J. Japan Statist. Soc.
Vol. 25 No. 2 1995 151-158

INFORMATION INEQUALITIES FOR THE MINIMAX RISK

Michikazu Sato* and Masafumi Akahira*

This paper presents lower bounds for the minimax risk under quadratic loss, derived from information inequalities for the Bayes risk obtained by Borovkov and Sakhanienko, Brown and Gajek. In addition, admissibility of a minimax estimator is discussed, and we provide examples which illustrate that they are good bounds.

1. Introduction

Borovkov and Sakhanienko [2] and Brown and Gajek [3] presented lower bounds for the Bayes risk under scaled quadratic loss assuming some regularity conditions. In cases where the regularity conditions of Brown and Gajek [3] do not necessarily hold, Sato and Akahira [8] have also obtained lower bounds.

The purpose of this paper is to obtain lower bounds for the minimax risk under quadratic loss. In Section 2, a lower bound for the minimax risk are given, using a family of prior distributions, which is an application of the results of Borovkov and Sakhanienko [2] and Brown and Gajek [3]. The result is shown to be useful in order to prove an estimator to be minimax, and its admissibility is also discussed. Since the assumptions in Section 2 are too strong to use in cases when the parameter space is bounded, we obtain lower bounds for the minimax risk under less restrictive assumptions in Section 3, which are asymptotically good bounds as is shown in the examples.

2. A lower bound for the minimax risk: fixed sample case

In this section we obtain a lower bound for the minimax risk for a fixed sample case.

Let X be an observable random variable with probability densities p_θ relative to some σ-finite measure ν. Assume $\theta \in \Theta$, where $\Theta \subset \boldsymbol{R}$ is a (possibly infinite) interval. It is desirable to estimate θ by $a \in \Theta$ under loss

$$L(\theta, a) = (a-\theta)^2 .$$

Let $R(\theta, T) = E_\theta[L(\theta, T)]$ denote the risk of the nonrandomized estimator $T = T(X)$. Define

$$R^*(T) := \sup_\theta R(\theta, T) , \qquad r^* := \inf_T R^*(T) ,$$

then r^* is called to be the minimax risk and also T_0 is said to be minimax if $R^*(T_0) = r^* < \infty$. Let Θ° denote the interior of Θ. Let $\bar{\Theta}$ denote the closure of Θ. Let $g(\cdot)$ be a probability density with respect to the Lebesgue measure

Received September, 1994, Revised October, 1994, Accepted November, 1994
* Institute of Mathematics, University of Tsukuba, Ibaraki 305, Japan.
This research was supported in part by Grant-in-Aid for Science Research, No. 00001777, Ministry of Education, Science and Culture, Japan.

on Θ. This is the prior density. For any estimator T, let $B(g, T)=\int R(\theta, T)g(\theta)d\theta$ and let $B(g)=\inf_T B(g, T)$. $B(g)$ is the Bayes risk under g. When we have independently and identically distributed random variables $X_1, \ldots, X_n$ instead of X, we say "the size of sample is n," and rewrite r^* by r_n^*.

Let $\theta_1=\theta_0-\delta$, $\theta_2=\theta_0+\delta$ for $\delta>0$ and define a prior density g by

$$g(\theta):=\frac{1}{\delta}\cos^2\frac{\pi}{2\delta}(\theta-\theta_0) \qquad \text{for} \quad |\theta-\theta_0|<\delta\,.$$

We denote this prior distribution by $\mathrm{Cos}^2\,(\theta_0, \delta)$. We now make the following conditions (2a) to (2d).

(2a) There exist $\theta_1, \theta_2 \in \bar{\Theta}$ such that $\theta_1<\theta_2$ and, for a.e. $\theta \in (\theta_1, \theta_2)$, the amount of Fisher information

$$I(\theta):=E_\theta\left[\left\{\frac{\partial}{\partial\theta}\log p_\theta(X)\right\}^2\right]$$

exists. Define $V(\theta):=1/I(\theta)$ and assume $0<V(\theta)\leq\infty$ for a.e. $\theta \in (\theta_1, \theta_2)$.

(2b) The Bayes estimator under g

$$T_g(x)=\frac{\int_{\theta_1}^{\theta_2}\theta p_\theta(x)g(\theta)d\theta}{\int_{\theta_1}^{\theta_2}p_\theta(x)g(\theta)d\theta}\,,$$

can be extended to an absolutely continuous function on $[\theta_1, \theta_2]$.

