PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 27, Fasc. 2 (2007), pp. 167–179

INFORMATION INEQUALITIES FOR THE BAYES RISK OF PREDICTORS

BY

MARAREK KAŁUSZKA (ŁÓDŹ)

Abstract. The paper provides several lower bounds and an upper bound for the Bayes risk in statistical prediction theory. The bounds depend on the Fisher information or the bias of the Bayes predictor. The results improve and extend the inequalities of Brown and Gajek (1990), Takada (1999) and Koike (1999). As an application we evaluate the minimax risk in a problem of sequential prediction.

2000 AMS Mathematics Subject Classification: Primary 62F10; Secondary 62F15.

Key words and phrases: Prediction, Bayes risk, information inequality.

1. INTRODUCTION

An information inequality for the Bayes risk provides a useful lower/upper bound for the estimation error when the conditional expectation is difficult to find explicitly. Unlike the classical Cramér–Rao bound, it usually requires mild regularity conditions. The first results of this kind were established by Van Trees (1968) and Borovkov and Sakhanienko (1980). Using a bound of local type and integrating it with respect to the prior, Brown and Gajek (1990) obtained some improvements of the Borovkov and Sakhanienko inequality. Sato and Akahira (1996) applied a method based upon the variational calculus which works in some non-regular cases. The multidimensional parameter case was treated by Bobrovsky et al. (1987), Brown and Gajek (1990) and Gill and Levit (1995) among others.

Several applications of inequalities for the Bayes risk both in parametric and nonparametric problems can be found in the literature. The bounds enable evaluating the rate of convergence of the Bayes risk (Borovkov and Sakhanienko (1980)) and lead to estimates for the minimax risk (Van Trees (1968); Borovkov and Sakhanienko (1980); Bobrovsky et al. (1987); Brown and Gajek (1990); Gajek and Kałuszka (1994), (1995); Gill and Levit

(1995); Mizera (1996); Sato and Akahira (1996); Kałuszka (1997); Koike (1999)). The bounds are also employed in deriving asymptotically minimax estimators of a regression function and variance function in nonparametric regression (Belitser and Levit (1996); Munk and Ruymgaart (2002)). The last (but not least) possible application is to derive optimal choice of observation window (Kutoyants and Spokoiny (1999)).

In recent years, prediction models have received considerable attention. Bjørnstad (1996) extended the concept of likelihood to prediction models. The classical UMVUE theory was generalized by Yatracos (1992), Miyata (2001) and others. Cramér–Rao type lower bounds were derived by Yatracos (1992) and Nayak (2002). Minimax and admissible prediction of the signal with known background was examined by Zhang and Woodroofe (2005). The explicit form of minimax predictors can also be found in Wilczyński (2001), Jokiel-Rokita (2002), Trybuła (2003) and in the papers cited therein. The first bound for the Bayes risk of predictor was established by Takada (1999) who extended Theorem 2.1 of Brown and Gajek (1990). A bound on the Bayes prediction risk was also obtained by Gajek and Lipińska (2006).

The aim of the paper is to provide new lower bounds in a problem of prediction for squared error loss. We start with a factorization of the Bayes risk different from that of Brown and Gajek (1990) which gives an improvement of known inequalities under weaker assumptions. In particular, we do not assume that the family of data distributions is weakly differentiable in L^2. An application of derived inequalities to evaluation of the minimax risk in a problem of sequential prediction is presented.

2. MAIN RESULTS

Let X be a random element from a probability space $(\Omega, \mathscr{F}, P)$ to a measurable Borel space $(S, \mathscr{S})$ and let Z be a real-valued random variable on Ω. We consider predicting Z based on the observed value x of X. Suppose X has a density, say p_θ, with respect to a σ-finite measure λ on S, where $\theta \in \Theta = (\theta_1, \theta_2)$ and $-\infty \leqslant \theta_1 < \theta_2 \leqslant \infty$. Suppose also that θ is a random variable having a density g with respect to a σ-finite measure τ on Θ. Put

$$\Theta_g = \{\theta;\, g(\theta) > 0\} \quad \text{and} \quad \mathscr{L}^2 = \{U(X, \theta);\, \int \mathrm{E}_\theta(U^2)\, h(\theta)\, d\theta < \infty\},$$

where $h(\theta) = m(\theta)\, g(\theta)$ with a given strictly positive weight function m. Throughout the paper, we will write $\int f d\theta$ and $\int f dx$ instead of $\int_\Theta f(x, \theta)\, \tau(d\theta)$ and $\int_S f(x, \theta)\, \lambda(dx)$, respectively. Given functions f, g on a measurable space, we also write $f = g$ provided $f = g$ almost surely.

