The Annals of Statistics
1993, Vol. 21, No. 1, 1–13

NONEXISTENCE OF INFORMATIVE UNBIASED ESTIMATORS IN SINGULAR PROBLEMS

BY RICHARD C. LIU[1] AND LAWRENCE D. BROWN[2]

Cornell University

In many nonparametric problems, such as density estimation, nonparametric regression and so on, all the existing informative estimators are biased (asymptotic or finite sample). There has long been a suspicion that either informative unbiased estimators do not exist for such problems or they must be quite complicated. In this paper, we clarify the nonexistence of informative unbiased estimators in all singular problems both for fixed sample size and asymptotically (this includes most problems with optimal rate of convergence slower than $n^{-1/2}$). We also discuss situations in regular problems where such nonexistences can occur.

1. Introduction. Unbiasedness has always been one of the most popular criteria for an estimator to be reasonably good in many studies (either in an asymptotic sense or in the finite sample case). Most problems in the classical theory—later called *regular problems* in parametric, semiparametric or nonparametric contexts—have been equipped with enough regularity implicitly or explicitly [e.g., a parametric family $\{P_\theta\}$ which is quadratic mean differentiable and so on], so that certain magic phenomena happen over and over in those problems. For example, the optimal rate of convergence for such problems happens to be the magic sequence $\{n^{-1/2}\}$ ([3], [9]); informative unbiased estimates frequently exist and they can also be asymptotically optimal under mild conditions. Hence, notable contributions in finding better/best estimates among those estimators, respectively, such as uniformly minimum variance unbiased estimator (UMVUE) ([2], [11], [12]) and MLE ([9]), are also well established.

In the past decade, the study of useful singular (for a definition see Section 2) nonparametric problems has become more extensive (e.g., density estimation, nonparametric regression). Remarkable progress has been made in the area of optimal rates of convergence; it was initially a surprise that the magic sequence $\{n^{-1/2}\}$ is mostly no longer available ([7], [8], [20], [23]). In spite of years of searching for informative estimators in an asymptotic sense, it appears that all existing informative estimators are biased (asymptotically, as well as in finite samples). Therefore, many interesting phenomena seem to be quite different from those appearing in regular problems. The question to ask now is "*Are informative unbiased estimators also possible for singular*

Received October 1990; revised March 1992.
[1]Partially supported by NSF Postdoctoral Fellowships and by Cornell MSI.
[2]Partially supported by NSF Grant DMS-88-09016.
AMS 1991 *subject classifications.* Primary 62F11; secondary 62F12, 62G05, 62A99.
Key words and phrases. Unbiasedness, modulus of continuity, Hellinger distance, singular problems.

problems?" In this paper, we will show the nonexistence of informative unbiased estimators for such problems. Moreover, similar phenomena in regular problems will also be addressed. This provides evidence that "*bias–variance trade-off*" is an *essential component* of estimation for singular problems. A special case for a similar phenomenon was demonstrated earlier by Doss and Sethuramon (1989). (See also Section 4.2.)

2. Singular problems. Let $\mathbf{F} \equiv \{F\}$ be a family of distributions. Let $T(F)$ be a $\mathbf{V}$-valued functional of interest in the estimation problem (where $\mathbf{V}$ is a normed linear space with a norm $\|\cdot\|_{\mathbf{V}}$). Even though most of the results continue to hold for pseudonorms, there is no loss of generality in practice to assume that $\|\cdot\|_{\mathbf{V}}$ is a norm. For example, if we are interested in estimating a parameter θ from some parametric family $\{P_\theta\}$; then the functional of interest is $T(P_\theta) = \theta \in \mathbf{R}^k$; or if we are interested in estimating the true density value at a point [say $f(x_0)$] when $f = dF/dx$, then the functional of interest is $T(F) = f(x_0) \in \mathbf{R}$. Let $[R(T)]$ denote the smallest linear subspace containing the range space of the functional T. In this paper, we will focus our attention on those cases when $\dim[R(T)]$ is finite. Even though the following approach can explain some of the phenomena in other cases, cases when $\dim[R(T)]$ is infinite will be discussed and characterized more precisely in a forthcoming paper in a different way. Now the estimation problem is to estimate the true functional value $T(F)$ (through an estimator T_n in a measurable way w.r.t. the Borel σ-field on $\mathbf{V}$) from n i.i.d. observations $X_1, \dots, X_n$ generated by some $F \in \mathbf{F}$. For the convenience of our discussion in all the sections, let

$$b(\varepsilon, F_0, T, \mathbf{F}) = \sup\{\|T(F_1) - T(F_0)\|_{\mathbf{V}}: F_1 \in \mathbf{F},\ H(F_1, F_0) \le \varepsilon\} \tag{2.1}$$

be the Hellinger modulus of continuity for the functional T over the family $\mathbf{F}$, where $H(F, G)$ is the Hellinger distance between two probabilities F, G (cf. [3]). We will abuse the notation later by using $b(\varepsilon)$ whenever there is no confusion.

