The Annals of Statistics
1976, Vol. 4, No. 5, 998–1002

ON THE ATTAINMENT OF THE CRAMÉR–RAO LOWER BOUND

BY V. M. JOSHI

University of Michigan

It is often stated that the variance of an unbiased estimator of a function of a real parameter can attain the Cramér–Rao lower bound only if the family of distributions is a one-parameter exponential family. A rigorous proof of this statement, subject to certain regularity conditions, has been given by Wijsman. However, in general, the statement is not true. Assuming a revised set of regularity conditions it is shown here that there exists a more general class of distributions for which the Cramér–Rao lower bound for the variance is attained for almost all or even all values of the parameter in an interval. The class reduces to the exponential class only by imposing a restriction requiring the absolute continuity in the parameter of a function involving the logarithm of the probability density.

1. Introduction. Let X be a random variable having a density $P_\theta(\cdot)$ which involves a real parameter θ. Let a real valued statistic $t(X)$ be an unbiased estimator of a real valued function $m(\theta)$. It is usually stated that the variance of $t(X)$ can attain the lower bound provided by the Cramér–Rao (abbreviated hereafter as C–R) inequality

$$\operatorname{Var}_\theta t(X) \geqq [m'(\theta)]^2 \Big/ \operatorname{Var}_\theta \frac{\partial}{\partial\theta} \log p_\theta(X)$$

if, and only if, the distribution of X belongs to the exponential family ([1], [2] and [5, page 187]). No rigorous proof of the assertion seems to have been given previously to that of Wijsman [4]. Wijsman shows the statement to be true provided certain regularity conditions are satisfied.

But the "only" part of the statement viz that the C–R lower bound is attained only if the density is exponential is not true in general. In Theorem 3.1 is determined a more general class of density functions for which the C–R lower bound is attained for almost all θ in an interval. Even requiring the attainment of the lower bound for all θ instead of almost all θ does not restrict the distributions to the exponential class. The restriction to the exponential class is secured by imposing a condition of absolute continuity in θ either of $\log p_\theta(x)$ for fixed x or of a linear combination of terms $\log p_\theta(x_i)$, $i = 1, 2, 3, 4$ for fixed x_i. This is proved in Corollary 3.1 and Note 3.1 below it.

The result proved in Corollary 3.1 corresponds to but is more general than the theorem of Wijsman [4]. A comparison of the two is made in Section 4.

2. Notation. The sample space is an arbitrary measure space $(\mathscr{X}, \mathscr{A}, \mu)$ with

Received January 1974; revised February 1976.

AMS 1970 *subject classification.* Primary 62F10.

Key words and phrases. Cramér–Rao lower bound, variance of unbiased estimate, attainment of lower bound, one-parameter exponential family.

μ sigma-finite. The parameter space is the measure space $(\Theta, \mathscr{B}, \nu)$ with Θ a Borel subset of the real line, $\mathscr{B}$ the Borel σ-field of subsets of Θ and ν 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 θ is denoted by D. Any integration with respect to μ will always be understood to be over the whole of $\mathscr{X}$. The following conventions are used: a.a. x means almost all x with respect to the measure μ on $\mathscr{X}$, a null set K of $\mathscr{X}$ means a set $K \in \mathscr{A}$, such that $\mu(K) = 0$, with corresponding conventions for θ.

3. Main results. The following is an example of a nonexponential density for which the C–R lower bound is attained for all $\theta \in (-\infty, \infty)$. The example is interesting as the density satisfies most of the usual regularity conditions.

EXAMPLE. Select any number α such that $0 < \alpha < 1$. Determine β by $\int_\alpha^\beta (t^2 - 1)\exp(-t^2/2)\, dt = 0$. Put $A(t) = 2$ if $\alpha \leqq |t| \leqq \beta$, and $A(t) = 1$ otherwise. Let

$$p_\theta(x) = CA(|x - \theta|)\exp\left[-\frac{(x-\theta)^2}{2}\right],$$

where C is the normalizing constant. For this density function X is an unbiased estimator of θ with variance equal to 1, which is also the value for each θ of the lower bound provided by the C–R inequality.

The example shows that a density for which the C–R lower bound is attained is not necessarily exponential. A general class of such densities subject to certain regularity conditions is determined in the following Theorem 3.1.

REGULARITY CONDITIONS.

(a) Θ is an interval which may be finite or infinite.

