〈Research Paper〉

Journal of the Korean Statistical Society
Vol. 19, No. 2, 1990

An Optimality Criterion for Median-unbiased Estimators+

Nae Kyung Sung*

ABSTRACT

Sung [1990] presented an analogue of the classical Cramér-Rao inequality for median-unbiased estimators with continuous multivariate densities depending upon a vector parameter. In the process, *diffusivity*, a new dispersion measure relevant to median-unbiased estimators, was defined to be a function of median-unbiased estimator's density height. In this paper we shall elaborate these ideas by defining a second kind of diffusivity and discuss the role of mode-unbiasedness in median-unbiased estimation in connection with this second kind of diffusivity. In addition, median-unbiased estimation will be compared to mean-unbiased estimation.

1. Introduction

Let μ be the Lebesgue measure on Euclidean n-space E^n and let the random vector $X=(X_1, X_2, \cdots, X_n)$ have a probability distribution which is absolutely continuous with respect to μ, with density function $f(x;\theta)$, where θ is a real parameter belonging to some parameter set Θ. Let $\tau(\theta)$ be a certain parametric function of interest on Θ which is real-valued and differentiable. Let $\delta(X)$ be an estimator of $\tau(\theta)$.

The following information inequality is well known : under certain regularity conditions (see, e.g., Wijsman [1973]),

$$Var_\theta\ \delta \geq [(\partial/\partial\theta)E_\theta\delta]^2/I_2(\theta) \quad \text{for all } \theta\in\Theta, \tag{1}$$

where

$$I_2(\theta)=E_\theta[(\partial/\partial\theta)\log f(X;\theta)]^2.$$

If $\delta(X)$ is unbiased for $\tau(\theta)$, then (1) reduces to the usual Cramér-Rao inequality :

$$Var_\theta\ \delta \geq [\tau'(\theta)]^2/I_2(\theta).$$

As an alternative to unbiasedness, Brown [1947] introduced the notion of *median-unbiasedness*. Recall that an estimator $\delta(X)$ or $\tau(\theta)$ is called median-unbiased if

$$\Pr_\theta[\delta(X)\leq\tau(\theta)]\geq 1/2,\ \ \Pr_\theta[\delta(X)\geq\tau(\theta)]\geq 1/2 \ \text{ for all } \theta\in\Theta. \tag{2}$$

If the distribution of δ is continuous, then (2) becomes

$$\Pr_\theta[\delta(X)\leq\tau(\theta)]=\Pr_\theta[\delta(X)\geq\tau(\theta)]=1/2.$$

The property of median-unbiasedness was discussed in detail in relation to other properties of point estimators by Birnbaum [1964]. To avoid ambiguity, the usual unbiasedness is called *mean-unbiasedness* below. That is, an estimator $\delta(X)$ of $\tau(\theta)$ is called mean-unbiased if $E_\theta\delta(X)=\tau(\theta)$ for all θ Θ.

Let $\delta(X)$ be any median-unbiased estimator of $\tau(\theta)$ with a continuous density function $g(\cdot\,;\theta)$. The univariate version of a generalized Cramér-Rao analogue for median-unbiased estimators introduced by Sung [1990], which is based on the expected absolute sample score and a measure of dispersion called *diffusivity*, has the following form under certain regularity conditions:

$$1/2g(\tau(\theta)\,;\theta)\geq|\tau'(\theta)|/I_1(\theta), \tag{3}$$

where

$$I_1(\theta)=E_\theta|(\partial/\partial\theta)\log f(X\,;\theta)|,$$

and $1/2g(\tau(\theta)\,;\theta)$ is identified as diffusivity of δ.

In mean-unbiased estimation, we normally prefer an estimator of $\tau(\theta)$ with small variance. Such an estimator's density is in a sense concentrated around $\tau(\theta)$. Diffusivity reflects the same idea: it also is a measure of concentration of an estimator's density around some point.

Mode-unbiasedness, in addition to median-unbiasedness, is relevant to our further discussion:

Definition 1. An estimator $\delta(X)$ of $\tau(\theta)$ is called *mode-unbiased* if $\text{mode}_\theta\,\delta(X)=\tau(\theta)$ for all θ Θ.

Mode-unbiasedness certainly is a further desirable property for a median-unbiased estimator δ, and estimators that are both median- and mode-unbiased are singled out below.

Definition 2. If an estimator $\delta(X)$ is both median-unbiased and mode-unbiased for $\tau(\theta)$, it is called a *mode-median-unbiased* estimator.

