

THE VINCZE INEQUALITY FOR THE BAYES RISK

Nao Ohyauchi*

In this paper, the Vincze inequality for the Bayes risk of an estimator with the unbiasedness at any two specific values of the parameter is derived using the Lagrange method. The lower bound for the Bayes risk is also shown to be attained. The Cramér-Rao inequality is derived from the information inequality. Some examples on non-regular distributions are also given.

Key words and phrases: Bayes estimator, Bayes risk, Cramér-Rao inequality, information inequality, Lagrange method, non-regular distributions.

1. Introduction

In statistical estimation it is well known that the information inequalities of Cramér-Rao, Bhattacharyya, etc. play an important role, under suitable regularity conditions. In non-regular cases when the regularity conditions do not always hold, the Hammersley-Chapman-Robbins inequality is well known (Hammersley (1950), Chapman and Robbins (1951)), and it is useful for the distribution with the support depending on the parameter. In such a distribution, it is shown in Akahira and Takeuchi (1995) that the infimum of the variance of unbiased estimator is zero at any specified value of the parameter.

In Vincze (1992), the lower bound for the convex combination of variances of any unbiased estimator at two arbitrary points of the parameter is derived in an indirect way based on the Cramér-Rao type inequality, and it is shown to be unattainable. In this paper, we have a grip on the convex combination as the Bayes risk, and derive the Vincze inequality for the Bayes risk of an estimator with unbiasedness at any two specified points of the parameter, using the Lagrange method (see also Akahira and Ohyauchi (2003)). We also show that the lower bound for the Bayes risk can be attained. Further, the Cramér-Rao inequality is derived from the information inequality. Some examples on non-regular distributions are given. Related results to the information inequalities for the Bayes risk are found in Ohyauchi (2002) and Akahira and Ohyauchi (2002).

2. Formulation

Suppose that the joint probability density function (with respect to a σ -finite measure μ) of a random vector $\mathbf{X} := (X_1, \dots, X_n)$ is $f_{\mathbf{X}}(\mathbf{x}, \theta)$, where $\theta \in \Theta$. Let θ_1 and θ_2 be any disjoint points in Θ , and π_α be a prior probability measure such that $\pi_\alpha(\{\theta_1\}) = \alpha$ and $\pi_\alpha(\{\theta_2\}) = 1 - \alpha$, where $0 \leq \alpha \leq 1$. Let \mathcal{X} be a sample space of \mathbf{X} . Then we consider the Bayes risk $r(\pi_\alpha, \hat{g})$ of an estimator $\hat{g}(\mathbf{X})$ of a

function $g(\theta)$ of θ based on $\mathbf{X} := (X_1, \dots, X_n)$ with respect to a quadratic loss and the prior π_α , i.e.,

$$(2.1) \quad r(\pi_\alpha, \hat{g}) := \int_{\Theta} R(\theta, \hat{g}) \pi_\alpha(d\theta) = \alpha R(\theta_1, \hat{g}) + (1 - \alpha) R(\theta_2, \hat{g})$$

under the condition

$$(2.2) \quad E_{\theta_i}(\hat{g}) = g(\theta_i) \quad (i = 1, 2),$$

where

$$R(\theta, \hat{g}) = E_{\theta} \left[\{\hat{g}(\mathbf{X}) - g(\theta)\}^2 \right] = \int_{\mathcal{X}} \{\hat{g}(\mathbf{x}) - g(\theta)\}^2 f_{\mathbf{X}}(\mathbf{x}, \theta) d\mu(\mathbf{x}).$$

Now we want to obtain the estimator $\hat{g} = \hat{g}^*(\mathbf{X})$ minimizing the Bayes risk (2.1) under the condition (2.2). In order to do so, we define an amount of information as

$$(2.3) \quad I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) := \int_{\mathcal{X}} \frac{\{f_{\mathbf{X}}(\mathbf{x}, \theta_1) - f_{\mathbf{X}}(\mathbf{x}, \theta_2)\}^2}{f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)} d\mu(\mathbf{x}),$$

where

$$(2.4) \quad f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) = \int_{\Theta} f_{\mathbf{X}}(\mathbf{x}, \theta) \pi_\alpha(d\theta) = \alpha f_{\mathbf{X}}(\mathbf{x}, \theta_1) + (1 - \alpha) f_{\mathbf{X}}(\mathbf{x}, \theta_2).$$