DEFINITION 2.1 (Singular point/singular problem). In a functional estimation problem (defined above), a distribution $F_0 \in \mathbf{F}$ is called a "*singular point*" if

$$\lim_{\varepsilon \to 0^+} \frac{b(\varepsilon)}{\varepsilon} = \infty. \tag{2.2}$$

An estimation problem is called "*singular*" if there is a singular point $F_0 \in \mathbf{F}$.

2.1. *Finite sample property.* It is known in the literature that an estimator T_n is unbiased at $F \in \mathbf{F}$, if $E_F(T_n) = T(F)$. For our convenience, we adapt the following terms:

DEFINITION 2.2 (Local unbiasedness). An estimator T_n for estimating $T(F)$ is "*locally unbiased*" (*l-unbiased*) *at a distribution* $F_0 \in \mathbf{F}$ if and only if it is unbiased over a neighborhood of F_0.

DEFINITION 2.3 (Local informative estimator in finite sample). An estimator T_n is "*locally informative* (*l-informative*) *at a distribution*" F_0 in the finite sample case if there is a positive number M_{F_0} such that

$$\limsup_{\varepsilon \to 0} \sup_{\substack{H(F, F_0) \le \varepsilon \\ F, F_0 \in \mathbf{F}}} E_F\left(\|T_n\|_{\mathbf{V}}^2\right) \le M_{F_0}. \tag{2.3}$$

Moreover, T_n is said to be "*l-informative*" if it is l-informative at every $F \in \mathbf{F}$. Furthermore, an estimator T_n is said to be "*globally informative*" (*g-informative*) if it is l-informative with $\sup_{F \in \mathbf{F}} M_F < \infty$.

NOTE 1. In fact, all the above definitions are independent of the choice of the norm $\|\cdot\|_{\mathbf{V}}$. Similar effects happen in all the other definitions of this paper.

NOTE 2. Since "unbiasedness" (over the entire family) implies "l-unbiasedness" and so on, *results in our context* will be stronger in terms of "l-unbiasedness"/"l-informativity" than those in terms of "unbiasedness"/"g-informativity" and so forth.

THEOREM 1. *In a singular estimation problem*, no *estimate* T_n *can be both l-unbiased and l-informative at* any *singular point.*

In a wide class of interesting singular problems, such as nonparametric density estimations, quite a few informative estimators have continuous second moment with respect to the topology induced by Hellinger distance (e.g., many kernel estimators). Thus, (2.3) is automatically satisfied. Consequently, the following corollary will have a wide range of applications.

COROLLARY 1. *If an estimator T_n has continuous second moment, then T_n is l-biased at every singular point F_0 in the sense of Definition* 2.2.

2.2. *Asymptotic property.* In respect to some facts in asymptotics, Donoho and Liu [3] established that $\{b(n^{-1/2})\}$ defines a lower bound for the optimal rates of convergence in estimating a functional $T(F) \in \mathbf{R}$. The same argument implies that it is also true in our settings. As an immediate consequence, (2.2) implies the optimal rate for estimating $T(F)$ over $\mathbf{F}$ is slower than $\{n^{-1/2}\}$. Second, it also implies that any attainable modulus rate $\{b(n^{-1/2})\}$ is optimal. As to this effect, it is known that modulus rates are frequently attainable and thus optimal in many classes of interesting problems. For example, in [4], it is shown that when the functional is linear and the family is convex, a modulus rate can be attainable under mild conditions. Other situations where the $\{b(n^{-1/2})\}$ rate can be attained are also discussed there (e.g., estimating the mode of a density). Hence, modulus rates are commonly existing and desirable. (For definitions of rates of convergence, see e.g., [3], [15], [23].)

DEFINITION 2.4 (Asymptotic unbiasedness). Let $\{f(n)\}$ be a rate for the estimator T_n. Let $N(r, F_0)$ denote the Hellinger ball around F_0 with radius r. The estimator T_n is said to be "*locally asymptotically unbiased*" (*l-a-unbiased*) at a distribution $F_0 \in \mathbf{F}$ if there is a $t > 0$ such that

$$(2.4)\quad \lim_{M\to\infty} \limsup_{n\to\infty} \sup_{F_n^t \in N(t/\sqrt{n}, F_0)} \left\| E_{F_n^t} l_M\left[f^{-1}(n)\left(T_n - T(F_n^t)\right)\right] \right\|_{\mathbf{V}} = 0,$$

where $l_M(v) = v$ if $\|v\|_{\mathbf{V}} < M$ or otherwise equals to $t \cdot v$ with $t = M/\|v\|_{\mathbf{V}}$. Moreover, T_n is said to be "*asymptotically unbiased*" if it is locally asymptotically unbiased at every distribution $F \in \mathbf{F}$.

DEFINITION 2.5 (l-asymptotically informative estimator). Let T_n be an estimator for estimating $T(F)$. Let $\{f(n)\}$ be a rate of convergence of T_n. Then, T_n is said to be "*locally asymptotically informative* (*l-a-informative*) *at a distribution*" $F_0 \in \mathbf{F}$ with the same rate if there is a number M_{F_0} and a positive number t such that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{F_n^t \in N(t/\sqrt{n}, F_0)} E_{F_n^t}\left\{ \left\| l_M\left[f^{-1}(n)\left(T_n - T(F_n^t)\right)\right] \right\|_{\mathbf{V}}^2 \right\} \le M_{F_0}.$$

The estimator T_n is said to be "*l-a-informative*" if it is l-a-informative at every point $F \in \mathbf{F}$.

