ON INEQUALITIES OF CRAMÉR-RAO TYPE AND ADMISSIBILITY PROOFS

COLIN R. BLYTH[1] and DONALD M. ROBERTS
UNIVERSITY OF ILLINOIS

1. Introduction and Summary

This paper is a discussion of the Hodges-Lehmann [7] method of proving admissibility for quadratic loss.

Section 2 compares the inequality $EU^2 \geqq (EU)^2$ with the best Schwarz inequality $EU^2 \geqq (EU)^2 + \{\text{Cov}(U, V)\}^2/\text{Var } V$ obtainable using linear functions of U and V, and considers invariance properties of these inequalities.

In Section 3, for a random variable X with possible distributions indexed by θ, we define a Cramér-Rao type inequality as one giving a lower bound on Var $T(X)$ in terms of $ET(X)$. Theorem 2 shows that for the best Schwarz inequality Var $T \geqq \{\text{Cov}(T, V)\}^2/\text{Var } V$ using $V = V(X, \theta)$ to be of Cramér-Rao type, it is necessary that V depend on X only through a minimal sufficient statistic; this condition is also sufficient when there is a sufficient statistic with a complete family of possible distributions. In this case of completeness, it follows that the Cramér-Rao and Bhattacharyya inequalities require no regularity conditions beyond existence and nonconstancy of the derivatives involved.

Section 4 describes the Hodges-Lehmann method of proving an estimator T^* admissible for quadratic loss. In this method, the inequality showing that T makes T^* inadmissible is relaxed using the Cramér-Rao type inequality Var $(T - T^*) \geqq \{\text{Cov}(T - T^*, T^*)\}^2/\text{Var } T^*$, and the relaxed inequality is shown to have no nontrivial solutions. In all examples known to us, this use of a Cramér-Rao type inequality can be replaced by a use of the weaker result Var $(T - T^*) \geqq 0$; we suppose there are examples in which this cannot be done, but we have no such example.

Section 5 consists of several examples illustrating this method of proving admissibility.

2. Schwarz's inequality

For real valued random variables U and V, Schwarz's inequality

$$\{EU^2\}\{EV^2\} \geqq \{EUV\}^2 \tag{2.1}$$

means that if EU^2 and EV^2 both exist, then EUV also exists and its **square does** not exceed their product.

[1] Work supported by NSF Grant GP 23835.

If V is a nonzero constant, Schwarz's inequality reduces to the Jensen inequality

$$EU^2 \geqq \{EU\}^2. \tag{2.2}$$

This inequality is invariant under nonzero constant multiplications, and under translations: replacing U by cU just multiplies each side of (2.2) by c^2; and replacing U by $U - a$ just adds $a^2 - 2aEU$ to each side of (2.2).

If V is not a constant, Schwarz's inequality can be written

$$EU^2 \geqq \frac{\{EUV\}^2}{EV^2}. \tag{2.3}$$

This inequality is invariant under nonzero constant multiplications, but is not invariant under translations: replacing U by cU, and V by dV, just multiplies each side by c^2; but replacing U by $U - a$, and V by $V - b$, changes the inequality to

$$E(U - a)^2 \geqq \frac{\{E(U - a)(V - b)\}^2}{E(V - b)^2}, \tag{2.4}$$

that is,

$$EU^2 \geqq \{EU\}^2 - (EU - a)^2 + \frac{\{E(U - a)(V - b)\}^2}{E(V - b)^2}. \tag{2.5}$$

THEOREM 1. *The strongest result obtainable by considering the Schwarz inequality under all possible translations is*

$$EU^2 \geqq \{EU\}^2 + \frac{\{\mathrm{Cov}\,(U, V)\}^2}{\mathrm{Var}\, V}. \tag{2.6}$$

PROOF. To show that the inequality

$$EU^2 \geqq \{EU\}^2 + \sup_{a,b}\left[-(EU - a)^2 + \frac{\{E(U - a)(V - b)\}^2}{E(V - b)^2}\right] \tag{2.6'}$$

reduces to (2.6), notice that the second term on the right in (2.6′) is translation invariant: translations on U and V inside the square brackets just change a and b, and supremum over all a and b is being taken. In particular, (2.6′) is left unchanged when we replace, inside the square brackets only, U by $U_1 = U - EU$ and V by $V_1 = V - EV$. The result of this, after some easy simplifications, is

$$EU^2 \geqq \{EU\}^2 + \sup_{a,b}\left[\frac{-a^2EV_1^2 + 2abEU_1V_1 + (EU_1V_1)^2}{EV_1^2 + b^2}\right]. \tag{2.7}$$

Here the quantity in square brackets has, for each b, its maximum value $(EU_1V_1)^2/EV_1^2$ when $a = b(EU_1V_1)/EV_1^2$. So the inequality (2.6′) becomes

$$EU^2 \geqq \{EU\}^2 + \frac{(EU_1V_1)^2}{EV_1^2}, \tag{2.8}$$

which, written in terms of U and V, is (2.6).

