Journal of the Korean
Statistical Society
Vol. 22, No. 1, 1993

Improving L_1 Information Bound in the Presence of a Nuisance Parameter for Median-unbiased Estimators†

Nae-Kyung Sung [1]

ABSTRACT

An approach to make the information bound sharper in median-unbiased estimation, based on an analogue of the Cramér-Rao inequality developed by Sung *et al.*(1990), is introduced for continuous densities with a nuisance parameter by considering information quantities contained both in the parametric function of interest and in the nuisance parameter in a linear fashion. This approach is comparable to that of improving the information bound in mean-unbiased estimation for the case of two unknown parameters. Computation of an optimal weight corresponding to the nuisance parameter is also considered.

KEYWORDS: Diffusivity, Information inequality, L_1 estimation, Lower bound, Median-unbaised estimator, Nuisance parameter.

[1] Statistics Department, Ewha University, Seoul, 120-750, Korea.
† Research was supported by Korea Science Foundation Grant 913-0105-003-2.

1. EXPONENTIAL FAMILIES AND INFORMATION INEQUALITY

We shall suppose that $X = (X_1, \cdots, X_n)$ is a random sample from a population with distribution function F characterized by an unknown parametric vector $\theta \in \Theta$, where the parameter space Θ is either the Euclidean r-space $\Re^r$ or a rectangle in $\Re^r$. We are interested in estimators for $\tau(\theta)$, a real-valued function on Θ, which is partial-differentiable with respect to any component of θ. Let $Y = \delta(X)$ be an estimator of $\tau(\theta)$.

To determine a good estimator δ(X) for $\tau(\theta)$, we need a well-defined criterion under which the perfomance of the estimator can be assessed. In choosing a criterion of estimation, one attempts to provide a measure of closeness of a parametric function of interest, usually restricted to a smaller class of estimators. An optimum estimator in the restricted class is determined by minimizing the measure of closeness.

A general approach given by Lehmann (1951) to this problem is to find a risk-unbiased estimator. The class of risk-unbiased estimators includes the class of L_p-unbiased estimators which minimizes the Minkowsky metric. Under the classical mean-unbiased estimation or L_2-unbiased estimation, the concentration of an estimator around $\tau(\theta)$ is measured by the variance. It is, however, well-known that many statisticians stressed the arbitrariness of restricting the class of estimators and of comparing efficiencies of estimators in terms of variance. It is therefore worthwhile to look for other conditions such as median-unbiasedness. See Pfanzagl (1970) for an application of comparisons of efficiency utilizing median-unbiased estimators. We remark that the class of median-unbiased estimator corresponds to that of L_1-unbiased estimators.

Definition. $\delta(X)$ is called *median-unbiased* for $\tau(\theta)$ if

$$\mathrm{Pr}_\theta\,[\delta(X) \le \tau(\theta)] = \mathrm{Pr}_\theta\,[\delta(X) \ge \tau(\theta)] = 1/2 \quad \text{for all } \theta \in \Theta.$$

The exponential families in mean-unbiased estimation share a number of optimal properties which enables a proper statistical analysis feasible. In the same connection, several approaches has been made as well to establish and identify the exponential families in median-unbiased estimation.

As Lehmann (1986) (p.95) pointed out for the single parameter case in that $\tau(\theta) = \theta$, some families of probability distributions with monotone likelihood ratios admit median-unbiased estimators which are optimum in the sense that among all median-unbiased estimators they minimize the expected loss for any monotone loss function, that is, any loss function which for fixed θ has a minimum of 0 for the true parameter value and is nondecreasing as the parameter moves away from the true value in either direction. The usual convex loss functions are necessarily monotone loss functions. In particular, it can be verified under suitable assumptions on the random loss that an optimal median-unbiased estimator minimizes the probability of differing from θ by more than any given amount. That is, an optimal median-unbiased estimator is stochastically closer to θ than any other median-unbiased estimators. This kind of stochastic ordering is a special case of the stochastic dominance given by Hwang (1985).

Pfanzagl (1971) extended Lehmann's result to the cases of the binomial families and the Poisson families by introducing randomized estimators.