In the discussion below, we rename the univariate version of diffusivity discussed in Sung [1990] diffusivity of the first kind, and view this diffusivity of the first kind as analogous to *mean square error* in mean-unbiased estimation. The discussion below also involves an analogue of *variance*, called diffusivity of the second kind. For the sake of completeness bothy diffusivities are defined:

Definition 3. Let Y be a random variable having a continuous density function $g(y\,;\theta)$ where θ is a real-valued parameter. Let $\tau(\theta)$ be the median of Y.

$$1/2g(\tau(\theta)\,;\theta)$$

is defined to be *diffusivity of the first kind* of Y, and

$$1/2\max_y g(y\,;\theta)$$

is defined to be *diffusivity of the second kind* of Y.

We note that diffusivity of the second kind reduces to diffusivity of the first kind in the case of mode-median-unbiased estimators.

We now develop a stronger inequality for median-unbiased estimators than (3), which leads in Secion 2 to a Cramér-Rao analogue for mode-median-unbiased estimates involving diffusivity of the second kind.

2. An Inequality for Diffusivity of the Second Kind

Theorem 1. Let $\mathcal{M}=\{y^*:\ g(y^*\,;\theta)=\max_y g(y\,;\theta)\}$. Let $\phi(\theta)$ belong to $\mathcal{M}$. If there exists $\phi_0(\theta)\in\mathcal{M}$ such that $\phi_0(\theta)\neq\phi(\theta)$, we assume that $\phi(\theta)=\phi_0(\theta)+c$, where c is a constant. Let $\tau(\theta)$ be a real-valued differentiable functiion on Θ. Let $Y\equiv\delta(X)$ be a median-unbiased estimator of $\tau(\theta)$ having that $\zeta(Y)$ is pivotal for θ (*i.e.*, $\zeta(Y)$ has a distribution which does not depend on the parameter θ). Then, under the regularity conditions :

(i) Θ is either the real line, or an interval on the real line.
(ii) $(\partial/\partial\theta) f(x;\theta)$ exists for every $\theta\in\Theta$. (4)
(iii) $(\partial/\partial\theta_0) \log f(x;\theta_0)$ is dominated by $G(x)$, where θ_0 belongs to a neighborhood of θ and $E_\theta G(X)<\infty$, $\theta\in\Theta$.

we have

$$\frac{1}{2\max_y g(y;\theta)} \geq \frac{|\phi'(\theta)|}{I_1(\theta)}. \tag{5}$$

Proof : Let $\Delta\theta$ belong to a neighborhood of 0. Note that in view of the existence of ζ, we can write

$$\int_{-\infty}^{\phi(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy = \int_{-\infty}^{\phi(\theta)} g(y;\theta)dy \tag{6}$$

for all $\theta\in\Theta$.

Consider the following probability :

$$\int_{\phi(\theta)}^{\phi(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy.$$

Using (6), one has

$$\int_{\phi(\theta)}^{\phi(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy = \int_{[x:\delta(x)\leq\phi(\theta)]} [f(x;\theta)-f(x;\theta+\Delta\theta)]\mu(dx), \tag{7}$$

and

$$\int_{\phi(\theta)}^{\phi(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy = \int_{[x:\delta(x)>\phi(\theta)]} [f(x;\theta+\Delta\theta)-f(x;\theta)]\mu(dx), \tag{8}$$

We divide both sides of (7) and (8) by $\Delta\theta$, take absolute values to both sides of resulting equalities and add them to obtain

$$\liminf_{\Delta\theta\to 0} \left| \frac{2\int_{\phi(\theta)}^{\phi(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy}{\Delta\theta} \right| \leq \liminf_{\Delta\theta\to 0} \int \left| \frac{f(x;\theta+\Delta\theta)-f(x;\theta)}{\Delta\theta} \right| \mu(dx),$$

which can be rewritten as

$$(2|\phi'(\theta)|g(\phi(\theta);\theta))^{-1} \geq \left(\liminf_{\Delta\theta\to 0} \int \left| \frac{f(x;\theta+\Delta\theta)-f(x;\theta)}{\Delta\theta} \right| \mu(dx)\right)^{-1}. \tag{9}$$

The right-hand side of (9) is bounded below, by the regularity condition (4iii), by

$$\left(\lim_{\Delta\theta\to 0} \int \left| \frac{f(x;\theta+\Delta\theta)-f(x;\theta)}{\Delta\theta f(x;\theta)} \right| f(x;\theta)\ \mu(dx)\right)^{-1} = \left(\int \left| \lim_{\Delta\theta\to 0} \frac{f(x;\theta+\Delta\theta)-f(x;\theta)}{\Delta\theta f(x;\theta)} \right| f(x;\theta)\ \mu(dx)\right)^{-1} = I_1(\theta)^{-1}.$$

