The Korean Communications
in Statistics Vol. 11 No. 1, 2004
pp. 13-20

A Bhattacharyya Analogue for Median-unbiased Estimation

Nae Kyung Sung [1)]

Abstract

A more general version of diffusivity based on total variation of density is defined and an information inequality for median-unbiased estimation is presented. The resulting information inequality can be interpreted as an analogue of the Bhattacharyya system of lower bounds for mean-unbiased estimation. A condition on which the information bound is achieved is also given.

Keywords: Bhattacharyya, median-unbiased estimation, diffusivity, total variation

1. Introduction

Diffusivity is a measure of spread for median-unbiased estimators defined by Sung *et al.* (1990*a*). It could be regarded as a local version of the risk curve introduced by Birnbaum (1961). Sung (1990*b*) extended this measure to the multivariate case by introducing a generalized definition of median-unbiasedness. A more general discussion about diffusivity in view of comparing its role in median-unbiased estimation with its counterpart in mean-unbiased estimation in relation to various versions of information inequalities can be found in Sung (1990*c*), Sung (1993) and Sung (1997).

The efficiency of mean-unbiased estimators under the usual quadratic loss may be assessed by the Cramer-Rao lower bound. One way to generalize the Cramer-Rao lower bound and provide a system of lower bounds in which the Cramer-Rao lower bound is a special case was first given by Bhattacharyya (1946).

The Bhattacharyya lower bound of order k, $k \geq 2$, requires more stringent regularity conditions than the Cramer-Rao lower bound. The conditions under which a mean-unbiased estimator achieves the Bhattacharyya lower bound of the kth order was given by Fend (1959).

In this paper we define another kind of diffusivity which is a function of total variation of density and give an information inequality for median-unbiased estimators based on the newly-defined total variation-type diffusivity, which can be regarded as an analogue of the Bhattacharyya system of information inequalities.

Let $X=(X_1,\cdots,X_n)$ be a random sample from a population with distribution function F char-

1) Senior Researcher, Institute of Statistics, Korea University, Seoul 136-701, Korea.
E-mail: hyeilsnk@hanafos.com

acterized by an unknown parameter $\theta \in \Theta$. Consider an estimator or a decision $\delta(X)$ for $\tau(\theta)$, a parametric function of interest.

Definition 1 $\delta(X)$ is called *median-unbiased* for $\tau(\theta)$ if

$$median_{\theta}\,\delta(X) = \tau(\theta) \text{ for all } \theta \in \Theta.$$

Definition 2 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. Then $1/2g(\tau(\theta);\theta)$ is defined to be the *diffusivity* of Y.

As was shown by Sung *et al.* (1990*a*), under certain regularity conditions, the following information inequality based on diffusivity holds:

$$\frac{1}{2g_{\delta}(\tau(\theta);\theta)} \geq \frac{|\tau'(\theta)|}{I_1(\theta)}$$

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

$$I_1(\theta) = E_{\theta}\left|\frac{\partial \log f(x;\theta)}{\partial \theta}\right|.$$

Before we introduce another kind of diffusivity, we first define the total variation of a density function.

Definition 3 We suppose that a random variable Y has a continuous density function $g(y;\theta)$. Assume that $g(y;\theta)$ is defined and finite-valued on the finite interval $[a,b]$. Let $a=y_0<y_1<\cdots<y_k=b$ be a finite partition of $[a,b]$. Let

$$T_Y[a,b] = \sup \sum_{i=1}^{k} |g(y_i) - g(y_{i-1})|,$$

where the supremum is taken over all partitions of $[a,b]$. $T_Y[a,b]$ is called the *total variation* (TV) of g on $[a,b]$. We put $T_Y[R]=\sup_{a,b} T_Y[a,b]$.

Definition 4 For a random variable Y with a continuous density function $g(y;\theta)$, we define $1/T_Y[S(\theta)]$ as *TV-type diffusivity* of Y, where $S(\theta)$ is defined by

$$g(y;\theta)>0 \text{ for } y \in S(\theta),$$
$$g(y;\theta)=0 \text{ for } y \in R-S(\theta),$$

and $T_Y[S(\theta)]$ is the total variation of Y on $S(\theta)$.

Suppose that $S(\theta)$ is a countable union of intervals. Then $T_Y[S(\theta)]$ has a simple form:

Lemma 1 For a random variable Y with a continuous and differentiable density function $g(y;\theta)$,

$$T_Y[S(\theta)]=E_\theta\left|\frac{\partial \log g(y;\theta)}{\partial y}\right|,$$

where $S(\theta)$ is the set of y's such that $g(y;\theta)>0$ and is a countable union of intervals.