DEFINITION 2.6 (Infinitesimal irregularity). $b(\varepsilon)$ is *irregular infinitesimally* if

$$(2.5)\quad \lim_{n\to\infty} \frac{b(an^{-1/2})}{b(n^{-1/2})} = h(a) \to 0 \quad \text{as } a \to 0^+ \quad \text{and} \quad \limsup_{a\to 0^+} \frac{h(a)}{a} = \infty,$$

where $h(a) > 0$ for $a > 0$, is a real valued function.

NOTE 1. The above definition is satisfied by all $b(\varepsilon)$ Hölderian with leading exponent less than 1 [i.e., $b(\varepsilon) = A\varepsilon^q + o(\varepsilon^q)$ and $0 < q < 1$].

NOTE 2. Most nonparametric problems in the literature are Hölderian with $q < 1$, and modulus rates are often attainable.

NOTE 3. Infinitesimal irregularity cannot be dropped [e.g., some nonparametric estimation problems with modulus rates $n^{-1/2}(\log(n))^{1/2}$ (cf. [13], [16]), which are singular but infinitesimally regular, can have l-a-unbiased and l-a-informative estimators with their modulus rates. We will not pursue this further here].

THEOREM 2. *In a singular problem, if $b(\varepsilon)$ is irregular infinitesimally at some singular point F_0, then there is no l-a-unbiased, and l-a-informative estimator with the modulus rate $\{b(n^{-1/2})\}$ at that singular point.*

3. Regular problems. It is worth noticing that, even though globally informative and unbiased estimators often exist for regular problems, the phenomenon of "nonexistence of informative unbiased estimators" can still occur in some nice regular problems. Many such examples share the same characteristic that they have some singular points in their closure. Therefore, the effect of a singular point (even as a limit point but not really in the family) cannot be easily ignored. In this section, we will explore such an effect through the following examples.

EXAMPLE 1 (Poisson case). Consider the family $\mathbf{F} = \{P_\lambda\text{: Poisson}(\lambda) \text{ with } \lambda > 0\}$. Let the functional of interest be $T(P_\lambda) = 1/\lambda$. This example has been considered in [6]. Then, by reparametrization, this is equivalent to the problem (denoted by Ξ) of estimating $T(F_\theta) = \theta$ from the parametric family $\mathbf{G} = \{F_\theta = \text{Poisson}(\lambda)\text{: } \theta = 1/\lambda,\ \theta > 0\}$. It is clear that Fisher information at each θ exists and is equal to $1/\theta^3$, hence it is regular. However, the Fisher information tends to zero as θ tends to infinity. In fact, by considering the extended parametric problem (denoted by Ψ) with

$$(3.1)\quad \tilde{\mathbf{G}} = \mathbf{G} \cup \{F_\infty = \text{Poisson}(0)\} \quad \text{and} \quad T(F_\infty) = c \quad (\text{some constant}),$$

the problem becomes singular because $b(\varepsilon) = \infty$ at F_∞. Therefore, problem Ξ can be obtained by deleting a singular point from problem Ψ. Moreover, it is also well known that problem Ξ has a similar "behavior" to the binomial problem (see Lehmann's book [12]); namely, there is no unbiased estimator for such problems. (Note: This behavior does not necessarily imply that the closure of the problem under Hellinger topology is singular.) Furthermore, by Theorem 2, there can be no l-a-informative and l-a-unbiased estimator in a neighborhood of $\theta = \infty$, even though such estimators do exist on $[0, b]$ for $b < \infty$.

However, the fact that Fisher information degenerates to zero (as in Example 1) is not a necessary characteristic. One interesting question to ask here is "*Can anything go wrong in a quadratic-mean-differentiable problem with Fisher informations totally bounded away from zero over the entire family*?" Even for such problems, the effect of a singular point as a limit point is still unavoidable. See Example 2.

EXAMPLE 2 (QMD family with Fisher informations totally bounded away from zero). Let us consider the parametric problem with the family of distributions as

$$\mathbf{F}_\Omega = \left\{F_\omega\text{: } \frac{dF_\omega}{d\lambda}(x) \equiv f_\omega(x) = (1 + \omega^2 \sin(x/\omega)),\ \omega \in \Omega\right\},$$

where λ denotes the Lebesgue measure, $|x| \le 1/2$, and $\Omega = \{\omega\text{: } 0 < \omega < a < 1\}$ (where a will be chosen later). It can be verified directly that every f_ω is a density. Now, since the Fisher information I_{F_ω} exists and is continuous in a small open interval $\Omega = \{\omega\text{: } 0 < \omega < a\}$, a straightforward calculation shows that $I_F(\omega)$ tends to $1/24$ as $\omega \to 0$. Thus, by choosing "a" sufficiently small, the Fisher informations of the entire family will then be totally bounded away from zero. In addition to that, readers may easily check that $\{f_\omega: 0 < \omega < a\}$ is quadratic-mean-differentiable.

