A Festschrift for Herman Rubin
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 45 (2004) 207–217

Notes on the bias-variance trade-off phenomenon

Jeesen Chen[1]

University of Cincinnati

Abstract: The main inequality (Theorem 1) here involves the Hellinger distance of a statistical model of an observation X, which imposes bounds on the mean of any estimator in terms of its variance. We use this inequality to explain some of the bias-variance trade-off phenomena studied in Doss and Sethuraman (1989) and Liu and Brown (1993). We provide some quantified results about how the reduction of bias would increase the variance of an estimator.

1. Introduction

In certain estimation problems the following "bias-variance trade-off" phenomenon might occur: the price of reducing the bias of an etimator T is the dramatic increase of its variance. For problems exhibiting this property, one shouldn't apply the bias reducing procedures blindly. Furthermore, any estimator having good mean square error performance should be biased, and there is a balance between the bias function and the variance function. It is desirable to study the scope of this phenomenon and how the variance and the bias of an estimator affect each other.

Doss and Sethuraman (1989) seem to have been the first to demonstrate the existence of the long suspected bias-variance trade-off phenomenon. However, this result requires stringent conditions, such as the nonexistence of unbiased estimators for the problem and the square integrability of relative densities for the statistical model, thus severely restricting its applicability.

Liu and Brown (1993) broadened the scope of, and brought a new element, the singular/regular property of an estimation problem, into the study of the trade-off phenomenon. Here the focus is on a special aspect of the trade-off phenomenon, the "nonexistence of informative (i.e. bounded variances) unbiased estimators" property, and its connection with the singular/regular property is studied. For singular estimation problems, the bias-variance trade-off phenomenon is an essential component since the "nonexistence of informative unbiased estimators" property always holds (see Theorem 1 of Liu and Brown (1993)). For regular estimation problems, however, the connection is not clear. On one hand, due to the effect of a singular point as a limiting point, the "nonexistence of informative unbiased estimators" property does occur in some regular estimation problems, even though those problems may be quadratic-mean-differentiable with Fisher information totally bounded away from zero. (See Example 2 of Liu and Brown (1993)). On the other hand, there are many known regular estimation problems having informative unbiased estimators. Therefore, focusing on the singular/regular property alone can't completely describe the scope of bias-variance trade-off phenomenon.

[1]Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221-0025, USA. e-mail: `swilly@math.uc.edu`

Keywords and phrases: Hellinger distance, Hellinger modulus, variance-mean relationship, bias-variance trade-off phenomenon, singular problems.

AMS 2000 subject classifications: primary 62F11; secondary 62F12, 62G05, 62A99.

It is intriguing to consider how the results of Liu and Brown (1993) may be perceived. The impression may be that Theorem 1 of Liu and Brown (1993), the "nonexistence of informative unbiased estimators" for a singular estimation problem, seems compatible with the well-known Rao-Cramér inequality. This inequality, under suitable regularity conditions, provides a lower bound of variances for unbiased estimators in terms of the reciprocal of the Fisher information number. For a singular point (or, a point with zero Fisher information number), the lower bound of variances for unbiased estimators becomes infinite; hence it is impossible to have an informative unbiased estimator (if the regularity conditions of Rao-Cramér inequality hold). With this impression, one might be surprised to see Example 4 of Liu and Brown (1993) which exhibits an unbiased estimator with finite variance at a singular point. This seems to contradict the Rao-Cramér inequality or Theorem 1 of Liu and Brown (1993). Of course, there is no contradiction here: first, Example 4 of Liu and Brown (1993) violates the required regularity conditions for the Rao-Cramér inequality; second, Theorem 1 of Liu and Brown (1993) only prevents the possibility of an unbiased estimator having a uniform finite upper bound for variances in any Hellinger neighborhood of a singular point, and not the possibility of an unbiased estimator with finite variance at *a* singular *point*. Nevertheless, the possible confusion indicates the need to find a framework in which we can put all the perception here into a more coherent view. One suggestion is to use an "appropriate variation" of the Rao-Cramér inequality to understand the bias-variance trade-off phenomenon. This modification of the Rao-Cramér inequality would place restrictions regarding the variances of unbiased estimators on the supremum of variances in any Hellinger neighborhood of a point, instead of restricting the variance of the point only. (We believe our results in this paper validate the above suggestion.)

Low (1995), in the context of the functional estimation of finite and infinite normal populations, studies possible bias-variance trade-off by solving explicitly constraint optimization problems: imposing a constraint on either the variance or the square of the bias, then finding the procedure which minimizes the supremum of the unconstrained performance measure. This approach, due to mathematical difficulties involved, seems very difficult to carry out for general estimation problems. However, the investigation of the "bias-variance trade-off" phenomenon in the framework of the study of quantitative restrictions between bias and variance is interesting.