For the problem of estimating a single parameter θ from a probability distribution with many nuisance parameters, Pfanzagl (1979) also showed that given a density of the type

$$C(\theta,\eta)h(x)H(T(x),\theta)G(S(x),\theta,\eta), \tag{1.1}$$

where η are nuisance parameters, there exists a median-unbiased estimator of θ of minimal risk, or, equivalently, of maximal concentration about θ, under the monotone loss functions. Pfanzagl's result applies to certain exponential families with density

$$C(\theta,\eta)h(x)\exp[a(\theta)T(x)+\sum_i a_i(\theta,\eta)S_i(x)]$$

for every sample size if a is increasing and continuous in θ. Though Pfanzagl's existence theorem is analogous to the Lehmann-Scheffé theorem (1950) developed in the class of median-unbiased estimators, it is not easy to find a median-unbiased estimator of minimal risk directly.

Sung *et al.* (1990) identified, in the course of establishing the information bound for median-unbiased estimators, the following form of the location family as an L_1 exponential family:

$$\exp[h(x-\theta)], \tag{1.2}$$

where h is strictly concave.

This family of distributions is a special case of (1.1) and gives optimal median-unbiased estimators in the sense that diffusivity defined by Sung (1990), a measure of dispersion, is minimized among all possible median-unbiased estimators. Noting that the problem of estimating a location parameter can be always converted to a scale problem and median-unbiasedness is invariant under strictly monotone transformations, we remark that (1.2) can be generalized to the following form:

$$q'(x)\exp[h(q(x)-\theta))],$$

where q is strictly monotone.

It is assumed from now on that the random sample were drawn from a population with continuous density function f. Let $\delta(X)$ be a median-unbiased estimator having a continuous density g_δ. Then, as was shown by Sung *et al.* (1990), under certain regularity conditions, the following information inequality holds:

$$1/2g_\delta(\tau(\theta);\theta) \geq |\tau'(\theta)|/I_1(\theta), \tag{1.3}$$

where I_1 is the first absolute moment of the sample score:

$$I_1(\theta) = E_\theta|(\partial/\partial\theta)\log f(x;\theta)|.$$

The left-hand side term in (1.3) is called diffusivity, which is the reciprocal of twice the median-unbiased estimator's density height evaluated at its median point.

Diffusivity could be regarded as a local version of the risk curve introduced by Birnbaum (1961):

$$a(u,\theta,\delta) = \begin{cases} \Pr_\theta[\delta(X) \leq u] = F_\theta(u,\delta) & \text{for } u < \theta, \\ \Pr_\theta[\delta(X) \geq u] = 1 - F_\theta(u-,\delta) & \text{for } u > \theta. \end{cases}$$

The risk curve $a(\cdot,\theta,\delta)$ is monotone increasing to the left of θ and monotone decreasing to the right. It measures tail size of the distribution of δ with respect to θ and the dispersion of δ about θ is measured by the elevation of this curve. That is, estimators are compared only on the basis of their complete distribution functions for each $\theta \in \Theta$ rather than on the basis of certain functionals such as mean squared error. That diffusivity is a local version of the risk curve was first observed by Stangenhaus and David (1977).

When δ is median-unbiased for θ and possesses a density $g_\delta(\cdot;\theta)$, then $a(\theta,\theta,\delta) = 1/2$ and $a'(\theta-,\theta,\delta) = |a'(\theta+,\theta,\delta)| = g_\delta(\theta,\theta)$, so that diffusivity is indeed a natural scalar summarization of the elevation of $a(\cdot,\theta,\delta)$.

Birnbaum's risk function is, in fact, a special form of stochastic dominance. Hence, it can be seen that diffusivity is a limiting measure of concentration of a median-unbiased estimator around the true parametric value of interest. Stangenhaus (1977) arrived the same conclusion in essence from a different point of view.

The L_1 exponential family given in (1.2) leads to optimal median-unbiased estimators which attains the information bound in (1.3) in terms of diffusivity. Typically, the normal family and the double-exponential family belong to the L_1 exponential family. Once one can show that a family of distributions of interest belongs to the L_1 exponential family given in (1.2), it is very easy to find an optimal median-unbiased estimator of θ. Further, it can be shown that such an optimal median-unbiased estimator is a maximum likelihood estimator (See Sung *et al.* (1990)). However, this exponential family does not cover many other frequently occurred distributions in statistics. Furthermore, the L_1 exponential family cannot be extended to the multi-parameter case.

In this paper we deal with a method of improving the L_1 information inequality in (1.3) and making the lower bound sharper in the presence of a nuisance parameter for distributions which do not belong to the L_1 exponential family, but to the Pfanzagl's family (1.1). Motivation of the method is summarized in Section 2.