Denote by $B(g)$ the Bayes risk with respect to the prior g, i.e.

$$B(g) = \min_{\delta \in \mathscr{L}^2_X} \int \mathrm{E}_\theta (\delta - Z)^2\, h(\theta)\, d\theta.$$

Herein, $\mathscr{L}_X^2 \subset \mathscr{L}^2$ means the set of all predictors. By δ_g we denote the Bayes predictor of Z with respect to the prior g, that is,

$$\delta_g(X) = \frac{\int Y(X, \theta) p_\theta(X) h(\theta) d\theta}{\int p_\theta(X) h(\theta) d\theta},$$

where $Y = \mathrm{E}_\theta [Z \mid X]$. Clearly, Y is the best mean-square predictor of Z if θ is known. We assume that there exists a random variable $L = L(X, \theta)$ such that $0 < \mathrm{Var}_\theta L < \infty$ and

$$(\mathrm{E}_\theta \delta_g)' = \mathrm{E}_\theta(\delta_g L) \tag{2.1}$$

for all $\theta \in \Theta_g$. The prime $'$ denotes differentiation with respect to θ. The variance $\mathrm{Var}_\theta L$ is called a *generalized Fisher information*. A standard choice of L is $L = (\ln p_\theta(X))'$ but other functions are possible, e.g.

$$L = (\ln p_\theta(X))' + \psi_1 - \psi_2 \mathrm{E}_\theta(\delta_g \psi_1)/\mathrm{E}_\theta(\delta_g \psi_2)$$

with $\psi_1, \psi_2 \in \mathscr{L}^2$ such that $\mathrm{E}_\theta(\delta_g \psi_2) \neq 0$ for all $\theta \in \Theta_g$.

Given a differentiable function, say $H(\theta)$, we set

$$A = \int (H(\theta) b(\theta))' d\theta, \quad B = \int y(\theta) H(\theta) d\theta, \tag{2.2}$$

$$C = \int \frac{H^2(\theta)}{h(\theta) V(\theta)} d\theta, \quad D = \int \frac{[H'(\theta) + H(\theta) \mathrm{E}_\theta L]^2}{h(\theta)} d\theta, \tag{2.3}$$

where $0/0$ is by convention 0. Here and below, $V(\theta) = 1/\mathrm{Var}_\theta L$ and

$$b(\theta) = \mathrm{E}_\theta(\delta_g - Y), \quad y(\theta) = (\mathrm{E}_\theta Y)' - \mathrm{E}_\theta(YL). \tag{2.4}$$

Clearly, $b(\theta) = \mathrm{E}_\theta \delta_g - \mathrm{E}_\theta Z$, so $b(\theta)$ is the bias of δ_g. Theorem 2.1 provides a lower bound on the Bayes risk which is expressed in the bias of the Bayes predictor and the generalized Fisher information.

THEOREM 2.1. *Assume that* $0 < C + D < \infty$. *Then*

$$\begin{aligned} B(g) \geqslant & \int \mathrm{E}_\theta (Z - Y)^2 h(\theta) d\theta + \frac{(A+B)^2}{C+D} \\ & + \int \left(b'(\theta) + y(\theta) - b(\theta) \mathrm{E}_\theta L - \frac{A+B}{C+D} \frac{H(\theta)}{h(\theta) V(\theta)} \right)^2 h(\theta) V(\theta) d\theta \\ & + \int \left(b(\theta) - \frac{A+B}{C+D} \frac{H'(\theta) + H(\theta) \mathrm{E}_\theta L}{h(\theta)} \right)^2 h(\theta) d\theta. \end{aligned} \tag{2.5}$$

Equality holds in (2.5) *iff there exists a function* $c(\theta)$ *such that*

$$\delta_g - Y = c(\theta)(L - \mathrm{E}_\theta L) + b(\theta).$$

Proof. Obviously,

$$B(g) = \int \mathrm{E}_\theta (Y - Z)^2 h(\theta) d\theta + \int \mathrm{E}_\theta (\delta_g - Y)^2 h(\theta) d\theta. \tag{2.6}$$

From the Cauchy–Schwarz inequality we get

$$\mathrm{E}_\theta(\delta_g - Y)^2 \geqslant [\mathrm{E}_\theta(\delta_g - Y)]^2 + \frac{\left(\mathrm{Cov}_\theta(\delta_g - Y, K)\right)^2}{\mathrm{Var}_\theta K} \tag{2.7}$$

for any $K \in \mathscr{L}^2$. Equality occurs in (2.7) iff $\delta_g - Y = c(\theta)(K - \mathrm{E}_\theta K) + \mathrm{E}_\theta(\delta_g - Y)$ with a real-valued function c. Putting

$$K = \left(H'(\theta) + H(\theta) L\right)/h(\theta)$$

in (2.7) and integrating with respect to $h(\theta)$, we arrive at

$$\int \mathrm{E}_\theta(\delta_g - Y)^2 h(\theta)\, d\theta \geqslant \frac{\alpha^2}{\beta} + \int \left[b(\theta) - \frac{\alpha}{\beta}\mathrm{E}_\theta K\right]^2 h(\theta)\, d\theta \tag{2.8}$$
$$+ \int \left[\mathrm{Cov}_\theta(\delta_g - Y, K) - \frac{\alpha}{\beta}\mathrm{Var}_\theta K\right]^2 \frac{h(\theta)}{\mathrm{Var}_\theta K}\, d\theta,$$

where

$$\alpha = \int \mathrm{E}_\theta\left((\delta_g - Y)K\right) h(\theta)\, d\theta = A + B < \infty,$$
$$\beta = \int \mathrm{E}_\theta(K^2)\, h(\theta)\, d\theta = \int \left(\mathrm{Var}_\theta K + (\mathrm{E}_\theta K)^2\right) h(\theta)\, d\theta = C + D < \infty,$$
$$\mathrm{Cov}_\theta(\delta_g - Y, K) = \frac{H(\theta)}{h(\theta)}\left(b'(\theta) + y(\theta) - b(\theta)\,\mathrm{E}_\theta L\right).$$