Proof: Let $S(\theta)$ be a countable union of disjoint intervals $J_i=(a_i,b_i)$, $i=1,2,\cdots$. Let $a_i=y_{i0}<y_{i1}<\cdots<y_{ik}=b_i$ be a partition of J_i. With the supremum taken over all partitions of J_i,

$$T_Y[J_i]=\sup\sum_{j=1}^{k}\left|\frac{g(y_{ij})-g(y_{i(j-1)})}{y_{ij}-y_{i(j-1)}}\right|(y_{ij}-y_{i(j-1)})=\int_{J_i}\left|\frac{\partial g(y;\theta)}{\partial y}\right|dy.$$

Hence, for $\cup_i J_i$, the right-hand side term becomes

$$T_Y[S(\theta)]=\int_{S(\theta)}\left|\frac{(\partial/\partial y)g(y;\theta)}{g(y;\theta)}\right|g(y;\theta)dy=\int_{S(\theta)}\left|\frac{\partial \log g(y;\theta)}{\partial y}\right|g(y;\theta)dy.\ \text{(Q.E.D.)}$$

We state without proof the following lemma:

Lemma 2 For a random variable Y with a continuous density function $g(y;\theta)$,

$$T_Y[R]=T_Y[S(\theta)]+\{\textit{total sum of jump sizes of } g\},$$

where $S(\theta)$ is the set of y's such that $g(y;\theta)>0$.

If the value of diffusivity is 0, then we say that diffusivity is not defined. In mean-unbiased estimation, the variance is not defined for, *e.g.*, the Cauchy distribution which has heavy tails. On the contrary, diffusivity is not defined for a distribution such as x_1^2 which approaches ∞ at a point.

Bickel and Lehmann (1979) defined a functional $\Delta(Y)$ for a random variable Y to be a measure of spread if Δ satisfies

(*i*) $\Delta(Y)\geq 0$
(*ii*) $\Delta(aY)=|a|\Delta(Y)$ for $a\neq 0$

(*iii*) $\Delta(Y+b)=\Delta(Y)$ for all b
(*iv*) $\Delta(-Y)=\Delta(Y)$

It is interesting that for continuous densities the TV-type diffusivity satisfies the Bickel and Lehmann's conditions so that it enjoys the defining properties of the measures of spread even though it is a completely different measure of dispersion from usual measures of dispersion for mean-unbiased estimation. It might also be noticed that the TV-type diffusivity is defined without reference to the parametric function of interest, which is analogous to the variance in mean-unbiased estimation. On the other hand, the original diffusivity given by Sung *et al.* (1990*a*) depends on the parametric function of interest by its definition, which may be regarded as an analogue of the mean square error for median-unbiased estimation, since the mean square error of an estimator, in general, depends on the parametric function of interest. However, We remark that a good estimator in terms of the original diffusivity is not necessarily a good estimator in terms of the TV-type diffusivity.

We also note that the TV-type diffusivity reduces to the original diffusivity if the median-unbiased estimator of $\tau(\theta)$ has a unimodal density function which is continuous over the real line and has a mode at $\tau(\theta)$.

2. Bhattacharyya-type Information Inequality

Let $X=(X_1,\cdots,X_n)$ be a random vector of n *iid* random variables having a joint density function $f(x;\theta)$, where θ is a real-valued parameter. Let τ be a real-valued function on θ, which is differentiable. Let $Y\equiv\delta(X)$ be median-unbiased for $\tau(\theta)$. We assume that Y has a known density $g(y;\theta)$ and both f and g are continuous.

We also assume that g has a global maximum at $\phi(\theta)$ and has at most countable number of local extreme points, with the convention that an interior point is taken in case that g is constant over an interval and all points belonging to that interval are local extreme points. We denote local extreme points of g as $\Psi_i(\theta)$ and assume further that $\Psi_i(\theta)=\phi(\theta)+c_i$, where c_i is a constant, $i=1,2,\cdots$. Of course, $c_i=0$ if $\Psi_i(\theta)=\phi(\theta)$ for some i. We assume that there exists a one-to-one transformation ξ such that given $\Psi_i(\theta)$, $i=1,2,\cdots$, $\xi(\Psi_i(\theta))$ is a constant for all $\theta\in\Theta$ and $\xi(Y)$ has a distribution which does not depend on the parameter θ.