(See Akahira (1975), pp. 20, 21, Akahira and Takeuchi (1981), p. 42 for $\alpha = 1/2$, and Vincze (1992) for $0 \leq \alpha \leq 1$.) Akahira (2003) also defined another amount of information by

$$(2.5) \quad J_{\mathbf{X}}(\theta_1, \theta_2) := \int_{\mathcal{X}} \frac{f_{\mathbf{X}}(\mathbf{x}, \theta_1) f_{\mathbf{X}}(\mathbf{x}, \theta_2)}{f_{\mathbf{X}}(\mathbf{x}, \theta_1) + f_{\mathbf{X}}(\mathbf{x}, \theta_2)} d\mu(\mathbf{x}),$$

which is easier to calculate than $I_{\mathbf{X}}^{(1/2)}(\theta_1, \theta_2)$. The relationship between (2.3) and (2.5) is easily seen to be

$$(2.6) \quad I_{\mathbf{X}}^{(1/2)}(\theta_1, \theta_2) + 8J_{\mathbf{X}}(\theta_1, \theta_2) = 4.$$

Since $I_{\mathbf{X}}^{(1/2)}(\theta_1, \theta_2) \geq 0$, it follows from (2.6) that

$$0 \leq J_{\mathbf{X}}(\theta_1, \theta_2) \leq \frac{1}{2}.$$

Letting $\theta_2 = \theta_1 + \Delta$, we have

$$\begin{aligned} J_{\mathbf{X}}(\theta_1, \theta_2) &= \frac{1}{2} - \frac{\Delta^2}{8} E_{\theta_1} \left[\left\{ \frac{\partial}{\partial \theta_1} \log f_{\mathbf{X}}(\mathbf{X}, \theta_1) \right\}^2 \right] + o(\Delta^2) \\ &=: \frac{1}{2} - \frac{\Delta^2}{8} I_{\mathbf{X}}(\theta_1) + o(\Delta^2) \end{aligned}$$

as $\Delta \rightarrow 0$, provided that the differentiation under the integral sign is allowed, where $I_{\mathbf{X}}(\theta_1)$ is the amount of the Fisher information on θ_1 based on \mathbf{X} . A non-regular Cramér-Rao type bound for the convex combination of variances of any unbiased estimator at the two points is obtained by Vincze (1992), and it is also shown that the bound can not be attained.

In this paper, we get the lower bound for the Bayes risk under the condition (2.2), using the Lagrange method, and show that the lower bound is attainable at the points θ_1 and θ_2 .

3. The Vincze inequality for the Bayes risk

First, we derive an information inequality for the Bayes risk of the estimator under the condition (2.2). From (2.2) we have

$$(3.1) \quad \int_{\mathcal{X}} \hat{g}(\mathbf{x}) f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) = \alpha g(\theta_1) + (1 - \alpha)g(\theta_2) =: \eta \quad (\text{say}).$$

Letting

$$h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) := \alpha \{f_{\mathbf{X}}(\mathbf{x}, \theta_1) - f_{\mathbf{X}}(\mathbf{x}, \theta_2)\}$$

for $0 \leq \alpha \leq 1$, we obtain

$$(3.2) \quad \int_{\mathcal{X}} \hat{g}(\mathbf{x}) h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) = \alpha \{g(\theta_1) - g(\theta_2)\} =: \delta \quad (\text{say}).$$

It is easily seen that (3.1) and (3.2) imply (2.2).