This best Schwarz inequality (2.6) is invariant under nonzero constant multiplications, under translations, and under replacement of U by $U - V$: replacing U by cU, and V by dV, just multiplies each side of (2.6) by c^2; replacing U by $U - a$, and V by $V - b$, just adds $a^2 - 2aEU$ to each side of (2.6); and replacing U by $U - V$ just adds $EV^2 - 2EUV$ to each side of (2.6).

Comparing the Jensen inequality (2.2) and the best Schwarz inequality (2.6), we see that (2.6) is a stronger result unless U and V are uncorrelated. Both inequalities reduce to the same equality when U is a constant. The Jensen inequality is an equality if and only if U is a constant. The best Schwarz inequality is an equality if and only if U and V are linearly related (which is true in particular if U is a constant).

3. Inequalities of Cramér-Rao type

Let X be a random variable with possible probability measures P_θ, $\theta \in \Omega$ on subsets of $\mathscr{X}$. There are no restrictions here on the space $\mathscr{X}$ or on the family of probability measures: any such restrictions will be stated where needed. For

$T = T(X)$ any real valued statistic, and
$V = V(X, \theta)$ any real valued random variable,

the best Schwarz inequality (2.6) gives, at every θ for which $\operatorname{Var} V > 0$,

$$ET^2 \geqq \{ET\}^2 + \frac{\{\operatorname{Cov}(T, V)\}^2}{\operatorname{Var} V}, \tag{3.1}$$

that is,

$$\operatorname{Var} T \geqq \frac{\{\operatorname{Cov}(T, V)\}^2}{\operatorname{Var} V}, \tag{3.2}$$

or equivalently, from the invariance properties of (2.6),

$$E\{T - g(\theta)\}^2 \geqq \{ET - g(\theta)\}^2 + \frac{\{\operatorname{Cov}(T, V)\}^2}{\operatorname{Var} V}, \tag{3.3}$$

for any real valued function g defined on Ω. This inequality appears to give a lower bound on the risk of T as an estimator of $g(\theta)$ for squared error loss $\{T - g(\theta)\}^2$, and obviously also for quadratic loss $a(\theta)\{T - g(\theta)\}^2$ with $a(\theta) > 0$. But the apparent bound is useless because it depends on T: rather than compute $\operatorname{Cov}(T, V)$ to get a lower bound on the risk, we would compute $\operatorname{Var} T$ to get the risk itself.

However, when V is such that $\operatorname{Cov}(T, V)$ depends on T only through ET; that is, when V has the property

$$ET_1 \underset{\theta}{\equiv} ET_2 \Rightarrow \operatorname{Cov}(T_1, V) \underset{\theta}{\equiv} \operatorname{Cov}(T_2, V), \tag{3.4}$$

then the best Schwarz inequality (3.1) takes the very useful form of a lower bound on the risk of T in terms of ET:

$$ET = m(\theta) \Rightarrow ET^2 \geqq \{m(\theta)\}^2 + b_m(\theta), \tag{3.5}$$

that is,

$$ET = m(\theta) \Rightarrow \text{Var } T \geqq b_m(\theta), \tag{3.6}$$

or equivalently,

$$ET = m(\theta) \Rightarrow E\{T - g(\theta)\}^2 \geqq \{m(\theta) - g(\theta)\}^2 + b_m(\theta), \tag{3.7}$$

where $b_m(\theta) = \{\text{Cov }(T, V)\}^2/\text{Var } V$. We will refer to (3.5) as an *inequality of Cramér-Rao type*. The question as to what functions V satisfy condition (3.4) and therefore give Cramér-Rao type inequalities has the following partial answer.

THEOREM 2. *A necessary condition for V to give a Cramér-Rao type inequality is that V depend on X only through a minimal sufficient statistic. This condition is also sufficient, when the minimal sufficient statistic has a complete family of possible distributions.*

PROOF. *Necessity.* For every statistic T and every sufficient statistic S, notice that $E(T|S)$ is a statistic and has the same expectation as T. The property (3.4) for V therefore implies

$$\text{Cov }\{E(T|S), V\} \equiv \text{Cov }(T, V), \tag{3.8}$$

that is,

$$E\{[E(T|S)]V\} - [E\{E(T|S)\}]EV \equiv E(TV) - (ET)(EV), \tag{3.9}$$

that is,

$$E\{[E(T|S)]V\} \equiv E(TV), \tag{3.10}$$

that is,

$$E[E\{[E(T|S)]V|S\}] \equiv E[E(TV|S)], \tag{3.11}$$

that is,

$$E[E(T|S)E(V|S)] \equiv E[E(TV|S)]. \tag{3.12}$$

This is true for every T. In particular, for $T = V(X, \theta_0)$ the above identity gives, at $\theta = \theta_0$,

$$E_{\theta_0}[\{E(V|S)\}^2] = E_{\theta_0}[E(V^2|S)]. \tag{3.13}$$

This shows, for every θ_0 in Ω, that the distribution of V given S must be concentrated on one point, that is, V must be a function of S.

