The Annals of Statistics
1977, Vol. 5, No. 1, 197–205

ON THE CRAMÉR–RAO INEQUALITY

BY VÁCLAV FABIAN[1] AND JAMES HANNAN[2]

Michigan State University

Sufficient conditions for the Cramér–Rao inequality are formulated which do not impose any requirement on the estimator and which are independent of the choice of the densities. A hereditary property is proved for these conditions and the attainability of the lower bound is studied.

1. Introduction. We are concerned with the original form of the Cramér–Rao inequality, a slight extension of Cramér's (1946, Section 32) formulation (see 15.6 of Savage (1954) for some history and references). This is, of course, only one among many possible inequalities, see Section 3.2 of Blyth and Roberts (1972) for a survey and for references. The Cramér–Rao inequality deserves some special interest if only because of its usefulness in asymptotic efficiency considerations.

The usual sufficient conditions for the inequality put restrictions on the estimator itself which is obviously highly undesirable. Barankin (1949), Corollary 5.1 shows that no requirement on the estimator is necessary if the derivative of the relative density exists in the sense of the strong convergence. In fact, the strong convergence can be replaced by weak convergence and one obtains a condition on a statistical problem under which the Cramér–Rao inequality holds for all estimators.

This result (Theorem 2.3 and the multidimensional Theorem 2.9) is very easy to prove but seems to us a considerable improvement over the form in which the Cramér–Rao inequality appears in current literature.

Our formulation of the Cramér–Rao inequality leads then to additional questions, which this paper solves. Section 3 shows that the conditions in Theorems 2.3 and 2.9 are hereditary in the sense that if they are satisfied for a problem in which 1 observation is taken they are also satisfied if n independent observations are taken. In Section 4 we show that if under our conditions the variance of an estimator is equal to its Cramér–Rao bound on an interval in R then the probabilities form an exponential family. This is an analoguous result to that which holds under the standard regularity conditions (see Wijsman (1973)).

The following assumptions and notation will be used throughout the paper.

We assume that $(X, \mathscr{A}, \mu)$ is a measure space, Θ a nonempty set, P_θ a probability on $\mathscr{A}$ with a density p_θ with respect to μ, for every θ in Θ. The symbols E_θ and Var_θ stand for the expectation and variance under P_θ. By $L_2(\theta)$ or by

Received September 1975.

[1] Research supported by NSF Grants no. GP-31123X2 and no. MPS 75-08256.

[2] Research supported by NSF Grants no. GP-33677X2 and no. MPS 75-07467.

AMS 1970 *subject classification.* Primary 62F10.

Key words and phrases. Estimation, Cramér–Rao inequality, weak derivative, attainment.

$L_2(P_\theta)$ we denote the set of all finite $\mathscr{A}$-measurable functions with a finite variance under P_θ. The pseudonorm $\|\ \|_\theta$ is called the norm for short. The inner product in $L_2(\theta)$ is denoted by $(\ ,\ )_\theta$. The weak convergence $w\lim g_n = g$ in $L_2(\theta)$ means that g_n, g are in $L_2(\theta)$ and $(g_n, h)_\theta \to (g, h)_\theta$ for every h in $L_2(\theta)$.

If $M \subset R$ and g is a function on M then $\lim_{y\to x} g(y)$ has an obvious meaning for an accumulation point x of M. The function has a derivative $\dot{g}(x)$ at x if $\dot{g}(x) = \lim_{y\to x}(y-x)^{-1}[g(y) - g(x)]$ exists and is finite; and a derivative $\dot{g}$ if it has a derivative at each point x in M. If γ is a function on M into $L_2(\theta)$ and if $w\lim_{y\to x}(y-x)^{-1}[\gamma(y)-\gamma(x)] = q$ in $L_2(\theta)$, we say that γ has at x a weak derivative in $L_2(\theta)$ equal to q and we shall write this as $\gamma'(x) = q$. Since weak limits in $L_2(\theta)$ are not unique, $\gamma'(x) = q_1$, $\gamma'(x) = q_2$ imply $q_1 = q_2$ a.e. (P_θ) only, not $q_1 = q_2$.

The indicator function of a set A will be denoted by χ_A.

2. The Cramér-Rao inequality.

2.1. Remark. In this section, θ will be a given element in Θ and we shall shorten the notation $(\ ,\ )_\theta$, $\|\ \|_\theta$ by omitting the subscript.