(b) The set $M = \{(x, \theta) \in \mathscr{X} \times \Theta : D \log p_\theta(x) \text{ exists}\} \in (\mathscr{A} \times \mathscr{B})$ and on M, $D \log p_\theta(x)$ is $(\mathscr{A} \times \mathscr{B})$-measurable.

(c) $0 < \operatorname{Var}_\theta D \log p_\theta(X) < \infty$ for a.a. θ.

(d) $\int p_\theta(x)\, d\mu(x)$ is differentiable with respect to θ under the integral sign for a.a. θ.

(e) $m(\theta) = \int t(x)p_\theta(x)\, d\mu(x)$ is differentiable under the integral sign for a.a. θ and the statistic $t(x)$ is not equal to some constant for a.a. x.

THEOREM 3.1. *Let the regularity conditions* (a) *to* (e) *be satisfied. The variance of* $t(X)$ *attains the* C–R *lower bound for* a.a. θ *if and only if, for* $\theta \in \Theta$, *and* $x \notin K$ *where* K *is a null set of* $\mathscr{X}$,

$$p_\theta(x) = c(\theta)h(x)\exp\{q(\theta)t(x)\}\exp\{S(\theta, x)\}$$

in which $h(x) > 0$ *and for* a.a. θ, $c(\theta) > 0$, $c'(\theta)$ *and* $q'(\theta)$ *exist and are finite and* $q'(\theta) \neq 0$, *and further for each* $x \notin K$, $DS(\theta, x) = 0$ *for* a.a. θ.

PROOF. Let L be the union of the null sets of Θ on which the regularity conditions (c) to (e) or the attainment of the C–R lower bound fail to hold. L is a (possibly empty) null set of Θ. The joint measurability of $D \log p_\theta(x)$ in x and θ on the set M in condition (b) implies that the set $M_1 = \{(x, \theta) \in \mathscr{X} \times \Theta,\ D \log p_\theta(x)$ is finite$\} \in \mathscr{A} \times \mathscr{B}$. Let M_1^c be the complement of M_1. For each $\theta \notin L$, the θ-section of M_1^c has zero μ-measure by condition (c). Hence M_1^c is $(\mu \times \nu)$-null. Define on $\mathscr{X} \times \Theta$ a $(\mathscr{A} \times \mathscr{B})$-measurable, finite function $v(\theta, x)$ by $v(x, \theta) = D \log p_\theta(x)$ if $(x, \theta) \in M_1$ and $v(x, \theta) = 0$ if $(x, \theta) \in M_1^c$. For each $\theta \notin L$, $D \log p_\theta(x)$ coincides with $v(x, \theta)$ outside a μ-null set of $\mathscr{X}$. Also for each $\theta \notin L$, $\operatorname{Var}_\theta D \log p_\theta(X) > 0$ and the C–R bound is attained which implies that $m'(\theta) \neq 0$. Hence there exist $\theta_1, \theta_2 \notin L$, such that $m(\theta_1) \neq m(\theta_2)$. Using this result the whole of the argument in [4] up to equation (8) remains applicable to our case for $\theta \notin L$. The argument is not repeated here. Borrowing the result derived in equation (8) of [4],

$$D \log p_\theta(x) = a(\theta)t(x) + b(\theta)\,, \qquad x \notin K,\ \theta \notin N^x\,, \tag{1}$$

where K and N^x are respectively null sets of $\mathscr{X}$ and Θ, for a.a. θ $a(\theta)$ and $b(\theta)$ are measurable and finite functions of θ, and $a(\theta) \neq 0$, for $\theta \notin L$. Note that the null set L of Θ defined in the beginning of this proof gets absorbed in the set N^x in (1).

Applying Lusin's theorem ([3], Theorem 2.3, page 217), there exist functions $q(\theta)$ and $f(\theta)$ such that

$$q'(\theta) = a(\theta) \qquad \text{and} \qquad f'(\theta) = b(\theta) \quad \text{for a.a.} \quad \theta\,. \tag{2}$$

(By Lusin's theorem, $q(\theta)$ and $f(\theta)$ could be restricted to be continuous functions. But no purpose is served by imposing this restriction as it does not survive in the end.)