THEOREM 3.1. *Under the condition (2.2), the information inequality for the Bayes risk is given by*

$$(3.3) \quad r(\pi_{\alpha}, \hat{g}) \geq \alpha(1 - \alpha) \{g(\theta_1) - g(\theta_2)\}^2 \left\{ \frac{1}{\alpha(1 - \alpha) I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2)} - 1 \right\} \\ =: B^{(\alpha)}(\theta_1, \theta_2) \quad (\text{say}).$$

The lower bound $B^{(\alpha)}(\theta_1, \theta_2)$ is attained by the estimator

$$(3.4) \quad \hat{g}^*(\mathbf{X}) := \eta + \frac{\{g(\theta_1) - g(\theta_2)\} \{f_{\mathbf{X}}(\mathbf{X}, \theta_1) - f_{\mathbf{X}}(\mathbf{X}, \theta_2)\}}{I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) f^{(\alpha)}(\mathbf{X}; \theta_1, \theta_2)}.$$

PROOF. Under the conditions (3.1) and (3.2), we consider the problem of minimizing

$$\int_{\mathcal{X}} \hat{g}^2(\mathbf{x}) f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}).$$

Letting

$$(3.5) \quad F(\lambda_1, \lambda_2; \hat{g}) := \int_{\mathcal{X}} \hat{g}^2(\mathbf{x}) f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) \\ - \lambda_1 \left\{ \int_{\mathcal{X}} \hat{g}(\mathbf{x}) f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) - \eta \right\} \\ - \lambda_2 \left\{ \int_{\mathcal{X}} \hat{g}(\mathbf{x}) h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) - \delta \right\},$$

we get $\hat{g} = \hat{g}^*$ such that \hat{g}^* minimizes F . Since

$$(3.6) \quad F(\lambda_1, \lambda_2; \hat{g}) = \int_{\mathcal{X}} \left[f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) \hat{g}^2(\mathbf{x}) \right. \\ \left. - \{ \lambda_1 f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) + \lambda_2 h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) \} \hat{g}(\mathbf{x}) \right] d\mu(\mathbf{x}) \\ + \lambda_1 \eta + \lambda_2 \delta,$$

in order to obtain \hat{g}^* , for almost all $\mathbf{x}[\mu]$, it is enough to take $\hat{g}(\mathbf{x})$ minimizing $[\dots]$ in (3.6). Hence we have

$$(3.7) \quad \hat{g}^*(\mathbf{x}) = \frac{\lambda_1}{2} + \frac{\lambda_2 h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)}{2f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)},$$

since the integrand of (3.6) is a quadratic expression of $\hat{g}(\mathbf{x})$. From the conditions (3.1) and (3.2) we have

$$(3.8) \quad \int_{\mathcal{X}} \left\{ \frac{\lambda_1}{2} + \frac{\lambda_2 h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)}{2f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)} \right\} f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) = \eta,$$

$$(3.9) \quad \int_{\mathcal{X}} \left\{ \frac{\lambda_1}{2} + \frac{\lambda_2 h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)}{2f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)} \right\} h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) = \delta.$$

Since

$$\int_{\mathcal{X}} h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) = 0,$$

it follows from (3.8), (3.9) and (2.3) that

$$(3.10) \quad \frac{\lambda_1}{2} = \eta, \quad \frac{\lambda_2}{2} = \frac{\delta}{\alpha^2 I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2)}.$$

From (3.7) and (3.10) we have

$$(3.11) \quad \hat{g}^*(\mathbf{X}) = \eta + \frac{\delta h^{(\alpha)}(\mathbf{X}; \theta_1, \theta_2)}{\alpha^2 I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) f^{(\alpha)}(\mathbf{X}; \theta_1, \theta_2)} \\ = \eta + \frac{\{g(\theta_1) - g(\theta_2)\} \{f_{\mathbf{X}}(\mathbf{X}, \theta_1) - f_{\mathbf{X}}(\mathbf{X}, \theta_2)\}}{I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) f^{(\alpha)}(\mathbf{X}; \theta_1, \theta_2)}.$$

It follows from (3.11) that the mean-squared error of \hat{g}^* is given by

$$\begin{aligned}
 (3.12) \quad E_{f^{(\alpha)}} \left[\{\hat{g}^*(\mathbf{X}) - \eta\}^2 \right] &= \int_{\mathcal{X}} \{\hat{g}^*(\mathbf{x}) - \eta\}^2 f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2) d\mu(\mathbf{x}) \\
 &= \frac{\delta^2}{\alpha^4 \left\{ I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) \right\}^2} \int_{\mathcal{X}} \frac{\{h^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)\}^2}{f^{(\alpha)}(\mathbf{x}; \theta_1, \theta_2)} d\mu(\mathbf{x}) \\
 &= \frac{\delta^2}{\alpha^2 I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2)} = \frac{\{g(\theta_1) - g(\theta_2)\}^2}{I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2)}.
 \end{aligned}$$

