INFORMATION INEQUALITIES FOR THE BAYES RISK

By Lawrence D. Brown[1] and Leslaw Gajek

Cornell University and Instytut Matematyki

This paper presents lower bounds, derived from the information inequality, for the Bayes risk under scaled quadratic loss. Some numerical results are also presented which give some idea concerning the precision of these bounds. An appendix contains a proof of the information inequality without conditions on the estimator. This result is a direct extension of an earlier result of Fabian and Hannan.

1. Introduction. The principal results of this paper are a series of lower bounds for the Bayes risk of an estimator under scaled quadratic loss. These bounds are all derived from the information inequality. The first explicit bound is in Corollary 2.3. Improved bounds are in Theorem 2.7 and Theorem 2.9; another is in Theorem 2.10.

The bound in Corollary 2.3 was also established earlier by Borovkov and Sakhanienko (1980) using a different method of proof. Their proof actually involves slightly milder regularity conditions. It does not, however, make explicit the connection to the information inequality, nor does it seem to lead to results like Theorems 2.7, 2.9 or 2.10.

Section 3 of the paper contains several examples which illustrate the applicability and the degree of precision of the bounds in Corollary 2.3 and Theorem 2.7. The bound in Corollary 2.3 is sharp in the case of estimating the expectation parameter of an exponential family under a conjugate prior. Generally speaking, the bounds become less precise as one moves away from such a situation. Hence, in general, the bounds become more nearly precise for larger sample sizes.

Conventional proofs of the information inequality generally contain an assumption about the estimator in addition to assumptions about the family of distributions. [See, for example, Cramér (1946) or, more recently, Lehmann (1983), Theorem II. 6.4.] Such assumptions are inconvenient or inappropriate for various rigorous applications of the information inequality such as those here or in the sequel to this paper, Brown and Low (1991). Furthermore, such assumptions are unnecessary. Fabian and Hannan (1977) and Simons and Woodroofe (1983) have proved versions of the information inequality which avoid such inappropriate conditions. [See also Pitman (1979).] The Appendix begins with the result of Fabian and Hannan and then provides more convenient regularity conditions which imply those of Fabian and Hannan. This yields conditions (A6) or (A7), which are of a familiar form. We note, however,

---

that the approach of Simons and Woodroofe, based on the Hellinger distance may be satisfactory for some cases where the approach of the Appendix, based on the Hilbert norm, fails to apply.

One possible application for lower bounds on the Bayes risk, such as those given here, is to obtain lower bounds on the minimax risk. This subject is explored in Brown and Low (1991).

Lower bounds on the Bayes risk are also a convenient avenue to certain classical asymptotic results such as the fact that (in standard situations) the set of superefficiency has Lebesgue measure zero. However, for the full statement of such results one really needs a bound on the Bayes risk under truncated quadratic risk. Such a bound is developed and used in Brown (1988).

2. Inequalities for quadratic loss. This section presents a lower bound on the Bayes risk of an estimator under quadratic loss. The bound is derived directly from the usual information inequality.

Setting. Let X be an observable random variable with probability densities p_θ relative to some σ-finite measure ν. Assume $\theta \in \Theta$, where $\Theta \in \mathbb{R}$ is a (possibly infinite) interval. It is desired to estimate θ by $a \in \mathbb{R}$ under loss $L(\theta, a) = m(\theta)(a - \theta)^2$, where $m > 0$ is a specified weight function. Let $R(\theta, \delta) = E_\theta(L(\theta, \delta(X)))$ denote the risk of the nonrandomized estimator δ. Let $\Theta°$ denote the interior of Θ.

Let $g(\cdot)$ be a probability density with respect to Lebesgue measure on Θ. This is the prior density. For any estimator δ, let $B(g, \delta) = \int R(\theta, \delta)g(\theta)\,d\theta$ and let $B(g) = \inf_\delta B(g, \delta)$. $B(g)$ is the Bayes risk under g. Let $\mathrm{sp}(g) = \{\theta: g(\theta) > 0\}$ and let $\mathrm{csp}(g)$ denote its closure.

