Hans Brøns and Inger Harder Hansen

A GENERALIZATION OF THE CRAMÉR-RAO INEQUALITY

Preprint 1971 No. 6

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1971

POLYTEKNISK FORLAG OG TRYKKERI, SØLVGADE 83, 1307 K. TLF. 11 98 22

<u>INTRODUCTION</u>

The famous Cramér-Rao inequality (Rao [1945], Cramér [1946]) gives a lower bound for the variance of unbiased estimators. By generalizing the concept of a mean-value one is led to consider new concepts of unbiasedness. It is shown that in these cases an inequality similar to the inequality of Cramér-Rao holds, providing a lower bound to a quantity which can be interpreted as the measure of dispersion corresponding to the particular choice of mean-value.

The joint results of the authors were first collected in Harder Hansen[1961].

1. Generalized Mean-Values.

In this section we list the properties of generalized mean-values which are needed in the following. For proofs see Brøns [1967], comp. Brøns, Brunk, Franck, and Hanson [1969].

Let g be a convex function on the <u>open</u> interval J on R (the set of real numbers), and consider the sets

$$V_g = \{\tau \in J \mid g(\tau') \geq g(\tau) \forall \tau' \leq \tau\},\ H_g = \{\tau \in J \mid g(\tau') \geq g(\tau) \forall \tau' \geq \tau\}\ .$$

V_g is empty or an interval with the same left endpoint as J, and H_g is empty or an interval with same right endpoint as J.

The set $V_g \cap H_g$ is either empty or an interval, and is the set of points in which g assumes its minimum. Its left endpoint is denoted $m_0 g$ and its right endpoint $m_1 g$. Both values can be infinite. If H_g is empty, $m_0 g$ is the right endpoint of J, and if V_g is empty $m_1 g$ is the left endpoint of J. The identity

$$\{\tau \in J \mid m_0 g \leq \tau\} = \{\tau \in J \mid D^{+} g(\tau) \geq 0\}$$

and the similar one for m_1 and $D^{-} g$ characterize m_0 and m_1 in terms of g's left and right derivatives $D^{-} g$ and $D^{+} g$.

Let $\mathcal{A}$ be a σ-field on a set X. A <u>C-function</u> on X is a function $g: X \times J \to R$ such that the partial function $x \mapsto g(x,\tau)$ is $\mathcal{A}$-measurable for all $\tau \in J$ and the partial function $\tau \mapsto g(x,\tau)$ is convex for all $x \in X$.

Let $D^- g(x,\tau)$ denote the left derivative and $D^+ g(x,\tau)$ the right derivative in τ of the convex function $\tau \mapsto g(x,\tau)$. Then the functions $x \mapsto D^- g(x,\tau)$ and $x \mapsto D^+ g(x,\tau)$ are $\mathcal{A}$-measurable functions for all $\tau \in J$, and the function $x \mapsto g(x,\tau) - g(x,\theta)$ is integrable with respect to a probability measure P on $\mathcal{A}$ for all $(\tau,\theta) \in J^2$ if and only if $x \mapsto D^- g(x,\tau)$ (or $x \mapsto D^+ g(x,\tau)$) are P-integrable for all $\tau \in J$. When the integrals exist the function

$$h(\tau) = \int (g(x,\tau) - g(x,\theta)) P(dx)$$

is convex for all $\theta \in J$ and

$$D^- h(\tau) = \int D^- g(x,\tau) P(dx), \quad D^+ h(\tau) = \int D^+ g(x,\tau) P(dx).$$

$m_0 h$ and $m_1 h$ are independent of θ.

P is said to have <u>unique g- mean-value</u> θ if $m_0 h = m_1 h = \theta$. P has <u>regular</u> g-mean-value , if h also is twice differentiable in θ with positive second derivative , and the function $x \mapsto D^+ g(x,\tau)$ further has positive second moment for all $\tau \in J$.

