

# SOME APPLICATIONS OF THE CRAMÉR-RAO INEQUALITY

J. L. HODGES, JR. AND E. L. LEHMANN  
UNIVERSITY OF CALIFORNIA

## 1. Summary and introduction

In 1945 and 1946, Cramér [1] and Rao [2] independently investigated the problem of obtaining a simple lower bound to the variance of point estimates. In 1947 Wolfowitz [3] simplified the conditions under which Cramér had obtained this bound and extended the result to sequential estimates. In the present paper, use is made of the Cramér-Rao result, in Wolfowitz's form, to investigate some problems of the minimax theory of estimation.

The Bayes method for obtaining minimax estimates developed by Wald since 1939 [4], [5], is completely satisfactory whenever the minimax estimate is the Bayes solution for some *a priori* distribution of the parameter. However, frequently minimax estimates are not Bayes solutions, but only limits of Bayes solutions. When this occurs, the possibility is left open that the minimax estimate is not admissible; that is, that there exists some other minimax estimate whose risk is never greater and is for some parameter value less than that of the given estimate.

In section 2 we consider certain estimation problems in which the loss is proportional to the square of the error of estimate, and use the Cramér-Rao bound to establish directly that certain estimates, which can be shown to be minimax by the Bayes method, are in addition admissible. In section 3 we consider several problems of sequential estimation, for some of which previously no minimax estimates have been known. In all of these cases it turns out that there are minimax estimates based on samples of fixed size.

Problems similar to those treated in the present paper were considered simultaneously by Girshick and Savage [6], the scope of whose work is much larger than ours. Portions of both papers were presented at the joint colloquium of the Stanford and California statistical groups, resulting in a fruitful exchange of ideas. The method introduced here has been employed in [6] to obtain extensions of some of our results.

## 2. Estimates based on samples of fixed size

Let  $X$  be a random variable with distribution  $P_\theta$ ,  $\theta \in D$ , so that the probability that  $X$  falls in a set  $A$  is given by

$$(2.1) \quad P_\theta(A) = \int_A p_\theta(x) d\mu(x).$$

Let  $f(X)$  be any estimate of  $\theta$  and let  $b_f(\theta) = E_\theta[f(X)] - \theta$  be its bias. Then the Cramér-Rao inequality states that the variance  $\sigma_f^2(\theta)$  of  $f(X)$  satisfies

$$(2.2) \quad \sigma_f^2(\theta) \geq \frac{[1 + b'_f(\theta)]^2}{E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log p_\theta(X) \right)^2 \right]}.$$

We shall now prove a theorem which will essentially reduce the problem of proving that certain estimates are admissible and minimax, to proving that there is a unique solution to a differential inequality related to (2.2). It will be convenient to associate with each bias function  $b(\theta)$  the function  $C_b(\theta)$  defined by

$$(2.3) \quad C_b(\theta) = b^2(\theta) + \frac{[1 + b'(\theta)]^2}{E_\theta \left[ \frac{\partial}{\partial \theta} \log p_\theta(X) \right]^2}.$$

If the loss is defined to be the square of the error of estimation, (2.3) has the significance of a lower bound on the risk of an estimate whose bias function is  $b(\theta)$ . Suppose now that  $g(X)$  is an estimate for which the risk everywhere attains this lower bound. We may then substitute, in a proof of the admissibility of  $g(X)$ , the bound (2.3) for the actual risk.

**THEOREM 1.** *If the loss is squared error, if  $g(X)$  is an estimate for which (2.2) becomes an equality, if the inequality (2.2) is satisfied for all estimates, and if, for every bias function  $b(\theta)$ ,*

$$(2.4) \quad C_b(\theta) \leq C_{b_\theta}(\theta)$$

*for all  $\theta \in D$  implies  $b(\theta) \equiv b_\theta(\theta)$ , then  $g(X)$  is admissible.*