(2c) For T_g, the Cramér-Rao inequality (or the C-R inequality for short)

$$\mathrm{Var}_\theta\, T_g\geq V(\theta)\left\{\frac{d}{d\theta}E_\theta[T_g]\right\}^2$$

holds for a.e. $\theta \in (\theta_1, \theta_2)$.

(2d) A constant v_* satisfies $0<v_*\leq\inf_{\theta_1<\theta<\theta_2}V(\theta)$ and $v_*<\infty$.

In the above, "a.e. θ" means almost all θ with respect to the Lebesgue measure.

Then it follows from Borovkov and Sakhanienko [2] and Brown and Gajek [3] that

$$B(g)>v_*\left(1+\frac{\pi^2 v_*}{\delta^2}\right)^{-1}. \tag{2.1}$$

The bound (2.1) is also a bound for r^*, but in order to get a good bound for r^*, we generally need to consider a family of prior distributions and a manipulation of limit. Here we use

$$v^*:=\sup V(\theta)\,,$$

where the supremum is taken over all values θ where $V(\theta)$ is defined.

We make the following conditions (2e) and (2f).

(2e) There exists θ_0 such that $(\theta_0, \infty)\subset\Theta$ and for a.e. $\theta>\theta_1$, $V(\theta)$ is defined as in (2a).

(2f) For any bounded estimator T, $\theta \mapsto E_\theta[T]$ is absolutely continuous on any bounded closed interval in (θ_0, ∞) and the C-R inequality holds for a.e. $\theta > \theta_0$.

THEOREM 2.1. *Assume that the conditions* (2e) *and* (2f) *hold. Then*

$$r^* \geq \liminf_{\theta \to \infty} V(\theta) =: v_\infty \quad \text{(say)}. \tag{2.2}$$

PROOF. Without loss of generality we assume $v_\infty > 0$. Fix M with $0 < M < \infty$. For a sufficiently large $k \in \boldsymbol{N}$, let

$$v_*(k) := \min\{\inf_{\theta > k} V(\theta), M\}.$$

Then v_* is monotone increasing in k and

$$\lim_{k \to \infty} v_*(k) = \min\{v_\infty, M\}.$$

Let a prior distribution be $\mathrm{Cos}^2(2k, k)$. From (2.1), we have

$$r^* > v_*(k)\left(1 + \frac{\pi^2 v_*(k)}{k^2}\right)^{-1}.$$

Letting $k \to \infty$ we obtain

$$r^* \geq \min\{v_*, M\}.$$

Since this holds for any $0 < M < \infty$, we get $r^* \geq v_\infty$. □

REMARKS. (i) The bound (2.2) is useful when $v_\infty = v^*$ (see Example 2.1). In the case when the size of sample is n, (2.2) can be written as

$$n r_n^* \geq v_\infty. \tag{2.3}$$

If $v_\infty < v^*$ and the assumptions of Theorem 3.1 hold, however, the bound (2.3) is not sharp for a sufficiently large n (see also Theorem 2.3).

(ii) If we let $\theta \to -\infty$ instead of $\theta \to \infty$, then a similar result holds.

Next, in order to consider admissibility, we make the following conditions (2g) to (2i).

(2g) $\Theta = \boldsymbol{R}$ and, for every θ, $V(\theta)$ in (2a) exists and $0 < V(\theta) < \infty$.

(2h) $\lim_{\theta \to \infty} V(\theta)$, $\lim_{\theta \to -\infty} V(\theta)$ exist and are finite.

(2i) For any estimator T which satisfies $E_\theta[T^2] < \infty$ for all θ, $E_\theta[T]$ is differentiable with respect to θ and the C-R inequality holds.