On the other hand, it follows from (2.1), (3.1) and (3.2) that for any estimator \hat{g} satisfying (2.2)

$$\begin{aligned}
 (3.13) \quad r(\pi_{\alpha}, \hat{g}) &= \alpha R(\theta_1, \hat{g}) + (1 - \alpha) R(\theta_2, \hat{g}) \\
 &= E_{f^{(\alpha)}} \left[(\hat{g} - \eta)^2 \right] - \alpha \{\eta - g(\theta_1)\}^2 - (1 - \alpha) \{\eta - g(\theta_2)\}^2 \\
 &= E_{f^{(\alpha)}} \left[(\hat{g} - \eta)^2 \right] - \alpha(1 - \alpha) \{g(\theta_1) - g(\theta_2)\}^2.
 \end{aligned}$$

From (3.12) and (3.13) we have

$$\begin{aligned}
 r(\pi_{\alpha}, \hat{g}) &\geq E_{f^{(\alpha)}} \left[(\hat{g}^* - \eta)^2 \right] - \alpha(1 - \alpha) \{g(\theta_1) - g(\theta_2)\}^2 \\
 &= \frac{\{g(\theta_1) - g(\theta_2)\}^2}{I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2)} - \alpha(1 - \alpha) \{g(\theta_1) - g(\theta_2)\}^2 \\
 &= \alpha(1 - \alpha) \{g(\theta_1) - g(\theta_2)\}^2 \left\{ \frac{1}{\alpha(1 - \alpha) I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2)} - 1 \right\}.
 \end{aligned}$$

Hence we have the lower bound (3.3), and see that the bound is attained by \hat{g}^* . Thus we complete the proof.

Remark that the lower bound is also derived by Vincze (1992) in an indirect way based on the Cramér-Rao type inequality, which is different from ours.

Next we shall show that the Cramér-Rao inequality can be derived from the information inequality (3.3). In particular, let $\Theta = \mathbf{R}^1$, $\theta_1 = \theta$ and $\theta_2 = \theta + \Delta$ ($\Delta > 0$). Suppose that the support $\{\mathbf{x} \mid f_{\mathbf{X}}(\mathbf{x}, \theta) > 0\}$ is independent of θ , and $g(\theta)$ and $f_{\mathbf{X}}(\mathbf{x}, \theta)$ are differentiable in θ . Then we have for small Δ

$$(3.14) \quad g(\theta_2) - g(\theta_1) = g'(\theta)\Delta + o(\Delta),$$

$$(3.15) \quad f_{\mathbf{X}}(\mathbf{x}, \theta_2) - f_{\mathbf{X}}(\mathbf{x}, \theta_1) = \Delta \frac{\partial f_{\mathbf{X}}(\mathbf{x}, \theta)}{\partial \theta} + o(\Delta),$$

which implies

$$\begin{aligned}
I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) &= I_{\mathbf{X}}^{(\alpha)}(\theta, \theta + \Delta) = \int_{\mathcal{X}} \frac{\{f_{\mathbf{X}}(\mathbf{x}, \theta) - f_{\mathbf{X}}(\mathbf{x}, \theta + \Delta)\}^2}{\alpha f_{\mathbf{X}}(\mathbf{x}, \theta) + (1 - \alpha) f_{\mathbf{X}}(\mathbf{x}, \theta + \Delta)} d\mu(\mathbf{x}) \\
&= \int_{\mathcal{X}} \frac{\Delta^2 \{\partial f_{\mathbf{X}}(\mathbf{x}, \theta) / \partial \theta\}^2}{f_{\mathbf{X}}(\mathbf{x}, \theta)} d\mu(\mathbf{x}) + o(\Delta^2) \\
&= \Delta^2 E_{\theta} \left[\left\{ \frac{\partial \log f_{\mathbf{X}}(\mathbf{X}, \theta)}{\partial \theta} \right\}^2 \right] + o(\Delta^2) \\
&=: \Delta^2 I_{\mathbf{X}}(\theta) + o(\Delta^2),
\end{aligned}$$