In this paper, we observe that the "nonexistence of informative unbiased estimators" phenomenon and the "bias-variance trade-off" phenomenon exemplify the mutual restrictions between mean functions and variance functions of estimators. These restrictions are described in our main inequality, Theorem 1. We are able to use this inequality to study, for finite sample cases, the "bias-variance trade-off" phenomenon and the "nonexistence of informative unbiased estimators" phenomenon for singular as well as regular estimation problems. A simple application of Theorem 1, Corollary 1, induces a sufficient condition for the "nonexistence of informative unbiased estimators" phenomenon. Corollary 1 is applicable to singular problems, (e.g. it implies Theorem 1 of Liu and Brown (1993)), as well as to regular problems (e.g. Example 2 of Section 4). Additional applications, such as Theorem 2 and Theorem 3, shed further light on the trade-off phenomenon by giving some quantified results. These results not only imply (and extend) Theorem 1 and Theorem 3 of Liu and Brown (1993), they also provide a general lower bound for constraint minimax performance. (See Corollary 3 and related comments.) We may summarize the idea conveyed by these results as: if the estimator we consider has variance less than the smallest possible variances for any unbiased estimators, then the range of the bias function is at least comparable to a fixed proportion of the range of the parameter function to be estimated.

We address the influence of a singular point as a limiting (parameter) point in Theorem 4. Although this is not a direct consequence of Theorem 1, the format of Theorem 1 facilitates results like Theorem 4.

We state our results in Section 2 and prove them in Section 3. In Section 4 we explain the meaning of Example 2 and Example 4 of Liu and Brown (1993) in our approach to the "bias-variance trade-off" phenomenon. We also argue that examples like Example 4 of Liu and Brown (1993) validate our version of the "mean-variance restriction," in which the restrictions imposing on the bias function of an estimator by its variance function are on the difference of biases at two points instead of the bias function at a point. Example 1 of Section 4, which has been considered by Low (1995) (and maybe others also), shows that our lower bound for minimax performance, Corollary 3, is sharp. The last example, Example 2, shows that the "nonexistence of informative unbiased estimator" phenomenon may occur even if the parameter space does not have any limiting point (with respect to Hellinger distance.)

2. Statements of results

We shall consider the following estimation problem. Let X be a random variable, which takes values in a measure space (Ω, μ), with distribution from a family of probability measures $\mathcal{F} = \{P_\theta : \theta \in \Theta\}$. Furthermore, it is assumed that every P_θ in $\mathcal{F}$ is dominated by the measure μ, and if $P_{\theta_1} = P_{\theta_2}$, then $\theta_1 = \theta_2$. For $\theta \in \Theta$, we denote the Radon-Nikodym derivative of P_θ with respect to the $\sigma-$measure μ as $f_\theta = dP_\theta/d\mu$. For $\theta_1, \theta_2 \in \Theta$, let

$$\rho(\theta_1, \theta_2) := \left\{ \int_\Omega \left[f_{\theta_1}(x)^{1/2} - f_{\theta_2}(x)^{1/2} \right]^2 \mu(dx) \right\}^{1/2} \tag{2.1}$$

denote the Hellinger distance between θ_1 and θ_2, on Θ, induced by the statistical model $\mathcal{F} = \{P_\theta : \theta \in \Theta\}$. Suppose $(V, \|\cdot\|)$ is a pseudo-normed linear space, and $q : \Theta \mapsto V$ is a function. We shall estimate $q(\theta)$ based on an observation X. The estimators $T : \Omega \mapsto \mathbf{V}$ we consider are well-behaved functions (satisfying the required measurability conditions) so that, for $\theta \in \Theta$

$$\psi_T(\theta) := \int_\Omega f_\theta(x) T(x) \mu(dx) \tag{2.2}$$

is meaningful and belongs to $\mathbf{V}$, and $v_T^*(\theta) := \int_\Omega f_\theta(x) \|T(x)\|^2 \mu(dx)$ is meaningful. We also adopt the following notaitons:

$$\beta_T(\theta) := E_\theta\big(T(x) - q(\theta)\big), \tag{2.3}$$

the bias function of T;

$$\gamma_T(\theta) := \left\{ E_\theta \|T(X) - q(\theta)\|^2 \right\}^{1/2} \tag{2.4}$$

the mean square risk function of T; and, for $\Theta_0 \subset \Theta$,

$$M_T(\Theta_0) := \sup \left\{ E_\theta \|(T(X) - q(\theta)\|^2 : \theta \in \Theta_0 \right\}. \tag{2.5}$$

The starting point of our study is the following inequality:

Theorem 1. *For $\theta, \theta_0 \in \Theta$, if $\rho(\theta, \theta_0) > 0$, then*

$$\begin{aligned}
&\big[\gamma_T(\theta) + \gamma_T(\theta_0)\big]\rho(\theta, \theta_0) \\
&\quad \geq \left\| \big(\beta_T(\theta) - \beta_T(\theta_0)\big) + \left(1 - \frac{1}{2}\rho^2(\theta, \theta_0)\right)\big(q(\theta) - q(\theta_0)\big) \right\| \\
&\quad \geq \left| \big\|\beta_T(\theta) - \beta_T(\theta_0)\big\| - \left(1 - \frac{1}{2}\rho^2(\theta, \theta_0)\right)\big\|q(\theta) - q(\theta_0)\big\| \right|. \qquad (2.6)
\end{aligned}$$

An easy consequence of (2.6) is:

Corollary 1. *Suppose Θ_1 is a non-empty subset of $\Theta - \{\theta_0\}$, then*

$$\begin{aligned}
&2 \sup\big\{\gamma_T(\theta) : \theta \in \Theta_1 \cup \{\theta_0\}\big\} \\
&\qquad + \sup\left\{ \frac{\|\beta_T(\theta) - \beta_T(\theta_0)\|}{\rho(\theta, \theta_0)} : \theta \in \Theta_1 \right\} \\
&\quad \geq \sup\left\{ \left[1 - \frac{1}{2}\rho^2(\theta, \theta_0)\right] \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)} : \theta \in \Theta_1 \right\}. \qquad (2.7)
\end{aligned}$$

Let us denote the value of the right-hand side of (2.7) as $Q_q(\theta_0; \Theta_1)$. We point out that the quantity $Q_q(\theta_0; \Theta_1)$ does not depend on the estimator T. It is easy to see that $Q_q(\theta_0; \Theta_1) = \infty$ is a sufficient condition for the "nonexistence of informative unbiased estimators" phenomenon. There are two ways to make $Q_q(\theta_0; \Theta_1) = \infty$: either $\inf_{\theta \in \Theta} \rho(\theta, \theta_0) > 0$ with $\sup_{\theta \in \Theta_1} [1 - \frac{1}{2}\rho^2(\theta, \theta_0)]\|q(\theta) - q(\theta_0)\| = \infty$ or $\inf_{\theta \in \Theta_1} \rho(\theta, \theta_0) = 0$ with $\limsup_{\rho(\theta, \theta_0) \to 0, \theta \in \Theta_1} \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)} = \infty$. See Example 2 of Section 4 for the first case and Examples 1 and 3 of Liu and Brown (1993) for the second case.

In the following, we focus on the case that θ_0 is a limit point of Θ_1 with respect to $\rho-$distance. Note that we may replace $Q_q(\theta_0; \Theta_1)$ in the right-hand side of (2.7) by an easily computable lower bound $\limsup_{\rho(\theta, \theta_0) \to 0, \theta \in \Theta_1} \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)}$.

For the convenience of our discussion let us introduce:

Definition 1 (Hellinger Information). Suppose $\Theta_1 \subset \Theta$ and θ_0 is a non-isolated point of Θ_1 with respect to $\rho-$metric on Θ. The Hellinger Information of θ_0 about the $q(\cdot)-$estimation problem and the (sub-)parameter space Θ_1 is defined as

$$J_q(\theta_0; \Theta_1) := 4 \left[\limsup_{\rho(\theta, \theta_0) \to 0+, \theta \in \Theta_1} \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)} \right]^{-2}. \qquad (2.8)$$

For the development of this notation and its relationship to Fisher Information, see Chen (1995). We mention here that this notation is related to "sensitivity" proposed by Pitman (1978). Also, it is equivalent to the "Geometric Information" in Donoho and Liu (1987), and, in terms of Hellinger modulus (see Liu and Brown (1993) (2.9) and (2.2)), it is $(\lim_{\varepsilon \to 0+} \frac{b(\varepsilon)}{\varepsilon})^{-2}$. When $J_q(\theta_0; \Theta) = 0$ (resp. > 0), we say that the $q(\cdot)-$estimation problem is singular (resp. regular) at point θ_0.

With the notation of Hellinger Information, an easy corollary of Theorem 1 is:

Corollary 2. *Suppose θ_0 is an accumulation point of $\Theta_0 \subset \Theta$. Then, for $J = J_q(\theta_0; \Theta_0)$*

$$2\left[M_T(\Theta_0)\right]^{1/2} + \sup\left\{ \frac{\|\beta_T(\theta) - \beta_T(\theta_0)\|}{\rho(\theta, \theta_0)} : \theta \in \Theta_0, \theta \neq \theta_0 \right\} \geq \frac{2}{\sqrt{J}}, \qquad (2.9)$$

or, equivalently,

$$\sup\left\{\frac{\|\beta_T(\theta)-\beta_T(\theta_0)\|}{\rho(\theta,\theta_0)}:\theta\in\Theta_0,\ \theta\neq\theta_0\right\}\geq\frac{2}{\sqrt{J}}\Big[1-\big(M_T(\Theta_0)J\big)^{1/2}\Big]. \quad (2.10)$$

A trivial implication of (2.10) is: if $M_T(\Theta_0) < 1/J_q(\theta_0;\Theta_0)$, then T is not unbiased on Θ_0. Moreover, (2.10) puts a restriction on the bias function $\beta_T(\theta)$ of T. We shall state this restriction more explicitly in the next theorem.