Write $r_\delta = p_\delta/p_\theta$ interpreting 0/0 as 0. If $P_\delta \ll P_\theta$, r_δ is a density of P_δ with respect to P_θ or, as we may say, a relative density of P_δ. By r_θ' we mean the weak in $L_2(\theta)$ derivative of $\delta \rightsquigarrow r_\delta$ at θ.

Suppose $t \in L_2(\theta)$, $q \in L_2(\theta)$, $\|q\| > 0$. We obtain

$$\|t - E_\theta t\| \geqq \frac{|(q, t - E_\theta t)|}{\|q\|} \tag{1}$$

because the right-hand side is the norm of a projection of $t - E_\theta t$ on the space generated by q, in $L_2(\theta)$. Since $\mathrm{Var}_\theta(t) = \|t - E_\theta t\|^2$, (1) becomes the usual Cramér–Rao inequality if $\|q\|^2$ is the Fisher information at θ (but we shall postpone this consideration until Remark 2.5) and if $(q, 1) = 0$, $(q, t) = \dot{E}_\theta t$, where $\dot{E}_\theta t$ is the derivative of $\delta \rightsquigarrow E_\delta t$ at θ. This is a condition involving t. But if required of all t, it can be formulated as

$$(q, t) = \dot{E}_\theta t \qquad \text{for all} \quad t \in L_2(\theta) \tag{2}$$

(since $1 \in L_2(\theta)$ and $\dot{E}_\theta 1 = 0$), and has an easy interpretation if (2.2.i) below is satisfied.

2.2. Lemma. *Suppose* $\theta \in \Theta \subset R$, $q \in L_2(\theta)$ *and*

$$P_\delta \ll P_\theta \qquad \textit{for every} \quad \delta \quad \textit{in} \quad \Theta\,. \tag{i}$$

Then

$$q = r_\theta' \tag{ii}$$

if and only if (2.1.2) *holds together with*

$$\{r_\delta;\ \delta \in \Theta\} \subset L_2(\theta)\,. \tag{iii}$$

Proof. Straightforward since under (i), $E_\delta t = (r_\delta, t)$ if the latter has sense.

2.3. THEOREM (Cramér–Rao inequality). *Suppose* $\theta \in \Theta \subset R$ *and*

$$\text{(i)} \qquad \textit{for every } \delta \text{ in } \Theta, \quad P_\delta \ll P_\theta ,$$

$$\text{(ii)} \qquad r_\theta' \ \textit{exists} ,$$

$$\text{(iii)} \qquad \operatorname{Var}_\theta r_\theta' > 0 .$$

Then for every t *in* $L_2(P_\theta)$ *the derivative* $\dot{E}_\theta t$ *exists and*

$$(1) \qquad \operatorname{Var}_\theta t \geqq \frac{(\dot{E}_\theta t)^2}{\operatorname{Var}_\theta r_\theta'} .$$

PROOF. From (2.1.1) and (2.1.2) by Lemma 2.2.

2.4. REMARK. We have not assumed that Θ is an interval. Notice that if the assumptions of the preceding theorem are not satisfied with the original Θ, they may be satisfied for a subset of Θ or perhaps for a subset of Θ mapped into another set. Then (2.3.1) holds, but the changes may affect the right-hand side in (2.3.1).

2.5. REMARK. Suppose $\dot{p}_\theta$ exists almost everywhere (P_θ) and r_θ' exists. It is easy to see that then $\dot{p}_\theta/p_\theta = r_\theta'$ a.e. (P_θ). On the other hand neither the existence of $\dot{p}_\theta$ nor the existence of r_θ' implies the other. That $\dot{p}_\theta$ may exist but not r_θ' is easy to see. In the following example, r_θ' exists but not $\dot{p}_\theta$.

Take $X = [0, 2\pi]$, λ the Lebesgue measure on X, $P_0 = (1/2\pi)\lambda$, $\theta = 0$, $\Theta = \{0, 1, \frac{1}{2}, \frac{1}{3}, \cdots\}$, $g(x) = x - \pi$, $h_n(x) = \sin nx$,

$$r_{1/n} = 1 + \frac{1}{5n}[h_n + g] .$$

Then

$$n(r_{1/n} - r_0) = \tfrac{1}{5}[h_n + g] .$$

As $n \to \infty$, this does not converge even in measure on $[0, 1]$. But the weak limit in $L_2(0)$ exists and is equal to $g/5$.