Combining (2.6) and (2.8) yields (2.5). ∎

Remark 2.1. Adopt assumptions of Theorem 2.1. For the sake of convenience, assume also that $\int \left(H(\theta) b(\theta)\right)' d\theta = 0$ and let $\mathrm{E}_\theta L = 0$. From (2.5) we get

$$B(g) \geqslant \int \mathrm{E}_\theta(Z - Y)^2 h(\theta)\, d\theta + \frac{B^2}{C+D} + \int \left(b(\theta) - \frac{B}{C+D}\frac{H'(\theta)}{h(\theta)}\right)^2 h(\theta)\, d\theta \tag{2.9}$$
$$+ \int \left(b'(\theta) + y(\theta) - \frac{B}{C+D}\frac{H(\theta)}{h(\theta)\, V(\theta)}\right)^2 h(\theta)\, V(\theta)\, d\theta,$$

where

$$B = \int y(\theta) H(\theta)\, d\theta, \qquad C = \int \frac{H^2(\theta)}{h(\theta)\, V(\theta)}\, d\theta, \qquad D = \int \frac{\left(H'(\theta)\right)^2}{h(\theta)}\, d\theta. \tag{2.10}$$

Observe that for $Z = Y = \theta$ and $L = \left(\ln p_\theta(X)\right)'$, we obtain a lower bound different from that in Theorem 2.1 of Brown and Gajek (1990). By (2.9) we obtain the Van Trees inequality for a prediction problem:

$$B(g) \geqslant \int \mathrm{E}_\theta(Z - Y)^2 h(\theta)\, d\theta + \frac{B^2}{C+D}. \tag{2.11}$$

Put $H(\theta) = y(\theta)\, V(\theta)\, h(\theta)$ in (2.11) to get inequality (2.3) of Takada (1999) under weaker assumptions; observe that Assumption 1 of Takada (1999) implies that $y(\theta) = \mathrm{E}_\theta\,[(Y)']$. See also Gajek and Lipińska (2006) for an improvement of (2.11).

We say that Van Trees' inequality (2.11) is *optimal* if it cannot be improved by (2.9). Clearly, (2.11) is optimal iff the bias b of the Bayes estimator δ_g satisfies the system of equations:

$$b = \frac{B}{C+D}\frac{H'}{h}, \qquad b'+y = \frac{B}{C+D}\frac{H}{hV}, \tag{2.12}$$

which gives the following necessary condition: $[(b'+y)\,hV]' = hb$ or, equivalently,

$$hVb''+(hV)'b'-hb+(yhV)' = 0. \tag{2.13}$$

To make the presentation clear, we drop the argument θ in h, V, b and y. For $h = g$ and $Y = \theta$, (2.13) is the Sato–Akahira differential equation (see Sato and Akahira (1996), formula (2.6)). Furthermore, from (2.12) it follows that if (2.11) is optimal, then H satisfies the equation

$$\left(\frac{H'}{h}\right)' - \frac{H}{hV} + cy = 0 \tag{2.14}$$

with a real c. Observe that (2.14) is the Euler–Lagrange condition for the problem of minimization of $C+D$ subject to $\int y(\theta)\, H(\theta)\, d\theta = B$, $H(\theta_0) = H(\theta_1) = 0$. If $H = mgV$ as in Brown and Gajek (1990), then

$$g(\theta) = \frac{I(\theta)}{m(\theta)} \exp\big(\hat{y}(\theta)+d\big),$$

where $\hat{y}$ is such that $\big((\hat{y}(\theta))'/I(\theta)\big)' = 1-cy(\theta)$ and d is such that $\int g\,d\theta = 1$. For instance, if $Y = \theta$ and if $I(\theta) = m(\theta) = 1$ for all $\theta \in \Theta_g = \boldsymbol{R}$, then an optimal prior distribution may be the normal one.

