*The Annals of Statistics*
1973, Vol. 1, No. 3, 538–542

# ON THE ATTAINMENT OF THE CRAMÉR–RAO LOWER BOUND

BY R. A. WIJSMAN
*University of Illinois at Urbana-Champaign*

A rigorous proof is given of the often stated fact that if the variance of an unbiased estimator of a function of a real parameter attains the Cramér–Rao lower bound then the family of distributions must be a one-parameter exponential family.

The classical Cramér–Rao inequality [1], [3] and [6] states that under suitable regularity conditions a lower bound for the variance of an unbiased estimator $t(X)$ of a real-valued function $m(\theta)$ of a real parameter $\theta$ is given by the inequality

$$\text{Var}_\theta\, t(X) \geqq [m'(\theta)]^2/\text{Var}_\theta \frac{\partial}{\partial\theta} \log p_\theta(X) \tag{1}$$

in which $p_\theta$ is the density of the random variable $X$. It is usually stated that the lower bound can be attained only if the family of distributions of $X$ is one-parameter exponential [2], [3] and [7, page 187]. That this is to be expected can be seen by realizing that (1) is nothing else but a statement that the square of the correlation between $t(X)$ and $(\partial/\partial\theta)\log p_\theta(X)$ cannot exceed 1, and equality is attained if and only if $(\partial/\partial\theta)\log p_\theta(x) = a(\theta)t(x) + b(\theta)$ for some functions $a$ and $b$. Then by integration over $\theta$ the desired exponential form of $p_\theta$ is obtained. However, this heuristic approach conveniently ignores the fact that the above affine relation between $t(x)$ and $(\partial/\partial\theta)\log p_\theta(x)$ may fail to hold on a null set which may depend on $\theta$ and that a priori nothing can be assumed about the functions $a$ and $b$, not even measurability, let alone integrability. Since a rigorous proof does not seem to have appeared in the literature it may be appropriate to produce one here, the more so since the proof seems to be neither completely trivial nor standard.

The following assumptions will be made. The sample space is an arbitrary measure space $(\mathscr{X}, \mathscr{A}, \mu)$, with $\mu$ sigma-finite. The parameter space is the measure space $(\Theta, \mathscr{B}, \nu)$, with $\Theta$ a Borel subset of the real line, $\mathscr{B}$ the Borel subsets of $\Theta$ and $\nu$ Lebesgue measure. There is given a random variable $X$ with values in $\mathscr{X}$ and distribution $P_\theta(dx) = p_\theta(x)\mu(dx)$, $\theta \in \Theta$. For convenience differentiation with respect to $\theta$ will be denoted by $D$. Any integration with respect to $\mu$ will always be understood to be over the whole of $\mathscr{X}$. We shall make the following.

REGULARITY CONDITIONS.

(i) $\Theta$ is an open interval (possibly infinite or semi-infinite);

Received December 1971; revised June 1972.
*AMS* 1970 *subject classification*. Primary 62F10.
*Key words and phrases*. Cramér–Rao lower bound, variance of unbiased estimate, attainment of lower bound, exponential family.

(ii) $p_\theta(x) > 0$ for every $\theta \in \Theta$, $x \in \mathscr{X}$, $p_\theta(\cdot)$ is $\mathscr{A}$-measurable for every $\theta \in \Theta$, and $p_\theta(x)$ is a continuously differentiable function of $\theta$ for every $x \in \mathscr{X}$;

(iii) $0 < \mathrm{Var}_\theta\, D \log p_\theta(X) < \infty$ for every $\theta \in \Theta$;

(iv) $\int p_\theta(x)\mu(dx)$ can be differentiated under the integral sign with respect to $\theta$;

(v) $\int t(x)p_\theta(x)\mu(dx)$ is finite and can be differentiated under the integral sign with respect to $\theta$.