Now, by adding one distribution with density $f_0(x) = 1$ (for $|x| \le 1/2$) to the family $\mathbf{F}_\Omega$ (call the new family $\mathbf{F}_{\Omega'}$, where $\Omega' = \Omega \cup \{0\}$), the modulus at $\omega = 0$ becomes

$$b(\varepsilon) = 8^{1/3}\varepsilon^{2/3} + o(\varepsilon^{2/3}).$$

Hence, f_0 is a singular point. Let us now assume that there is a globally informative unbiased estimator T_n for the parametric problem $\mathbf{F}_\Omega$. Then, due to Fatou's lemma as well as the fact that $f_\omega(x) \to f_0(x)$ pointwisely as $\omega \to 0$, we have that T_n is also a globally informative unbiased estimator for the parametric problem $\mathbf{F}_{\Omega'}$, which certainly contradicts Theorem 1. Hence, there is no globally informative unbiased estimator for the problem $\mathbf{F}_\Omega$ even though it is QMD with Fisher informations totally bounded away from 0. An asymptotic result follows from Theorem 2.

4. Discussion. Readers should notice that having $\mathbf{V}$ being a finite dimensional space does not mean the statistical problem has to be finite dimensional (e.g., in a nonparametric problem of estimating the density at a point, say x_0, $\mathbf{V} = \mathbf{R}$ is in fact one-dimensional). Also, our results in the asymptotic sense are in terms of the behavior of the asymptotic distributions. Such results are more delicate than results in terms of the limiting behaviors of the bias and second moments. Results in terms of limiting behaviors of bias and second moment can be derived simply by following an argument similar to the proof of Theorem 1 without too much extra effort and are consequently easier to prove than the results in our Theorem 2.

4.1. *Examples on singular problems.* Since interesting singular problems naturally arise in all areas of statistics regardless of the problem being finite dimensional (e.g., parametric) or infinite dimensional (e.g., nonparametric), we will take a look at some interesting examples below.

EXAMPLE 3 (Mixing normal). In spite of the fact that most interesting singular problems are nonparametric, mixing normal is one of the many interesting singular problems in parametrics. For simplicity, we will use a simple version of mixing normal, which will directly lead to the consequence that many interesting mixing normal problems (e.g., [24]) are in fact singular problems. Let $X_1, \dots, X_n$ be i.i.d. observations from some distribution with density of the form

$$f_\mu(x) = \tfrac{1}{2}\phi(x-\mu) + \tfrac{1}{2}\phi(x+\mu),$$

where $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ is the density for $N(0,1)$, and w.l.o.g. we can assume $\mu \ge 0$. The problem is to estimate the functional $T(f_\mu) = \mu$ from those observations. A direct computation shows that

$$b(\varepsilon) = A\varepsilon^{1/2} + o(\varepsilon^{1/2})$$

at $\mu = 0$, where A is a positive constant. Therefore, the problem is singular. Consequently, there will be no locally informative unbiased estimator in either the finite sample sense or the asymptotic sense over any neighborhood of $\mu = 0$. However, we can use some *biased* estimator such as the maximum likelihood estimator. This estimator is informative in our sense both in finite samples and asymptotically at the modulus rate $b_{\mu=0}(n^{-1/2}) \approx n^{-1/4}$.

On the other hand, singular problems are widely common in nonparametrics. Hence, phenomena in Section 2 can be easily seen in this area. For example, Sacks and Ylvisaker [20] considered a density estimation problem over their family in estimating $T(f) = f(0)$ (density value at 0). For this problem, the modulus at every density f in the family has been computed (see [5], [14]) as

$$b(\varepsilon) = \tilde{A}_f\varepsilon^{4/5} + o(\varepsilon^{4/5}),$$

where $\tilde{A}_f$ is a constant. Hence, it is a singular problem. It is quite obvious, from the asymptotic minimax kernel estimator found by Sacks and Ylvisaker (which has also been reconfirmed in [5] to be the best linear estimator with the modulus rate for this problem), the asymptotic second moments (normalized by the factor $n^{-4/5}$) are nicely bounded by the asymptotic minimax risk $(3/4)15^{-1/5}M^{4/5}$ (computed in [20]). Thus, this kernel estimator gives useful information. On the other hand, it has to be asymptotically biased at every singular point as required in Theorem 2.

Due to the effect (e.g., from Example 1) that there is no unbiased estimator for such a problem (see also [15]), and due to the fact that the second moment of an unbiased estimator can be arbitrarily bad over an arbitrarily small Hellinger neighborhood of a singular point, it is easy to suspect that:

1. an l-unbiased estimator *never* exists at a singular point or
2. simply the second moment of an l-unbiased estimator should be infinity at a singular point. But that is *not* precisely the case, as shown by the following example.

EXAMPLE 4 (Unbiased estimator with finite variance at a singular point). Without loss of generality, we can consider the case when a single observation X is obtained from a distribution in the parametric family $\mathbf{F} = \{\text{Uniform}(\theta^3 - 1/2, \theta^3 + 1/2): |\theta| < 1/2\}$. Let

$$T(X) = \operatorname{sign}(X)\Big/\Big\{3\big((X - \operatorname{sign}(X)/2)^2\big)^{1/3}\Big\} \quad \text{for } |X| > 1/2,$$

or otherwise be equal to 0. Then, a direct computation shows that $T(X)$ is indeed an unbiased estimator for the entire parameter space with $\text{Var}_0(T(X)) = 0 < \infty$. On the other hand, since

$$b(\varepsilon) = (1/2)^{1/3}\varepsilon^{2/3}$$

at $\theta_0 = 0$, $F_{\theta_0} = \text{Uniform}(-1/2, 1/2)$ is in fact a singular point.