Theorem 1 Let $\tau(\theta)$ be a real-valued function on Θ, which is differentiable. Let $Y\equiv\delta(X)$ be a median-unbiased estimator having a continuous density g_δ. We assume that g_δ has a global maximum at $\phi(\theta)$. Then under the following regularity conditions:

(*i*) Θ is either the real line, or an interval on the real line.
(*ii*) $(\partial/\partial\theta)f(x;\theta)$ exists *a.e.* for every $\theta\in\Theta$.
(*iii*) $0<E_\theta|(\partial/\partial\theta)\log f(x;\theta)|<\infty$ for every $\theta\in\Theta$,

we have

$$\frac{1}{T_Y[S(\theta)]} \geq \frac{|\phi'(\theta)|}{I_1(\theta)} ,$$

where $g_\delta(y;\theta) > 0$ for $y \in S(\theta)$.

Proof: Let $A_i = [x : \Psi_i(\theta) < \delta(X) \leq \Psi_{i+1}(\theta)]$. Note that A_i's are disjoint and $\cup_i A_i = R$. Consider $P_{\theta+\Delta\theta}[\Psi_i(\theta) < Y \leq \Psi_i(\theta+\Delta\theta)]$, where $\Delta\theta$ is in the neighborhood of 0. We can write this probability as follows:

$$\int_{\psi_i(\theta)}^{\psi_i(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy = \int_{A_i} f(x;\theta+\Delta\theta)dx - \int_{A_i} f(x;\theta)dx + \int_{\psi_{i+1}(\theta)}^{\psi_{i+1}(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy.$$

Taking absolute values to both sides after moving the last integral term of the right-hand side of the above to the left-hand side, we have

$$\left| \int_{\psi_i(\theta)}^{\psi_i(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy - \int_{\psi_{i+1}(\theta)}^{\psi_{i+1}(\theta+\Delta\theta)} g(y;\theta+\Delta\theta)dy \right| \leq \int_{A_i} |f(x;\theta+\Delta\theta) - f(x;\theta)|\, dx .$$

Let $\Delta\theta$ be in the neighborhood of 0. Dividing both sides by $\Delta\theta$ and letting $\Delta\theta \to 0$, and using the assumption that $\Psi_i'(\theta) = \phi'(\theta)$, one finds

$$|\phi'(\theta)|\, |g(\psi_i(\theta);\theta) - g(\psi_{i+1}(\theta);\theta)| \leq \lim_{\Delta\theta\to 0} \int_{A_i} \left| \frac{f(x;\theta+\Delta\theta) - f(x;\theta)}{\Delta\theta} \right| dx .$$

Adding above inequality over all i's, we have

$$|\phi'(\theta)| \sum_i |g(\psi_i(\theta);\theta) - g(\psi_{i+1}(\theta);\theta)| \leq \lim_{\Delta\theta\to 0} \int \left| \frac{f(x;\theta+\Delta\theta) - f(x;\theta)}{\Delta\theta} \right| dx .$$

Since $|(\partial / \partial\theta)\log f(x;\theta)|$ exists and integrable for all θ,

$$|\phi'(\theta)| \sum_i |g(\psi_i(\theta);\theta) - g(\psi_{i+1}(\theta);\theta)| \leq \int \left| \frac{\partial \log f(x;\theta)}{\partial\theta} \right| f(x;\theta)dx .$$

Note that $\sum_i |g(\psi_i(\theta);\theta) - g(\psi_{i+1}(\theta);\theta)|$ is nothing but the total variation of g. Hence,

$$|\phi'(\theta)|T_Y[S(\theta)] \le \int \left|\frac{\partial \log f(x;\theta)}{\partial\theta}\right| f(x;\theta)dx. \quad \text{(Q.E.D.)}$$

When we take a sample from a strictly monotone density, it is often possible to find median-unbiased estimators for which the original diffusivity is smaller than the lower bound. Such a density has a discontinuity point at an endpoint of its support so that the conditions for a regular case are usually not satisfied. But Theorem 1 still holds in such cases due to the regularity condition (ii).

Example Let X be a random variable with density function

$$f(x;m) = c[-(x-m)^4 + 2(x-m)^2 + 8], \qquad |x-m| < 2,$$

where $c = 15/448$. Since f is symmetric about m, then X itself is a median-unbiased estimator of m. In addition, f is a bimodal density. f has maxima at $m \pm 1$, and has a local minimum at m. Also, $f(m-1) = f(m+1) = 9c$, $f(m) = 8c$. Therefore the values of diffusivity of the original and the TV-type are $1/16c$ and $1/20c$, respectively. Since $(\partial/\partial m)\log f(x;m) = 4(x-m)^3 - 4(x-m)$, then

$$I_1(m) = 4c\int_{m-2}^{m+2} |4(x-m)^3 - 4(x-m)|\,dx = 20c.$$

Hence the lower bound is achieved only with the TV-type diffusivity.