Slight variations on these conditions appear in the literature. In particular, differentiability of $p_\theta$ with respect to $\theta$ is often assumed to hold for all $x$ except for $x$ in a $\mu$-null set. If this null set were allowed to depend on $\theta$ then one gets into trouble with (iv) and (v). So usually it is assumed that the null set does not depend on $\theta$. We may then as well remove this sèt from $\mathscr{X}$, and implied in (ii) is that we have taken this liberty.

THEOREM. *Let $m$ be a real-valued function on $\Theta$, not identically constant; let $t(X)$ be an unbiased estimator of $m(\theta)$ and let Regularity Conditions* (i)—(v) *be satisfied. Then the inequality* (1) *is an equality for all $\theta \in \Theta$ if and only if there exists $K \in \mathscr{A}$ with $\mu K = 0$ such that for $x \in \mathscr{X} - K$, $\theta \in \Theta$,*

$$p_\theta(x) = c(\theta)h(x)e^{q(\theta)t(x)} \tag{2}$$

*in which $c(\theta)$ and $h(x)$ are* $> 0$, *$q$ is strictly monotonic, and both $c$ and $q$ are continuously differentiable.*

PROOF. The "if" part is of course well known and easily verifiable so that we shall proceed to prove the "only if" part. In the following we shall need the measurability of $p_\theta(x)$ and $D \log p_\theta(x)$ as functions on $(\mathscr{X} \times \Theta, \mathscr{A} \times \mathscr{B})$. This follows from the fact that these functions are $\mathscr{A}$-measurable for each fixed $\theta$ and continuous on $\Theta$ for each $x$ [5, Chapter IV, Theorem T 47] (I am indebted to L. L. Helms for this reference). Also note that $p_\theta > 0$ on $\mathscr{X}$ implies that for any $A \in \mathscr{A}$, $P_\theta A = 0$ for some $\theta$ implies $\mu A = 0$.

It is given that (1) is an equality for each $\theta \in \Theta$. This implies, for each $\theta$, that $t(X) - m(\theta)$ and $D \log p_\theta(X)$ are linearly dependent elements in the space $L_2(\mathscr{X}, \mathscr{A}, P_\theta)$; hence there exist constants $c_1(\theta)$ and $c_2(\theta)$ with $c_1^2 + c_2^2 > 0$, and a $P_\theta$-null (and therefore $\mu$-null) set $N_\theta$, such that

$$c_1(\theta)D \log p_\theta(x) + c_2(\theta)[t(x) - m(\theta)] = 0 \qquad \text{for} \quad x \notin N_\theta\,. \tag{3}$$

If $c_1(\theta) = 0$, then $c_2(\theta) \neq 0$ and (3) implies $t(x) = m(\theta)$ a.e. $\mu$ and therefore a.e. $P_{\theta_1}$, for any $\theta_1 \in \Theta$. Taking expectation with respect to $P_{\theta_1}$ we get then $m(\theta_1) = m(\theta)$, hence $m$ is constant on $\Theta$. Since this is excluded by hypothesis, we must have $c_1(\theta) \neq 0$. It follows that (3) can be rewritten in the form

$$D \log p_\theta(x) = a(\theta)t(x) + b(\theta)\,, \qquad x \notin N_\theta\,. \tag{4}$$

We cannot have $a(\theta) = 0$ for any $\theta$, otherwise for that $\theta$ $D \log p_\theta(x)$ would be constant a.e. $\mu$ on $\mathscr{X}$, so that $\mathrm{Var}_\theta\, D \log p_\theta(X) = 0$, contradicting Regularity