THEOREM 2.2. *Assume that the conditions* (2g) *to* (2i) *hold. If T_0 is an unbiased estimator of θ and the equality of the C-R inequality holds for all θ, T_0 is admissible.*

PROOF. Assume that T_0 is improved by T_1 and denote $b(\theta) = E_\theta[T_1] - \theta$. Then, from the assumption of T_0 and (2i), we have

$$V(\theta) = R(\theta, T_0) \geq R(\theta, T_1) \geq b^2(\theta) + V(\theta)\{1 + b'(\theta)\}^2, \tag{2.4}$$

and by (2g), we get $b'(\theta) \leq 0$, hence b is monotone, and by (2.4), it is bounded. Therefore, $b(\pm\infty) = \lim_{\theta \to \pm\infty} b(\theta)$ exists and is finite, and by applying the mean value theorem for each $i \in \boldsymbol{Z}$, we get $\theta_i \in (i, i+1)$ satisfying $\lim_{i \to \pm\infty} b'(\theta_i) = 0$. Substituting t_i to (2.4) and taking limits, we have $b(\pm\infty) = 0$. Since b is monotone, we have $b \equiv 0$, hence (2.4) is an equality for all θ. This is a contradiction. □

Note that this proof is similar to that of Problem 1 of Hodges and Lehmann [8].

THEOREM 2.3 *Assume that the conditions* (2g) *to* (2i) *hold. If a minimax estimator exists and* $v_\infty=\inf_\theta V(\theta)$ *and* $V(\theta)$ *is not constant, then the bound* (2.2) *is not sharp.*

PROOF. Assume that the bound is sharp. Then, for a minimax estimator T_0,

$$R(\theta, T_0)\leq R^*(T_0)=r^*=v_\infty\leq V(\theta)\,. \tag{2.5}$$

On the other hand, by a similar way to the proof of Theorem 2.2, we have $b(\theta):=E_\theta[T_0]-\theta\equiv 0$, and $R(\theta, T_0)\geq V(\theta)$. Hence (2.5) becomes an equality for all θ and it contradicts the assumption that $V(\theta)$ is not constant. □

EXAMPLE 2.1. If in Theorem 2.1, $v_\infty=v^*<\infty$, and for an unbiased estimator T_0 the equality of the C-R inequality holds for all θ, then T_0 is minimax. For example, if X is normally distributed as $N(\theta, 1)$ and $\Theta=\boldsymbol{R}$, then X is minimax and from Theorem 6.2, X is admissible. In the case when $\Theta=(0, \infty)$, although X is minimax it is not admissible since it can be improved by $\max\{X, 0\}$.

3. Lower bounds for the minimax risk: asymptotic case

In Section 2, we considered a fixed sample case. If Θ is bounded, however, this method does not work well for a fixed sample, as is illustrated in Example 3.1 and Example 3.2. In such cases, under some regularity conditions, however, we can get an asymptotically good bound.

Suppose that $T_n=T_n(X_1, \ldots, X_n)$ is an estimator when the size of sample is n. A sequence of estimators $\{T_n\}$ (or T_n for short) is said to be asymptotically minimax if

$$\lim_{n\to\infty}\frac{r_n^*}{R^*(T_n)}=1\,,$$

where 0/0 is defined by 1.

THEOREM 3.1. *Assume that the conditions* (2a) *to* (2d) *hold. Then*

$$\frac{nr_n^*}{v_*}>\left(1+\frac{\pi^2 v_*}{\delta^2 n}\right)^{-1}>1-\frac{\pi^2 v_*}{\delta^2 n}\,, \tag{3.1}$$

where $\delta=(\theta_2-\theta_1)/2$.

PROOF. From (2.1), when the size of sample is 1, we get

$$r^*>v_*\left(1+\frac{\pi^2 v_*}{\delta^2}\right)^{-1}.$$

When the size of sample is n, then r^* and v_* are replaced by r_n^* and v_*/n, respectively, and we obtain (3.1). □

EXAMPLE 3.1. If X_j's are $N(\theta, \sigma^2)$ random variables and $\Theta=(\theta_0-\delta_0, \theta_0+\delta_0)$, then

$$\frac{nr_n^*}{\sigma^2}>1-\frac{\pi^2\sigma^2}{\delta_0^2 n}\,.$$

This is the best bound of all bounds of the form

$$\frac{nr_n^*}{\sigma^2} \geq 1 - Cn^{-\rho} \qquad \text{for a sufficiently large } n,$$

where C and ρ are independent of n. Indeed we have

$$\frac{nr_n^*}{\sigma^2} = 1 - \frac{\pi^2\sigma^2}{\delta_0^2 n} + o\left(\frac{1}{n}\right).$$

This follows from Bickel [1] and the fact that $\bar{X} = \sum_{j=1}^n X_j/n$ is a sufficient statistic and distributed as $N(\theta, \sigma^2/n)$. For a small n, the bound from Donoho, Liu and MacGibbon [5]

$$\frac{nr_n^*}{\sigma^2} \geq \frac{d}{1 + (\sigma^2/\delta_0^2 n)}$$

with $d \approx (1.247)^{-1}$ is an improvement (numerical comparison is given by Brown and Low [4]), but it is not a good bound for a sufficiently large n.