Remark 2.2. Given $L = L(X, \theta) \in \mathscr{L}^2$, we have

$$\int \mathrm{E}_\theta (\delta_g - Y)^2\, h(\theta)\, d\theta \geqslant \frac{\big(\int\int (\delta_g - Y)\big(H'(\theta)+H(\theta)\, L\big)\, p_\theta(x)\, d\theta dx\big)^2}{\int\int \big[\big(H'(\theta)+H(\theta)\, L\big)^2/h(\theta)\big]\, p_\theta(x)\, d\theta dx}$$

by the Cauchy–Schwarz inequality; we do not assume that (2.1) holds. Hence, the following Van Trees type inequality is valid:

$$B(g) \geqslant \int \mathrm{E}_\theta (Z-Y)^2\, h(\theta)\, d\theta + \frac{\big(\int \mathrm{E}_\theta\big[\big(H'(\theta)+H(\theta)\, L\big)\, Y\big]\, d\theta\big)^2}{C+D}, \tag{2.15}$$

where L is such that

$$\int\int \delta_g(x)\big(H'(\theta)+H(\theta)\, L\big)\, p_\theta(x)\, d\theta dx = 0$$

and C, D are defined by (2.3). This becomes equality iff $(\delta_g - Y - c_1 L - c_2) \times p_\theta(X) = 0$, with c_i being a function on Θ for $i = 1, 2$. A standard choice is

$$H = 1 \quad \text{and} \quad L = \sum_{i=1}^{d} \big(a_i(\theta)/p_\theta(X)\big)(\partial^i/\partial\theta^i)\, p_\theta(X).$$

The bound (2.15) with $H = 1$ and $L = \big(Q(X, \theta)p_\theta(X)g(\theta)\big)'/p_\theta(X)$ is due to Gajek and Lipińska (2006). Under some assumptions on g, Q, and p_θ which ensure that $\int \big(Q(X, \theta)p_\theta(X)g(\theta)\big)' d\theta = 0$, they have shown that (2.15) becomes equality in a few prediction problems. Clearly, the inequality (2.15) is the same as (2.11) if $\int \big(H(\theta)\mathrm{E}_\theta Y\big)' d\theta = 0$ but they hold under different assumptions on L and H.

Remark 2.3. Put $\mathrm{E}X = \int \mathrm{E}_\theta X h(\theta)\, d\theta$. Since $\mathrm{E}\delta_g = \mathrm{E}Y$, we have

$$\begin{aligned} B(g) &= \mathrm{E}(Z - Y)^2 + \mathrm{E}(\delta_g - Y)^2 \\ &= \mathrm{E}(Z - Y)^2 + \mathrm{E}(Y - \mathrm{E}Y)^2 - \mathrm{E}(\delta_g - \mathrm{E}\delta_g)^2. \end{aligned} \tag{2.16}$$

Let $L \in \mathscr{L}^2$ be such that (2.1) is satisfied. By the Cauchy–Schwarz inequality we get

$$\mathrm{E}(\delta_g - \mathrm{E}\delta_g)^2 \geqslant \frac{\big(\int (\mathrm{E}_\theta(\delta_g L) - \mathrm{E}_\theta \delta_g \mathrm{E}_\theta L) h(\theta)\, d\theta\big)^2}{\int \mathrm{E}_\theta(L^2) h(\theta)\, d\theta}. \tag{2.17}$$

Combining (2.16) and (2.17) one obtains the *upper bound* for the Bayes risk involving only the bias of the Bayes estimator and the Fisher information:

$$\begin{aligned} B(g) &\leqslant \mathrm{E}(Z - Y)^2 + \mathrm{E}(Y - \mathrm{E}Y)^2 \\ &\quad - \frac{\big(\int (b'(\theta) - b(\theta)\mathrm{E}_\theta L + y(\theta)) h(\theta)\, d\theta\big)^2}{\int \big(\mathrm{Var}_\theta L + (\mathrm{E}_\theta L)^2\big) h(\theta)\, d\theta} \end{aligned} \tag{2.18}$$

with equality iff $\delta_g - \mathrm{E}\delta_g = cL$, where c is real.

The right-hand side of (2.5) depends on the bias of predictor δ_g which may be difficult to calculate. We will get rid of this dependence. To simplify the notation, we assume $Y = Z$. Recall that $y(\theta) = (\mathrm{E}_\theta Y)' - \mathrm{E}_\theta(YL)$. Put

$$l(\theta) = \exp\Big(-\int_{\theta_0}^{\theta} \mathrm{E}_t L\, dt\Big), \tag{2.19}$$

$$\begin{aligned} u_1(\theta) &= \int_{\theta_0}^{\theta} \left(y(t) - \frac{H(t)\, l(t)}{V(t)\, h(t)}\right) l(t)\, dt \\ &\quad + \frac{B}{C + D}\, \frac{\big(H'(\theta) + H(\theta)\mathrm{E}_\theta L\big)\, l(\theta)}{h(\theta)}, \end{aligned} \tag{2.20}$$

$$u_2(\theta) = \frac{1}{C + D}\left(\frac{\big(H'(\theta) + H(\theta)\mathrm{E}_\theta L\big)\, l(\theta)}{h(\theta)} - \int_{\theta_0}^{\theta} \frac{H(t)\, l(t)}{V(t)\, h(t)}\, dt\right), \tag{2.21}$$

$$(2.22)\qquad v(\theta) = -\left(\frac{V(\theta)h(\theta)}{l^2(\theta)}\right)' \frac{1}{\theta-\theta_0} \geqslant 0,$$

$$(2.23)\qquad w(\theta) = \frac{v(\theta)h(\theta)}{v(\theta)l^2(\theta)+h(\theta)},$$

$$(2.24)\qquad \bar{u}_i = \frac{\int u_i(\theta)w(\theta)\,d\theta}{\int w(\theta)\,d\theta} \quad \text{for } i = 1, 2,$$

where B, C, D are given by (2.2) and (2.3). The following result provides a counterpart of Theorem 2.7 of Brown and Gajek (1990).