2. CRAMER-RAO BOUND IN THE PRESENCE OF A NUISANCE PARAMETER

In this section an approach to make the Cramér-Rao bound sharper in the presence of a nuisance parameter is introduced by inserting the information quantity involving the nuisance parameter in a linear fashion. The result is, of course, identical to the classical Cramér-Rao inequality for the 2-parameter case (see Cramér (1946)).

Assume that $\theta = (\theta_1, \theta_2)$ and we are interested in estimating $\tau(\theta_1)$. In this case, θ_2 plays a role of the nuisance parameter. Let $f(x;\theta)$ be the joint continuous density function, where x represents the observations $x_1, \cdots, x_n$. This restriction will be required in the sequel. Let $\delta(X)$ be mean-unbiased for $\tau(\theta_1)$, *i.e.*,

$$\int \delta(x) f(x;\theta) dx = \tau(\theta_1).$$

Under the usual regularity conditions, we have

$$\text{cov}\ (\delta(X), \frac{\partial \log f(x;\theta)}{\partial \theta_1}) = \tau'(\theta_1), \tag{2.1}$$

and

$$\text{cov}\ (\delta(X), \frac{\partial \log f(x;\theta)}{\partial \theta_2}) = 0. \tag{2.2}$$

Let k be a fixed value. The relations (2.1) and (2.2) lead to

$$\text{cov}\ (\delta(X), \frac{\partial \log f(x;\theta)}{\partial \theta_1} - k\frac{\partial \log f(x;\theta)}{\partial \theta_2}) = \tau'(\theta_1). \tag{2.3}$$

From the equation (2.3), the following inequality can be obtained:

$$\text{var}\ \delta(X) \geq [\tau'(\theta_1)]^2 / \mathrm{E}\left[\frac{\partial \log f}{\partial \theta_1} - k\frac{\partial \log f}{\partial \theta_2}\right]^2. \tag{2.4}$$

It can be easily verified that the denominator of (2.4) is minimized for

$$k^* = \mathrm{E}\left[\frac{\partial \log f}{\partial \theta_1}\frac{\partial \log f}{\partial \theta_2}\right] / \mathrm{E}\left[\frac{\partial \log f}{\partial \theta_2}\right]^2,$$

since it is a quadratic function of k.

Remark. If the covariance between $(\partial \log f/\partial\theta_1)$ and $(\partial \log f/\partial\theta_2)$ is zero, we cannot make any improvement on the lower bound.

3. IMPROVING L_1 INFORMATION BOUND

Following similar steps to those in Section 2, we develop a method to improve the L_1 information bound when a nuisance parameter is present.

For convenience sake, let $\delta(X)$ be a median-unbiased estimator of θ_1. As was in Section 2, θ_2 is considered to be a nuisance parameter. Let g_δ be the continuous density function of δ.

By the definition of median-unbiasedness, we have

$$\int_{-\infty}^{\theta_1} g_\delta(y;\theta)dy = \int_{\theta_1}^{\infty} g_\delta(y;\theta)dy, \tag{3.1}$$

or, equivalently, in terms of f,

$$\int_{[\delta(x)\le\theta_1]} f(x;\theta)dx = \int_{[\delta(x)\ge\theta_1]} f(x;\theta)dx. \tag{3.2}$$

Applying the mean value theorem to (3.1) on the domain $[x : \delta(x) \le \theta_1]$ and $[x : \delta(x) \ge \theta_1]$ separately and combining them, we obtain

$$\begin{aligned} &2g_\delta(\theta_1 + h\lambda_h; \theta_1 + h, \theta_2) \\ &= \int \operatorname{sign}(\delta(x) - \theta_1)\frac{f(x;\theta_1 + h, \theta_2) - f(x;\theta_1,\theta_2)}{h}dx \end{aligned} \tag{3.3}$$

for some $h > 0$ and $0 \le \lambda_h \le 1$.