Note that the bound (3.1) is not generally an asymptotically good bound. In order to get better bounds in the cases where V is not constant, we make the following conditions (3a) and (3b).

(3a) There exist θ_0 and θ_1 such that $-\infty \leq \theta_0 < \theta_1 < \infty$, $(\theta_0, \theta_1) \subset \Theta$ and for a.e. $\theta > \theta_0$, $V(\theta)$ is defined and $v^* = \lim_{\theta \downarrow \theta_0} V(\theta)$.

(3b) For any bounded estimator T_n, $\theta \mapsto E_\theta[T_n]$ is absolutely continuous on any bounded closed interval in (θ_0, θ_1) and the C-R inequality holds for a.e. $\theta \in (\theta_0, \theta_1)$.

THEOREM 3.2. *Assume that the conditions* (3a) *and* (3b) *hold. Then*

$$\liminf_{n\to\infty} nr_n^* \geq v^* .$$

PROOF. If $\theta_0 = -\infty$, we have, from Theorem 2.1 and its Remarks, $nr_n^* \geq v^*$. So we will only consider the other cases. Without loss of generality we assume that $\theta_0 = 0$ and $v^* > 0$. Fix M with $0 < M < \infty$ and let

$$V_+(\delta) := \min\{\inf_{0<\theta<\delta} V(\theta), M\} .$$

Then, from (3a),

$$V_+(\delta) \to \min\{v^*, M\} \qquad \text{as} \quad \delta \downarrow 0 .$$

So, for a sufficiently small $\delta > 0$, let $\mathrm{Cos}^2(2\delta, \delta)$ be a prior distribution. Then, from (2.1),

$$nr_n^* > V_+(3\delta)\left\{1 + \frac{\pi^2 V_+(3\delta)}{\delta^2 n}\right\}^{-1} .$$

Hence the assertion follows from taking $\liminf_{n\to\infty}$ of both sides of the above and letting $\delta \downarrow 0$ and $M \uparrow \infty$. □

REMARKS. (i) A similar result can be obtained in the case $\theta \uparrow \theta_0$ instead of $\theta \downarrow \theta_0$ in (3a).

(ii) From the theorem above, it follows that (3.1) is not an asymptotically better bound in cases when $v_* < v^*$. But, for a fixed n, Theorem 3.2 is meaningless. In Theorem 3.3, we obtain a bound which is meaningful for a fixed n and is asymptotically better in such cases.

THEOREM 3.3. *(I) Assume that the following conditions (3c), (3d) and (3e) hold.*

(3c) $$0<v^*<\infty\,.$$

(3d) *There exist* $\theta_0\in\Theta^\circ$, δ_0, d, $k>0$ *such that* $(\theta_0-\delta_0,\theta_0+\delta_0)\subset\Theta$, *and for a.e.* $\theta\in(-\delta_0,\delta_0)$,

$$V(\theta_0+\theta)\geq v^*-k|\theta|^d\,.$$

(3e) *For any bounded estimator* T_n, $\theta\mapsto E_\theta[T_n]$ *is absolutely continuous in any closed interval in* $(\theta_0-\delta_0,\theta_0+\delta_0)$ *and the C-R inequality holds for a.e.* $\theta\in(\theta_0-\delta_0,\theta_0+\delta_0)$.

Then

$$\frac{nr_n^*}{v^*}>1-Cn^{-d/(d+2)} \tag{3.2}$$

holds for $n>2(\pi v^*)^2\delta_0^{-d-2}(kd)^{-1}$, *where*

$$C:=(2\pi^2d^{-1})^{d/(d+2)}k^{2/(d+2)}(d/2+1)(v^*)^{(d-2)/(d+2)}$$

which is independent of n. *In particular, if* $d=2$,

$$\frac{nr_n^*}{v^*}>1-2\pi\sqrt{\frac{k}{n}}$$

holds for $n>(\pi v^*)^2\delta_0^{-4}k^{-1}$.