Sufficiency. If S is a sufficient statistic and $V = V(S, \theta)$ we have

$$\text{(3.14)} \qquad \begin{aligned} \operatorname{Cov}(T, V) &= E\{TV(S, \theta)\} - (ET)(EV) \\ &= E\{[E(T|S)]V(S, \theta)\} - (ET)(EV). \end{aligned}$$

When S has a complete family of possible distributions (making S minimal), we see that if $ET_1 \equiv ET_2$ then $E(T_1|S)$, $E(T_2|S)$ are two functions of S which have the same expectation and are therefore (with probability one) the same function, giving us $\operatorname{Cov}(T_1, V) \equiv \operatorname{Cov}(T_2, V)$.

REMARK. When the minimal sufficient statistic S does not have a complete family of possible distributions, $ET_1 \equiv ET_2$ does not imply $\operatorname{Cov}\{T_1, V(S, \theta\} \equiv \operatorname{Cov}\{T_2, V(S, \theta)\}$. A counterexample is provided by the usual example X rectangular $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $-\infty < \theta < \infty$, in which X is minimal sufficient but not complete. Take $V = X$, $T_1 = 0$, $T_2 = \pm 1$ according as the integer nearest to X is above or below X. Then $ET_1 \equiv 0$ and $ET_2 \equiv 0$ but $\operatorname{Cov}(T_1, V) \not\equiv \operatorname{Cov}(T_2, V)$.

COROLLARY 1. *When X has a complete family of possible measures, X is itself a minimal sufficient statistic, and the theorem reduces to this simple form: every best Schwarz inequality is of Cramér-Rao type; that is, every random variable V has property* (3.4).

Without appealing to the theorem, it is clear that ET determines T and therefore determines $\operatorname{Cov}(T, V)$.

COROLLARY 2. *Every Cramér-Rao type inequality is invariant under a sufficiency reduction. If instead of working with X, we work only with the sufficient statistic S, then any Cramér-Rao type bound that was available is still available, and unchanged, because the V involved is a function of X through S only.*

For the standard particular Cramér-Rao type inequalities, this invariance under a sufficiency reduction is usually proved using the factorization theorem for densities.

3.1. *Families of variables V.* In all of this section, we could have taken a family $V_\alpha = V_\alpha(X, \theta)$, $\alpha \in A$ of such V and replaced b_m by $\sup_\alpha[\{\operatorname{Cov}(T, V_\alpha)\}^2/\operatorname{Var} V_\alpha]$. This is a commonly used way of getting a good bound. We have not done this here, because we are considering applications in which an obvious best V is available.

3.2. *Particular inequalities.* Particular choices of the V give the following particular Cramér-Rao type inequalities:

(a) the Cramér-Rao inequality [4] uses

$$\text{(3.15)} \qquad V = \frac{1}{p_\theta(X)}\left\{\frac{\partial}{\partial\theta} p_\theta(X)\right\};$$

(b) the kth Bhattacharyya inequality [2] uses

$$\text{(3.16)} \qquad V_\alpha = \frac{1}{p_\theta(X)}\left\{\sum_{i=1}^{k} c_{i\alpha} \frac{\partial^i}{\partial\theta^i} p_\theta(X)\right\},$$

where the $c_{i\alpha}$ range over all real numbers,

(c) the Barankin inequality [1] uses

$$V_\alpha = \frac{1}{p_\theta(X)}\left\{\sum_{i=1}^{n_\alpha} c_{i\alpha} p_{\theta_i}(X)\right\}, \tag{3.17}$$

where n_α ranges over the positive integers, the $c_{i\alpha}$ range over the reals, and the θ_i range over Ω;

(d) the Chapman-Robbins inequality [3] uses

$$V_\alpha = \frac{1}{p_\theta(X)}\{p_{\theta_\alpha}(X)\}, \tag{3.18}$$

where θ_α ranges over Ω;

(e) the Kiefer inequality [9] uses a continuous mixture, where Barankin uses a discrete mixture, of $p_{\theta_1}(X)$ over points θ_1 of Ω.

In all of these, the family of distributions is taken to be dominated, with P_θ having density p_θ relative to a fixed measure μ. The V used are all of the form $p_{\theta_1}(X)/p_\theta(X)$ where θ indexes the true distribution and θ_1 is any other point of Ω, or linear combinations over values of θ_1 in Ω, or limits of such linear combinations.

Existence of $V = p_{\theta_1}(X)/p_\theta(X)$ requires only that P_θ dominate P_{θ_1}; we can use $\frac{1}{2}(P_\theta + P_{\theta_1})$ for μ and will have $p_{\theta_1}(x) = 0$ whenever $p_\theta(x) = 0$. Such a V satisfies property (3.4) because for $ET = m(\theta)$ we have

$$\begin{aligned} \operatorname{Cov}(T, V) &= E(TV) - (ET)(EV) \\ &= \int t(x)\frac{p_{\theta_1}(x)}{p_\theta(x)} p_\theta(x)\, d\mu(x) - m(\theta) \int \frac{p_{\theta_1}(x)}{p_\theta(x)} p_\theta(x)\, d\mu(x) \\ &= m(\theta_1) - m(\theta), \end{aligned} \tag{3.19}$$

which depends on T only through ET. And such a V will not be a constant, provided P_{θ_1} and P_θ are not the same measure. Thus, for the Barankin and Chapman-Robbins inequalities, all that is needed is that the P_{θ_1} measures used differ from and be dominated by P_θ.