The preceding phenomenon does not occur in the dominated location parameter situation: if r_θ' exists then there is a choice of the density such that $\dot{p}_\theta$ exists a.e., and then $r_\theta' = \dot{p}_\theta/p_\theta$. In more detail, assume $\Theta = R$, $\theta = 0$, μ is the Lebesgue measure on Θ, $p_\delta(x) = f(x + \delta)$ with an f for all $\delta \in \Theta$, $\mu \ll P_0$. Assume that r_0' exists. Then

$$\dot{E}_0 h = \int r_0' h f \, d\mu \qquad \text{for every} \quad h \in L_2(P_0) .$$

Take $h = \chi_{(-\infty, a)}$, set $F(a) = \int_{-\infty}^{a} f \, d\mu$. Since $F(a + \delta) = E_\delta h$, F has a derivative, say f_0, and $f_0(a) = \int_{-\infty}^{a} r_0' f \, d\mu$ for all a. We obtain that $f = f_0$ a.e., f_0 is absolutely continuous and has a derivative a.e. (This also follows from V.6.1 in Stein (1970); we are indebted to our colleague Clifford Weyl for the reference.)

2.6. REMARK. Bounds other than the Cramér–Rao bound can be and were obtained from (2.1.1). This inequality does not have much use unless the right-hand side has some independence of t. For example we may require (as Blyth and Roberts (1972) do) that

$$(1)\qquad (q, t_1 - E_\theta t_1) = (q, t_2 - E_\theta t_2) \quad \text{if} \quad t_i \in L_2(P_\theta)\,,$$
$$E_\delta t_1 = E_\delta t_2 \quad \text{for every} \quad \delta \in \Theta\,.$$

Suppose $P_\delta \ll P_\theta$ and $r_\delta \in L_2(P_\theta)$ for every $\delta \in \Theta$. Notice that q satisfies (1) if and only if q belongs to the orthogonal complement $M^\perp$ of $M = \{\tau;\ \tau \in L_2(\theta),\ E_\delta \tau = 0$ for all $\delta\}$. If N is the closed linear subspace spanned by $\{r_\delta;\ \delta \in \Theta\}$ then $M = N^\perp$ and the set Q of all q satisfying (1) is given by $N^{\perp\perp}$, i.e.,

$$(2)\qquad Q = N\,.$$

2.7. Remark. The absolute continuity assumption (2.3.i) is rather strong (it is sometimes incorrectly omitted as in Kiefer (1952), relation (1), Rao (1965), Assertion (5a.2.(v)) and Blyth (1974), Corollary to Theorem 1). It can be somewhat relaxed. For example if, in Theorem 2.3, the assertion is restricted to all t in $T = \{h;\ h \in L_2(\theta),\ \sup_{\delta \in \Theta} E_\delta h^2 < +\infty\}$ then (2.3.i) can be weakened to

$$(\text{i})\qquad (\delta - \theta)^{-2} P_\delta\{p_\theta = 0\} \to 0 \quad \text{as} \quad \delta \to \theta\,.$$

Indeed, if $h \in T$, then

$$\left(\frac{1}{\delta - \theta} E_\delta h \chi_{\{p_\theta = 0\}}\right)^2 \leqq (E_\delta h^2) \frac{1}{(\delta - \theta)^2} P_\delta(p_\theta = 0) \to 0$$

and, with $q = r_\theta'$,

$$(q, h) = \lim_{\delta \to \theta} \frac{1}{\delta - \theta} [E_\delta h - E_\theta h - E_\delta h \chi_{\{p_\theta = 0\}}] = \dot{E}_\theta h\,.$$

Thus (2.1.2) holds with t restricted to T and (2.3.1) follows from (2.1.1).

2.8. Notation. In the next two theorems we shall consider the case of $\Theta \subset R^k$. We shall then interpret $\dot{E}_\theta t$ as the vector of the first partial derivatives of $\delta \rightsquigarrow E_\delta t$ at θ. We shall also denote by $\Theta^{(i)}$ the intersection of Θ with $\{\theta + \varepsilon e_i;\ \varepsilon \in R\}$ where e_i denotes the ith unit vector in R^k. If we denote by $r^{(i)}$ the restriction of $\delta \rightsquigarrow r_\delta$ to $\Theta^{(i)}$, the weak derivative of $r^{(i)}$ at θ, denoted by $(r_\theta^{(i)})'$ has the meaning of the weak partial derivative, at θ.