2. The Generalized Cramér-Rao Inequality.

Let λ be a measure on $\mathcal{A}$, and $(P_\theta)_{\theta \in \Theta}$ a family of probability measures on $\mathcal{A}$ with the open interval $\Theta \subset J$ as parameter space. P_θ is supposed to be absolutely continuous with respect to λ with finite density $x \mapsto f(x,\theta)$.
We put $l(x,\theta) = -\log f(x,\theta)$.
The family $(P_\theta)_{\theta \in \Theta}$ is said to be <u>smooth</u> if

i) The function $\theta \mapsto l(x,\theta)$ is absolutely continuous with respect to

Lebesgue measure and has Radon-Nikodym derivative $D1(x,\theta)$ for all $x \in X$

ii) $0 < \int (D\,1(x,\theta))^2 f(x,\theta)\,\lambda(dx) < +\infty$ for all $\theta \in \Theta$.

iii) For τ tending decreasingly to θ

$$\frac{f(x,\theta) - f(x,\tau)}{(\tau-\theta)\, f(x,\theta)}$$

will tend to $D\,1(x,\theta)$ in quadratic mean with respect to P_θ .

Theorem. Assume that P_θ has the regular g-mean-value θ for all $\theta \in \Theta$, and let $\gamma(\theta)$ denote the second derivative of the function $h_\theta : J \to R$ $\tau \mapsto \int (g(x,\tau) - g(x,\theta))\, f(x,\theta)\lambda\,(dx)$ taken in the point θ. If the family $(P_\theta)_{\theta \in \Theta}$ is smooth

$$(1) \qquad \frac{\int (D^+ g(x,\theta))^2 f(x,\theta)\,\lambda(dx)}{\gamma^2(\theta)} \geqq \frac{1}{\int (D^+ 1(x,\theta))^2 f(x,\theta)\lambda(dx)}$$

for all $\theta \in \Theta$. Here the equality sign is valid if and only if there exists a function $k: \Theta \to R$ such that $\tau \mapsto D^+ g(x,\tau)k(\tau)$ is locally integrable for all $x \in X$ and such that

$$(2) \qquad f(x,\tau) = f(x,\theta)e^{-\int_\theta^\tau D^+ g(x,\xi)\,k(\xi)\,d\xi}$$

for τ and θ in Θ .

Proof. Since $D^- h_\theta(\tau) > 0$ for $\tau < \theta$, $D^+ h_\theta(\tau) < 0$ for $\tau < \theta$, $D^- h_\theta(\theta) \leqq D^+ h_\theta(\theta)$ and $D^- h_\theta$ and $D^+ h_\theta$ are continuous in the point θ it follows that $D^- h_\theta(\theta) = D^+ h_\theta(\theta) = 0$.

Therefore

$$\int D^- g(x,\theta) P_\theta(dx) = \int D^+ g(x,\theta)\, P_\theta(dx)$$

and so $D^{-}g(x,\theta) = D^{+}g(x,\theta)$ with P_θ-probability one.
Since $\tau \mapsto D^{+}g(x,\tau)$ is decreasing $(D^{+}g(x,\tau) - D^{+}g(x,\tau))^2$ will tend decreasingly to 0 for $\tau \downarrow \theta$, i.e. $D^{+}g(x,\tau)$ tends to $D^{+}g(x,\theta)$ in quadratic mean with respect to P_θ. Since

$$\frac{1}{\tau-\theta}\int D^{+}g(x,\tau)\, f(x,\theta)\,\lambda\,(dx) = \int D^{+}g(x,\tau)\,\frac{f(x,\theta)-f(x,\tau)}{(\tau-\theta)\,f(x,\theta)}\; f(x,\theta)\lambda(dx)$$

it follows that

$$\gamma\,(\theta) = \int D^{+}g(x,\theta)\; D\,1(x,\theta)\; f\,(x,\theta)\lambda(dx)\;.$$

Applying Cauchy-Schwarz's inequality one finds that

$$(3)\quad \gamma^2(\theta) \leqq \int (D^{+}g(x,\theta))^2 f(x,\theta)\lambda(dx)\;\; \int (D\,1(x,\theta))^2 f(x,\theta)\lambda(dx)$$

which proves (1).