4.2. *Relation to other works.* Rosenblatt [19] has shown that there is no unbiased estimator in a density estimation problem. However, his results do not establish the nonexistence of asymptotically unbiased and informative estimators.

Even though *bias* and *variance* are both important in measuring the accuracy of an estimator, and are both essential in the informativity of a confidence statement given by an estimator, it has been a trend in estimation theory that bias is a primary concern. Besides those successful findings in constructing informative unbiased estimators, methods for improving existing informative estimators by reducing their biases are very popular. In the study of regular problems, many methods (e.g., jacknife and so on) are available for such a purpose in different situations. However, in singular problems, little has been known whether reducing bias towards zero is a smart thing to do. Doss and Sethuraman [6] have shown that whenever (1) "the family has common support"; (2) "the relative density functions dP_θ/dP_{θ_0} belong to $L^2(P_{\theta_0})$"; and (3) "there exists no unbiased estimator for the problem"; the variance tends to infinity as the bias is reduced towards zero in a fixed sample size setting. Only a handful of examples in the literature state such misbehavior concerning unbiased estimators. They usually serve as examples and still leave behind the mysteries about "*When will the misbehavior occur*?" and "*Will this phenomenon affect a large area of statistical problems*?" Since most of the existing examples come from exponential families [in cases where conditions (1) and (2) are satisfied], analyticity can then be used to explain certain phenomena like this. However, conditions (1) and (2) are rather restrictive in the sense that they are not satisfied in many interesting problems where closely related misbehaviors may occur (e.g., nonparametric problems). Furthermore, condition (3)—the crucial assumption which results that improving bias towards zero is undesirable—is quite a strong and abstract assumption in the following senses:

1. There is no concrete way of knowing that your problem satisfies this condition (consequently, you do not know when to apply it).
2. The nondesirability of reducing bias too far is in fact a consequence of a weaker condition namely "there is no informative unbiased estimator" instead of condition (3).

It is understandable that if all unbiased estimators are not informative estimators (or do not exist), then your estimator may turn bad if you try to reduce its bias towards zero. In many cases, the result can be reflected in terms of unbounded variance (or second moment). Readers may also notice that Doss and Sethuraman's result does not really mean that bias reduction is bad. It merely indicates that, for certain problems, the reduction should not be pushed too far (i.e., reducing bias towards zero). In fact, for all singular problems, the informativity will indeed be arbitrarily bad over any neighborhood of a singular point when the bias is reduced towards zero. We will explore one theorem below. Other results of the same nature (such as those in the asymptotic sense) can be proved by a similar technique; we will not pursue this here.

THEOREM 3. *Let n be a fixed sample size. Let $\{T_n^m\}_{m=1}^{\infty}$ be a sequence of estimators such that* $\lim_{m\to\infty}\|E_F(T_n^m) - T(F)\|_{\mathbf{V}} = 0$ *for all F in a neighborhood of a singular point F_0. Then, for every $\varepsilon > 0$,*

$$\lim_{m\to\infty} \sup_{F\in\mathbf{F},\, H(F, F_0)\le\varepsilon} E_F\big(\|T_n^m\|_{\mathbf{V}}^2\big) = \infty.$$

5. Proofs.

PROOF OF THEOREM 1. Let $F_0 \in \mathbf{F}$ be a singular point. Without loss of generality, let $F_{\varepsilon,1} \in \mathbf{F}$ be the extremal distribution which gives the $b(\varepsilon)$ for $\varepsilon > 0$.

NOTE. In cases where $F_{\varepsilon,1}$ does not exist in $\mathbf{F}$, we can choose any distribution $G_\varepsilon \in \mathbf{F}$ such that $H(G_\varepsilon, F_0) \le \varepsilon$ and

$$b(\varepsilon) - |T(G_\varepsilon) - T(F_0)| \le \varepsilon^3.$$

By such choices, the following argument will still go through just by replacing $b(\varepsilon)$ with $\tilde{b}(\varepsilon) = |T(G_\varepsilon) - T(F_0)|$ and replacing $F_{\varepsilon,1}$ with G_ε. (Since all cases in the proofs of the theorems in this section can be argued in a similar way, without loss of generality, we will use $F_{\varepsilon,1}$ as the extremal distribution in *all* the proofs in Section 5.) For our convenience in the following proof, let $f_0^{(n)}$, $f_{\varepsilon,1}^{(n)}$ be the densities for the n-fold product probability distributions $F_0^{(n)}$ and $F_{\varepsilon,1}^{(n)}$, respectively, w.r.t. a dominating measure—say, $\mu = (F_0 + F_{\varepsilon,1})/2$ [where $G^{(n)}$ denotes the n-fold product probability measure for n i.i.d. random variables generated from the distribution G]. Notations of this kind will be used throughout Section 5.