For the Cramér-Rao and Bhattacharyya inequalities, θ must be real valued (easily extended to linear spaces [2]) and the derivatives involved must exist and not be constants. In addition property (3.4), obvious for the Barankin V, must survive the limit operation: this is assured by requiring differentiability under the expectation sign. However, this additional regularity condition is unnecessary when there is sufficient statistic S with a complete family of possible measures, because the Barankin variables V and therefore limits of them depend on X only through S, and *all* functions of S, θ have property (3.4) by Theorem 2.

Barankin [1] shows that when $p_\theta(x) = 0$ implies $p_{\theta_1}(x) = 0$ for all $\theta_1 \in \Omega$, it is enough to consider variables V of his form, because they give an achievable bound. It is unnecessary to use any other V because they cannot give a better bound; but it may be convenient to use other V that give the best bound more easily.

Cramér-Rao type inequalities have two uses in proving something good about T as an estimator of $g(\theta)$ with quadratic loss: (i) in proving minimum risk, either locally or uniformly, for given expectation, and (ii) in proving admissibility.

The second use (ii) is the subject of the rest of this paper; the first use (i) is as follows.

3.3. *Local use* (*i*). Suppose T_0, with $ET_0 = m(\theta)$, achieves equality in (3.5) at $\theta = \theta_0$:

$$E_{\theta_0}\{T_0 - g(\theta_0)\}^2 = \{m(\theta_0) - g(\theta_0)\}^2 + b_m(\theta_0). \tag{3.20}$$

Then every other estimator T with the same expectation $ET = m(\theta)$ has

$$E_{\theta_0}\{T - g(\theta_0)\}^2 \geqq E_{\theta_0}\{T_0 - g(\theta_0)\}^2. \tag{3.21}$$

Therefore among all estimators of $g(\theta)$ having expectation $m(\theta)$, the unique (with P_{θ_0} probability one) one with minimum risk for $\theta = \theta_0$ is T_0; this for every $g(\theta)$ and every quadratic loss. Uniqueness because if T_1 were another estimator with expectation $m(\theta)$ and achieving equality in (3.5) for $\theta = \theta_0$, then $\frac{1}{2}(T_0 + T_1)$ would also have expectation $m(\theta)$ and would violate (3.5) at θ_0 unless $P_{\theta_0}(T = T_0) = 1$.

In particular, taking $g(\theta) = m(\theta)$, we see that T_0 is the unique (with P_{θ_0} probability one) unbiased estimator of $m(\theta)$ with minimum risk (minimum variance) at $\theta = \theta_0$.

3.4. *Uniform use* (i). Suppose T^*, with $ET^* = m^*(\theta)$, achieves equality in (3.5) for all θ:

$$E\{T^* - g(\theta)\}^2 \underset{\theta}{\equiv} \{m^*(\theta) - g(\theta)\}^2 + b_{m^*}(\theta). \tag{3.22}$$

Then every other estimator T with the same expectation $ET = m^*(\theta)$ has $E\{T - g(\theta)\}^2 \geqq E\{T^* - g(\theta)\}^2$, all θ. Therefore among all estimators of $g(\theta)$ having expectation $m^*(\theta)$, the unique (with probability one, all θ) one with uniformly minimum risk (or variance) is T^*; this for every $g(\theta)$ and every quadratic loss.

In particular, taking $g(\theta) = m^*(\theta)$, we see that T^* is the unique (with probability one) unbiased estimator of $m^*(\theta)$.

This use (i) is vacuous when there is a sufficient statistic S with a complete family of measures. The Rao-Blackwell theorem enables us to restrict attention to functions of S, and if T^* has expectation $m^*(\theta)$, then no other T has this same expectation and so the stated uniformly minimum risk properties are obvious.

4. Admissibility proofs for quadratic loss

An estimator T^* is inadmissible for estimating $g(\theta)$, with quadratic loss, if there is an estimator T that is a nontrivial solution of the inadmissibility inequality

$$E\{T - g(\theta)\}^2 \leqq E\{T^* - g(\theta)\}^2, \tag{4.1}$$

or equivalently, writing $m(\theta)$ for ET,

$$\operatorname{Var} T + \{m(\theta) - g(\theta)\}^2 \leqq E\{T^* - g(\theta)\}^2, \tag{4.1'}$$

or, again equivalently,

$$E\{T - T^*\}^2 + 2E\{T - T^*\}\{T^* - g(\theta)\} \leqq 0. \tag{4.1''}$$

By a nontrivial solution (there is always the trivial solution $T = T^*$) we mean one for which the inequality is strict for at least one θ.

Hodges and Lehmann [7] introduced the following method, also used in [6] and [10], of proving admissibility. Using any Cramér-Rao type inequality (3.5) we see that if T is a nontrivial solution of the inadmissibility inequality (4.1) then T is also a nontrivial solution of the relaxed inequality

$$b_m(\theta) + \{m(\theta) - g(\theta)\}^2 \leqq E\{T^* - g(\theta)\}^2. \tag{4.2}$$

This inequality is a relaxation of (4.1), the left side of (4.1) having been replaced by something at least as small. If it can be shown that this relaxed inequality (4.2) has no nontrivial solution m that is the expectation of some T, it therefore follows that (4.1) can have no nontrivial solutions, and so T^* is proved admissible for every quadratic loss.