**PROOF.** Since loss is squared error,  $R_f(\theta) = b_f^2(\theta) + \sigma_f^2(\theta) \geq C_{b_f}(\theta)$ . Suppose for some estimate  $f(X)$ ,  $R_g(\theta) \geq R_f(\theta)$  for all  $\theta \in D$ . Since by assumption  $C_{b_g}(\theta) = R_g(\theta)$ , we have  $C_{b_g}(\theta) \geq C_{b_f}(\theta)$  and from (2.4) conclude  $b_g(\theta) = b_f(\theta)$ . From this follows  $b'_g(\theta) = b'_f(\theta)$ ,  $C_{b_g}(\theta) = C_{b_f}(\theta)$ ,  $R_f(\theta) \geq C_{b_f}(\theta) = C_{b_g}(\theta) = R_g(\theta)$ , and hence  $R_f(\theta) = R_g(\theta)$ .

**COROLLARY 1.** *Theorem 1 remains valid if the loss function is squared error divided by a function  $q(\theta)$  which is everywhere positive and finite.*

**PROOF.** Admissibility is not affected when the risk function is divided by such a function.

**COROLLARY 2.** *If in addition to the assumptions of theorem 1 we assume that  $g(X)$  is a constant risk estimate, then  $g(X)$  is an admissible minimax estimate.*

**PROOF.** Any constant risk admissible estimate is minimax.

**Remarks.** (i) A statistical problem is not completely specified until the loss function has been stated. Squared error is the classical loss function for estimation, primarily for reasons of convenience [7, p. 516]. An alternative loss function is obtained if we divide squared error by the variance of  $X$ , thus measuring the seriousness of errors in terms of the difficulty of estimation as reflected by  $\sigma_X^2$  as a function of  $\theta$ . This alternative approach is particularly desirable in those problems for which, when the loss function is squared error itself, the minimax risk is infinite. For, when this happens, every estimate is minimax and the minimax principle provides no basis for choice.

## CRAMÉR-RAO INEQUALITY

Those loss functions obtained by dividing squared error by a function of  $\theta$  have been termed "quadratic loss functions" by Girshick and Savage [6].

(ii) In all of the problems considered below, the family of distributions is complete in the sense of [8], and hence every estimate is uniquely determined by its bias function. Since in the proof of theorem 1 we have established the uniqueness of the admissible minimax bias function, it follows that the estimates shown below to be admissible minimax estimates are in fact the unique minimax estimates.

(iii) In the statistical applications which we shall make of this theorem, we sometimes replace a sample  $X_1, X_2, \dots, X_n$  by a single sufficient statistic, say  $X$ . It is known that nothing is lost by this simplification, since from the risk point of view one may restrict oneself to estimates which are nonrandomized functions of a sufficient statistic [9].

Actually, it is not necessary to work with the single sufficient statistic, since the Cramér-Rao inequality may be applied directly to a sample, but the regularity conditions are easier to check when dealing with a single variable.

As an application of theorem 1 we shall now consider five specific problems, showing in each case that a given estimate is admissible and minimax. To apply theorem 1 we must check the validity of (2.2) for all estimates. By the method of proof given by Wolfowitz [3], (2.2) can be shown to be valid under the following assumptions:

- (i) The parameter  $\theta$  lies in an open interval  $D$  of the real line, which may be infinite or semi-infinite;
- (ii) For almost all  $x$ ,  $\frac{\partial p_\theta(x)}{\partial \theta}$  exists for all  $\theta \in D$ ;
- (iii) The expression  $\int p_\theta(x) d\mu(x)$  may be differentiated under the integral sign;
- (iv) For every  $\theta \in D$ ,  $E_\theta \left[ \frac{\partial \log p_\theta(X)}{\partial \theta} \right]^2 > 0$ ;
- (v) The expression  $\int f(x) p_\theta(x) d\mu(x)$  may be differentiated under the integral sign.

The problems we treat concern the binomial, Poisson, normal, and chi-square distributions, and we now check the validity of (2.2) for these distributions.