THEOREM 2.2. *Under the assumptions of Theorem* 2.1, *suppose*

$$\theta \to V(\theta)h(\theta)/l^2(\theta)$$

is absolutely continuous and unimodal with maximum at $\theta_0 \in \Theta$. *Then*

$$(2.25)\qquad B(g) \geqslant \frac{B^2}{C+D} + \int w(\theta)\big(u_1(\theta)-\bar{u}_1\big)^2\,d\theta - \frac{\big(B(C+D)^{-1} + \int w(\theta)\big(u_1(\theta)-\bar{u}_1\big)\big(u_2(\theta)-\bar{u}_2\big)\,d\theta\big)^2}{(C+D)^{-1} + \int w(\theta)\big(u_2(\theta)-\bar{u}_2\big)^2\,d\theta} > 0.$$

Moreover, if $\int (H(\theta)b(\theta))'\,d\theta = 0$, *then*

$$(2.26)\qquad B(g) \geqslant \frac{B^2}{C+D} + \int w(\theta)\big(u_1(\theta)-\bar{u}_1\big)^2\,d\theta.$$

Proof. Put

$$\beta(\theta) = b(\theta)l(\theta) + \int_{\theta_0}^{\theta} \left(y(t) - \frac{A+B}{C+D}\frac{H(t)}{V(t)h(t)}\right) l(t)\,dt.$$

Theorem 2.1 implies

$$B(g) \geqslant \frac{(A+B)^2}{C+D} + Z,$$

where

$$Z = \int \big[\beta(\theta) - \big(u_1(\theta) + Au_2(\theta)\big)\big]^2 \frac{h(\theta)}{l^2(\theta)}\,d\theta + \int \big(\beta'(\theta)\big)^2 \frac{V(\theta)h(\theta)}{l^2(\theta)}\,d\theta.$$

Applying the inequality (2.17) in Brown and Gajek (1990), we obtain

$$Z \geqslant \int \left[\big(\beta(\theta) - \big(u_1(\theta) + Au_2(\theta)\big)\big)^2 \frac{h(\theta)}{l^2(\theta)} + \big(\beta(\theta) - \beta(\theta_0)\big)^2 v(\theta)\right] d\theta.$$

Taking the infimum over all $\beta(\theta)$, we get

$$Z \geqslant \int w(\theta)\big(u_1(\theta) + Au_2(\theta) - \beta(\theta_0)\big)^2\,d\theta.$$

Minimizing the quadratic in $\beta(\theta_0)$ leads to

$$\int \mathrm{E}_\theta (\delta - Y)^2 h d\theta \geqslant \frac{(A+B)^2}{C+D} + \int w(\theta)\big(u_1(\theta) - \bar{u}_1 + A\big(u_2(\theta) - \bar{u}_2\big)\big)^2 d\theta. \tag{2.27}$$

Clearly, if $\int \big(H(\theta) b(\theta)\big)' d\theta = 0$, then $A = 0$. Otherwise, we minimize the quadratic in (2.27) with respect to A to obtain (2.25). ∎

We now derive an analogous result to that of Theorem 2.9 in Brown and Gajek (1990). Let B, C, D, l and u_1 be defined by (2.2), (2.3), (2.19) and (2.20), respectively, with a real $\theta_0 \in \Theta$. Recall that $b(\theta) = \mathrm{E}_\theta(\delta_g - Y)$ and $y(\theta) = (\mathrm{E}_\theta Y)' - \mathrm{E}_\theta(YL)$. Put

$$\beta(\theta) = b(\theta) l(\theta) + \int_{\theta_0}^{\theta} \left(y(t) - \frac{B}{C+D} \frac{H(t)}{h(t) V(t)} \right) l(t) dt,$$

$$A_1 = \int u_1(\theta) M'(\theta) d\theta, \qquad A_2 = \int \frac{\big(M(\theta) l(\theta)\big)^2}{h(\theta) V(\theta)} d\theta,$$

$$A_3 = \int \frac{\big(M'(\theta) l(\theta)\big)^2}{h(\theta)} d\theta,$$

where M is an absolutely continuous function on Θ.

THEOREM 2.3. *Suppose* $\int \big(H(\theta) b(\theta)\big)' d\theta = 0$ *and* $\int \big(\beta(\theta) M(\theta)\big)' d\theta = 0$. *Under the assumptions of Theorem* 2.1 *we have*

$$B(g) \geqslant \frac{B^2}{C+D} + \frac{A_1^2}{A_2 + A_3}. \tag{2.28}$$

Proof. From (2.5) we have

$$\begin{aligned} B(g) &\geqslant \frac{B^2}{C+D} + \int (\beta - u_1)^2 \frac{h}{l^2} d\theta + \int (\beta')^2 \frac{hV}{l^2} d\theta \\ &= \frac{B^2}{C+D} + \frac{A_1^2}{A_2 + A_3} + \int \left(\beta - u_1 + \frac{A_1}{A_2 + A_3} \frac{M' l^2}{h} \right)^2 \frac{h}{l^2} d\theta \\ &\quad + \int \left(\beta' + \frac{A_1}{A_2 + A_3} \frac{M l^2}{hV} \right)^2 \frac{hV}{l^2} d\theta. \ \blacksquare \end{aligned}$$