where $I_{\mathbf{X}}$ is called the amount of Fisher information. Then

$$\begin{aligned}
(3.16) \quad B^{(\alpha)}(\theta_1, \theta_2) &= \alpha(1 - \alpha) \{g(\theta) - g(\theta + \Delta)\}^2 \left\{ \frac{1}{\alpha(1 - \alpha) I_{\mathbf{X}}^{(\alpha)}(\theta, \theta + \Delta)} - 1 \right\} \\
&= \alpha(1 - \alpha) \{g'(\theta)\}^2 \Delta^2 (1 + o(1)) \left[\frac{1}{\alpha(1 - \alpha) \Delta^2 \{I_{\mathbf{X}}(\theta) + o(1)\}} - 1 \right] \\
&= \frac{\{g'(\theta)\}^2}{I_{\mathbf{X}}(\theta)} + o(1).
\end{aligned}$$

On the other hand, we have

$$E_{\theta}[\hat{g}(\mathbf{X})] = g(\theta), \quad E_{\theta + \Delta}[\hat{g}(\mathbf{X})] = g(\theta + \Delta).$$

We also have

$$(3.17) \quad r(\pi_{\alpha}, \hat{g}) = \alpha V_{\theta}(\hat{g}) + (1 - \alpha) V_{\theta + \Delta}(\hat{g}).$$

Since by (3.14) and (3.15)

$$\begin{aligned}
(3.18) \quad V_{\theta + \Delta}(\hat{g}) &= E_{\theta + \Delta} [\{\hat{g}(\mathbf{X}) - g(\theta + \Delta)\}^2] \\
&= E_{\theta + \Delta} [\{\hat{g}(\mathbf{X}) - g(\theta) - g'(\theta)\Delta + o(\Delta)\}^2] \\
&= \int_{\mathcal{X}} \{\hat{g}(\mathbf{x}) - g(\theta) - g'(\theta)\Delta + o(\Delta)\}^2 \\
&\quad \times \left\{ f_{\mathbf{X}}(\mathbf{x}, \theta) + \Delta \frac{\partial f_{\mathbf{X}}(\mathbf{x}, \theta)}{\partial \theta} + o(\Delta) \right\} d\mu(\mathbf{x}) \\
&= \int_{\mathcal{X}} \{\hat{g}(\mathbf{x}) - g(\theta)\}^2 f_{\mathbf{X}}(\mathbf{x}, \theta) d\mu(\mathbf{x}) + o(1) \\
&= E_{\theta} [\{\hat{g}(\mathbf{X}) - g(\theta)\}^2] + o(1) \\
&= V_{\theta}(\hat{g}) + o(1),
\end{aligned}$$

it follows from (3.17) and (3.18) that

$$(3.19) \quad r(\pi_{\alpha}, \hat{g}) = V_{\theta}(\hat{g}) + o(1).$$

Letting $\Delta \rightarrow 0$, from (3.3), (3.14), (3.16) and (3.19) we obtain for any unbiased estimator of $g(\theta)$

$$(3.20) \quad V_{\theta}(\hat{g}) \geq \frac{\{g'(\theta)\}^2}{I_{\mathbf{X}}(\theta)},$$

which is the Cramér-Rao inequality. From (3.4), (3.14) and (3.15) it follows that

$$\hat{g}^*(\mathbf{X}) - g(\theta) = \frac{g'(\theta)}{I_{\mathbf{X}}(\theta)} \frac{\partial \log f_{\mathbf{X}}(\mathbf{X}, \theta)}{\partial \theta},$$

which is a necessary and sufficient condition for the equality in (3.20) to hold.

4. Bayes estimation

In the formulation in Section 2, we assume that the condition (2.2) holds. However, the interest is focused on the ordinary Bayes estimator without the condition (2.2). We consider the Bayes estimator $\hat{g}_B(\mathbf{X})$, that is the estimator minimizing the Bayes risk (2.1). Then the Bayes estimator is

$$\hat{g}_B^{(\alpha)}(\mathbf{X}) = \frac{\alpha g(\theta_1) f_{\mathbf{X}}(\mathbf{X}, \theta_1) + (1 - \alpha) g(\theta_2) f_{\mathbf{X}}(\mathbf{X}, \theta_2)}{f^{(\alpha)}(\mathbf{X}, \theta_1, \theta_2)}.$$