**LEMMA 1.** *If  $p_\theta(x)$  is any of the following:*

- (a)  $\binom{n}{x} \theta^x (1-\theta)^{n-x}$ ,  $x = 0, 1, \dots, n$ ;  $0 < \theta < 1$ ;  $\mu = \text{counting measure}$ ;
- (b)  $\frac{\theta^x}{x!} e^{-\theta}$ ,  $x = 0, 1, \dots$ ;  $0 < \theta < \infty$ ;  $\mu = \text{counting measure}$ ;
- (c)  $\frac{1}{\sqrt{2\pi}\sigma} e^{-1/2(x-\theta)^2/\sigma^2}$ ,  $-\infty < x < \infty$ ;  $-\infty < \theta < \infty$ ;  $\mu = \text{Lebesgue measure}$ ;
- (d)  $\frac{x^{\nu/2-1} e^{-x/2\theta}}{2^{\nu/2} \Gamma(\frac{\nu}{2}) \theta^{\nu/2}}$ ,  $0 < x < \infty$ ;  $0 < \theta < \infty$ ;  $\mu = \text{Lebesgue measure}$ ;

*then (2.2) is satisfied.*

**PROOF.** Conditions (i)-(iv), none of which involve the estimate  $f(X)$ , are obviously satisfied. In checking condition (v), there is no loss of generality in assuming that  $f(X)$  has finite variance, since otherwise (2.2) certainly holds.

For distribution (a), condition (v) is obvious, since we are dealing with a finite sum. In cases (c) and (d) the result follows immediately from well known properties of the bilateral and unilateral Laplace transform, respectively. In case (b) our assumption guarantees the absolute convergence of the power series

$\sum_{x=0}^{\infty} f(x) \theta^x/x!$  in the open interval  $0 < \theta < \infty$ , and hence the series may be differentiated term by term in that interval.

In the examples below we need in each case only check (2.4), the remaining conditions of theorem 1 obviously being satisfied.

*Problem 1.* Let  $X_1, X_2, \dots, X_n$  be a sample from the normal distribution with unknown expectation  $\theta$  and known variance which we may without loss of generality take to be 1. Let the loss be squared error. It has been known for some time that  $\bar{X} = \sum X_i/n$  is a minimax estimate for  $\theta$ . This result was obtained by Stein and Wald [10] for a different loss function for the much harder sequential problem, and was proved explicitly for the loss function here employed by Wolfowitz [11]. We shall now use theorem 1 to prove both the admissibility and minimaxity of  $\bar{X}$ .

Since  $\bar{X}$  is sufficient we need only consider estimates of the form  $f(\bar{X})$ , and since  $\bar{X}$  is normally distributed we may apply (c) of lemma 1. We need only check (2.4) which now becomes

$$(2.5) \quad b^2(\theta) + \frac{1}{n} [1 + b'(\theta)]^2 \leq \frac{1}{n}$$

for every  $\theta \in D$  implies  $b(\theta) \equiv 0$ . Since neither term on left side of (2.5) can be negative,  $|b(\theta)|$  is bounded and  $b'(\theta)$  is never positive. Consequently there exists a sequence  $\{\theta_i\}$  for which  $b'(\theta_i)$  approaches 0 as  $|\theta_i|$  approaches  $\infty$ , and hence by the hypothesis of (2.5),  $b(\theta_i)$  does likewise. But since  $b(\theta)$  is monotone, it must always be 0.

It is interesting to observe that if we assume certain additional information about  $\theta$ , the estimate  $\bar{X}$  may continue to be minimax without any longer being admissible. This is the case, for example, if we assume it known that  $\theta > \theta_0$ . For,  $b^2(\theta) + [1 + b'(\theta)]^2/n$  is still a lower bound for the risk, and by an argument analogous to the one just given it is easily seen that  $\sup_{\theta > \theta_0} b^2(\theta) + [1 + b'(\theta)]^2/n \geq 1/n$ . Hence the minimax risk is still 1 and  $\bar{X}$  is minimax; its inadmissibility follows from the fact that  $P(\bar{X} < \theta_0) > 0$ , [9].