Now set for $x \notin K$, $\theta \in \Theta$,

$$\log p_\theta(x) = q(\theta)t(x) + f(\theta) + T(\theta, x) \tag{3-i}$$

$$T(\theta, x) = T(\theta_0, x) + S(\theta, x) \tag{3-ii}$$

where θ_0 is an arbitrary fixed point of Θ,

$$\log h(x) = T(\theta_0, x)\,, \tag{3-iii}$$

$$\log c(\theta) = f(\theta)\,, \tag{3-iv}$$

so that for $x \notin K$ and $\theta \in \Theta$,

$$p_\theta(x) = c(\theta)h(x) \exp\{q(\theta)t(x)\} \exp\{S(\theta, x)\}\,. \tag{4}$$

In (4) $c(\theta)$ and $h(x) > 0$ by (3-iii) and (3-iv) and for each $x \notin K$, $DS(\theta, x) = 0$ for a.a. θ by (1) and (2) combined with (3-i) and (3-ii). Further, since for each $x \notin K$, $D \log p_\theta(x)$ is finite for a.a. θ, $q'(\theta)$ and $c'(\theta)$ exist for a.a. θ. Lastly since $a(\theta) \neq 0$ for a.a. θ, it follows from (2) that $q'(\theta) \neq 0$ for a.a. θ. This completes the proof of Theorem 3.1.

NOTE 3.1. For any fixed $x \notin K$, since $DS(\theta, x) = 0$ for a.a. θ, if $S(\theta, x)$ does not vanish for all θ then $S(\theta, x)$ is either a singular function of θ or is of unbounded variation in θ. This follows from Theorem 7.8 (2°), page 121 in [3].

The class of density functions (4) reduces to the exponential class if the function $S(\theta, x)$ is eliminated. This is secured in the following corollary by introducing a restriction regarding the absolute continuity of $\log p_\theta(x)$ on Θ.

COROLLARY 3.1. *If in Theorem* 3.1 *the function* $\log p_\theta(x)$ *satisfies the restriction that for each* $x \notin H$ *where* H *is a null set of* $\mathscr{X}$, $\log p_\theta(x)$ *is an absolutely continuous function of* θ *on* Θ, *then*

$$p_\theta(x) = c_1(\theta)h(x)\exp\{q_1(\theta)t(x)\}\,; \qquad x \notin K_1\,, \quad \theta \in \Theta\,, \tag{5}$$

where K_1 *is a null set of* $\mathscr{X}$, $c_1(\theta)$ *and* $h(x) > 0$ *for* $\theta \in \Theta$, $x \notin K_1$, $\log c_1(\theta)$ *and* $q_1(\theta)$ *are absolutely continuous in* θ *and for* a.a. θ, $c_1(\theta)$ *and* $q_1(\theta)$ *are differentiable with respect to* θ *and* $q_1'(\theta) \neq 0$.

PROOF. Let $K_1 = H \cup K$ where K is the null set of in (4). Select arbitrarily $x_1, x_2, x_3, x_4 \notin K_1$ such that $t(x_2) \neq t(x_1)$. Set $\mathbf{x} = \{x_1, x_2, x_3, x_4\}$ and

$$\begin{aligned} F(\mathbf{x}, \theta) &= [t(x_2) - t(x_1)]\{[\log p_\theta(x_4) - T(\theta_0, x_4)] - [\log p_\theta(x_3) - T(\theta_0, x_3)]\} \\ &\quad - [t(x_4) - t(x_3)]\{[\log p_\theta(x_2) - T(\theta_0, x_2)] \\ &\quad - [\log p_\theta(x_1) - T(\theta_0, x_1)]\} \\ &= [t(x_2) - t(x_1)][S(\theta, x_4) - S(\theta, x_3)] \\ &\quad - [t(x_4) - t(x_3)][S(\theta, x_2) - S(\theta, x_1)] \end{aligned} \tag{6}$$

by (4). Because of the assumed absolute continuity of $\log p_\theta(x)$, for each fixed set of $x_1, x_2, x_3, x_4 \notin K$, $F(\mathbf{x}, \theta)$ is absolutely continuous on Θ. Also for each $x \notin K_1$, $DS(\theta, x) = 0$ for a.a. θ. Hence $DF(\mathbf{x}, 0) = 0$ for a.a. θ. It follows that for fixed $\mathbf{x}$, $F(\mathbf{x}, \theta)$ has a constant value on the interval Θ. But by (3-ii) for $x \notin K$, $S(\theta_0, x) = 0$. Hence $F(\mathbf{x}, \theta_0) = 0$ so that $F(\mathbf{x}, \theta)$ vanishes identically for all $\theta \in \Theta$, and any $x_1, x_2, x_3, x_4 \notin K_1$.