Suppose T_n is l-unbiased and l-informative at a singular point F_0. Then, by a direct application of the Cauchy–Schwarz inequality (similar to Pitman [17], page 35),

$$\begin{aligned}
b^2(\varepsilon) &\equiv \|T(F_0) - T(F_{\varepsilon,1})\|_{\mathbf{V}}^2\\
&= \|E_{F_0}(T_n) - E_{F_{\varepsilon,1}}(T_n)\|_{\mathbf{V}}^2\\
&\le \left(\int \|T_n\|_{\mathbf{V}} |f_0^{(n)} - f_{\varepsilon,1}^{(n)}|\, d\mu\right)^2\\
&= \left(\int \|T_n\|_{\mathbf{V}} \left|\{f_0^{(n)}\}^{1/2} + \{f_{\varepsilon,1}^{(n)}\}^{1/2}\right| \left|\{f_0^{(n)}\}^{1/2} - \{f_{\varepsilon,1}^{(n)}\}^{1/2}\right| d\mu\right)^2 \qquad (5.1)\\
&\le \left(\int \|T_n\|_{\mathbf{V}}^2 \cdot \left(\{f_0^{(n)}\}^{1/2} + \{f_{\varepsilon,1}^{(n)}\}^{1/2}\right)^2 d\mu\right) \cdot H^2\big(F_0^{(n)}, F_{\varepsilon,1}^{(n)}\big)\\
&= O\big(H^2\big(F_0^{(n)}, F_{\varepsilon,1}^{(n)}\big)\big)\\
&= O(n\cdot\varepsilon^2).
\end{aligned}$$

Therefore,

$$\left(\frac{b(\varepsilon)}{\varepsilon}\right)^2 \le O(1). \tag{5.2}$$

This contradicts that F_0 is a singular point. □

Proof of Theorem 2. For simplicity, the following illustrations will give a clear picture of the proof of Theorem 2 (an alternative proof is also available in [15]). Let $F_0 \in \mathbf{F}$ be a singular point with infinitesimal irregularity. Let T_n be an estimator for estimating the functional $T(F)$, which is *l-a-unbiased* and *l-a-informative* with modulus rate $f(n) = b_{F_0}(n^{-1/2})$ at F_0. And let $\mu_n = (F_0 + F_{an^{-1/2},1})/2$. Due to the frequent usage of the following notations, the proof will be neater by redefining these notations as follows:

$$f^{-1}(n)(T_n - T(F_0)) \equiv W_{n,0}, \qquad f^{-1}(n)(T_n - T(F_{an^{-1/2},1})) \equiv W_{n,1,a},$$

$$f^{-1}(n)(T(F_0) - T(F_{an^{-1/2},1})) \equiv Z_{n,a}.$$

First of all, it is clear that, for any $x, y \in \mathbf{V}$,

$$\|l_M(x+y)\|_{\mathbf{V}} \le \|l_M(x)\|_{\mathbf{V}} + \|y\|_{\mathbf{V}}.$$

Therefore,

$$\|l_M(W_{n,1,a})\|_{\mathbf{V}} \le \|l_M(W_{n,0})\|_{\mathbf{V}} + \|Z_{n,a}\|_{\mathbf{V}}. \tag{5.3}$$

Second, by l-a-unbiasedness, and l-a-informativity (through an argument with Chebyshev inequality), we know that

$$\begin{aligned} P_{F_0}(\|W_{n,0}\|_{\mathbf{V}} \ge M-1) &= P_{F_0}(\|l_M(W_{n,0})\|_{\mathbf{V}} \ge M-1) \\ &\le E_{F_0}(\|l_M(W_{n,0})\|_{\mathbf{V}}^2)/(M-1)^2 \\ &\to 0 \quad \text{as } n \to \infty, \text{ then } M \to \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} &\Big| \|Z_{n,a}\|_{\mathbf{V}} - \big\| E_{F_0}\{l_M(W_{n,1,a})\}\big\|_{\mathbf{V}} \Big| \\ &\quad \le \big\| E_{F_0}\{Z_{n,a} + l_M(W_{n,0})\} - E_{F_0}\{l_M(W_{n,1,a})\}\big\|_{\mathbf{V}} + \big\| E_{F_0}(l_M(W_{n,0}))\big\|_{\mathbf{V}} \\ &\quad \le \Big(E_{F_0}\Big\{\big\|Z_{n,a} + l_M(W_{n,0}) - l_M(W_{n,1,a})\big\|_{\mathbf{V}}^2\Big\}\Big)^{1/2} \cdot P_{F_0}(\|W_{n,0}\|_{\mathbf{V}} \ge M-1) \\ &\qquad + \big\| E_{F_0}(l_M(W_{n,0}))\big\|_{\mathbf{V}} \\ &\quad \le O(1)\cdot o(1) + o(1) \quad \text{as } n \to \infty, \text{ then } M \to \infty \\ &\quad = o(1) \quad \text{as } n \to \infty, \text{ then } M \to \infty. \end{aligned}$$