*Problem 2.* Let  $X_1, \dots, X_n$  be a sample from a Poisson distribution with unknown mean  $\theta$ . For the loss we take squared error divided by  $\theta$ ; see remark (i) above. Since  $X = \sum_{i=1}^n X_i$  is sufficient for  $\theta$  we may restrict consideration to estimates  $f(X)$ . Taking  $g(X) = X/n$ , we shall prove admissibility of  $X/n$  by checking (2.4), which now becomes the condition

$$(2.6) \quad \frac{n b^2(\theta)}{\theta} + [1 + b'(\theta)]^2 \leq 1 \quad \text{for all } \theta \in D$$

implies  $b(\theta) = 0$  for all  $\theta \in D$ .

CRAMÉR-RAO INEQUALITY

Since neither term on the left can be negative,  $|b(\theta)|$  is bounded by  $\sqrt{\theta/n}$  and  $b'(\theta)$  satisfies the inequality

$$b'(\theta) \leq \sqrt{1 - \frac{nb^2(\theta)}{\theta}} - 1 \leq -\frac{nb^2(\theta)}{2\theta}.$$

Thus  $\lim_{\theta \rightarrow 0} b(\theta) = 0$ ,  $b'(\theta) \leq 0$ , and hence  $b(\theta) \leq 0$  for all  $\theta$ . But if for some  $\theta_0$ ,  $b(\theta_0)$  were negative, it would thereafter always be less than or equal to the function  $c(\theta)$  for which  $c'(\theta) = -nc^2(\theta)/(2\theta)$  and  $c(\theta_0) = b(\theta_0)$ . This latter function may be obtained explicitly by solving the differential equation, and is easily seen not to be absolutely bounded by  $\sqrt{\theta/n}$ .

We justified the choice of loss functions for the present problem in part by the remark that there exists no estimate with bounded risk function when the loss is squared error. That this is so is easily seen. For let  $f(X)$  be such an estimate. Then

$$c_{b_f}(\theta) = b_f^2(\theta) + \frac{\theta}{n} [1 + b'_f(\theta)]^2$$

is bounded. But boundedness of the second term of  $c_{b_f}(\theta)$  implies that  $b'_f(\theta) \leq -1 + \epsilon$  for all sufficiently large  $\theta$ , and hence the unboundedness of the first term. An analogous remark applies to the  $\chi^2$ -problem which we treat next and, in general, whenever the range of  $\theta$  and the function  $E_\theta \left[ \frac{\partial}{\partial \theta} \log p_\theta(X) \right]^2$  are both unbounded and (2.2) holds for all estimates.

*Problem 3.* We next consider the estimation of the parameter in the chi-square distribution. This problem arises, for example, if we have a sample  $X_1, X_2, \dots, X_n$  from a normal distribution of known expectation but unknown variance. Then the

statistic  $\sum_{i=1}^n [X_i - E(X_i)]^2$  is sufficient for the variance and has a chi-square distribution. Suppose now that  $X$  has a chi-square distribution of  $n$  degrees of freedom with expectation  $\theta$ , and take the loss to be squared error divided by  $\theta^2$ . This loss function is chosen according to the principle discussed in remark (i) above. We shall now show, by means of theorem 1, that the estimate  $g(X) = X/(n+2)$  is the unique admissible minimax estimate for  $\theta$ . It is interesting that this estimate is biased, while the minimum variance unbiased estimate has constant risk but is neither minimax nor admissible.

Condition (2.4) of theorem 1 now becomes the condition

$$(2.7) \quad \frac{b^2(\theta)}{\theta^2} + \frac{2}{n} [1 + b'(\theta)]^2 \leq \frac{2}{n+2} \text{ for every } \theta \in D$$

implies  $b(\theta) \equiv -2\theta/(n+2)$ .

Since neither term on the left of the hypothesis of (2.7) can be negative, we have  $b'(\theta) < 0$  and  $|b(\theta)| < \theta$ . It follows that  $b(0+) = 0$ . If for any  $0$ ,  $b'(\theta) = b(\theta)/\theta$ , their common value must be  $-2/(n+2)$ , since the expression  $r^2 + 2[1+n]r/n$  has a minimum of  $2/(n+2)$  when  $r = -2/(n+2)$ .