Differentiating (3.2) with respect to θ_2 and allowing interchanging the integral and the derivative signs, we have

$$\int \operatorname{sign}(\delta(x) - \theta_1)\frac{\partial f(x;\theta_1,\theta_2)}{\partial\theta_2}dx = 0. \tag{3.4}$$

Let k be an arbitrary constant. After multiplying k to (3.4), we subtract the result from (3.2), take the absolute values to both sides, and take the limit as $h \to 0$ to obtain

$$2g_\delta(\theta_1;\theta) \le \int \left|\frac{\partial \log f(x;\theta)}{\partial\theta_1} - k\frac{\partial \log f(x;\theta)}{\partial\theta_2}\right| f(x;\theta)dx. \tag{3.5}$$

Suppose that

$$\int \left|\frac{\partial \log f}{\partial \theta_2}\right| f(x;\theta)dx = c < \infty.$$

Then, the right-hand side of (3.5) could be written as

$$\int \left|\left(\frac{\partial \log f(x;\theta)}{\partial \theta_1} / \frac{\partial \log f(x;\theta)}{\partial \theta_2}\right) - k\right| \left|\frac{\partial \log f(x;\theta)}{\partial \theta_2}\right| f(x;\theta)dx. \quad (3.6)$$

Therefore, the right-hand side in (3.5) is minimized for

$$k^* = \text{median}\left(\frac{\partial \log f(x;\theta)}{\partial \theta_1} / \frac{\partial \log f(x;\theta)}{\partial \theta_2}\right) \quad (3.7)$$

with respect to the density of the form

$$\left|\frac{\partial \log f(x;\theta)}{\partial \theta_2}\right| f(x;\theta)/c. \quad (3.8)$$

Such a k^* always exists since the integrand of (3.6) is strictly convex in k.

When $\delta(X)$ is, in general, a median-unbiased estimator for $\tau(\theta_1)$, the L_1 information inequality in the presence of a nuisance parameter can be expressed as follows:

$$1/2g_\delta(\tau(\theta_1;\theta) \geq |\tau'(\theta_1)|/\mathrm{E}\left|\frac{\partial \log f(x;\theta)}{\partial \theta_1} - k^* \frac{\partial \log f(x;\theta)}{\partial \theta_2}\right| \quad (3.9)$$

The equality in (3.9) holds if and only if both (3.10) and (3.11) below hold:

$$\left|\int_{[\delta(x) \leq \tau(\theta_1)]} \frac{\partial \log f}{\partial \theta_1} f dx\right| = \int_{[\delta(x) \leq \tau(\theta_1)]} \left|\frac{\partial \log f}{\partial \theta_1} - k^* \frac{\partial \log f}{\partial \theta_2}\right| f dx, \quad (3.10)$$

and

$$\left|\int_{[\delta(x) \geq \tau(\theta_1)]} \frac{\partial \log f}{\partial \theta_1} f dx\right| = \int_{[\delta(x) \geq \tau(\theta_1)]} \left|\frac{\partial \log f}{\partial \theta_1} - k^* \frac{\partial \log f}{\partial \theta_2}\right| f dx. \quad (3.11)$$

Remark. Besides utilizing information contained in a nuisance parameter to improve the lower bound, one may consider an information inequality on a general L_p space. Barankin (1949) dealt with the problem of minimizing the sth absolute central moment to give a best L_p-unbiased estimator, where $1/s + 1/p = 1$, and $p > 1$. In this case the Fisher information I_2 is changed to I_p. Ignoring the nuisance parameter and applying the Hölder's inequality to (3.3), we have

$$\frac{1}{2g_\delta(\tau(\theta_1);\theta)} \geq \frac{|\tau'(\theta_1)|}{[\mathrm{E}\,|\partial \log f/\partial\theta_1|^p]^{1/p}}$$

for a given value of p. In particular, one can observe that squared diffusivity cannot exceed the Fisher's bound in median-unbiased estimation when $p = 2$. This fact can be seen as a special case of the result given by Pfanzagl (1970) for median-unbiased estimation with squared diffusivity as a measure of concentration.

4. EVALUATION OF k

In Section 3 we derived a formula for k, given in (3.7) and (3.8), which makes the lower bound, that is, the denominator of the right-hand side of (3.9), sharper. Unfortunately, computation of (3.7) may in fact be problematical. Analytical solutions cannot be expected except for a few cases.

In order to obtain an optimum value of k, we must identify the form of the density given in (3.8) and find the median value of the statistic given in (3.7). In many cases evaluation of k seems to be not easy.