The Hodges-Lehmann method works directly with the definition of admissibility, using the standard mathematical technique of replacing a complicated expression by a simple bound for it. This replaces an integral inequality in T by an easier inequality in m. When the Cramér-Rao inequality is used, as in the examples of [6], [7], [10], we have $b_m(\theta) = [m'(\theta)]^2/\text{Var } V$, so that (4.2) is always a differential inequality, which in those examples is easily shown to have no nontrivial solutions. (Any best Schwarz inequality could be used on (4.1) in the same way, but failing property (3.4) the relaxed inequality would still involve T so would be no improvement on (4.1).)

4.1. *Which Cramér-Rao type inequality to use.* The Hodges-Lehmann method cannot succeed unless T^* achieves equality in the Cramér-Rao type inequality (3.5) for all θ: otherwise $m^*(\theta) = ET^*$ would be a nontrivial solution of the relaxed inequality (4.2). So for the method to succeed the V used in (3.5) must be a linear function $\alpha(\theta)T^* + \beta(\theta)$ of T^*, or equivalently $V = T^*$ because of the invariance of (2.6) under linear transformations; there is no point in trying special cases of (3.5) such as the Cramér-Rao inequality.

For a Hodges-Lehmann proof of the admissibility of T^*, we check that T^* has property (3.4) [no check is needed when the family of distributions of X is complete], and write down the relaxed inequality (4.2) that results from using $V = T^*$ in (3.5):

$$\frac{\{\text{Cov }(T, T^*)\}^2}{\text{Var } T^*} + \{m(\theta) - g(\theta)\}^2 - E\{T^* - g(\theta)\}^2 \leqq 0. \tag{4.3}$$

Because of the invariance of (2.6) under replacement of U by $U - V$, the application of (3.5) with $V = T^*$ to $E(T - T^*)^2$ in (4.1″) gives the same inequality as (4.3):

$$\frac{\{\text{Cov } T - T^*, T^*)\}^2}{\text{Var } T^*} + \{m(\theta) - m^*(\theta)\}^2 + 2E\{T - T^*\}\{T^* - g(\theta)\} \leqq 0. \tag{4.3'}$$

NOTE 1. The relaxed inequality (4.3′) can be further relaxed to

$$\{m(\theta) - m^*(\theta)\}^2 + 2E\{T - T^*\}\{T^* - g(\theta)\} \leqq 0. \tag{4.4}$$

This amounts to using the Jensen inequality (2.2) instead of the Cramér-Rao type inequality (3.5) on $E(T - T^*)^2$ in (4.1″). If it can be shown that the further relaxed inequality (4.4) has no nontrivial solution m that can be the expectation function of some T, then the inadmissibility inequality (4.1) can have no nontrivial solutions, and T^* is proved admissible. All the examples of [6], [7], [10] can be worked using (4.4) instead of (4.3).

NOTE 2. Can the Hodges-Lehmann method fail to work? That is, can it happen that T^* is admissible so (4.1) has no nontrivial solutions, but (4.3) does have nontrivial solutions? And can it happen that neither (4.1) nor (4.3) has nontrivial solutions, but (4.4) does have nontrivial solutions? We do not yet have answers to these questions, except that it is easy to give examples (see Example 4) in which T^* is a constant and admissible, but (4.4) does have nontrivial solutions (the relaxed inequality (4.3) is not available when T^* is a constant). Negative answers for a particular T^* or class of T^* would permit use of the Hodges-Lehmann method in proving such a T^* inadmissible.

NOTE 3. When X, or a sufficiency reduction of X, has a complete family of distributions, why use a relaxed inequality? After all, the reason given for using (4.3) instead of (4.1) was to get an inequality in m instead of T, and under completeness m determines T so that (4.1) itself can be written in terms of m. The answer is that (4.1) in terms of m may involve an integral that can be eliminated by changing to (4.3). This is what happens in Example 1, where we begin by trying to work with (4.1) without relaxations. Inequality (4.4) has no such essential advantage over (4.3), but may be somewhat simpler and easier to work with than (4.3).

NOTE 4. It is often convenient to carry out these proofs in terms of $Z = T - T^*$ and $\zeta(\theta) = EZ = m(\theta) - m^*(\theta)$. In terms of Z and ζ, the inadmissibility inequality (4.1) is

$$\{\zeta(\theta)\}^2 + 2\zeta(\theta)\{m^*(\theta) - g(\theta)\} + 2\operatorname{Cov}(Z, T^*) + \operatorname{Var} Z \leqq 0, \tag{4.5}$$

and the relaxed inequality (4.3) is

$$\{\zeta(\theta)\}^2 + 2\zeta(\theta)\{m^*(\theta) - g(\theta)\} + 2\operatorname{Cov}(Z, T^*) + \frac{\{\operatorname{Cov}(Z, T^*)\}^2}{\operatorname{Var} T^*} \leqq 0, \tag{4.6}$$

and the further relaxed inequality (4.4) is

$$\{\zeta(\theta)\}^2 + 2\zeta(\theta)\{m^*(\theta) - g(\theta)\} + 2\operatorname{Cov}(Z, T^*) \leqq 0, \tag{4.7}$$

and the Hodges-Lehmann proof consists of showing that $\zeta(\theta) \equiv 0$ is the only solution of (4.6) or (4.7) that can be the expectation of some T, so that no nontrivial solution exists, and T^* must be admissible.