Theorem 2. *Suppose θ_0 is an accumulation point of $\Theta_0\subset\Theta$, and M is a positive number such that $M<[J_q(\theta_0;\Theta_0)]^{-1}$. Let $d_M := 1-(MJ_q(\theta_0;\Theta_0))^{1/2}$. Suppose that T is an estimator with $E_\theta\|T(X)-q(\theta)\|^2\leq M$ for all $\theta\in\Theta_0$. Then, for any $\lambda>0$, there exist $\theta_\lambda\in\Theta_0$, not dependent on T, such that $0<\rho(\theta_\lambda,\theta_0)\leq(2\lambda d_M)^{1/2}$, $\|q(\theta_\lambda)-q(\theta_0)\|>0$, and*

$$\big\|\beta_T(\theta_\lambda)-\beta_T(\theta_0)\big\|\geq(1-\lambda)d_M\cdot\big\|q(\theta_\lambda)-q(\theta_0)\big\|. \quad (2.11)$$

Applying Theorem 2, it is easy to obtain a lower bound for constrained minimax performance.

Corollary 3. *Let θ_0 be an accumulation point of Θ, $J=J_q(\theta_0;\Theta)>0$. Let M and τ be positive numbers, and*

$$B(M;\tau):=\inf_T\sup_\theta\big\{\|\beta_T(\theta)-\beta_T(\theta_0)\|^2\big\}$$

where θ is over $\|q(\theta)-q(\theta_0)\|\leq\tau$ and T is over $E_\theta\|T(X)-q(\theta)\|^2\leq M$. Then

$$B(M;\tau)\geq\Big\{\big[1-(MJ)^{1/2}\big]\wedge 0\Big\}^2\tau^2. \quad (2.12)$$

In the restriction normal mean case (see Example 1), the lower bound (2.12) is sharp.

Now, let us turn to the case in which θ_0 is a singular point, i.e., $J_q(\theta_0;\Theta_0)=0$. From (2.9) or (2.10), we have either $M_T(\Theta_0)=\infty$ or $\sup\{\frac{\|\beta_T(\theta)-\beta_T(\theta_0)\|}{\rho(\theta,\theta_0)}:\theta\in\Theta_0,\ \theta\neq\theta_0\}=\infty$. This implies the non-existence of an informative unbiased estimator for such Θ_0. Therefore, Theorem 1 of Liu and Brown (1993) is a weaker version of Corollary 2.

From Theorem 2 (or Corollary 3), it is easy to see that there exists no sequence of asymptotically unbiased estimators (based on the same finite number of observations) that would have uniformly bounded variance in any small Hellinger neighborhood of a singular point θ_0. Hence, Theorem 2 above implies Theorem 3 of Liu and Brown (1993). For singular estimation problems, those estimators achieving good mean square error performance must balance bias and variance, and (2.11) gives a quantitative result about its bias function $\beta_T(\theta)$. Furthermore, we are able to describe the "rate" of $\|\beta_T(\theta)-\beta_T(\theta_0)\|$ as follows.

Theorem 3. *Suppose $J_q(\theta_0;\Theta_0)=0$. Let $\Theta_1=\{\theta_1,\theta_2,\ldots\}\subset\Theta_0-\{\theta_0\}$ be a slow sequence of θ_0 in the sense that $\lim_{j\to\infty}\rho(\theta_j,\theta_0)=0$ and $\lim_{j\to\infty}\frac{\|q(\theta_j)-q(\theta_0)\|}{\rho(\theta_j,\theta_0)}=\infty$. If T is an estimator with $\sup\{E_{\theta_j}\|T(X)-q(\theta_j)\|^2:j=0,1,2,\ldots\}<\infty$, then*

$$\lim_{j\to\infty}\frac{\|\beta_T(\theta_j)-\beta_T(\theta_0)\|}{\|q(\theta_j)-q(\theta_0)\|}=1. \quad (2.13)$$

One of the important observations of Liu and Brown (1993) is that the bias-variance trade-off phenomenon might occur on a set Θ_1 due to the effect of a singular point θ_0 as a limit point of Θ_1. The next result states it more explicitly.

Theorem 4. *Suppose* $\mathbf{V}$ *is a subspace of* $d-$*dimensional Euclidean space* $\mathbf{R}^d$ *with the usual Euclidean norm* $\|\cdot\|$. *Let* θ_0 *be a singular point,* $\Theta_1 = \{\theta_1, \ldots\} \subset \Theta - \{\theta_0\}$ *be a slow sequence of* θ_0 *and* T *be an unbiased estimator on* Θ_1. *Then,* $\sup\{E_\theta \|T(X) - q(\theta)\|^2 : \theta \in \Theta_1\} = \infty$.