Note that $g(\phi(\theta);\theta)=\max_y g(y;\theta)$. (Q.E.D.)

$$| \int_A [(\partial/\partial\theta)\log f(x\,;\,\theta)]\, f(x\,;\,\theta)\mu(dx) |$$
$$= \int_A | (\partial/\partial\theta)\log f(x\,;\,\theta) |\ \ f(x\,;\,\theta)\mu(dx), \tag{10}$$

and

$$| \int_{A^c} [(\partial/\partial\theta)\log f(x\,;\,\theta)]\, f(x\,;\,\theta)\mu(dx) |$$
$$= \int_{A^c} | (\partial/\partial\theta)\log f(x\,;\,\theta) |\ \ f(x\,;\,\theta)\mu(dx), \tag{11}$$

where $A=[x : \delta(x) \leq \phi(\theta)]$, or, equivalently, if and only if the function $\log f(x\,;\,\theta)$ is monotone in θ on A, and the same holds as well on A^c. It seems that finding such a family of distributions satisfying (10) and (11) is not trivial.

We now identify a special location family of distributions for which a median-unbiased estimator of the location parameter exists and attains the lower bound. If we assume that $X=(X_1, X_2, \cdots, X_n)$ is a random vector from a continuous density of the form $f(x_i\,;\theta)=c\,\exp[h(x_i-\theta)]$, where c is a constant, and h is strictly concave, then strict concavity of h implies that $(\partial/\partial\theta)\ \log f(X\,;\,\theta) = -\Sigma_{i=1}^n h'(X_i-\theta)$ is strictly decreasing in θ. Hence if we take a mode-median-unbiased estimator $\delta(X)$ satisfying $\Sigma_{i=1}^n h'(X_i-\delta(X))=0$, then $\delta(X)\geq\theta$ for $[x : (\partial/\partial\theta)\log f(x\,;\,\theta)\geq 0]$, and $\delta(X)\leq\theta$ for $[x : (\partial/\partial\theta)\log f(x\,;\,\theta) \leq 0]$. Hence we proved :

Theorem 2. Let $X=(X_1, X_2, \cdots, X_n)$ be a sample of n iid random variables from a continuous density of the form $f(x_i\,;\,\theta)=c\,\exp h(x_i-\theta)$, where c is a constant, and h is strictly concave. Assume that all conditions given in Theorem 1 are satisfied. If we take a median-unbiased estimator $\delta(X)$ of θ such that $\Sigma_{i=1}^n h'(X_i-\delta(X))=0$, and if δ is also mode-unbiased for θ, then such a median-unbiased estimator δ attains the lower bound in relation (5) and is the maximum likelihood estimator of θ.

Generally, of course, the two types of diffusivity coincide in the case of mode-median-unbiased estimators. The following example illustrates such a case, with both diffusivities meeting the bound.

Example 1. Let $X_1, X_2, \cdots, X_n$ be a sample of n independent observations from $N(\mu, 1)$. The sample mean is a mode-median-unbiased estimator of μ and satisfies assumptions in Theorem 2. Hence it is optimal for μ. Similarly, if the underlying density is a double exponential distribution : $f(x_i\,;\,\theta)=(1/2)\exp(-|x_i-\theta|)$, with $n=2k+1$, where k is a positive integer, then the sample median is amode-median-unbiased estimator of θ, with both diffusivities meeting the bound.

In general, median-unbiasedness is invariant under strictly monotone transformations. Moreover, median-unbiased *optimality* is preserved under such transformations for diffusivity of the first kind. However, as the following example shows, and in keeping with Theorem 2, diffusivity of the *second* kind requires, in addition, mode-unbiased invariance for this sort of invariance of optimality.

Example 2. Let x be a single observation from a lognormal distribution :

$$f(x\,;\,\xi)=\frac{1}{\sqrt{2\pi}\sigma x}\exp[-\frac{1}{2}(\frac{\log x-\xi}{\sigma})^2],\ x>0,\ \sigma>0,$$

where σ is known and ξ is unknown. Let $\tau(\xi)=e^\xi$. Since the median of X is e^ξ, X is median-unbiased for $\tau(\xi)$. Noting that f is unimodal and has the mode at $\exp(\xi-\sigma^2)$, we see

$$\frac{1}{2f(\exp(\xi-\sigma^2))} = \frac{\sqrt{2\pi}\sigma\exp(\xi-\sigma^2)}{2}\exp(\sigma^2/2).$$

The right-hand side of the inequality is given by

$$\frac{|(\partial/\partial\xi)\exp(\xi-\sigma^2)|}{I(\xi)} = \frac{\sqrt{2\pi}\sigma\exp(\xi-\sigma^2)}{2}.$$

kind converges to its lower bound as $\sigma \downarrow 0$. Even though $\log X$ has a normal distribution and is a median-unbiased estimator of ξ, meeting the lower bound, X does not achieve the lower bound for e^{ξ}. This is because e^{ξ} is not the mode of X while the exponential function is a strictly monotone transformation.