Next in (6) keep x_1, x_2, x_3 fixed. Writting x for x_4 and recalling that $t(x_2) \neq t(x_1)$ we obtain from (6)

$$\begin{aligned} S(\theta, x) &= S(\theta, x_3) + [t(x) - t(x_3)][t(x_2) - t(x_1)]^{-1}\{S(\theta, x_2) - S(\theta, x_1)\} \\ &= u(\theta)t(x) + v(\theta)\,, \qquad x \notin K_1 \end{aligned} \tag{7}$$

since x_1, x_2, x_3 are constants.

Since $t(x)$ is not constant a.e. on $\mathscr{X}$ and for each $x \notin K_1$, $DS(\theta, x) = 0$ for a.a. θ, the functions $u(\theta)$ and $v(\theta)$ in (7) satisfy

$$Du(\theta) = Dv(\theta) = 0 \qquad \text{for a.a. } \theta\,. \tag{8}$$

Substituting in (4) for $S(\theta, x)$ by (7) and putting $q_1(\theta) = q(\theta) + u(\theta)$, $c_1(\theta) = c(\theta)\exp\{v(\theta)\}$ we obtain (5). The restrictions on $c_1(\theta)$, $h(x)$ and $q_1(\theta)$ follow from the restrictions in (4) and from (8) and the assumed absolute continuity in θ of $\log p_\theta(x)$. This completes the proof of Corollary 3.1.

NOTE 3.2. The restrictions of absolute continuity in θ of $\log c_1(\theta)$ and $q_1(\theta)$ in Corollary 3.1 can be removed if instead of assuming the absolute continuity in θ of $\log p_\theta(x)$ the weaker assumption is made that for any quadruple of points

$x_1, x_2, x_3, x_4 \notin K_1$, the function $F(\mathbf{x}, \theta)$ in (6) is for fixed $\mathbf{x}$ absolutely continuous in θ. This is obvious from the proof of the corollary.

4. Comparison with Wijsman's theorem. The result proved by Wijsman is similar to that in Corollary 3.1. The difference between the two is that under Wijsman's theorem (i) the derivatives $c_1'(\theta)$, $q_1'(\theta)$ exists for all $\theta \in \Theta$ and $q_1(\theta)$ is strictly monotonic, and (ii) $c_1'(\theta)$ and $q_1'(\theta)$ are continuous functions of θ.

The item (i) arises because the regularity conditions (c), (d), (e) and the attainment of the C–R lower bound hold only for a.a. θ instead of for all θ as in Wijsman's theorem.

The item (ii) is more important and arises because Wijsman requires $Dp_\theta(x)$ to be continuous in θ for every x. This continuity condition is imposed to secure the joint-measurability in x and θ of $Dp_\theta(x)$. But as the argument of Theorem 3.1 shows, this restriction is not necessary for securing the joint-measurability of $Dp_\theta(x)$.

Suppose that (as in Wijsman's theorem) regularity conditions (c) to (e) hold for all $\theta \in \Theta$ and the C–R lower bound is attained for all θ. Then in Corollary 3.1 since the C–R lower bound is defined for all θ, it follows from (5) that for $x \notin K_1$, $c_1'(\theta)$ and $q_1'(\theta)$ exist for all θ, and $q_1'(\theta) \neq 0$ for any θ. Hence by Rolle's theorem $q_1(\theta)$ is strictly monotonic in (θ). Thus the difference (i) between Corollary 3.1 and Wijsman's theorem disappears entirely. But the difference (ii) would still remain as the functions $q_1'(\theta)$ and $c_1'(\theta)$ need not be continuous in θ. There exist functions which are monotonic and everywhere differentiable but for which the differential coefficient is not continuous as for instance the following:

$$\Theta = (-1, 1), \qquad q_1(\theta) = 4\theta + \theta^2 \sin\frac{1}{\theta} \quad \text{if} \quad \theta \neq 0, \quad q_1(0) = 0.$$

Then q_1' exists and is positive but not continuous at $\theta = 0$.

Thus in a material respect the result proved in Corollary 3.1 is more general than Wijsman's theorem.

Acknowledgment. I would like to thank the referee, the associate editor and the editor for their valuable comments.

REFERENCES

[1] 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.

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

[3] Saks, S. (1937). *Theory of the Integral* (2nd revised ed.). Hafner, New York.

[4] Wijsman, R. A. (1973). On the attainment of the Cramér–Rao lower bound. *Ann. Statist.* **1** 538–542.

[5] Zacks, Shelemyahu (1971). *The Theory of Statistical Inference.* Wiley, New York.

Department of Statistics
1447 Mason Hall
University of Michigan
Ann Arbor, Michigan 48104