Equality holds in (3) if and only if for each $\theta \in \Theta$ there exists a constant $k(\theta)$ such that $D\,1(x,\theta) = D^{+}g(x,\theta)k(\theta)$.

Since $\theta \mapsto 1(x,\theta)$ is absolutely continuous it follows that

$$1(x,\tau) - 1(x,\theta) = \int_\theta^\tau D^{+}g(x,\xi)k(\xi)\lambda(d\xi)$$

which is the same as (2).

3. Unbiased Estimation.

Let g be a C-function on R and $t : X \to R$ an $\mathcal{A}$- measurable function. The function g_t: $(x,\theta) \mapsto g(t(x),\theta)$ is a C-function on X.

t is said to be a <u>g-unbiased estimator</u> if P_θ has the unique g_t-meanvalue θ for all $\theta \in \Theta$. t is a <u>regular g-unbiased estimator</u> if θ is a regular g_t - mean - value for all $\theta \in \Theta$.

The theorem shows that if t is a regular g-unbiased estimator and the family $(P_\theta)_{\theta \in \Theta}$ is smooth then

$$(4) \quad V_\theta(t) = \frac{\int (D^+ g(t(x),\theta))^2 f(x,\theta)\lambda(dx)}{\gamma_{g_t}(\theta)^2} \geqq \frac{1}{\int (Dl(x,\theta))^2 f(x,\theta)\lambda(dx)}$$

with equality if and only if

$$f(x,\tau) = f(x,\theta) e^{-\int_\theta^\tau D^+ g(t(x),\xi) k(\xi)\, d\xi} .$$

Example 1. For

$$g(x,\theta) = \frac{(x-\theta)^2}{2} \quad , \quad (x,\theta) \in R^2$$

the probability P on $\mathcal{A}$ has a unique g_t-mean-value θ if and only if $\int t(x)P(dx)$ exsists,is finite and equal to θ. This meanvalue is regular when t has finite variance and

$$V_\theta(t) = \int (t(x) - \theta)^2 f(x,\theta)\lambda(dx).$$

(4) reduces in this way to the well-known Cramér-Rao inequality. Equality holds if and only if

$$f(x,\tau) = f(x,\theta) e^{-\int_\theta^\tau k(\xi)d\xi + t(x)\int_\theta^\tau k(\xi)d\xi} ,$$

in which case the family is exponential.

Example 2. For

$$g(x,\theta) = \begin{cases} \frac{1}{p}(\theta - x) & \text{for} \quad x \leqq \theta \\ \frac{1}{q}(x - \theta) & \text{for} \quad x \geqq \theta \end{cases}$$

where $0 < p = 1 - q < 1$, P has a unique g_t-mean-value if and only if t has the unique p-quantile θ. This mean-value is regular when the distribution function for t is differentiable in θ with positive derivative $k(\theta)$ and

$$V_\theta(t) = \frac{pq}{k^2(\theta)} \quad .$$

(4) provides a lower bound for $V_\theta(t)$ for p-quantile unbiased estimators. This bound is equivalent to an upper bound to the height $k(\theta)$ of the density of t in θ. $p = \frac{1}{2}$ corresponds to median-unbiasedness.

The examples illustrate how V(t) in general provides a natural measure of dispersion corresponding to a specific concept of unbiasedness. The theorem chap. IX, p.9-10 in Brøns [1967] shows that the asymptotic variance of empirical g-mean-values are proportional to V(t).

References.

Hans Brøns [1967] : Forelæsningsnoter i matematisk statistik og sandsynlighedsregning 9-15. Institut for matematisk statistik, Københavns universitet.

H.K. Brøns, H.D. Brunk, W.E. Franck, And D.L. Hanson [1969]: Generalized means and associated families of distributions. Ann. Math. Statist. 40, 339-355.

H. Cramér [1946] : Mathematical Methods of Statistics. Princeton.

C.R. Rao [1945] : Information and Accuracy attainable in Estimation of Statistical Parameters, Bull. Cal. Math. Soc 37, 81-91.