(II) *Assume that (3c) and the following conditions (3f) and (3g) hold.*

(3f) *There exist* $\theta_0\in\bar{\Theta}$, δ_0, d, $k>0$ *such that* $(\theta_0,\theta_0+\delta_0)\subset\Theta$, *and for a.e.* $\theta\in(0,\delta_0)$

$$V(\theta_0+\theta)\geq v^*-k\theta^d\,.$$

(3g) *For any bounded estimator* T_n, $\theta\mapsto E_\theta[T_n]$ *is absolutely continuous on any closed interval in* $(\theta_0,\theta_0+\delta_0)$ *and the C-R inequality holds for a.e.* $\theta\in(\theta_0,\theta_0+\delta_0)$.

Then

$$\frac{nr_n^*}{v^*}>1-Dn^{-d/(d+2)} \tag{3.3}$$

holds for $n>2^{d+3}(\pi v^*)^2\delta_0^{-d-2}(kd)^{-1}$, *where*

$$D:=2^{3d/(d+2)}(\pi^2d^{-1})^{d/(d+2)}k^{2/(d+2)}(d/2+1)(v^*)^{(d-2)/(d+2)}$$

which is independent of n.

PROOF. Without loss of generality we assume $\theta_0=0$.

(I) For $0<\delta<\delta_0$, let $\mathrm{Cos}^2(0,\delta)$ be the prior distribution. If the size of sample is 1, from (2.1) we have, for $v^*>k\delta^d$,

$$r^*>(v^*-k\delta^d)\left\{1+\frac{\pi^2}{\delta^2}(v^*-k\delta^d)\right\}^{-1},$$

hence, by using $1/(1+x)>1-x$ for $x>0$, we get

$$\frac{r^*}{v^*}>1-\frac{k\delta^d}{v^*}-\frac{\pi^2v^*}{\delta^2}\,.$$

This holds even if $v^*\leq k\delta^d$. If the size of sample is n then, by replacing r^*, v^*

and k with r_n^*, v^*/n and k/n, respectively, we obtain

$$\frac{nr_n^*}{v^*}>1-\frac{k\delta^d}{v^*}-\frac{\pi^2 v^*}{\delta^2 n}.$$

Letting $\delta=cn^{-\lambda}$ for $c, \lambda>0$, we have $0<\delta<\delta_0$ for a sufficiently large n. In particular, let $\lambda=1/(d+2)$, $c=\{2(\pi v^*)^2(kd)^{-1}\}^{1/(d+2)}$. These values are chosen here in order to get the asymptotically best bound of all λ's and c's. Then we get (3.2) and the range of n by solving $\delta<\delta_0$ for n.

(II) Let $\mathrm{Cos}^2(1+\epsilon)\delta/2, (1-\epsilon)\delta/2$ be the prior distribution, where $0<\delta<\delta_0$ and ϵ be a sufficiently small positive number. Since

$$r^*>(v^*-k\delta^d)\left\{1+\frac{\pi^2}{\{(1-\epsilon)\delta/2\}^2}(v^*-k\delta^d)\right\}^{-1}$$

letting $\epsilon\downarrow 0$ and by a similar way to (I), we get

$$\frac{r^*}{v^*}>1-\frac{2^d k}{v^*}\left(\frac{\delta}{2}\right)^d-\frac{\pi^2 v^*}{(\delta/2)^2}.$$

Hence we obtain (3.3) by replacing δ and k with $\delta/2$ and $2^d k$, respectively, in (I), and the range of n by replacing δ_0 with $\delta_0/2$ in (I). □

In this theorem, the larger we take d which satisfies the assumption, the better we get the bound asymptotically. If we fix d, then the smaller we take k which satisfies the assumption, the better we get the bound asymptotically. We also see that δ_0 disappears in (3.2) and (3.3), but appears in the range of n. If we fix d and k, then the larger we take δ_0 which satisfies the assumption, the wider the range of n will be. Generally however, the smaller we take δ_0, the better we can get the bound asymptotically by changing k.