3. Mode-unbiasedness in Median-unbiased Estimation

In this section we show that mode-unbiasedness needs to be added to median-unbiasedness in order to establish a Cramér-Rao analogue based on diffusivity of the second kind. In contrast, only mean-unbiasedness need be invoked for the usual Cramér-Rao inequality.

Let $b_2(\theta)=E_\theta(\delta)-\tau(\theta)$. $b_2(\theta)$ the *mean-bias* of δ when δ is an estimator of $\tau(\theta)$. We assume that the mean-bias function $b_2(\theta)$ is differentiable.

Let $E_\theta(\delta)=\phi_2(\theta)$. The information inequality (1) also yields a lower bound for the mean square error of estimators of $\tau(\theta)$ which are not necessarily unbiased, but, with present notations, takes the following form in terms of the variance:

$$Var_\theta\delta \geq \frac{[\phi'_2(\theta)]^2}{I_2(\theta)} = \frac{[\tau'(\theta)+b'_2(\theta)]^2}{I_2(\theta)}. \tag{12}$$

We now consider an equivalent situation in median-unbiased estimation. Let δ be a median-unbiased estimator of $\tau(\theta)$. We assume that δ has a continuous density function g. Note that we have the following relationship between diffusivity of the first kind and diffusivity of the second kind:

$$\frac{1}{2g(\tau(\theta)\,;\,\theta)} \geq \frac{1}{2\max_y g(y\,;\,\theta)},$$

in analogy to the fact that the mean square error exceeds the variance.

It might also be noticed that diffusivity of the second kind is defined without reference to the parametric function of interest, in analogy to variance in mean-unbiased estimation. On the other hand, diffusivity of the first kind depends on the parametric function of interest by its definition, in analogy to mean square error. Accordingly diffusivity of the first kind is comparable to mean square error, and diffusivity of the second kind is comparable to variance.

We now define a concept of bias naturally associated with median-unbiased estimation, based on the diffusivity of the second kind.

Definition 4. Let δ be an estimator of $\tau(\theta)$. Then $b_1(\theta)=\text{mode}_\theta(\delta)-\tau(\theta)$ is defined to be the *mode-bias* of δ.

Let $\phi_1(\theta)$ be a mode of δ. Then $\phi_1(\theta)=\tau(\theta)+b_1(\theta)$. Therefore the inequality (5) can be expressed as follows:

$$\frac{1}{2\max_y g(y\,;\,\theta)} \geq \frac{|\phi'_1(\theta)|}{I_1(\theta)} = \frac{|\tau'(\theta)+b'_1(\theta)|}{I_1(\theta)} \tag{13}$$

The ineqaulity (13) is seen to be analogous to (12). The numerator of the right-hand side of (12) reduces to $[\tau'(\theta)]^2$ if the mean-bias is zero. Analogously, when δ is mode-median-unbiased for $\tau(\theta)$, then (13) reduces to the inequality (3).

The discussion of this section may be summarized by saying that, in the context of Cramér-Rao type inequalities, mode-median-unbiasedness is to median-unbiased estimation as mean-unbiasedness is to "estimation". With this in mind, we say that a median-unbiased estimator of $\tau(\theta)$ is *optimal* if it attains the lower bound in the class of mode-median-unbiased estimators of $\tau(\theta)$.

estimator of the unknown mean and so is the sample median. They are also mode-unbiased estimators. But the sample mean is the optimal median-unbiased estimator and the sample median is less efficient than the sample mean, in the sense of this paper.

References

1. Birnbaum, A.(1964). Median-unbiased Estimators. *Bulletin of Mathematical Statistics*, 1, 25-34.
2. Brown, G.W.(1947). On Small Sample Estimation. *Annals of Mathematical Statistics*, 18, 582-585.
3. Sung, N.K.(1990). A Generalized Cramér-Rao Analogue for Median-unbiased Estimators. *Journal of Multivariate Analysis*, 32, 204-212.
4. Wijsman, R.A.(1973). On the Attainment of the Cramér-Rao Bound. *Annals of Statistics*, 1, 538-542.