3. Proofs

Theorem 1 is a simple application of the following inequality.

Lemma 1. *For points* $\eta_1, \eta_2 \in \mathbf{V}; \theta_1, \theta_2 \in \Theta$ *with* $\rho(\theta_1, \theta_2) > 0$, *we have*

$$\begin{aligned} &\Big\{E_{\theta_1}\big\|T(X) - \eta_1\big\|^2\Big\}^{1/2} + \Big\{E_{\theta_2}\big\|T(X) - \eta_2\big\|^2\Big\}^{1/2} \\ &\qquad \geq \left\|\psi_T(\theta_1) - \psi_T(\theta_2) - \frac{1}{2}\rho^2(\theta_1, \theta_2)(\eta_1 - \eta_2)\right\| \Big/ \rho(\theta_1, \theta_2). \end{aligned} \tag{3.1}$$

Proof. Without loss of generality, we assume that $E_{\theta_i}\|T(X) - \eta_i\|^2 < \infty$ for $i = 1, 2$. Define $\alpha_i(x) = f_{\theta_i}(x)^{1/2}(T(x) - \eta_i)$ for $i = 1, 2$, and $\beta(x) = f_{\theta_1}(x)^{1/2} - f_{\theta_2}(x)^{1/2}$. Then

$$\begin{aligned} &\int_\Omega \beta(x)\big[\alpha_1(x) + \alpha_2(x)\big]\mu(dx) \\ &\quad = \int_\Omega \Big[f_{\theta_1}(x)\big(T(x) - \eta_1\big) - f_{\theta_2}(x)\big(T(x) - \eta_2\big) \\ &\qquad\qquad + \big[f_{\theta_1}(x) f_{\theta_2}(x)\big]^{1/2}(\eta_1 - \eta_2)\Big]\mu(dx) \\ &\quad = E_{\theta_1}\big(T(X) - \eta_1\big) - E_{\theta_2}\big(T(X) - \eta_2\big) \\ &\qquad + \int_\Omega \big[f_{\theta_1}(x) f_{\theta_2}(x)\big]^{1/2}\mu(dx)(\eta_1 - \eta_2) \\ &\quad = \big(\psi_T(\theta_1) - \eta_1\big) - \big(\psi_T(\theta_2) - \eta_2\big) + \left[1 - \frac{1}{2}\rho^2(\theta_1, \theta_2)\right](\eta_1 - \eta_2) \\ &\quad = \big(\psi_T(\theta_1) - \psi_T(\theta_2)\big) - \frac{1}{2}\rho^2(\theta_1, \theta_2)(\eta_1 - \eta_2). \end{aligned} \tag{3.2}$$

On the other hand, by the triangle inequality and the Cauchy–Schwarz inequality,

$$\begin{aligned} &\left\|\int_\Omega \beta(x)\big[\alpha_1(x) + \alpha_2(x)\big]\mu(dx)\right\| \\ &\qquad \leq \sum_{i=1}^{2}\left\|\int_\Omega \beta(x)\alpha_i(x)\mu(dx)\right\| \\ &\qquad \leq \sum_{i=1}^{2}\int_\Omega |\beta(x)|\big\|\alpha_i(x)\big\|\mu(dx) \end{aligned}$$

$$\leq \sum_{i=1}^{2} \left[\int_\Omega \beta^2(x)\mu(dx) \right]^{1/2} \cdot \left[\int_\Omega \left\|\alpha_i(x)\right\|^2 \mu(dx) \right]^{1/2}$$
$$= \rho(\theta_1, \theta_2) \sum_{i=1}^{2} \left[E_{\theta_i} \left\| T(X) - \eta_i \right\|^2 \right]^{1/2}. \tag{3.3}$$

Combining (3.2) and (3.3), we obtain (3.1). □

Proof of Theorem 1. Applying Lemma 1, we have

$$\begin{aligned}
&\big[\gamma_T(\theta) + \gamma_T(\theta_0)\big]\rho(\theta, \theta_0) \\
&\quad \geq \left\| \psi_T(\theta) - \psi_T(\theta_0) - \frac{1}{2}\rho^2(\theta, \theta_0)\big(q(\theta) - q(\theta_0)\big) \right\| \\
&\quad = \left\| \beta_T(\theta) - \beta_T(\theta_0) + \left(1 - \frac{1}{2}\rho^2(\theta, \theta_0)\right)\big(q(\theta) - q(\theta_0)\big) \right\|,
\end{aligned} \tag{3.4}$$

this proves the first inequality of (2.6).