Condition (iii). Let $\theta_1$ be any point of $\Theta$. The right-hand side of (4) is integrable with respect to $p_{\theta_1}\, d\mu$ and therefore so is the left-hand side, and the two integrals are equal since the integrands differ only on a null set. Define

$$g(\theta, \theta_1) = \int (D \log p_\theta) p_{\theta_1}\, d\mu \,. \tag{5}$$

Since $D \log p_\theta$ is jointly measurable, $g(\cdot, \theta_1)$ is $\mathscr{B}$-measurable. By the assumption on $m$ it is possible to choose $\theta_1, \theta_2 \in \Theta$ such that $m(\theta_1) \neq m(\theta_2)$. Integrate (4) on both sides with respect to $p_{\theta_i}\, d\mu$, $i = 1, 2$: $g(\theta, \theta_i) = a(\theta)m(\theta_i) + b(\theta)$. Then subtract the second of these two equations from the first and divide by $m(\theta_1) - m(\theta_2)$:

$$a(\theta) = \frac{g(\theta, \theta_1) - g(\theta, \theta_2)}{m(\theta_1) - m(\theta_2)} \,. \tag{6}$$

It follows from (6) that $a$ is $\mathscr{B}$-measurable. Also integrate (4) on both sides with respect to $p_\theta\, d\mu$. The left-hand side yields 0, as is well known, using the Regularity Conditions. Thus we get $0 = a(\theta)m(\theta) + b(\theta)$. $m$, as an integral of a jointly measurable function, is certainly $\mathscr{B}$-measurable (even differentiable), and it follows then that $b$ is also $\mathscr{B}$-measurable. Now define the set

$$N = \{(x, \theta) \in \mathscr{X} \times \Theta : D \log p_\theta(x) \neq a(\theta)t(x) + b(\theta)\} \tag{7}$$

then $N \in \mathscr{A} \times \mathscr{B}$ and $N_\theta$ is its $\Theta$-section at $\theta$ [4, Section 34]. Let $N^x$ be the $\mathscr{X}$-section of $N$ at $x$. Since $\mu N_\theta = 0$ for every $\theta$, by a version of Fubini's theorem [4, Section 36, Theorem A] there exists $K \in \mathscr{A}$ with $\mu K = 0$ such that $x \notin K$ implies $\nu N^x = 0$. In the remainder of the proof it will be assumed that $x \notin K$. Then (7) can be written in the form

$$D \log p_\theta(x) = a(\theta)t(x) + b(\theta) \,, \qquad \theta \notin N^x \,. \tag{8}$$

Let $\theta_0 \in \Theta$ be chosen arbitrarily and put $p_{\theta_0}(x) = h(x)$. Since $\Theta$ is an interval we may write

$$\log p_\theta(x) = \log h(x) + \int_{\theta_0}^{\theta} D \log p_\vartheta(x)\, d\vartheta \,. \tag{9}$$

Replacing the integrand in (9) with the right-hand side of (8) we obtain

$$\log p_\theta(x) = \log h(x) + \int_{\theta_0}^{\theta} [a(\vartheta)t(x) + b(\vartheta)]\, d\vartheta \,. \tag{10}$$

Thus, $[a(\theta)t(x) + b(\theta)]$ is integrable over every finite interval (but we do not know yet that $a$ and $b$ separately have that property). Since $t$ is not constant a.e. $\mu$, we can choose $x_1$, $x_2$ such that $t(x_1) \neq t(x_2)$. Evaluate (10) at $x_1$ and $x_2$ and take the difference. Since the integral on the right hand side of (10) is finite for both $x_1$ and $x_2$, the difference of the integrals equals the integral of the difference of the integrands. We get

$$\begin{aligned} \log p_\theta(x_1) - \log p_\theta(x_2) &= \log h(x_1) - \log h(x_2) \\ &\quad + [t(x_1) - t(x_2)] \int_{\theta_0}^{\theta} a(\vartheta)\, d\vartheta \,. \end{aligned} \tag{11}$$

Since $t(x_1) - t(x_2) \neq 0$ it follows from (11) that $\int_{\theta_0}^{\theta} a(\vartheta)\, d\vartheta$ is finite and the same is then true for $\int_{\theta_0}^{\theta} b(\vartheta)\,d\vartheta$, by (10). Put

$$q(\theta) = \int_{\theta_0}^{\theta} a(\vartheta)\,d\vartheta\,, \qquad \log c(\theta) = \int_{\theta_0}^{\theta} b(\vartheta)\,d\vartheta\,, \tag{12}$$

then (10) becomes (2) and the only thing left to prove is the assertion in the conclusion of the theorem about the functions $q$ and $c$.