Transposition of matrices will be denoted by superscript T.

2.9. Theorem. *Suppose $\theta \in \Theta \subset R^k$ for a positive integer k. Suppose*

$$(\text{i})\qquad P_\delta \ll P_\theta \quad \textit{for every} \quad \delta \in \textstyle\bigcup_{i=1}^k \Theta^{(i)}\,,$$
$$(\text{ii})\qquad (r_\theta^{(i)})' \quad \textit{exists for every} \quad i\,.$$

Then, with Φ the matrix with the (i, j) element equal to $E_\theta((r_\theta^{(i)})'(r_\theta^{(j)})')$, the equation

$$(1)\qquad \Phi a = \dot{E}_\theta t$$

has at least one solution a and

$$(2)\qquad \operatorname{Var}_\theta t \geqq (\dot{E}_\theta t)^T a$$

with the right-hand side independent of the choice of the solution a.

PROOF. Applying Lemma 2.2 to $\Theta^{(i)}$ we obtain that with $q_i = (r_\theta^{(i)})'$

$$(q_i, t - E_\theta t) = (\dot{E}_\theta t)_i \,. \tag{3}$$

Project $t - E_\theta t$ on the linear subspace spanned by $\{q_1, \cdots, q_k\}$. A projection exists and is of the form $a^T\langle q_1, \cdots, q_k\rangle$ with an a satisfying the equation (1) and the squared norm of the projection is equal to the right-hand side in (2).

2.10. REMARK. The Cramér–Rao inequality for a vector estimate $t = \langle t_1, t_2, \cdots, t_r\rangle$ with $t_i \in L_2(\theta)$, is a direct consequence of the inequality for one-dimensional estimates, although it is often treated as a separate problem. Assume that the conditions of Theorem 2.9 are satisfied (they generalize conditions of Theorem 2.3); for simplicity assume that Φ is nonsingular so that (2.9.2) can be written with the right-hand side equal to $(\dot{E}_\theta t)^T\Phi^{-1}(\dot{E}_\theta t)$. Let V be the covariance matrix of $\langle t_1, \cdots, t_r\rangle$ under P_θ, and Δ the matrix with the ith column equal to $\dot{E}_\theta t_i$. Let $u \in R^r$ and form $\tau = u^T t$. Then $\mathrm{Var}_\theta\, \tau = u^T V u$, $\dot{E}_\theta \tau = \Delta u$ and Theorem 2.9, applied to τ, gives

$$u^T V u \geqq u^T \Delta^T \Phi^{-1} \Delta u$$

for every $u \in R^k$. This is the usual result, sometimes equivalently formulated by stating that

$$V - \Delta^T\Phi^{-1}\Delta$$

is positive semidefinite.

3. Heredity.

3.1. DEFINITION. Call the family $\{P_\delta;\ \delta \in \Theta\}$ *regular at* θ if it satisfies the conditions of Theorem 2.3.

3.2. NOTATION. We shall consider n families $\{P_{\delta,i};\ \delta \in \Theta\}$ $(i = 1, 2, \cdots, n)$ instead of one. The corresponding symbols will be used with a subscript, e.g., X_i, μ_i, $P_{\theta,i}$, etc. In addition we shall consider the family $\{P_\delta;\ \delta \in \Theta\}$ where P_δ is the product measure $P_{\delta,1} \times \cdots \times P_{\delta,n}$.

3.3. THEOREM. *If* $\Theta \subset R$, $\theta \in \Theta$ *and if each* $\{P_{\delta,i};\ \delta \in \Theta\}$ *is regular at* θ *then also* $\{P_\delta;\ \delta \in \Theta\}$ *is regular at* θ; *in addition*

$$r_\theta'(x_1, x_2, \cdots, x_n) = \textstyle\sum_1^n r'_{\theta,i}(x_i) \,. \tag{1}$$

PROOF. Condition (2.3.i) is easy to verify, and Condition (2.3.iii) will follow from (2.3.ii) and (1). To prove (1) we can assume that $n = 2$ and then complete the argument by induction.