We next observe that  $b'(\theta) \leq b(\theta)/\theta$ . For, suppose that for some  $\theta$ ,  $b'(\theta) > b(\theta)/\theta$ . Then we should have

$$\left[ \frac{b(\theta)}{\theta} \right]^2 + \frac{2}{n} [1 + b'(\theta)]^2 > \left[ \frac{b(\theta)}{\theta} \right]^2 + \frac{2}{n} \left[ 1 + \frac{b(\theta)}{\theta} \right]^2,$$



which, by the previous paragraph, is not less than  $2/(n+2)$ . But this contradicts (2.7). Observing  $\theta^2[b(\theta)/\theta]' = \theta b'[\theta - b(\theta)]$ , we conclude that  $b(\theta)/\theta$  is a nonincreasing function of  $\theta$ .

We shall now prove that  $b'(\theta)$  is not bounded away from  $b(\theta)/\theta$  for large  $\theta$ . For suppose  $b'(\theta) \leq b(\theta)/\theta - \epsilon$  for all  $\theta \geq \theta_0$ . Then for  $\theta \geq \theta_0$ ,  $b(\theta)$  will lie below that function  $c(\theta)$  for which  $c(\theta_0) = b(\theta_0)$  and  $c'(\theta) = c(\theta)/\theta - \epsilon$ . But  $c(\theta) = \epsilon \cdot \theta \log \theta + k \cdot \theta$ , which violates  $-\theta < b(\theta) \leq c(\theta)$ . Analogously,  $b'(\theta)$  cannot be bounded away from  $b(\theta)/\theta$  as  $\theta \rightarrow 0$ . For otherwise  $b(\theta)/\theta \geq c(\theta)/\theta > 0$  for  $\theta$  sufficiently small, while we know that  $b(\theta) \leq 0$ .

We next see that if for some sequence  $\{\theta_i\}$ ,  $b'(\theta) - b(\theta_i)/\theta_i \rightarrow 0$  then  $b(\theta_i)/\theta_i \rightarrow -2/(n+2)$ . For, the hypothesis of (2.7) may be written

$$(2.8) \quad \left\{ \left[ \frac{b(\theta)}{\theta} \right]^2 + \frac{2}{n} \left[ 1 + \frac{b(\theta)}{\theta} \right]^2 \right\} + \frac{2}{n} \left\{ \left[ b'(\theta) - \frac{b(\theta)}{\theta} \right]^2 - 2 \left[ b'(\theta) - \frac{b(\theta)}{\theta} \right] \left[ 1 + \frac{b(\theta)}{\theta} \right] \right\} \leq \frac{2}{n+2}$$

and our hypothesis implies that the second term on the left side of (2.8) approaches 0; consequently the first term must approach its minimum which implies our statement.

Combining the results of the two preceding paragraphs, and using the monotonicity of  $b(\theta)/\theta$ , we see that  $b(\theta)/\theta \rightarrow -2/(n+2)$  as  $\theta \rightarrow 0$  or  $\infty$ , whence our result (2.7) follows.

*Problem 4.* Suppose that  $X$  has the binomial distribution (a) of lemma 1, and that the loss is squared error divided by  $\theta(1-\theta)$ . Condition (2.4) now becomes the condition

$$(2.9) \quad b^2(\theta) + \frac{\theta(1-\theta)}{n} [1 + b'(\theta)]^2 \leq \frac{\theta(1-\theta)}{n} \text{ implies } b(\theta) \equiv 0.$$

Letting  $\theta$  tend to 0 and 1 yields  $b(0) = b(1) = 0$ , while  $b'(\theta) \leq 0$  since  $b^2(\theta)$  is nonnegative.

*Problem 5.* As a final example, we consider the previous problem, using however the classical squared error loss function. It is already known [9] that in this case  $\frac{X}{\sqrt{n}(\sqrt{n}+1)} + \frac{1}{2(\sqrt{n}+1)}$  is an admissible minimax estimate for  $\theta$ , but we shall now establish this fact as a consequence of theorem 1. The verification of (2.4) involved will in any case be needed when considering the sequential problem in the next section.