Remark 2.4. Suppose $H(\theta) b(\theta)$ is an absolutely continuous function and suppose τ is the Lebesgue measure. It is evident that the assumption

$$\int \big(H(\theta) b(\theta)\big)' d\theta = 0 \tag{2.29}$$

holds if $H(\theta) b(\theta)$ has a compact support. Moreover, if $\int \big(H(\theta)^2 / h(\theta)\big) d\theta < \infty$, then

$$\left(\int |H(\theta) b(\theta)| d\theta\right)^2 \leqslant \int \frac{H^2(\theta)}{h(\theta)} d\theta \int b(\theta)^2 h(\theta) d\theta \leqslant \int \frac{H^2(\theta)}{h(\theta)} d\theta B(g) < \infty.$$

This implies (2.29) provided $\Theta = (\theta_1, \infty)$ or $\Theta = (-\infty, \theta_2)$. The condition (2.29) also holds if

$$\int |H'(\theta) b(\theta)| \, d\theta < \infty \quad \text{and} \quad \limsup_{\theta \to \theta_i} |H(\theta)| = \infty, \; i = 1, 2$$

(see Gajek and Kałuszka (1995), Remark 2.3, for a proof).

We now show that the inequalities (2.5), (2.25), (2.26), and (2.28) improve several bounds which can be found in the literature.

EXAMPLE 2.1. Let $Z = Y = \theta$, let $H(\theta) = h(\theta) V(\theta)$, and let the following assumptions be fulfilled:

A1. $(\mathrm{E}_\theta \delta_g)' = \int \delta_g(x) \big(p_\theta(x)\big)' dx, \; \theta \in \Theta_g$;

A2. $0 < I(\theta) < \infty, \; \theta \in \Theta_g$;

A3. $\int \big(H(\theta) b(\theta)\big)' d\theta = 0$;

A4. $\theta \to H(\theta)$ is absolutely continuous and unimodal with maximum at $\theta_0 \in \Theta$,

where $I(\theta)$ stands for the Fisher information and $b(\theta) = \mathrm{E}_\theta \delta_g - \theta$. Put $L = \big(\ln p_\theta(X)\big)'$. Under the assumptions A1–A4, from (2.26) it follows that

$$B(g) \geqslant \frac{C^2}{C+D} + \left(\frac{CD}{C+D}\right)^2 E^{-1}, \tag{2.30}$$

where

$$C = \int h(\theta) V(\theta) \, d\theta, \qquad D = \int \frac{\big((Vh)'(\theta)\big)^2}{h(\theta)} d\theta, \tag{2.31}$$

$$E^{-1} = \int w(\theta) \left(u(\theta) - \frac{\int u(\theta) w(\theta) \, d\theta}{\int w(\theta) \, d\theta} \right)^2 d\theta,$$

with

$$u(\theta) = \frac{\theta}{C} + \frac{(Vh)'(\theta)}{Dh(\theta)}, \qquad v(\theta) = -\frac{(Vh)'(\theta)}{\theta - \theta_0}, \qquad w(\theta) = \frac{v(\theta) h(\theta)}{v(\theta) + h(\theta)}.$$

By Theorem 2.7 of Brown and Gajek (1990),

$$B(g) \geqslant C - \left(\frac{1}{C} + \frac{1}{D} + \frac{1}{E}\right)^{-1}. \tag{2.32}$$

Since for any positive reals C, D, E we have

$$\frac{C^2}{C+D} + \left(\frac{CD}{C+D}\right)^2 E^{-1} = C - \left(\frac{1}{C} + \frac{1}{D} + \frac{1}{E}\right)^{-1} + \left(\frac{1}{C} + \frac{1}{D}\right)^{-2} \left[1 + \left(\frac{1}{C} + \frac{1}{D}\right) E\right]^{-1} E^{-1}, \tag{2.33}$$

the inequality (2.30) is an improvement of the Brown and Gajek inequality (2.32).

EXAMPLE 2.2. Put $M(\theta) = q(\theta)H(\theta)$, where q is an absolutely continuous function. Under the notation and the assumptions A1–A3 of Example 2.1, we infer from Theorem 2.3 that

$$B(g) \geqslant \frac{C^2}{C+D} + \left(\frac{CD}{C+D}\right)^2 E_1^{-1}, \tag{2.34}$$

where

$$E_1^{-1} = \frac{\left(\int (\theta/C + ((Vh)'(\theta))/Dh)(qVh)'\,d\theta\right)^2}{\int q^2 Vh d\theta + \int ([(qVh)']^2/h)\,d\theta}.$$

From Theorem 2.9 of Brown and Gajek (1990) it follows that

$$B(g) \geqslant C - \left(\frac{1}{C} + \frac{1}{D} + \frac{1}{E_1}\right)^{-1}. \tag{2.35}$$

By (2.33), the inequality (2.34) improves (2.35).