Let $\alpha = 1/2$. Then it follows from (2.5) that the Bayes risk of $\hat{g}_B^{(1/2)}$ is given by

$$(4.1) \quad r\left(\pi_{1/2}, \hat{g}_B^{(1/2)}\right) = \frac{1}{2} \{g(\theta_1) - g(\theta_2)\}^2 \int_{\mathcal{X}} \frac{f_{\mathbf{X}}(\mathbf{x}, \theta_1) f_{\mathbf{X}}(\mathbf{x}, \theta_2)}{f_{\mathbf{X}}(\mathbf{x}, \theta_1) + f_{\mathbf{X}}(\mathbf{x}, \theta_2)} d\mu(\mathbf{x}) \\ = \frac{1}{2} \{g(\theta_1) - g(\theta_2)\}^2 J_{\mathbf{X}}(\theta_1, \theta_2).$$

Hence, for any estimator $\hat{g} = \hat{g}(\mathbf{X})$ of $g(\theta)$

$$(4.2) \quad r(\pi_{1/2}, \hat{g}) \geq r\left(\pi_{1/2}, \hat{g}_B^{(1/2)}\right) = \frac{1}{2} \{g(\theta_1) - g(\theta_2)\}^2 J_{\mathbf{X}}(\theta_1, \theta_2) =: B^*(\theta_1, \theta_2).$$

On the other hand, it follows from (3.3) that, under the condition (2.2), the lower bound for the Bayes risk is given by

$$(4.3) \quad B^{(1/2)}(\theta_1, \theta_2) = \{g(\theta_1) - g(\theta_2)\}^2 \left\{ \frac{1}{I_{\mathbf{X}}^{(1/2)}(\theta_1, \theta_2)} - \frac{1}{4} \right\}.$$

Comparing $B^*(\theta_1, \theta_2)$ with $B^{(1/2)}(\theta_1, \theta_2)$, we obtain from (2.6) and (4.3)

$$\frac{B^*(\theta_1, \theta_2) - B^{(1/2)}(\theta_1, \theta_2)}{\{g(\theta_1) - g(\theta_2)\}^2} \\ = -\frac{1}{16I_{\mathbf{X}}^{(1/2)}(\theta_1, \theta_2)} \left\{ I_{\mathbf{X}}^{(1/2)}(\theta_1, \theta_2) - 4 \right\}^2 \leq 0.$$

Note that $B^*(\theta_1, \theta_2)$ is less than $B^{(1/2)}(\theta_1, \theta_2)$ since B^* is the lower bound for the Bayes risk of any estimator of $g(\theta)$ without the condition (2.2).

5. Examples

In this section we give two examples of Theorem 3.1 and its related result.

Example 5.1. Suppose that X_1, \dots, X_n are independent and identically distributed (i.i.d.) random variables with the uniform distribution on the interval $[\theta - (1/2), \theta + (1/2)]$. Let $g(\theta) = \theta$, and $\theta_1 < \theta_2 < \theta_1 + 1$. Define $x_{(1)} := \min_{1 \leq i \leq n} x_i$, $x_{(n)} := \max_{1 \leq i \leq n} x_i$, and $S_j := \{\mathbf{x} : x_{(n)} - 1/2 \leq \theta_j \leq x_{(1)} + 1/2\}$ ($j = 1, 2$). Then the amount of information is given by

$$\begin{aligned} I_{\mathbf{X}}^{(\alpha)}(\theta_1, \theta_2) &= \int \cdots \int \frac{\{f_{\mathbf{X}}(\mathbf{x}, \theta_1) - f_{\mathbf{X}}(\mathbf{x}, \theta_2)\}^2}{\alpha f_{\mathbf{X}}(\mathbf{x}, \theta_1) + (1 - \alpha) f_{\mathbf{X}}(\mathbf{x}, \theta_2)} dx_1 \cdots dx_n \\ &= \int \cdots \int_{S_1 - (S_1 \cap S_2)} \frac{1}{\alpha} dx_1 \cdots dx_n + \int \cdots \int_{S_2 - (S_1 \cap S_2)} \frac{1}{1 - \alpha} dx_1 \cdots dx_n \\ &= \frac{1}{\alpha} \left\{ 1 - \left(1 + \frac{\delta}{\alpha} \right)^n \right\} + \frac{1}{1 - \alpha} \left\{ 1 - \left(1 + \frac{\delta}{\alpha} \right)^n \right\} \\ &= \frac{1}{\alpha(1 - \alpha)} \{ 1 - (1 + (\theta_1 - \theta_2))^n \}, \end{aligned}$$