If g is a function on X_1, denote by $\tilde{g}$ the function $\langle x_1, x_2\rangle \rightsquigarrow g(x_1)$ and use the same convention for functions on X_2. Write $q_i = r'_{\theta,i}$ and

$$r_{\delta,i} = 1 + (\delta - \theta)q_{\delta,i} \tag{2}$$

so that

$$w \lim_{\delta\to\theta} q_{\delta,i} = q_i \qquad \text{in} \quad L_2(P_{\theta_i}) \,. \tag{3}$$

Let $h \in L_2(P_\theta)$. Then

$$(4) \qquad (\tilde{q}_{\delta,1}, h) = \int (q_{\delta,1}, h(\cdot, x_2))_1 \, dP_{\theta,2}(x_2)$$

by the Fubini theorem. For almost all x_2, $h(\cdot, x_2) \in L_2(P_{\theta,1})$ and the integrand in (4) converges to $(q_1, h(\cdot, x_2))_1$. But the integrands are bounded by $||h(\cdot, x_2)||_1 \sup_\delta ||q_{\delta,1}||_1$ which is $P_{\theta,2}$ integrable since weak convergence implies boundedness of the norms. By the dominated convergence theorem, (4) converges to $\int (q_1, h(\cdot, x_2)) \, dP_{\theta,2}(x_2) = (\tilde{q}_1, h)$. Thus, extending the result to $i = 2$ by symmetry,

$$(5) \qquad w \lim_{\delta \to \theta} \tilde{q}_{\delta,i} = \tilde{q}_i \quad \text{in} \quad L_2(P_\theta) .$$

We have $r_\delta = \tilde{r}_{\delta,1} \tilde{r}_{\delta,2}$ and, from (2),

$$(6) \qquad (\delta - \theta)^{-1}[r_\delta - 1] = \tilde{q}_{\delta,1} + \tilde{q}_{\delta,2} + (\delta - \theta)\tilde{q}_{\delta,1}\tilde{q}_{\delta,2} .$$

The last term has the norm converging to 0. This and (5) imply (1).

4. The attainment of the lower bound.

4.1. Notation and remarks. The set Θ will be a nondegenerate interval in R, the Lebesgue integral over Θ will be denoted by $\int$.

When all P_θ are mutually absolutely continuous, then a set A is P_θ null for some θ if and only if for all θ and we shall say that A is null. Similarly we shall use the term a.e., and we shall have $\mathrm{Var}_\theta \, t > 0$ for some if and only if for all θ.

By Et we shall mean the function $\theta \rightsquigarrow E_\theta t$; similarly for Var, P. In Section 2 we worked with the weak derivative at θ of p_δ/p_θ; denote it by $q(\theta)$ here and denote by q the corresponding function $\theta \rightsquigarrow q(\theta)$. Also write $\mathrm{Var}\, q$ for $\theta \rightsquigarrow \mathrm{Var}_\theta \, q(\theta)$. Again, properly speaking, $q(\theta)$ is an equivalence class and if we write $h = q(\theta)$ for a function h, we mean that h is in the class.

We shall say that an estimator u attains the CR bound at θ if θ is in Θ and (2.3.1) holds with an equality sign; we shall say that u attains the CR bound if it does so at every θ in Θ.

4.2. Theorem. *Suppose Θ is a nondegenerate interval in R, $\delta \in \Theta$, t a random variable with $\mathrm{Var}_\delta \, t > 0$ and $\mathrm{Var}_\theta \, t < +\infty$ for every θ in Θ. Then*

(i) *$\{P_\theta; \theta \in \Theta\}$ is regular at every θ in Θ and t attains the Cramér–Rao bound*, i.e.,

$$(1) \qquad \mathrm{Var}\, t = (\dot{E}t)^2 / \mathrm{Var}\, q ,$$

if and only if there is a ψ such that

(ii) *ψ is a strictly monotone differentiable function on Θ and for every θ,*

$$(2) \qquad C(\theta) e^{\psi(\theta) t}$$

is a density of P_θ with respect to P_δ for a suitable number $C(\theta)$.

In addition: If (ii) *holds for a ψ then*

$$(3) \qquad \psi(\theta) = \int_\delta^\theta \alpha \quad \text{for all} \quad \theta \quad \text{in} \quad \Theta , \quad \dot{\psi} = \alpha , \quad q = \alpha(t - Et)$$

with $\alpha = \dot{E}t / \mathrm{Var}\, t$ and, if u is an estimator, then the following three conditions are *equivalent*: (a) *u attains the* CR *bound at a* θ *in* Θ, (b) *u attains the* CR *bound*, (c) *there are numbers* c_1, c_2 *such that* $u = c_1 t + c_2$ a.e.