EXAMPLE 2.3. Let $X = (X_1, X_2, \ldots)$ be a discrete time stochastic process with values in $S \subset \boldsymbol{R}^\infty$. Assume $X^n = (X_1, \ldots, X_n)$ has a density p_θ^n relative to the Lebesgue measure on $\boldsymbol{R}^n$, with $\theta \in \Theta = (\theta_1, \theta_2)$. We want to predict $Y = Y(X, \theta)$ by

$$\delta^N(X) = \sum_{n=1}^{\infty} \delta_n(X^n)\mathbf{1}(N = n),$$

where N is a proper stopping time (i.e. $P_\theta(N < \infty) = 1$ for all θ) and $Y\mathbf{1}(N = n) = Y_n(X^n, \theta)$ for all n. Hereafter, $\mathbf{1}(c)$ stands for the indicator function, i.e. $\mathbf{1}(c) = 1$ if the condition c is true and $\mathbf{1}(c) = 0$ if it is not so. We assume that $\theta \to \mathrm{E}_\theta Y$ is absolutely continuous. The risk of a sequential predictor $(N, \delta^N(X))$ is given by

$$m(\theta)\mathrm{E}_\theta(\delta^N(X) - Y)^2 + \gamma \mathrm{E}_\theta[(\ln p_\theta^N(X))']^2 \frac{H^2(\theta)}{h^2(\theta)} m(\theta), \tag{2.36}$$

where γ is a given positive real and $p_\theta^N(X) = \sum_{n=1}^{\infty} p_\theta^n(X^n)\mathbf{1}(N = n)$. The second term of the right-hand side of (2.36) is a penalty based upon the Fisher information.

Let $(N^*, \delta_g^{N^*})$ be a sequential Bayes predictor in a class $\mathcal{N}$, that is,

$$\delta_g^{N^*}(X) = \sum_{n=1}^{\infty} \frac{\int Y_n(X^n, \theta)\, p_\theta^n(X^n)\, h(\theta)\, d\theta}{\int p_\theta^n(X^n)\, h(\theta)\, d\theta} \mathbf{1}(N^* = n),$$

where N^* is a minimizer of the risk (2.36) in $\mathcal{N}$ with $\delta^N = \delta_g^N$. Put $L = (\ln p_g^{N^*}(X))'$ and assume that for all $\theta \in \Theta_g$

C1. $(\mathrm{E}_\theta \delta_g^{N^*})' = \mathrm{E}_\theta(\delta_g^{N^*} L)$ and $\mathrm{E}_\theta L = 0$;

C2. $0 < \mathrm{Var}_\theta L < \infty$;

C3. $\int (H(\theta)\mathrm{E}_\theta(\delta_g^{N^*} - Y))' d\theta = 0$.

From (2.11) we get the Koike type bound:

$$B(g) \geqslant \frac{B^2}{C+D} + \gamma C \geqslant \begin{cases} B^2/D & \text{for } B/D < \sqrt{\gamma}, \\ 2B\sqrt{\gamma} - \gamma D & \text{otherwise}, \end{cases} \tag{2.37}$$

where

$$B = \int (\mathrm{E}_\theta Y)' H(\theta)\, d\theta, \qquad C = \int \frac{H^2(\theta)}{h(\theta)} \mathrm{Var}_\theta L d\theta, \qquad D = \int \frac{[H'(\theta)]^2}{h(\theta)} d\theta.$$

Suppose (X_i) are i.i.d. random variables and suppose $0 < I_1(\theta) < \infty$ for all θ, where $I_1(\theta)$ denotes the Fisher information of X_1. If $H(\theta) = g(\theta)/I_1(\theta)$ and if $m(\theta) = 1$, then from (2.37) we obtain Theorem 1 of Koike (1999). The bound (2.37) is also valid under the assumptions different from C1–C3 (see Remark 2.2). Clearly, applying Theorem 2.2 or 2.3 one can improve the Koike inequality.

Suppose $\mathcal{N}$ is a class of sequential rules such that for every $N \in \mathcal{N}$ the function $\theta \to \mathrm{E}_\theta \delta_g^N$ is absolutely continuous, $\mathrm{Var}_\theta[(\ln p_\theta^N(X))'] = I_1(\theta)\mathrm{E}_\theta N$, and C1 is satisfied. Putting $H(\theta) = g(\theta)$ and $m(\theta) = I_1(\theta)$ in (2.37), we obtain the following bound for the minimax risk:

$$\inf_{N,\delta^N} \sup_{\theta\in\Theta} [I_1(\theta)\mathrm{E}_\theta(\delta^N - Y)^2 + \gamma \mathrm{E}_\theta N] \geqslant \max(B_1^2/D_1,\ 2B_1\sqrt{\gamma} - \gamma D_1) \tag{2.38}$$

for every $g \in C^1(\Theta)$ with a compact support, where

$$B_1 = \int (\mathrm{E}_\theta Y)' g(\theta)\, d\theta, \qquad D_1 = \int \frac{(g'(\theta))^2}{g(\theta) I_1(\theta)} d\theta.$$