5. Examples

EXAMPLE 1. *Uniform* $(0, \theta)$. For $Y_1, \cdots, Y_n$ independent, each uniformly distributed on $(0, \theta)$, the sufficient statistic $X = \max Y_i$ has the complete family of possible densities

$$\frac{n}{\theta^n} x^{n-1}, \qquad 0 \leqq x \leqq \theta, \theta > 0. \tag{5.1}$$

For estimating θ with quadratic loss, the uniformly best constant multiple of X is

$$T^* = \frac{n+2}{n+1} X, \tag{5.2}$$

which has expectation

$$m^*(\theta) = ET^* = \frac{n(n+2)}{(n+1)^2} \theta \tag{5.3}$$

and risk

$$E(T^* - \theta)^2 = \frac{\theta^2}{(n+1)^2}. \tag{5.4}$$

This estimator T^* was proved admissible by Karlin [8]. For an estimator $T(X)$ with expectation $m(\theta)$ we have $T(X) = m(X) + (X/n)m'(X)$. Because of this simple inversion, T^* provides an easy illustration of the advantage of working with a relaxation instead of with the inadmissibility inequality itself (see Note 3, Section 4).

We begin by trying to work directly with the inadmissibility inequality (4.1) in the estimator T whose expectation is m:

$$E\{T(X) - m(\theta)\}^2 + \{m(\theta) - \theta\}^2 \leqq \frac{\theta^2}{(n+1)^2}. \tag{5.5}$$

Because of the simple inversion, this can easily be written as an inequality in m:

$$E\{m(X) + \frac{X}{n} m'(X) - m(\theta)\}^2 + \{m(\theta) - \theta\}^2 \leqq \frac{\theta^2}{(n+1)^2}. \tag{5.5'}$$

Multiplying out the square in the first term and integrating one of the resulting terms, $EXm(X)m'(X)$, by parts, this inequality simplifies to

$$\frac{1}{n\theta^n} \int_0^\theta [m'(x)]^2 x^{n+1}\, dx + \{m(\theta) - \theta\}^2 \leqq \frac{\theta^2}{(n+1)^2}. \tag{5.5''}$$

We want to prove that $m = m^*$ is the unique solution of this inequality, which is hard to work with because of the integral in the first term. An application of Jensen's inequality to that integral gives a relaxed inequality

$$\frac{n+2}{n} \left\{m(\theta) - \frac{n+1}{\theta^{n+1}} \int_0^\theta m(x) x^n\, dx\right\}^2 + \{m(\theta) - \theta\}^2 \leqq \frac{\theta^2}{(n+1)^2}. \tag{5.6}$$

A check shows that this is exactly the same relaxed inequality (4.3) as results from applying to Var T the Cramér-Rao type inequality using T^*. This inequality (5.6) still contains an integral, but written in terms of $r(\theta) = (1/\theta^n)\int_0^\theta m(x)x^n\,dx$, for which $m(\theta) = (n/\theta)r(\theta) + r'(\theta)$, it does not:

$$(5.6') \qquad \frac{n+2}{n}\left\{r'(\theta) - \frac{r(\theta)}{\theta}\right\}^2 + \left\{\frac{n}{\theta}r(\theta) + r'(\theta) - \theta\right\}^2 \leqq \frac{\theta^2}{(n+1)^2}.$$

We want to show that $r(\theta) = \theta^2 n/(n+1)^2$, corresponding to $m = m^*$, is the unique solution of (5.6′). For convenience, we now write $\theta^2 s(\theta) = r(\theta) - \theta^2 n/(n+1)^2$ and use a typical Hodges-Lehmann argument to show that $s(\theta) \equiv 0$ is the unique solution of (5.6′) written in terms of $s(\theta)$ and slightly rearranged

$$(5.6'') \qquad \frac{n+2}{n}\{\theta s'(\theta) + (n+1)s(\theta)\}^2 + \left\{\theta s'(\theta) + \frac{1}{n+1}\right\}^2 \leqq \frac{1}{(n+1)^2}.$$

This inequality shows that $\theta s'(\theta) \leqq 0$; and that $\theta s'(\theta)$ and $\theta s'(\theta) + (n+1)s(\theta)$ are both bounded, making $s(\theta)$ also bounded. Now $\theta s'(\theta)$ cannot be bounded away from zero as $\theta \to 0$ or as $\theta \to \infty$, because either of these would make $s(\theta)$ unbounded. So there must be a sequence of θ values tending to zero and a sequence of θ values tending to infinity along which $\theta s'(\theta) \to 0$, and (5.6″) shows that $s(\theta) \to 0$ along these sequences. This together with $s'(\theta) \leqq 0$ implies $s(\theta) \equiv 0$ and proves T^* admissible.