Condition (2.4) now becomes

$$(2.10) \quad b^2(\theta) + \frac{\theta(1-\theta)}{n} [1 + b'(\theta)]^2 \leq \frac{1}{4(\sqrt{n}+1)^2} \text{ implies } b(\theta) \equiv \frac{\frac{1}{2} - \theta}{\sqrt{n}+1}.$$

**PROOF.** Since the second term on the left side of the hypothesis of (2.10) cannot be negative, we have  $|b(\theta)| \leq \frac{1}{2(\sqrt{n}+1)}$ . We next observe that  $b(\theta) \geq \frac{\frac{1}{2} - \theta}{\sqrt{n}+1}$  for  $\frac{1}{2} \leq \theta < 1$ . For if on the contrary we had, for some  $\frac{1}{2} \leq \theta_0 < 1$ ,

$b(\theta_0) < \frac{\frac{1}{2} - \theta}{\sqrt{n+1}}$ , we should be able to find a point  $\theta_0 < \theta_1 < 1$  at which  $b'(\theta_1) = \frac{b(1) - b(\theta_0)}{1 - \theta_0} > -\frac{1}{\sqrt{n+1}}$ , and at which  $b(\theta_1) < \frac{\frac{1}{2} - \theta}{\sqrt{n+1}}$ . It is clear from the identical satisfaction of (2.10) by  $b(\theta) = \frac{\frac{1}{2} - \theta}{\sqrt{n+1}}$  that this would imply the violation of (2.10) at  $\theta_1$ .

By the symmetrical argument in the interval  $0 < \theta \leq \frac{1}{2}$ , we find that in that interval  $b(\theta) \leq \frac{\frac{1}{2} - \theta}{\sqrt{n+1}}$ , and hence that  $b(\frac{1}{2}) = 0$ . It also follows that  $b'(\frac{1}{2}) \geq -\frac{1}{\sqrt{n+1}}$ ; but on substituting  $\frac{1}{2}$  for  $\theta$  in the hypothesis of (2.10) and using  $b(\frac{1}{2}) = 0$ , we find  $b'(\frac{1}{2}) = -\frac{1}{\sqrt{n+1}}$ .

We can now conclude that (2.10) is satisfied. For suppose  $b(\theta)$  satisfies the hypothesis of (2.10). By symmetry we need only consider the interval  $\frac{1}{2} \leq \theta < 1$ , and need only show that  $b(\theta_0) > \frac{\frac{1}{2} - \theta}{\sqrt{n+1}}$  for  $\frac{1}{2} < \theta_0 < 1$  leads to a contradiction. Consider the function  $c(\theta) = b(\theta) - \frac{\frac{1}{2} - \theta}{\sqrt{n+1}}$ .  $c(\theta)$  is continuous and has a continuous derivative, is nonnegative for  $\frac{1}{2} \leq \theta < 1$ , and  $c(\frac{1}{2}) = c'(1) = 0$ . Hence for every  $\epsilon > 0$  and every  $k > 0$  we can find  $\frac{1}{2} < \theta_1 < \theta_0$  for which  $|c'(\theta_1)| < \epsilon$  and  $c'(\theta_1) > kc(\theta_1)$ . This is easily seen by considering  $[\log c(\theta)]' = c'(\theta)/c(\theta)$ , and using the continuity of  $c'(\theta)$ .

Since  $b(\theta)$  is assumed to satisfy the hypothesis of (2.10) we can subtract to obtain  $c(\theta) \left[ b(\theta) + \frac{\frac{1}{2} - \theta}{\sqrt{n+1}} \right] + \frac{\theta(1-\theta)}{n} c'(\theta) \left[ 2 + b'(\theta) - \frac{1}{\sqrt{n+1}} \right] \leq 0$ . Take now  $\epsilon < 2 - \frac{2}{\sqrt{n+1}}$  and  $k > \frac{-n}{2\theta_0(1-\theta_0)(\sqrt{n}-\epsilon)}$  to obtain a contradiction.

