counter-example.

It would be interesting to know whether one can cause the Cramér-Rao inequality to be violated at more than one value of  $\theta$ . This seems likely. The author suspects that "violation sets" must be of Lebesgue measure zero (by analogy with what happens with "super-efficiency," cf., LeCam (1953)), but knows of no argument which rules out larger sets.

#### References

- [1] Blackwell, D. and Girshick, M.A. (1947). A lower bound for the variance of some unbiased sequential estimates. Ann. Math. Statist., 18, 277-280.
- [2] Chapman, D.G. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. Ann. Math. Statist. 22, 581-586.
- [3] Chow, Y.S., Robbins, H. and Teicher, H. (1965). Moments of randomly stopped sums. Ann. Math. Statist. 36, 789-799.
- [4] Ibragimov, I.A. and Khas'minskii, R.Z. (1974). On sequential estimation of the location parameter for families of distributions with discontinuous densities. Theor. Probability Appl. 19, 669-682.
- [5] LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. University of California Publ. in Statist. 1, 277-330.
- [6] Lehmann, E.L. and Stein, C. (1950). Completeness in the sequential case. Ann. Math. Statist. 21, 376-385.
- [7] Stein, C. (1950). Unbiased estimates with minimum variance. Ann. Math. Statist. 21, 406-415.
- [8] Wolfowitz, J. (1947). The efficiency of sequential estimates and Wald's equation for sequential processes. Ann. Math. Statist. 18, 215-230.