If $V(\theta)$ takes its maximum value at $\theta=\theta_0$ and is C^d for even d in $[\theta_0-\delta_0, \theta_0+\delta_0]$ and $V'(\theta_0)=V''(\theta_0)=\cdots=V^{(d-1)}(\theta_0)=0$, $V^{(d)}(\theta_0)\neq 0$, then, for each $|\theta|<\delta_0$, there exists $0<\eta<1$ such that

$$V(\theta_0+\theta)=v^*+V^{(d)}(\eta\theta)\theta^d\geq v^*-k|\theta|^d$$

where $k:=-\inf_{|\theta|<\delta_0}V^{(d)}(\theta_0+\theta)>0$. If we let $\delta_0\downarrow 0$, then $k\to -V^{(d)}(\theta_0)$. Hence we get

$$\liminf_{n\to\infty} n^{d/(d+2)}(1-(nr_n^*/v^*))\leq(2\pi^2 d^{-1})^{d/(d+2)}|V^{(d)}(\theta_0)|^{d/(d+2)}(d/2+1)(v^*)^{(d-2)/(d+2)}.$$

In particular, if $d=2$, we have

$$\limsup_{n\to\infty}\sqrt{n}\left(1-\frac{nr_n^*}{v^*}\right)\leq 2\pi\sqrt{|V''(\theta_0)|}.$$

This is meaningless, however, for a fixed n.

EXAMPLE 3.2. We consider the case of the Bernoulli trials, that is, X_1, X_2, ... is a sequence of i.i.d. random variables with a binomial distribution $B(1, \theta)$, where $0<\theta<1$. Then we have

$$V(\theta)=\theta(1-\theta), \quad V\left(\frac{1}{2}+\theta\right)=\frac{1}{4}-\theta^2.$$

Letting $v^*=1/4$, $\theta_0=\delta_0=1/2$, $d=2$, $k=1$ in Theorem 3.3 (I), we get

$$\frac{nr_n^*}{v^*}>1-\frac{2\pi}{\sqrt{n}} \qquad \text{for} \quad n\geq 10.$$

This is a non-trivial bound (that is, the right-hand side of the above is positive), however, if and only if $n \geq 40$. Since, for an unbiased estimator $U_n := \sum_{j=1}^{n} X_j/n$ of θ, the equality of the C-R inequality holds, it follows from Theorem 3.3 (or 3.2) that U_n is asymptotically minimax. But U_n is not minimax. In this case,

$$T_n := \frac{\sum_{j=1}^{n} X_j + \sqrt{n}/2}{n+\sqrt{n}}$$

is minimax and

$$r_n^* = R^*(T_n) = R(\theta, T_n) = \frac{1}{4(\sqrt{n}+1)^2}$$

(see Hodges and Lehmann [6]). Then we get

$$\frac{nr_n^*}{v^*} = 1 - \frac{2\sqrt{n}+1}{n+2\sqrt{n}+1} = 1 - a_n ,$$

where $a_n \sim 2/\sqrt{n}$. When the assumptions of Theorem 3.3 (I) for $d=2$ hold, then by considering a bound of the form

$$\frac{nr_n^*}{v^*} \geq 1 - Cn^{-\rho} \qquad \text{for a sufficiently large } n ,$$

where C and ρ are independent of n, then, although the value $\rho = 1/2$ in Theorem 3.3 cannot be improved, C may be improved.

Acknowledgments

The authors wish to thank the referee for useful comments.

REFERENCES

[1] Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9**, 1301–1309.

[2] Borovkov, A. A. and Sakhanienko, A. U. (1980). On estimates of the expected quadratic risk. *Probab. Math. Statist.* **1**, 185–195. (In Russian.)

[3] Brown, L. D. and Gajek, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18**, 1578–1594.

[4] Brown, L. D. and Low, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann. Statist.* **19**, 329–337.

[5] Donoho, D. L., Liu, R. C. and MacGibbon, B. (1991). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18**, 1416–1437.

[6] Hodges, J. L. Jr. and Lehmann, E. L. (1950). Some problems in minimax point estimation. *Ann. Math. Statist.* **21**, 182–197.

[7] Hodges, J. L. Jr. and Lehmann, E. L. (1951). Some applications of the Cramér-Rao inequality. *Proc. Second Berkeley Symposium Math. Statist. Probab.* 13–22.

[8] Sato, M. and Akahira, M. (1995). An information inequality for the Bayes risk. *To appear in Ann. Statist.*