Applying the triangle inequality and the fact that $1 - \frac{1}{2}\rho^2(\theta, \theta_0) \geq 0$, we obtain the second inequality of (2.6). □

Proof of Corollary 1. Notice that (2.6) implies

$$\begin{aligned}
&2\max\big(\gamma_T(\theta), \gamma_T(\theta_0)\big) + \frac{\|\beta_T(\theta) - \beta_T(\theta_0)\|}{\rho(\theta, \theta_0)} \\
&\quad \geq \left[1 - \frac{1}{2}\rho^2(\theta, \theta_0)\right] \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)}.
\end{aligned} \tag{3.5}$$

Letting θ vary over Θ_1 in inequality (3.5), we obtain (2.7). □

Proof of Corollary 2. It is easy to prove that

$$\begin{aligned}
Q_q(\theta_0; \Theta_0) \quad &\geq \quad \limsup_{\rho(\theta,\theta_0)\to 0, \theta\in\Theta_0} \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)} \\
&= \quad \left[\frac{1}{4} J_q(\theta_0; \Theta_0)\right]^{-1/2}.
\end{aligned} \tag{3.6}$$

This, together with Corollary 1, proves (2.9). □

Proof of Theorem 2. We use J to replace $J_q(\theta_0; \Theta_0)$ in this proof.

Applying Theorem 1 and the condition $\gamma_T(\theta) + \gamma_T(\theta_0) \leq 2M^{1/2}$, we have, for all $\theta \in \Theta_0$, that

$$\frac{\|\beta_T(\theta) - \beta_T(\theta_0)\|}{\rho(\theta, \theta_0)} \geq \left[1 - \frac{1}{2}\rho^2(\theta, \theta_0)\right] \frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)} - 2M^{1/2} \tag{3.7}$$

and

$$\begin{aligned}
&\frac{\|\beta_T(\theta) - \beta_T(\theta_0)\|}{\|q(\theta) - q(\theta_0)\|} \\
&\quad \geq 1 - \frac{1}{2}\rho^2(\theta, \theta_0) - 2M^{1/2} \cdot \left[\frac{\|q(\theta) - q(\theta_0)\|}{\rho(\theta, \theta_0)}\right]^{-1}.
\end{aligned} \tag{3.8}$$

For $\varepsilon > 0$, let $\Theta_0(\varepsilon) := \{\theta : q(\theta) \neq q(\theta_0), 0 < \rho(\theta, \theta_0) \leq \varepsilon\} \cap \Theta_0$. By (3.8), for $\varepsilon = (2\lambda d_M)^{1/2}$, we have

$$\begin{aligned}\sup_{\theta\in\Theta_0(\varepsilon)} \frac{\|\beta_T(\theta) - \beta_T(\theta_0)\|}{\|q(\theta) - q(\theta_0)\|} &\geq 1 - \frac{1}{2}\varepsilon^2 - 2M^{1/2} \cdot \left(\frac{1}{4}J\right)^{1/2} \\ &= (1-\lambda)d_M. \end{aligned} \tag{3.9}$$

This proves Theorem 2. □

Proof of Corollary 3. Let $d_M := 1 - (MJ)^{1/2}$. For the case $d_M \leq 0$, we use the trivial inequality $B(M;\tau) \geq 0$ and for the case $d_M > 0$, we use Theorem 2 to obtain $B(M;\tau) \geq (d_M\tau)^2$. This proves (2.12). □

Proof of Theorem 3. Let M be a positive number such that $E_{\theta_j}\|T(X) - q(\theta_j)\|^2 \leq M^2$ for $j = 0, 1, 2, \ldots$. Then, by (2.6),

$$2M \geq \frac{\|q(\theta_j) - q(\theta_0)\|}{\rho(\theta_j, \theta_0)} \left| \frac{\|\beta_T(\theta_j) - \beta_T(\theta_0)\|}{\|q(\theta_j) - q(\theta_0)\|} - \left(1 - \frac{1}{2}\rho^2(\theta_j, \theta_0)\right)\right|. \tag{3.10}$$

Hence,

$$\begin{aligned}&1 + \rho^2(\theta_j, \theta_0)/2 - 2M\big[\big\|q(\theta_j) - q(\theta_0)\big\|/\rho(\theta_j, \theta_0)\big]^{-1} \\ &\quad \leq \big\|\beta(\theta_j) - \beta(\theta_0)\big\|/\big\|q(\theta_j) - q(\theta_0)\big\| \\ &\quad \leq 1 + \rho^2(\theta_j, \theta_0)/2 + 2M\big[\big\|q(\theta_j) - q(\theta_0)\big\|/\rho(\theta_j, \theta_0)\big]^{-1}. \end{aligned} \tag{3.11}$$

Let $j \to \infty$, we have the desired (2.13). □

In order to prove Theorem 4, we need the following lemma.