Admissibility of T^* can also be proved using only the further relaxed inequality (4.4):

$$(5.7) \qquad \{\theta s'(\theta) + (n+2)s(\theta)\}^2 + \frac{2}{n+1}\theta s'(\theta) \leqq 0.$$

This inequality requires $\theta s'(\theta) \leqq 0$, and at this point we could use the inadmissibility inequality (5.5) with its first term omitted to see that $s(\theta)$ must be bounded, and admissibility would follow as above. But a proof can be given as below using only the inequality (5.7). (Note that $s(\theta) = a/\theta^{n+2}$ is a nontrivial solution of (5.7), but this solution is ruled out as not corresponding to any m, because $s(\theta) = (1/\theta^{n+2})\int_0^\theta m(x)x^n\,dx - n(n+1)^2$ which requires $\theta^{n+2}s(\theta) \to 0$ as $\theta \to 0$.)

First show that $\theta s'(\theta)$ cannot be bounded away from zero as $\theta \to \infty$; for if it were we would have $s(\theta) \to -\infty$, and so would have for all θ sufficiently large $s(\theta) < 0$ and therefore $\{\theta s'(\theta)\}^2 + \{2/(n+1)\}\theta s'(\theta) \leqq 0$, which requires $\theta s'(\theta)$ bounded as $\theta \to \infty$, which together with $s(\theta) \to -\infty$ as $\theta \to \infty$ would violate (5.7).

Thus there must be a sequence of θ values tending to infinity along which $\theta s'(\theta) \to 0$, and (5.7) shows that $s(\theta) \to 0$ along this sequence. From $s'(\theta) \leqq 0$ we can now conclude $s(\theta) \to 0$ as $\theta \to \infty$, and $s(\theta) \geqq 0$ for all θ.

Next show that $\theta s'(\theta)$ cannot be bounded away from zero as $\theta \to 0$; for if it were we would have $s(\theta) \to \infty$ as $\theta \to 0$. And in terms of $t(\theta) = \theta^{n+2}s(\theta)$, the inequality (5.7) is

$$(5.7') \qquad \left\{\frac{t'(\theta)}{\theta^{n+1}}\right\}^2 + \frac{2}{n+1}\left\{\frac{t'(\theta)}{\theta^{n+1}}\right\} - 2\frac{n+2}{n+1}\frac{t(\theta)}{\theta^{n+2}} \leqq 0.$$

We saw above that $t(\theta) \to 0$ as $\theta \to 0$, so $s(\theta) = t(\theta)/\theta^{n+2} \to \infty$ would imply (by L'Hospital's rule) $t'(\theta)/\theta^{n+1} \to \infty$, and therefore for all θ sufficiently small $t'(\theta)/\theta^{n+1} > 0$ and so

$$(5.8) \qquad \left\{\frac{t'(\theta)}{\theta^{n+1}}\right\}^2 \leqq 2\frac{n+2}{n+1}\frac{t(\theta)}{\theta^{n+1}}.$$

This gives, for all x sufficiently small,

$$(5.9) \qquad \frac{t'(x)}{2(t(x))^{1/2}} \leqq \text{const } x^{n/2}.$$

Integrating both sides from 0 to θ now gives, for all θ sufficiently small,

$$(5.10) \qquad (t(\theta))^{1/2} \leqq \text{const } \theta^{n/2+1}$$

that is,

$$(5.11) \qquad t(\theta) \leqq \text{const } \theta^{n+2}.$$

which would contradict the fact that $t(\theta)/\theta^{n+2} \to \infty$ as $\theta \to 0$.

Thus there must be a sequence of θ values $\to 0$ along which $\theta s'(\theta) \to 0$, and (5.7) shows that $s(\theta) \to 0$ along this sequence. We can now conclude that $s(\theta) \equiv 0$, and the admissibility of T^* is proved.

EXAMPLE 2. *Exponential family.* Let X be a real valued random variable with the following exponential family (which is complete) of possible densities relative to a fixed σ-finite measure μ:

$$(5.12) \qquad \beta(\omega) \cdot e^{\omega x}, \qquad -\infty < x < \infty, \qquad a < \omega < b.$$

Here the interval (a, b) must be contained in the interval Ω of ω values for which $\int_{-\infty}^{\infty} e^{\omega x}\, d\mu(x) < \infty$. Consider $T^* = (X + k\lambda)/(1 + \lambda)$, where $\lambda > -1$ and k are constants, as an estimator of $EX = -\beta'(\omega)/\beta(\omega)$ with quadratic loss.

Hodges and Lehmann [7] proved admissibility for $(a, b) = \Omega$ of particular estimators T^* for the specific exponential families binomial, Poisson, normal $(\omega, 1)$, and gamma with scale parameter ω, given by specific choices of μ. Their proofs use the relaxed inequality (4.3); different proofs can be given using only (4.4).

Girshick and Savage [6] proved $T^* = X$ admissible for $(a, b) = \Omega = (-\infty, \infty)$. Their proof uses (4.3) which they further relax to (4.4), so can be carried out using only (4.4).

Karlin [8], using the limiting Bayes method, gave sufficient conditions for T^*, with $k = 0$, to be admissible for $(a, b) = \Omega$.