which yields the lower bound

$$(5.1) \quad B^{(\alpha)}(\theta_1, \theta_2) = \alpha(1 - \alpha)(\theta_1 - \theta_2)^2 \left\{ \frac{1}{1 - (1 + (\theta_1 - \theta_2))^n} - 1 \right\}.$$

Then we have

$$(5.2) \quad \frac{f_{\mathbf{X}}(\mathbf{x}, \theta_1) - f_{\mathbf{X}}(\mathbf{x}, \theta_2)}{\alpha f_{\mathbf{X}}(\mathbf{x}, \theta_1) + (1 - \alpha) f_{\mathbf{X}}(\mathbf{x}, \theta_2)} = \frac{1}{\alpha} \chi_{S_1 - (S_1 \cap S_2)}(\mathbf{x}) + \frac{1}{1 - \alpha} \chi_{S_2 - (S_1 \cap S_2)}(\mathbf{x}),$$

where $\chi_A(\mathbf{x})$ denotes the indicator of a set A of \mathcal{X} . Then it follows from (3.4) and (5.2) that

$$(5.3) \quad \begin{aligned} \hat{\theta}^*(\mathbf{X}) &= \alpha\theta_1 + (1 - \alpha)\theta_2 \\ &\quad + \frac{\alpha(1 - \alpha)(\theta_1 - \theta_2)}{1 - (1 + \theta_1 - \theta_2)^n} \left\{ \frac{1}{\alpha} \chi_{S_1 - (S_1 \cap S_2)}(\mathbf{X}) + \frac{1}{1 - \alpha} \chi_{S_2 - (S_1 \cap S_2)}(\mathbf{X}) \right\}, \end{aligned}$$

which attains the lower bound $B^{(\alpha)}(\theta_1, \theta_2)$.

Now, let $\theta_2 = \theta_1 + (\Delta/n)$ ($\Delta > 0$). From (5.1) we have for large n

$$\begin{aligned} B^{(\alpha)}(\theta_1, \theta_2) &= \alpha(1 - \alpha) \frac{\Delta^2}{n^2} \left\{ \frac{1}{1 - (1 - \Delta/n)^n} - 1 \right\} \\ &\doteq \frac{\alpha(1 - \alpha)}{n^2} \Delta^2 \left(\frac{1}{1 - e^{-\Delta}} - 1 \right) = \frac{\alpha(1 - \alpha)}{n^2} \frac{\Delta^2}{e^{\Delta} - 1}. \end{aligned}$$

Then we obtain

$$\sup_{\Delta > 0} B^{(\alpha)} \left(\theta, \theta + \frac{\Delta}{n} \right) \doteq \frac{\alpha(1-\alpha)}{n^2} \frac{\Delta_0^2}{e^{\Delta_0} - 1} \doteq 0.65 \frac{\alpha(1-\alpha)}{n^2},$$

where $\Delta_0 \doteq 1.59$. The estimator with $\theta_2 = \theta_1 + (\Delta_0/n)$ in (5.3) attains the supremum lower bound.

Example 5.2. Suppose that a random variable X is distributed with a density

$$p(x, \theta) = \begin{cases} r & \text{for } 0 < x < \theta, \\ s := 2 - r & \text{for } \theta \leq x \leq \theta + \frac{1}{2}, \\ r & \text{for } \theta + \frac{1}{2} < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta < 1/2$, $0 < r < s < 1$. Let $0 < \theta_1 < \theta_2 < 1/2$. Then we consider the Bayes risk of an estimator of $g(\theta)$. First we have for $0 \leq \alpha \leq 1$,