In (11) the left-hand side is a continuously differentiable function of $\theta$ and therefore so is the right-hand side. Since $t(x_1) - t(x_2) \neq 0$ and the integral on the right-hand side of (11) is $q(\theta)$, we conclude that $q$ is continuously differentiable. By (10) and (12) the same is then true for $\log c(\theta)$, so for $c$. Substitute (2) into (5) and obtain

$$g(\theta, \theta_1) = Dq(\theta)m(\theta_1) + D\log c(\theta) \tag{13}$$

and it follows that $g(\cdot, \theta_1)$ is continuous on $\Theta$. Then consulting (6) we see that $a(\cdot)$ is continuous. Using the first of definitions (12) it follows then that $a(\theta) = Dq(\theta)$. Since the function $a$ is continuous and never 0 we must have $Dq(\theta) > 0$ everywhere or $< 0$ everywhere. Therefore, $q$ is strictly monotonic. This concludes the proof.

In the course of the proof we have seen that $a(\theta)m(\theta) + b(\theta) = 0$. Since both $a$ and $m$ are continuous ($m$ is even differentiable by the Regularity Conditions), it follows that $b$ is continuous. Consequently, the right-hand side of (8) is continuous on $\Theta$. Since this is also true for the left-hand side, the two sides must be equal everywhere on $\Theta$, so that the $\nu$-null set $N^x$ is in fact empty.

Note that since $q$ is strictly monotonic and continuously differentiable, it is a one-one bicontinuously differentiable function of $\theta$.

The following observation was made by the referee of this paper. Suppose $t(X)$ attains the Cramér–Rao lower bound as an unbiased estimator of $m(\theta)$. Then the only parametric functions whose unbiased estimator attains the Cramér–Rao lower bound are of the form $\alpha m(\theta) + \beta$. More precisely: suppose the non-constant function $m(\theta)$ has unbiased estimator $t(X)$ achieving equality in (1) at some $\theta_0 \in \Theta$, and another function $n(\theta)$ has unbiased estimator $u(X)$ also achieving equality in (1) at $\theta_0$ (with $t$ and $m$ replaced by $u$ and $n$). Suppose further that Regularity Conditions (i)—(v) (the latter for both $t$ and $u$) are satisfied. Then

$$u(x) = \alpha t(x) + \beta \quad \text{a.e.} \quad \mu\,, \tag{14}$$

with some constants $\alpha$, $\beta$; hence $n(\theta) = \alpha m(\theta) + \beta$. To see this, suppose first $n(\theta) =$ constant. Then by (1) for $u$ and $n$, and evaluated at $\theta = \theta_0$, $\mathrm{Var}_{\theta_0} u(X) = 0$ so $u(x) =$ const. a.e. $u$ which is (14) with $\alpha = 0$. Next, suppose $n$ is not constant, then an equation analogous to (4) evaluated at $\theta_0$ also holds for $u$. Equating the right-hand sides of these equations and observing the sentence following (4) yields (14).

**Acknowledgment.** I would like to thank the referee for many helpful comments.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[2] FEND, A. V. (1959). On the attainment of Cramér–Rao and Bhattacharyya bounds for the variance of an estimate. *Ann. Math. Statist.* **30** 381–388.

[3] FRÉCHET, M. (1943). Sur l'extension de certaines evaluations statistiques au cas de petits échantillons. *Rev. Inst. Internat. Statist.* **11** 182–205.

[4] HALMOS, PAUL R. (1950). *Measure Theory.* Van Nostrand, Princeton.

[5] MEYER, PAUL A. (1966). *Probability and Potentials.* Blaisdell Publ. Co., Waltham, Mass.

[6] RAO, C. RADHAKRISHNA (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37** 81–91.

[7] ZACKS, SHELEMYAHU (1971). *The Theory of Statistical Inference.* Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801