4.3. REMARK. The condition $\mathrm{Var}_\theta\, t < +\infty$ for all θ is implied by Condition (4.2.i). The inequality $\mathrm{Var}_\theta\, t < +\infty$ follows from Condition (4.2.ii) for every interior point θ of Θ. Thus, e.g., the condition can be omitted from the theorem if Θ is open. The constant $C(\theta)$ in (4.2.2) is, of course, the reciprocal of $E_\delta e^{\psi(\theta)t}$.

PROOF. Suppose (i) holds. Relation (1) and $\mathrm{Var}_\theta\, t > 0$ imply that $q(\theta)$ equals its projection on $t - E_\theta t$ and thus

$$(4) \qquad q = \alpha(t - Et) \qquad \text{with} \quad \alpha = \dot{E}t/\mathrm{Var}\, t\,.$$

We shall prove that (ii) holds with $\psi(\theta) = \int_\delta^\theta \alpha$.

If $\theta_n \to \theta$ in Θ then the existence of $q(\theta)$ implies the vague convergence of the probability distribution $P_{\theta_n}^t$ to $P_\theta{}^t$ and

$$\liminf \mathrm{Var}_{\theta_n} t = \liminf \tfrac{1}{2} \iint (x - y)^2\, d[P_{\theta_n}^t \times P_{\theta_n}^t](x, y) \geqq \mathrm{Var}_\theta\, t\,.$$

We have shown that $\mathrm{Var}\, t$ is lower semicontinuous.

The intermediate value property of derivatives (cf. e.g., Theorem V.5 in Graves, 1946) applied to $\dot{E}t$ yields that it is always positive or always negative, since by (1) it does not attain the value 0. Thus Et is strictly isotone and hence $\dot{E}t$ is integrable over finite subintervals of Θ; because of the lower semicontinuity and positiveness of $\mathrm{Var}\, t$, we obtain that α inherits the integrability property of $\dot{E}t$.

If B is in $\mathscr{A}$ then the properties of q, the Schwarz inequality and (4) imply

$$(5) \qquad \dot{P}B = E\chi_B q\,, \qquad |\dot{P}B| \leqq |\alpha|[P(B)\,\mathrm{Var}\, t]^{\frac{1}{2}}\,.$$

If θ is in Θ, $\overline{\delta, \theta}$ denotes the closed interval with indicated endpoints, and if B is nonnull, then the continuity of PB and (5) imply that $\dot{\log} PB$ is integrable over $\overline{\delta, \theta}$ and

$$(6) \qquad P_\theta B/P_\delta B = \exp \textstyle\int_\delta^\theta \dot{\log} PB\,.$$

Set

$$(7) \qquad f = \exp \textstyle\int_\delta^\theta q\,.$$

Since α and αEt are integrable over $\overline{\delta, \theta}$, the integral in (7) exists and f is of form (2).

Suppose that on B, $|t - c| < d/2$ for two numbers c, d. We obtain from (5) and (4) that on B

$$|q - \dot{\log} PB| = \left| q - \frac{E\chi_B q}{PB} \right| \leqq d|\alpha|\,.$$

Using (6) and (7), we have

$$E_\delta \chi_B f = \frac{P_\theta B}{P_\delta B} E_\delta \chi_B \exp \textstyle\int_\delta^\theta (q - \dot{\log} PB)\,.$$

If ε denotes $d \int_\delta^\theta |\alpha|$, we obtain

$$e^{-\varepsilon} P_\theta B \leqq E_\delta \chi_B f \leqq e^\varepsilon P_\theta B . \tag{8}$$

If (8) holds for a sequence of disjoint sets B_i, it holds also for their union. Relation (8) holds also for null sets, and therefore for all sets, and for all $\varepsilon > 0$. This shows that f is a density of P_θ with respect to P_δ.

The strict monotonicity of ψ follows since $|\alpha| > 0$ everywhere on Θ. To complete the proof of (ii) it is enough to show that $\dot\psi = \alpha$.