In some cases, however, a simple computational algorithm given below for the univariate density may help. Consider (3.6). Given an interval $(-M, M)$ on the domain, we partition the interval to m subintervals with indices $j_1, \cdots, j_m$. The value of M depends on the density given in (3.8). Let p_i be the value of

$$\frac{\partial \log f(x;\theta)}{\partial\theta_1} \Big/ \frac{\partial \log f(x;\theta)}{\partial\theta_2}$$

evaluated at a point on the ith subinterval. Also, let h_i be the area of curve

$$\left|\frac{\partial f(x;\theta)}{\partial\theta_2}\right|$$

evaluated on the ith subinterval, where $\sum_i h_i = 1/c$.

Then (3.6) can be represented approximately as follows:

$$(3.6) \approx c \sum_{-M}^{M} |p_i - k| h_i. \tag{4.1}$$

Assume, for simplicity, that $p_i - k \leq 0$ for $i = j_1, \cdots, j_l$, and $p_i - k > 0$ for $i = j_{l+1}, \cdots, j_m$. Then (4.1) can be written as

$$c[-\sum_{j_1}^{j_l} (p_i - k) h_i + \sum_{j_{l+1}}^{j_m} (p_i - k) h_i]. \tag{4.2}$$

Therefore, one can find an optimum k which makes (4.2) to be 0 approximately by varying input values of k.

In case that evaluation of multiple integrls is required, one may use a numerical integration technique such as the Monte Carlo importance sampling scheme (see, e.g., Rubinstein (1981)).

Example 1. Let X be a random variable with a density of the form

$$f(x; \xi, \sigma) = 2(x - \xi)/\sigma^2, \ \ \xi \leq x \leq \xi + \sigma.$$

$X - \sigma/\sqrt{2}$ is median-unbiased for ξ. Diffusivity at ξ is $.354\sigma$. Without utilizing the nuisance parameter σ, the L_1 lower bound is $.25\sigma$ and it does not attain the equality. But, if we consider the information contained in the nuisance parameter, the bound turns out to be the same as the value of diffusivity. In this case $k^* = 1/\sqrt{2}$.

Example 2. Let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$. It is well-known that $\bar{X}$ is an optimum median-unbiased estimator of μ. In this case the optimum value of k is zero, that is, no more improvement is possible even though we consider the information quantity contained in the nuisance parameter σ.

ACKNOWLEDGEMENTS

The author would like to thank Sookhyung Lee for her computational assistance.

REFERENCES

(1) Barankin, E. W. (1949). Locally best unbiased estimates. *Annals of Mathematical Statistics*, 20, 477-501.

(2) Birnbaum, A. (1961). A unified theory of estimators *I. Annals of Mathematical Statistics*, 32, 112-135.

(3) Cramér, R. (1946). *Mathematical methods of statistics*, Princeton University Press, Princeton.

(4) Hwang, J. T. (1985). Universal domination and stochastic domination: Estimation simultaneously under a broad class of loss functions. *Annals of Statistics*, 13, 295-314.

(5) Lehmann, E. L. (1951). A general concept of unbiasedness. *Annals of Mathematical Statistics*, 22, 587-592.

(6) Lehmann, E. L. (1986). *Testing statistical hypotheses*, 2nd edition, Wiley, New York.

(7) Lehmann, E. L. and Scheffé, H. (1950). Completeness, similar regions and unbiased estimation. *Sankhyā*, 10, 305-340.

(8) Pfanzagl, J. (1970). On the asymptotic efficiency of median unbiased estimates. *Annals of Mathematical Statistics*, 41, 1500-1509.

(9) Pfanzagl, J. (1971). On median unbiased estimates. *Metrika*, 18, 154-173.

(10) Pfanzagl, J. (1979). On optimal median unbiased estimators in the presence of nuisance parameters. *Annals of Statistics*, 7, 187-193.

(11) Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, Wiley, New York.

(12) Stangenhaus, G. (1977). *Optimum estimation under generalized unbiasedness*, Unpublished Ph.D. dissertation, Department of Statistics, Iowa State University, Ames, Iowa.

(13) Stangenhaus, G., and David, H. T. (1978). *A robust Cramér-Rao analogue*, Unpublished paper, Department of Statistics, Iowa State University, Ames, Iowa.

(14) Sung, N. K. (1990). A generalized Cramér-Rao analogue for median-unbiased estimators. *Journal of Multivariate Analysis*, 32, 204-212.

(15) Sung, N. K., Stangenhaus, G., and David, H. T. (1990). A Cramér-Rao analogue for median-unbiased estimators. *Trabajos de Estadística*, 5, 83-94.