Consequently,

$$\begin{aligned} &\big\| E_{F_0}\{l_M(f^{-1}(n)(T_n - T(F_{an^{-1/2},1})))\}\big\|_{\mathbf{V}} \\ &\quad = \|Z_{n,a}\|_{\mathbf{V}} + o(1) \quad \text{as } n \to \infty, \text{ then } M \to \infty \\ &\quad = h(a) + o(1) \quad \text{as } n \to \infty, \text{ then } M \to \infty, \text{ then } a \to 0^+. \end{aligned} \tag{5.4}$$

On the other hand,

$$
\begin{aligned}
&\left\| E_{F_0}\{l_M(f^{-1}(n)(T_n - T(F_{an^{-1/2},1})))\}\right\|_{\mathbf{V}} \\
&\quad \le \left\| E_{F_{an^{-1/2},1}}\{l_M(W_{n,1,a})\}\right\|_{\mathbf{V}} \\
&\qquad + \left\| \left[E_{F_0}\{l_M(W_{n,1,a})\} - E_{F_{an^{-1/2},1}}\{l_M(W_{n,1,a})\}\right]\right\|_{\mathbf{V}} \\
&\quad \le \left\| E_{F_{an^{-1/2},1}}\{l_M(W_{n,1,a})\}\right\|_{\mathbf{V}} + \int \|l_M(W_{n,1,a})\|_{\mathbf{V}}\left(f_0^{(n)} - f_{an^{-1/2},1}^{(n)}\right) d\mu_n \qquad (5.5)\\
&\quad = \left\| E_{F_{an^{-1/2},1}}\{l_M(W_{n,1,a})\}\right\|_{\mathbf{V}} \\
&\qquad + \int \|l_M(W_{n,1,a})\|_{\mathbf{V}}\left(\left(f_0^{(n)}\right)^{1/2} + \left(f_{an^{-1/2},1}^{(n)}\right)^{1/2}\right) \\
&\qquad \times \left(\left(f_0^{(n)}\right)^{1/2} - \left(f_{an^{1/2},1}^{(n)}\right)^{1/2}\right) d\mu_n.
\end{aligned}
$$

Now, by (5.3), l-a-unbiasedness, l-a-informativity, an application of Cauchy–Schwarz inequality on the second term of (5.5), and a well known fact that $H(F_0^{(n)}, F_{an^{-1/2},1}^{(n)}) = O(a)$ as $n \to \infty$, we have that, for large M and large n with $a < a_0$ such that $h(a)$ [defined in (2.5)] is less than 1,

the second term of (5.5)

$$
\begin{aligned}
&\le \left\{ \int \|l_M(W_{n,1,a})\|_{\mathbf{V}}^2 \left(\left(f_0^{(n)}\right)^{1/2} + \left(f_{an^{-1/2},1}^{(n)}\right)^{1/2}\right)^2 d\mu_n \right\}^{1/2} \\
&\quad \times H\left(F_0^{(n)}, F_{an^{-1/2},1}^{(n)}\right) \\
&\le \left\{ 2E_{F_0}\left(\|l_M(W_{n,1,a})\|_{\mathbf{V}}^2\right) + 2E_{F_{an^{-1/2},1}}\left(\|l_M(W_{n,1,a})\|_{\mathbf{V}}^2\right)\right\} \\
&\quad \times H\left(F_0^{(n)}, F_{an^{-1/2},1}^{(n)}\right) \\
&\le \left\{ 2E_{F_0}\left(\|l_M(W_{n,0})\|_{\mathbf{V}}^2\right) + 4E_{F_0}\left(\|l_M(W_{n,0})\|_{\mathbf{V}}\|Z_{n,a}\|_{\mathbf{V}}\right)\right. \\
&\qquad \left. + 2E_{F_0}\left(\|Z_{n,a}\|_{\mathbf{V}}^2\right) + 2E_{F_{an^{-1/2},1}}\left(\|l_M(W_{n,1,a})\|_{\mathbf{V}}^2\right)\right\}^{1/2} \\
&\quad \times H\left(F_0^{(n)}, F_{an^{-1/2},1}^{(n)}\right) \\
&\le \left(4M_{F_0} + 4M_{F_0}^{1/2} + 2\right)^{1/2} \cdot O(a) \quad \text{as } n \to \infty, \text{ then } M \to \infty, \text{ then } a \to 0^+
\end{aligned}
$$

Therefore,

$$
(5.5) = o(1) + O(a), \quad \text{as } n \to \infty, \text{ then } M \to \infty, \text{ then } a \to 0^+. \qquad (5.6)
$$

Hence, by (5.4) and (5.6), $h(a) \le O(a) + o(1)$ as $a \to 0$, which contradicts to the infinitesimal irregularity. □

PROOF OF THEOREM 3. Let F_0 be a singular point. Let n be the fixed sample size. Let $\{T_n^m\}_{m=1}^{\infty}$ be a sequence of estimators as given in Theorem 3. Then $\|E_F(T_n^m) - T(F)\|_{\mathbf{V}} \to 0$ as $m \to \infty$ for every F in a neighborhood of F_0 (say, an ε_1-Hellinger ball around F_0). Now, suppose that there is an

$\varepsilon_0 > 0$ such that

$$\lim_{m\to\infty} \sup_{F\in\mathbf{F},\, H(F,F_0)\le\varepsilon_0} E_F\left(\|T_n^m\|_{\mathbf{V}}^2\right) \neq \infty.$$