For $s \in \psi[\Theta]$, e^{st} is one of the densities in (2), multiplied by a constant. Since Var t is finite,

$$E_\delta t^2 e^{st} < +\infty \qquad \text{for every} \quad s \quad \text{in} \quad \psi[\Theta] . \tag{9}$$

Since $|(\partial^{(i)}/\partial s^i) e^{st}| \leqq |t|^i [e^{s_0 t} + e^{s_1 t}]$ for every $s \in [s_0, s_1]$ and every $i = 1, 2$, if s_0, s_1 are in $\psi[\Theta]$, it follows from the dominated convergence theorem that the function g, defined as $s \in \psi[\Theta] \rightsquigarrow E_\delta e^{st}$, has the first two derivatives $\dot g$ and $\ddot g$ given by

$$\dot g(s) = E_\delta t e^{st} , \qquad \ddot g(s) = E_\delta t^2 e^{st} .$$

Hence $\gamma = \log g$ satisfies

$$\dot\gamma \circ \psi = Et , \qquad \ddot\gamma \circ \psi = \operatorname{Var} t . \tag{10}$$

Since ψ is continuous and strictly monotone we obtain, indicating a difference in an obvious way, that

$$h^{-1} \psi]_\theta^{\theta+h} = h^{-1} Et]_\theta^{\theta+h} \frac{\psi]_\theta^{\theta+h}}{\dot\gamma \circ \psi]_\theta^{\theta+h}} \to \dot E_\theta t / \operatorname{Var}_\theta t = \alpha(\theta) .$$

We have shown that (i) implies (ii) with ψ as in (3).

Suppose (ii) holds for a ψ. Notice that (10) holds since its proof used only the fact that the densities in (2) have finite variances. Since $\dot\psi$ exists by assumption, (10) implies the existence of $\dot E t$ and, again, $\dot\psi = \alpha$.

Notice that the density p_θ in (2) has $\log p_\theta$ equal to $\psi(\theta) t - \gamma(\psi(\theta))$. Thus, with $q_0 = \alpha(t - Et)$ and as $\omega \to \theta$, we obtain at every point in X that

$$(\omega - \theta)^{-1} \frac{p_\omega - p_\theta}{p_\theta} \to \dot{\log}\, p_\theta = \dot\psi(\theta)[t - (\dot\gamma \circ \psi)(\theta)] = q_0(\theta) . \tag{11}$$

We shall prove that the above convergence holds for the norms in $L_2(\theta)$ which will imply the strong convergence.

The $L_2(\theta)$ square norm of the left-hand side in (11) is $(\omega - \theta)^{-2}[E_\theta(p_\omega/p_\theta)^2 - 1]$,

$$E_\theta \left(\frac{p_\omega}{p_\theta}\right)^2 = \exp\{\gamma(\psi(\theta) + 2\Delta) - 2\gamma(\psi(\theta) + \Delta) + \gamma(\psi(\theta))\} \tag{12}$$

with $\Delta = \psi]_\theta^\omega$. Subtract 1 from (12) and multiply by $(\omega - \theta)^{-2}$, then take the limit to obtain

$$\dot\psi(\theta)^2 \ddot\gamma(\psi(\theta)) = ||q_0||_\theta^2 .$$

We have completed the proof of (i), but we have also proved (3) except possibly the first relation in (3). However we had proved that if (i) holds then α is integrable over $\overline{\delta, \theta}$ and thus even the first relation in (3) holds.

To complete the proof consider an estimator u. If (a) holds for a θ then, for a number c, $u - E_\theta u = c\alpha(\theta)(t - E_\theta t)$ a.e. and $u = c_1 t + c_2$ a.e. with $c_1 = c\alpha(\theta)$, $c_2 = E_\theta u - c\alpha(\theta)E_\theta t$. The implications (c) $\Rightarrow$ (b) $\Rightarrow$ (a) are obvious.

This completes the proof.

REFERENCES

BARANKIN, E. W. (1949). Locally best unbiased estimates. *Ann. Math. Statist.* **20** 477–501.

BLYTH, COLIN R. and ROBERTS, DONALD M. (1972). On inequalities of Cramér-Rao type and admissibility proofs. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 17–30, Univ. of California Press.

CRAMÉR, HARALD. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

GRAVES, LAWRENCE M. (1946). *The Theory of Functions of Real Variables.* McGraw-Hill, New York.

KIEFER, J. (1952). On minumum variance estimators. *Ann. Math. Statist.* **23** 627–629.

RAO, C. RADHAKRISHNA (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

SAVAGE, LEONARD J. (1954). *The Foundations of Statistics.* Wiley, New York.

STEIN, ELIAS M. (1970). *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press.

WIJSMAN, R. A. (1973). On the attainment of the Cramér-Rao lower bound. *Ann. Statist.* **1** 538–542.

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
WELLS HALL
EAST LANSING, MICHIGAN 48824