### 3. Estimates based on sequential procedures

In the previous section we have considered only the class of estimates based on a sample of fixed size, and have shown that certain estimates are optimum within this class. However, as is well known [12], the efficiency of statistical procedures can often be improved by taking the observations sequentially. Various definitions of optimum sequential procedures are possible within the minimax theory. For example, one may try to minimize the maximum expectation of a linear combination of loss and cost, measuring the latter by the number of observations. Alternatively, one may place a bound on the expected number of observations and try to minimize the maximum expected loss. Both of these formulations have been considered in the literature [5], [10], [11]. We have applied the method of the present paper to obtain optimum sequential procedures only under the second definition; it seems doubtful that our method would give easy results under the first definition. Although the first definition of an optimum estimate has theoretical advantages, in practical applications the second is sometimes more reasonable. This may happen, for example, when cost and loss cannot be measured on a common scale of value, or when budgetary considerations compel one to place a separate bound on the average cost of experimentation.

It is interesting that in a number of problems it turns out that a procedure of fixed sample size  $n$  is optimum among all sequential procedures for which the expected value of the number of observations  $N$  never exceeds  $n$ . For example, this is the case in problems treated by Stein and Wald and by Wolfowitz. We shall now show that the same holds true for the five problems treated in section 2.

The basis of our results in the present section is the extension by Wolfowitz [3] of the Cramér-Rao inequality to the sequential case. Wolfowitz proved under certain regularity conditions that

$$(3.1) \quad \sigma_f^2(\theta) \geq \frac{[1 + b_f'(\theta)]^2}{E_\theta(N) E_\theta \left[ \frac{\partial}{\partial \theta} \log p_\theta(X) \right]^2}$$

where  $p_\theta(x)$  is the density of an individual observation and  $N$  is the (random) number of observations taken.

It is clear that theorem 1, with the obvious modifications, remains valid for the class of all sequential estimates, if we replace inequality (2.2) by inequality (3.1). Further, if we consider only those sequential procedures for which, for some integer  $n$ ,

$$(3.2) \quad E_\theta(N) \leq n \text{ for all } \theta \in D,$$

theorem 1 will be valid if in (3.1) we replace  $E_\theta(N)$  by  $n$ . To extend the results of section 2 to the sequential case, we must verify the satisfaction of (2.4) and of the regularity conditions under which Wolfowitz proved (3.1). We carry out these checks not for all sequential procedures satisfying (3.2), but only for bounded procedures; that is, for procedures such that

$$(3.3) \quad P(N \leq m) = 1 \text{ for all } \theta \in D$$

for some finite number  $m$ .

Since our results will be independent of the value of the bound  $m$ , provided only that it is sufficiently large, the restriction (3.3) is not serious from a practical point of view: any actual experiment does have a bound on the number of observations. However, the restriction is theoretically undesirable. We shall show below that the solutions obtained retain their minimax character when the restriction is removed. On the other hand, our argument does not establish the admissibility of the estimate within the unrestricted class of sequential procedures.

The Wolfowitz regularity conditions are contained in section 3 of his paper [3]. If the sequential procedure is bounded, and if the density is one of those considered in our lemma 1, all of these conditions are trivial, except for his (3.4). An examination of the proof shows that this condition is used only to permit a certain differentiation under the integral sign. We shall assure the applicability of the inequality by checking this differentiability directly.

**LEMMA 3.** *If the sequential procedure is bounded, and if the density is one of those considered in lemma 1, then (3.1) holds.*

**PROOF.** Let  $R_j$  be the set of points  $(x_1, x_2, \dots, x_j)$  for which  $N = j$ . Then

$$(3.4) \quad E_\theta [f(X_1, X_2, \dots, X_N)] = \sum_{j=1}^m \int_{R_j} f(x_1, x_2, \dots, x_j) p_\theta(x_1) p_\theta(x_2) \dots \\ \times p_\theta(x_j) d\mu(x_1) d\mu(x_2) \dots d\mu(x_j).$$