The explicit form of the bound (2.38) can be obtained by solving a variational problem. For example, if $\Theta = (\theta_1, \infty)$ and if there exist finite limits $\beta := \lim_{\theta\to\infty}(\mathrm{E}_\theta Y)' > 0$ and $\alpha := \lim_{\theta\to\infty} I(\theta) > 0$, then

$$\inf_{N,\delta^N} \sup_{\theta\in\Theta} [I_1(\theta)\mathrm{E}_\theta(\delta^N - Y)^2 + \gamma \mathrm{E}_\theta N] \geqslant \max\left(\frac{\beta^2}{4\pi^2}\alpha,\ 2\beta\sqrt{\gamma} - \gamma\frac{4\pi^2}{\alpha}\right). \tag{2.39}$$

To obtain (2.39) put $g(\theta) = g_0(\theta - n)$, where g_0 has the support $[0, 1]$, take the limit of the right-hand side of (2.38) as $n \to \infty$, and use the well-known fact that the infimum of the functional $\int_0^1 \left((g_0'(t))^2/g_0(t)\right) dt$ over the set of probability densities g_0 which are twice differentiable with $g_0(0) = g_0(1) = 0$ is equal to $4\pi^2$.

REFERENCES

[1] E. N. Belitser and B. Y. Levit, *Asymptotically minimax nonparametric regression in* L_2, Statistics 28 (1996), pp. 105–132.

[2] J. F. Bjørnstad, *On the generalization of the likelihood function and the likelihood principle*, J. Amer. Statist. Assoc. 91 (1996), pp. 791–806.

[3] B. Z. Bobrovsky, E. Mayer-Wolf and M. Zakai, *Some classes of global Cramér–Rao bounds*, Ann. Statist. 15 (1987), pp. 1421–1438.

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

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

[6] L. Gajek and M. Kałuszka, *Lower bounds for the asymptotic Bayes risk in the scale model* (*with an application to the second order minimax estimation*), Ann. Statist. 22 (1994), pp. 1831–1839.

[7] L. Gajek and M. Kałuszka, *Nonexponential applications of a global Cramér–Rao inequality*, Statistics 26 (1995), pp. 111–122.

[8] L. Gajek and V. Lipińska, *Sharp inequality for the Bayes prediction risk*, Inequality Theory and Applications 5 (2006), pp. 32–45.

[9] R. D. Gill and B. Y. Levit, *Applications of the Van Trees inequality: a Bayesian Cramér–Rao bound*, Bernoulli 1 (1995), pp. 59–79.

[10] A. Jokiel-Rokita, *Minimax prediction under random sample size*, Appl. Math. 29 (2002), pp. 127–134.

[11] M. Kałuszka, *Influence of censorship on the minimax risk of decision procedures*, Statistics 29 (1997), pp. 169–178.

[12] K. Koike, *A lower bound for the Bayes risk in the sequential case*, Comm. Statist. Theory Methods 28 (1999), pp. 857–871.

[13] Y. A. Kutoyants and V. Spokoiny, *Optimal choice of observation window for Poisson observations*, Statist. Probab. Lett. 44 (1999), pp. 291–298.

[14] Y. Miyata, *The lower bound for MSE in statistical prediction theory*, J. Japan Statist. Soc. 31 (2001), pp. 111–127.

[15] B. Mizera, *Lower bounds on the minimax risk of sequential estimators*, Statistics 28 (1996), pp. 123–129.

[16] A. Munk and F. Ruymgaart, *Minimax rates for estimating the variance and its derivatives in nonparametric regression—an application of the Van Trees inequality*, Aust. N.Z.J. Stat. 44 (2002), pp. 479–488.

[17] T. K. Nayak, *Rao–Cramér type inequalities for mean squared error of prediction*, Amer. Statist. 56 (2002), pp. 102–106.

[18] M. Sato and M. Akahira, *An information inequality for the Bayes risk*, Ann. Statist. 24 (1996), pp. 2288–2295.

[19] Y. Takada, *Lower bounds on the Bayes risk for statistical prediction problems*, Comm. Statist. Theory Methods 28 (1999), pp. 693–703.

[20] H. Van Trees, *Detection, Estimation and Modulation Theory*, Wiley, New York 1968.

[21] S. Trybuła, *Minimax mutual prediction of multinomial random variables*, Appl. Math. 30 (2003), pp. 371–377.

[22] M. Wilczyński, *Minimax nonparametric prediction under random sample size*, Sci. Math. Jpn. 53 (2001), pp. 401–409.

[23] Y. G. Yatracos, *On prediction and mean squared error*, Canad. J. Statist. 20 (1992), pp. 187–200.

[24] T. Zhang and M. Woodroofe, *Admissible minimax estimation of the signal with known background*, Statist. Sinica 15 (2005), pp. 59–72.

Institute of Mathematics
Technical University of Łódź
ul. Wólczańska 215
93-005 Łódź, Poland
E-mail: kaluszka@p.lodz.pl

Received on 7.12.2005