Lemma 2. *Suppose* $\mathbf{V}$ *is a subspace of d-dimensional Euclidean space* $\mathbf{R}^d$ *with the usual Euclidean norm* $\|\cdot\|$. *Let* $\Theta_1 = \{\theta_1, \ldots\} \subset \Theta - \{\theta_0\}$ *be a sequence with limit point* θ_0 *and* $\lim_{j\to\infty} q(\theta_j) = q(\theta_0)$. *Then, for any estimator* T,

$$E_{\theta_0}\big\|T(X) - q(\theta_0)\big\|^2 \leq M_T(\Theta_1). \tag{3.12}$$

Proof. (3.12) is automatically true if $M_T(\Theta_1) = \infty$. Let us consider the case that $M_T(\Theta_1) < \infty$. Since $\lim_{j\to\infty} \rho(\theta_j, \theta_0) = 0$ as $j \to \infty$, the distribution of T under $\theta = \theta_j$ converges to the distribution of T under $\theta = \theta_0$. Let us write

$$\begin{aligned} T &= (T_1, T_2, \ldots, T_d), \\ q(\theta) &= \big(q_1(\theta), q_2(\theta), \ldots, q_d(\theta)\big), \\ \psi_T(\theta) &= E_\theta T(X) = \big(\psi_1(\theta), \psi_2(\theta), \ldots, \psi_d(\theta)\big), \qquad \text{and} \\ \nu_T(\theta) &= \big(\mathrm{var}_\theta(T_1), \mathrm{var}_\theta(T_2), \ldots, \mathrm{var}_\theta(T_d)\big). \end{aligned}$$

Notice that

$$\begin{aligned}&E_\theta\big\|T(X) - q(\theta)\big\|^2 \\ &\quad = E_\theta\big\|T(X) - \psi_T(\theta)\big\|^2 + \big\|\psi_T(\theta) - q(\theta)\big\|^2 \\ &\quad = \sum_{i=1}^d \mathrm{var}_\theta(T_i) + \sum_{i=1}^d \big(\psi_i(\theta) - q_i(\theta)\big)^2, \end{aligned} \tag{3.13}$$

and, since $M_T(\Theta_1) < \infty$,

$$\lim_{j\to\infty} \psi_i(\theta_j) = \psi_i(\theta_0) \qquad \text{for } i = 1, 2, \ldots, d. \tag{3.14}$$

By Problem 4.4.9, page 150 of Bickel and Doksum (1977), we have

$$\liminf_j \operatorname{var}_{\theta_j}(T_i) \geq \operatorname{var}_{\theta_0}(T_i) \qquad \text{for } i = 1, 2, \ldots, d. \tag{3.15}$$

With the assumption $\lim_{j\to\infty} q(\theta_j) = q(\theta_0)$, and (3.13) $\sim$ (3.15), we have

$$\liminf_j E_{\theta_j} \big\| T(X) - q(\theta_j) \big\|^2 \geq E_{\theta_0} \big\| T(X) - q(\theta_0) \big\|^2. \tag{3.16}$$

This proves (3.12). □

Proof of Theorem 4. First, if $\lim_{j\to\infty} q(\theta_j) = \infty$, then it is easy to prove that $\sup\{E_\theta \|T(X) - q(\theta)\|^2 : \theta \in \Theta_1\} = \infty$. Next, if $\lim_{j\to\infty} q(\theta_j)$ exists, we simply change the definition of $q(\theta_0)$ to be equal to $\lim_{j\to\infty} q(\theta_j)$. Under this new definition of q, θ_0 is still a singular point and Θ_1 is still a slow sequence of θ_0. If $M_T(\Theta_1) < \infty$, then (3.12) implies $M_T(\Theta_1 \cup \{\theta_0\}) = M_T(\Theta_1) < \infty$ and (2.9) implies $2[M_T(\theta_1 \cup \{\theta_0\})] = \infty$, a contradiction. This proves $M_T(\Theta_1) = \infty$. □

4. Comments and examples

Example 2 of Liu and Brown (1993) shows that the "nonexistence of informative unbiased estimators" phenomenon might occur in a quadratic-mean-differentiable (QMD) problem with Fisher Information totally bounded away from zero. This statement is true if we replace the term "Fisher Information" by "Hellinger Information" since it is well-known that Fisher Information and Hellinger Information are equal in QMD problems. Due to the fact that the Hellinger Information number $J(\theta)$ is not necessarily continuous with respect to the Hellinger distance $\rho(\theta;\theta_0)$, the condition that Fisher Information (or Hellinger Information) be totally bounded away from zero does not exclude the possibility of a singular point as a limiting point. If such a singular limiting point exists, by Theorem 4, the "nonexistence of informative unbiased estimators" phenomenon could occur.