## CRAMÉR-RAO INEQUALITY

In view of the remarks just made we need only check that the right side of (3.4) may be differentiated under the integral sign. For density (a) of lemma 1 there is no difficulty, since we have simply a polynomial in  $\theta$ . With (b) we have a convergent multiple series of nonnegative terms. This can be rearranged as a convergent series of powers of  $\theta$ , which may be differentiated termwise. The normal cases are somewhat more involved. We may assume  $E_\theta |f(X_1, X_2, \dots, X_N)| < \infty$  for all  $\theta \in D$ , and hence the finiteness of each integral on the right of (3.4). Let  $\phi_j(x_1, x_2, \dots, x_j)$  be the characteristic function on  $R_j$ ; we must show the differentiability under the integral of

$$(3.5) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \phi_j(x_1, x_2, \dots, x_j) |f(x_1, x_2, \dots, x_j)| p_\theta(x_1) p_\theta(x_2) \dots \\ \times p_\theta(x_j) d(x_1) d(x_2) \dots d(x_j).$$

Recalling  $p_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2(x-\theta)^2}$ , collecting the exponents, and making an orthogonal transformation with  $y_1 = x_1 + x_2 + \dots + x_j$ , we see that (3.5) may be rewritten as

$$(3.6) \quad e^{-(j/2)\theta^2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} K(y_1, y_2, \dots, y_j) e^{\theta y_1} dy_1 dy_2 \dots dy_j.$$

Using the Fubini theorem, we see that the integral in (3.6) is a convergent Laplace transform and may therefore be differentiated.

A similar argument applies to the chi-square situation, using a unilateral instead of a bilateral Laplace transform.

We can now conclude that the estimates found to be admissible minimax estimates in problems 1-5 of the preceding section continue to have this property in the class of all estimates based on sequential procedures satisfying (3.2) and (3.3). We shall not have to recheck the differential inequalities which result from (2.4), since they are in each case unchanged.

Finally, we observe that condition (3.3) may be removed in all of these problems without effecting the conclusion that all estimates considered are minimax. For, if there were a sequential estimate  $\delta$  not satisfying (3.3) having a maximum risk  $r$  less by  $\epsilon > 0$  than the minimax risk for bounded sequential procedures, then we could construct a bounded sequential estimate with maximum risk  $< r + \frac{\epsilon}{2}$ . To see this, notice that in each of the cases treated there exists an estimate  $\delta_0$  of  $\theta$ , based on a single observation, whose risk is bounded, say by a constant  $k$ . Since  $E(N) \leq n$ ,  $P_\theta(N_\delta > m) \rightarrow 0$  uniformly in  $\theta$ . Let the estimate  $\delta$  be defined as follows. If  $N_\delta \leq m$ , let  $\delta'$  agree with  $\delta$ . If  $N_\delta > m$ , take an  $(m+1)$ -st observation and let  $\delta'$  agree with  $\delta_0(x_{n+1})$ . It is clear that  $P(N_{\delta'} \leq m+1) = 1$ , that  $E(N_\delta) \leq n$ , and that  $\sup_{\theta \in D} R_\theta(\delta')$  can be made less than  $r + \frac{\epsilon}{2}$  by taking  $m$  sufficiently large.

## REFERENCES

- [1] H. CRAMÉR, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85-94.
- [2] C. R. RAO, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37, No. 3 (1945), pp. 81-91.

SECOND BERKELEY SYMPOSIUM: HODGES AND LEHMANN

- [3] J. WOLFOWITZ, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 215–230.
- [4] A. WALD, "Contributions to the theory of statistical estimation and testing hypothesis," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 299–326.
- [5] ———, *Statistical Decision Functions*, Wiley, New York, 1950.
- [6] M. A. GIRSHICK and L. J. SAVAGE, "Bayes and minimax estimates arising from quadratic risk functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 53–73.
- [7] C. F. GAUSS, *Abhandlungen zur Methode der Kleinsten Quadrate*, Berlin, 1887.
- [8] E.L. LEHMANN and H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation —Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.
- [9] J. L. HODGES, JR. and E. L. LEHMANN, "Some problems in minimax point estimation," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 182–197.
- [10] C. STEIN and A. WALD, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 427–433.
- [11] J. WOLFOWITZ, "Minimax estimates of the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 218–230.
- [12] A. WALD, *Sequential Analysis*, Wiley, New York, 1947.