3. On the Attainment of the Lower Bound

We now identify the family of distributions for which the lower bound is achieved. In order to achieve the lower bound the following equality should hold:

$$\left|\int_{A_i} \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\,dx\right| = \int_{A_i}\left|\frac{\partial \log f(x;\theta)}{\partial\theta}\right| f(x;\theta)\,dx$$

for all i. In general consider the following inequality:

$$\int_E |h|\,dx \ge \left|\int_E h\,dx\right|,$$

where h is a real-valued function and $E \subset R$. Obviously the equality holds if and only if $h \ge 0$ *a.e.* on E or $h \le 0$ *a.e.* on E. Therefore we can deduce that the equality holds if and only if either $(\partial/\partial\theta)\log f(x;\theta)\geq 0$ or $(\partial/\partial\theta)\log f(x;\theta)\leq 0$ on A_i for all i, or, equivalently, $\log f(x;\theta)$ is non-increasing or non-decreasing in θ on A_i for all i.

Noting that for strictly convex h, $(\partial/\partial\theta)f(x;\theta)$ is strictly increasing in θ and that the regularity condition (*ii*) permits strictly convex h, we can slightly generalize Theorem 3 in Sung *et al.* (1990*a*) to find an optimal median-unbiased estimator for a certain density belonging to a location family.

Theorem 2 Let $X=(X_1,\cdots,X_n)$ be a sample of n *iid* random variables from a density of the form $f(x_1;\theta)=c\exp h(x_1-\theta)$, where c is a constant and h is strictly concave (convex). Assume that the regularity conditions in Theorem 1 are satisfied. If we take a median-unbiased estimator $\delta(X)$ of θ such that $\sum_i h'(X_i-\delta(X))=0$, then such an estimator δ attains the lower bound. Conversely, if a median-unbiased estimator δ of θ attains the lower bound, then δ satisfies $\sum_i h'(X_i-\delta(X))=0$.

4. A Comparison with the Bhattacharyya Lower Bound

The Bhattacharyya system of lower bounds, given by Bhattacharyya (1946) is a method of improving the Cramer-Rao lower bound by considering higher order derivatives of $f(x;\theta)$. That is, in the Bhattacharyya inequality, we increase the right-hand side of the Cramer-Rao inequality. The inequality given in Theorem 1 can be considered to be an analogue of the Bhattacharyya lower bound in median-unbiased estimation in a reverse way.

Let us consider a median-unbiased estimator δ of $\tau(\theta)$ and assume that δ has a continuous density function g. g could be a multi-modal density function. The TV-type diffusivity considers all local extreme points and measures density height changes of g, and accordingly it may be regarded as an analogous procedure to considering higher order derivatives in the Bhattacharyya system of lower bounds. The crucial difference compared to the Bhattacharyya inequality is that we decrease the left-hand side of the information inequality with the TV-type diffusivity.

References

[1] Bhattacharyya, A. (1946). On some analogues to the amount of information and their uses in statistical estimation. *Sankhya* 8: 1-14.

[2] Bickel, P. J. and Lehmann, E. L. (1979). Descriptive Statistics for Nonparametric Models IV. Spread. *Contributions to Statistics: Jaroslav Hajek memorial volume*. D. Reidel Pub. Co., Boston.

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

[4] Fend, A. V. (1959). On the attainment of Cramer-Rao and Bhattacharyya bounds for the

variances of an estimate. *Annals of Mathematical Statistics* 30: 381-388.

[5] Sung, N. K., Stangenhaus, G., and David, H. T. (1990*a*). A Cramer-Rao analogue for median-unbiased estimators. *Trabajos de Estadistica* 5: 83-94.

[6] Sung, N. K. (1990*b*). A generalized Cramer-Rao analogue for median-unbiased estimators. *Journal of Multivariate Analysis* 32: 204-212.

[7] Sung, N. K. (1990*c*). An optimality criterion for median-unbiased estimators. *Journal of the Korean Statistical Society* 19: 176-181.

[8] Sung, N. K. (1993). Improving L_1 information bound for median-unbiased estimators in the presence of a nuisance parameter. *Journal of the Korean Statistical Society* 22 (1): 1-12.

[9] Sung, N. K. (1997). Chapman-Robbins-type and Bayesian lower bound based on diffusivity. *Journal of the Korean Statistical Society* 26(4) 445-452.

[Received September 2003, Accepted November 2003]