Example 4 of Liu and Brown (1994) exhibits an unbiased estimator with finite variance at a singular point. The spirit of this example does not contradict the impression left by the "mean-variance restriction" described in Theorem 1 or Corollary 1. Obviously, one can modify an estimator so as to obtain an unbiased estimator at any predescribed point. The requirement that an estimator have finite variance at a predescribed point does not pose any conflict because the "mean-variance restriction" (Theorem 1) places a lower bound on the *sum* of variances at two points, instead of on variances at each point. Further, one could even view this example as a validation of the form of "mean-variance restriction" (Theorem 1), in which the restriction imposed by sums of variances (or, rather, sums of root mean-square risks) is on the difference of the bias function $(\beta_T(\theta) - \beta_T(\theta_0))$ and not on the bias function $(\beta_T(\theta_0))$ itself.

The following example shows that in the bounded normal case, the lower bound of Corollary 3 is sharp. This example has been considered by Low (1995).

Example 1. If $X \sim N(\theta, \sigma^2)$ and $q(\theta) = \theta$, then $J = J_q(\theta_0; \Theta) = \frac{1}{\sigma^2}$ for any open interval Θ which contains θ_0. By (2.9),

$$B(M;\tau) \geq \big\{ \big[1 - M^{1/2} \cdot \sigma^{-1}\big] \wedge 0 \big\}^2 \tau^2. \tag{4.1}$$

Let T_M be the affine procedure studied in Low (1995), (2.4),

$$T_M(X) = \big(M^{1/2} \cdot \sigma^{-1} \wedge 1\big)(X - \theta_0) + \theta_0. \tag{4.2}$$

It is easy to show that $E_\theta \|T_M(X) - \theta\|^2 \le M \wedge \sigma^2$ and that

$$\sup\big\{\big\|\beta_{T_M}(\theta) - \beta_{T_M}(\theta_0)\big\|^2 : |\theta - \theta_0| \le \tau\big\} = \big\{\big[1 - M^{1/2} \cdot \sigma^{-1}\big] \wedge 0\big\}^2 \tau^2.$$

This, together with (2.12) proves

$$B(M;\tau) = \big\{\big[1 - M^{1/2} \cdot \sigma^{-1}\big] \wedge 0\big\}^2 \tau^2. \tag{4.3}$$

If we compare $B(M;\tau)$ with $\beta^2(\nu,\sigma,\tau)$ in (2.1) and (2.3) of Low (1995), we find that $B(M;\tau) = \beta^2(M,\sigma,\tau)$ in the above Example 2. It is interesting to point out that Low's argument to obtain a lower bound on $\beta^2(\nu,\sigma,\tau)$ is an application of the Rao-Cramér Inequality. This approach, if extended to a general case, would require conditions to guarantee the differentiability of the bias function of T. Our method, which is based on Theorem 1, does not require the differentiability of the bias function of T.

Finally, let us exhibit an example of the "nonexistence of informative unbiased estimator" phenomenon for discrete Θ without any limiting point with respect to ρ-distance.

Example 2. Let $X \sim$ Poisson (θ) with $\theta \in N = \{1, 2, 3, \ldots\}$, and $r > 1$. Suppose we want to estimate $q(\theta) = e^{r\theta}$. The square of Hellinger distance is

$$\begin{aligned} \rho^2(\theta, 1) &= 2 - 2 \sum_{x=0}^{\infty} \left\{ \frac{e^{-\theta} \cdot \theta^x}{x!} \cdot \frac{e^{-1} \cdot 1^x}{x!} \right\}^{1/2} \\ &= 2 - 2 \exp\left\{ -\frac{1}{2} \left(\sqrt{\theta} - 1\right)^2 \right\}. \end{aligned}$$

It is easy to verify that $[1 - \frac{1}{2}\rho^2(\theta, 1)]\|q(\theta) - q(1)\|/\rho(\theta, 1) \to \infty$ for $\theta \to \infty$. According to Corollary 1, there exists no informative unbiased estimator for $q(\theta)$.

Acknowledgements

The author wishes to thank Professors Larry Brown and Richard Liu, for letting him read their work (Liu and Brown (1993)) prior to publication. Thanks to Professor Siva Sivaganesan and Ms. Clarissa Howison for their suggestions on the presentation of this paper.

References

[1] Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics, Basic Ideas and Selected Topics.* Prentice Hall, Englewood Cliffs, New Jersey. MR443141

[2] Chen, J. (1995). A Hellinger Distance Inequality and its Applications to Variance-Mean Relationship in Estimation Problems. Unpublished.

[3] Donoho, D. L. and Liu, R. C. (1987). Geometrizing rates of convergence I. Technical Report, Department of Statistics, University of California, Berkeley.

[4] Doss, H. and Sethuraman, J. (1989). The price of bias reduction when there is no unbiased estimate. *Ann. Statist.* **17**, 440–442. MR981462

[5] Liu, R. C. and Brown, L. D. (1993). Nonexistence of informative unbiased estimators in singular problems. *Ann. Statist.* **Vol. 21**, 1–13. MR1212163

[6] Low, M. (1995). Bias-Variance Tradeoffs in Functional Estimation Problems. *Ann. Statist.* **Vol. 23**, 824–835. MR1345202