Ping [10] gave a Hodges-Lehmann type proof of the following extension of Karlin's theorem: T^* is admissible provided

$$\lim_{\omega \to a} \int_{\omega}^{\omega_0} [\beta(\xi)]^{-\lambda} e^{-k\lambda\xi} \, d\xi = \infty = \lim_{\omega \to b} \int_{\omega_0}^{\omega} [\beta(\lambda)]^{-\lambda} e^{-k\lambda\xi} \, d\xi. \tag{5.13}$$

Ping writes down the relaxed inequality (his formula 1.4) but then replaces this by the further relaxed inequality (his formula 1.5), so that his proof actually uses only (4.4).

EXAMPLE 3. *Gamma with scale parameter.* For $X_1, \cdots, X_n$ independent, each with density $\lambda e^{-\lambda x}$, $x \geqq 0$, $\lambda > 0$, the sufficient statistic $X = \Sigma X_i$ has exponential family (complete) of possible densities

$$\frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}, \qquad x \geqq 0, \qquad \lambda > 0. \tag{5.14}$$

For this family with $n > 2$ (not necessarily integer valued), for estimating $\lambda = 1/EX$ with quadratic loss, the uniformly best constant multiple of X is

$$T^* = \frac{n-2}{X}, \tag{5.15}$$

which has expectation

$$m^*(\lambda) = ET^* = \frac{n-2}{n-1} \lambda \tag{5.16}$$

and risk

$$E(T^* - \lambda)^2 = \frac{\lambda^2}{n-1}. \tag{5.17}$$

This extimator T^* was proved admissible by Ghosh and Singh [5] using a limiting Bayes argument. Here it provides an example, for an exponential family, of using the Hodges-Lehmann method to prove admissibility of an estimator of something other than EX. (Example 2 is restricted to estimation of EX.)

For an estimator $T(X)$ with expectation $m(\lambda)$ and finite variance, Schwarz's inequality for T, $1/X$ proves the existence of

$$r(\lambda) = E\frac{T}{X} = \frac{\lambda^n}{\Gamma(n)} \int_0^{\infty} \frac{t(x)}{x} x^{n-1} e^{-\lambda x} \, dx, \tag{5.18}$$

and we have

$$m(\lambda) = \frac{n}{\lambda} r(\lambda) - r'(\lambda). \tag{5.19}$$

Written in terms of

$$s(\lambda) = \frac{r(\lambda)}{\lambda^2} - \frac{1}{n-1}, \tag{5.20}$$

the relaxed inequality (4.3) is

$$(5.21)\qquad (n-1)\{\lambda s'(\lambda)\}^2 + 2\{\lambda s'(\lambda)\} + (n-1)(n-2)\{s(\lambda)\}^2 \leqq 0,$$

and the further relaxed inequality (4.4) is

$$(5.22)\qquad \{(n-2)s(\lambda) - \lambda s'(\lambda)\}^2 + 2\lambda s'(\lambda) \leqq 0.$$

For each of these inequalities, it is easy to prove that $s(\lambda) \equiv 0$, corresponding to $m(\lambda) = m^*(\lambda)$, is the only solution that corresponds to some $m(\lambda) = ET$, so either inequality can be used to prove T^* admissible. The proofs are the same as those given for the corresponding inequalities in Example 1, except that the roles of $\lambda \to 0$ and $\lambda \to \infty$ are reversed.

EXAMPLE 4. *Normal* $(\theta, 1)$. For X normal $(\theta, 1)$, $-\infty < \theta < \infty$, the estimator $T^* \equiv 0$ is obviously admissible for estimating θ with quadratic loss, but the Hodges-Lehmann method fails to prove this. Because T^* is a constant, the relaxed inequality (4.3) is not available; the further relaxed inequality (4.4) is

$$(5.23)\qquad [m(\theta)]^2 - 2\theta m(\theta) \leqq 0.$$

This inequality has the nontrivial solution $m(\theta) = \theta$, corresponding to $T(X) = X$. The same happens for any constant T^* in the exponential family problem of Example 2.

REFERENCES

[1] E. W. BARANKIN, "Locally best unbiased estimates," *Ann. Math. Statist.*, Vol. 20 (1949), pp. 477–501.

[2] A. BHATTACHARYYA, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā*, Vol. 8 (1946–1948), pp. 1–14, 201–218, 315–328.

[3] D. G. CHAPMAN and H. ROBBINS, "Minimum variance estimation without regularity assumptions," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 581–586.

[4] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[5] J. K. GHOSH and R. SINGH, "Estimation of the reciprocal of scale parameter of a Gamma density," *Ann. Inst. Statist. Math.*, Vol. 22 (1970), pp. 51–55.

[6] M. A. GIRSHICK and L. J. SAVAGE, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 53–73.

[7] J. L. HODGES and E. L. LEHMANN, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 13–22.

[8] S. KARLIN, "Admissibility for estimation with quadratic loss,' *Ann. Math. Statist.*, Vol. 29 (1958), pp. 406–436.

[9] J. KIEFER, "On minimum variance estimates," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 627–629.

[10] C. PING, "Minimax estimates of parameters of distributions belonging to the exponential family," *Chinese Math.*, Vol. 5 (1964), pp. 277–299.