$$(5.4) \quad p^{(\alpha)}(x; \theta_1, \theta_2) = \alpha p(x, \theta_1) + (1 - \alpha)p(x, \theta_2) \\ = \begin{cases} r & \text{for } 0 < x < \theta_1, \\ \alpha s + (1 - \alpha)r & \text{for } \theta_1 \leq x < \theta_2, \\ s & \text{for } \theta_2 \leq x \leq \theta_1 + \frac{1}{2}, \\ \alpha r + (1 - \alpha)s & \text{for } \theta_1 + \frac{1}{2} < x \leq \theta_2 + \frac{1}{2}, \\ r & \text{for } \theta_2 + \frac{1}{2} < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since

$$(5.5) \quad p(x, \theta_1) - p(x, \theta_2) = \begin{cases} s - r & \text{for } \theta_1 \leq x < \theta_2, \\ r - s & \text{for } \theta_1 + \frac{1}{2} < x \leq \theta_2 + \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

it follows from (2.3) that

$$I_X^{(\alpha)}(\theta_1, \theta_2) = \int_0^1 \frac{\{p(x, \theta_1) - p(x, \theta_2)\}^2}{p^{(\alpha)}(x; \theta_1, \theta_2)} dx \\ = \frac{(s-r)^2}{\alpha(s-r) + r}(\theta_2 - \theta_1) + \frac{(s-r)^2}{\alpha(r-s) + s}(\theta_2 - \theta_1) \\ = \frac{(s-r)^2(\theta_2 - \theta_1)}{(\alpha(s-r) + r)(\alpha(r-s) + s)},$$

which yields the lower bound

$$(5.6) \quad B^{(\alpha)}(\theta_1, \theta_2) = \alpha(1-\alpha) \{g(\theta_1) - g(\theta_2)\}^2 \left\{ \frac{(\alpha(s-r) + r)(\alpha(r-s) + s)}{\alpha(1-\alpha)(s-r)^2(\theta_2 - \theta_1)} - 1 \right\}.$$

From (3.4), (5.4) and (5.5) we have

$$\hat{g}^*(X) = \alpha g(\theta_1) + (1 - \alpha)g(\theta_2) + \frac{(\alpha(s-r) + r)(\alpha(r-s) + s)}{(s-r)^2(\theta_2 - \theta_1)} \{g(\theta_1) - g(\theta_2)\} \\ \times \left\{ \frac{s-r}{\alpha(s-r) + r} \chi_{[\theta_1, \theta_2)}(X) + \frac{r-s}{\alpha(r-s) + s} \chi_{(\theta_1+1/2, \theta_2+1/2]}(X) \right\},$$

which attains the lower bound (5.6).

Acknowledgements

I would like to express my sincere gratitude to Professor M. Akahira for his kind guidance and constant encouragement. I also thank the referees for their kind comments.

REFERENCES

- Akahira, M. (1975). Asymptotic theory for estimation of location in non-regular cases, I: Order of convergence of consistent estimators, *Rep. Stat. Appl. Res., JUSE*, **22**, 8–26.
- Akahira, M. (2003). Personal communication.
- Akahira, M. and Ohyauchi, N. (2002). Information inequalities for the Bayes risk for a family of non-regular distributions, *Ann. Inst. Statist. Math.*, **54**, 806–815.
- Akahira, M. and Ohyauchi, N. (2003). An information inequality for the Bayes risk applicable to non-regular cases, *Proc. Sympos., Res. Inst. Math. Sci.*, **1334**, Kyoto University, 183–191.
- Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*, Lecture Notes in Statistics 7, Springer, New York.
- Akahira, M. and Takeuchi, K. (1995). *Non-Regular Statistical Estimation*, Lecture Notes in Statistics 107, Springer, New York.
- Chapman, D. G. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions, *Ann. Math. Statist.*, **22**, 581–586.
- Hammersley, J. M. (1950). On estimating restricted parameters, *J. Roy. Statist. Soc., Ser. B*, **12**, 192–240.
- Ohyauchi, N. (2002). Comparison of the Bayes risks of estimators for a family of truncated normal distributions, *Commun. Statist. -Theory and Meth.*, **31** (5), 699–718.
- Vincze, I. (1992). On nonparametric Cramér-Rao inequalities, *Order Statistics and Nonparametrics: Theory and Applications*, (eds. Sen, P. K. and Salama, I. A.), Elsevier Science Publishers B. V., 439–454.