Then, there must be a positive number M, and a subsequence $\{T_n^{m_k}\}_{k=1}^{\infty}$ such that

$$\sup_{F\in\mathbf{F},\, H(F,F_0)\le\varepsilon_0} E_F\left(\|T_n^{m_k}\|_{\mathbf{V}}^2\right) \le M \quad \text{for all } k. \tag{5.7}$$

Let ε_0' be a number such that $0 < \varepsilon_0' < \min(\varepsilon_0, \varepsilon_1)$. Then, (5.7) implies

$$\sup_{F\in\mathbf{F},\, H(F,F_0)\le\varepsilon_0'} E_F\left(\|T_n^{m_k}\|_{\mathbf{V}}^2\right) \le M \quad \text{for all } k. \tag{5.8}$$

Now, for every sufficiently large positive integer t, we can select a number $m_{k(t)}$ from the index set $\{m_k\}_{k=1}^{\infty}$ *inductively* such that $m_{k(t)} < m_{k(t')}$ if $t < t'$ and

$$\begin{aligned}\max\Big\{\big\| E_{F_0}(T_n^{m_{k(t)}}) - T(F_0)\big\|_{\mathbf{V}}, \big\| E_{F_{t^{-1/2},1}}(T_n^{m_{k(t)}}) - T(F_{t^{-1/2},1})\big\|_{\mathbf{V}}\Big\}\\ \le b(t^{-1/2})/6.\end{aligned} \tag{5.9}$$

Thus, by the same argument as in Theorem 1, by (5.9), and by the fact that $H^2(F_0^{(n)}, F_{t^{-1/2},1}^{(n)}) = nt^{-1}(1 + o(1))$, we have

$$\frac{4}{9}\frac{b^2(t^{-1/2})}{nt^{-1}(1+o(1))} \le 2\Big(E_{F_0}\big(\|T_n^{m_{k(t)}}\|_{\mathbf{V}}^2\big) + E_{F_{t^{-1/2},1}}\big(\|T_n^{m_{k(t)}}\|_{\mathbf{V}}^2\big)\Big). \tag{5.10}$$

Since F_0 is a singular point, the l.h.s. of (5.10) will blow up to infinity when $t \to \infty$ while the r.h.s. remains bounded. This is certainly a contradiction. □

Acknowledgments. The authors wish to thank David Donoho, Peter Hall and Iain Johnstone for useful communications; and wish to thank Naomi Altman for suggesting the word "informative" in our definition.

The authors wish to thank both referees for their constructive suggestions in making this paper clearer. Especially, we wish to thank one of the referees, who kindly suggested a very nice way of presenting the proof of Theorem 2 here, which is different than the one originally in the technical report.

REFERENCES

[1] BICKEL, P. J. and LEHMANN, E. L. (1969). Unbiased estimation in convex families. *Ann. Math. Statist.* **40** 1523–1535.

[2] BLACKWELL, D. (1947). Conditional expectation and unbiased sequential estimates. *Ann. Math. Statist.* **18** 105–110.

[3] DONOHO, D. L. and LIU, R. C. (1988). Geometrizing rates of convergence I. Technical report, Dept. Statistics, Univ. California, Berkeley.

[4] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence II. *Ann. Statist.* **19** 633–667.

[5] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19** 668–701.

[6] DOSS, H. and SETHURAMAN, J. (1989). The price of bias reduction when there is no unbiased estimate. *Ann. Statist.* **17** 440–442.
[7] FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.
[8] HASMINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contributions to Statistics: J. Hájek Memorial Volume* (J. Jurečková, ed.) 91–97. Academia, Prague.
[9] LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Publ. in Statistics* **1** 277–330.
[10] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, Berlin.
[11] LEHMANN, E. L. and SCHEFFÉ, H. (1950, 1955, 1956). Completeness similar regions and unbiased estimation, I, II, Corrigenda. *Sankhyā* **10** 305–340; **15** 219–236; **17** 250.
[12] LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.
[13] LEVIT, B. Y. (1990). Approximately integrable linear statistical models in nonparametric estimation. Technical Report 90-37C, Dept. Statistics, Purdue Univ.
[14] LIU, R. C. (1987). Geometry in robustness and nonparametrics. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.
[15] LIU, R. C. and BROWN, L. D. (1990). Nonexistence of good unbiased estimators in singular problems. Technical Report 90-31, Cornell.
[16] PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 52–68.
[17] PITMAN, E. J. G. (1979). *Some Basic Theory for Statistical Inference*. Chapman and Hall, London.
[18] RAO, C. R. (1948). Sufficient statistics and minimum variance estimates. *Proc. Cambridge Philos. Soc.* **45** 213–218.
[19] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.
[20] SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* **9** 334–346.
[21] SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
[22] STEIN, C. (1950). Unbiased estimates of minimum variance. *Ann. Math. Statist.* **21** 406–415.
[23] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.
[24] TAN, W. Y. and CHANG, W. C. (1972). Comparisons of method of moments and method of maximum likelihood in estimating parameters of a mixture of two normal densities. *J. Amer. Statist. Assoc.* **67** 702–708.

DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853