Information inequality. To state the information inequality, define $e(\theta) = E_\theta(\delta)$ whenever it exists. Let $I(\theta)$ denote the Fisher information. Usually,

$$I(\theta) = E_\theta\!\left(\!\left(\frac{\partial}{\partial\theta}\ln p_\theta(X)\right)^2\right).$$

[In unusual cases it may be sensible to define I even when $(\partial/\partial\theta)\ln p_\theta(x)$ fails to exist. See (A.2) of the Appendix.] For notational convenience, define $V(\theta) = I^{-1}(\theta)$ and assume $0 < V(\theta) \leq \infty$, for all $\theta \in \Theta$. The following conclusions of the information inequality will be needed for later application. These will be referred to as Assumption 1.

ASSUMPTION 1. (Information inequality).   Let $\theta_0 \in \Theta$. If $\mathrm{Var}_{\theta_0}(\delta) < \infty$, then $e(\theta)$ exists on a neighborhood of θ_0, is differentiable at θ_0 and

$$(2.1) \qquad\qquad \mathrm{Var}_{\theta_0}(\delta) \geq V(\theta_0)\big(e'(\theta_0)\big)^2.$$

The validity of Assumption 1 requires some regularity conditions on the family $\{p_\theta\}$; no conditions are needed on the estimator, δ, other than the

condition $\text{Var}_{\theta_0}(\delta) < \infty$ present in the statement of the assumption. The Appendix discusses this fact and Theorem A1 asserts that Assumption 1 is valid if $\{p_\theta\}$ satisfies condition (A.3) and one of the conditions (A.4)–(A.7).

The main results of this section follow from the formula contained in Theorem 2.1.

THEOREM 2.1. *Assume that $V(\cdot)$, $m(\cdot)$ and $g(\cdot)$ are absolutely continuous on Θ, that $\text{csp}(g)$ is compact with $\text{csp}(g) \subset \Theta°$ and that Assumption 1 is valid for all $\theta \in \Theta°$. Then*

$$(2.2) \qquad B(g, \delta) \geq \frac{C^2}{C + D} + Q_1(b) + Q_2(b) + Q_3(b),$$

where $h(\theta) = m(\theta)g(\theta)$, $b(\theta) = e(\theta) - \theta$ and

$$(2.3) \qquad C = \int V(\theta)h(\theta)\,d\theta,$$

$$(2.4) \qquad D = \int \frac{[(Vh)'(\theta)]^2}{h(\theta)}\,d\theta,$$

$$(2.5) \qquad \gamma = \gamma(b) = \int b(\theta)(Vh)'(\theta)\,d\theta = -\int b'(\theta)V(\theta)h(\theta)\,d\theta,$$

$$(2.6) \qquad Q_1(b) = \frac{C + D}{CD}\left(\gamma - \frac{CD}{C + D}\right)^2 \geq 0,$$

$$(2.7) \qquad Q_2(b) = \int \left(b'(\theta) + \frac{\gamma}{C}\right)^2 V(\theta)h(\theta)\,d\theta \geq 0,$$

$$(2.8) \qquad Q_3(b) = \int \left(b(\theta) - \frac{\gamma}{D}\frac{(Vh)'(\theta)}{h(\theta)}\right)^2 h(\theta)\,d\theta \geq 0.$$

PROOF. The final equality in (2.5) follows from integration by parts. By virtue of Assumption 1,

$$(2.9) \quad R(\theta, \delta) = m(\theta)\{\text{Var}_\theta(\delta) + b^2(\theta)\} \geq m(\theta)\{V(\theta)(1 + b'(\theta))^2 + b^2(\theta)\}.$$

Hence

$$B(g, \delta) \geq \int \{(1 + b'(\theta))^2 V(\theta) + b^2(\theta)\}m(\theta)g(\theta)\,d\theta$$

$$= \int V(\theta)h(\theta)\,d\theta + 2\int b'(\theta)V(\theta)h(\theta)\,d\theta$$

$$(2.10) \qquad + \int (b'(\theta))^2 V(\theta)h(\theta)\,d\theta + \int b^2(\theta)h(\theta)\,d\theta$$

$$= C - 2\gamma + \int (b'(\theta))^2 V(\theta)h(\theta)\,d\theta + \int b^2(\theta)h(\theta)\,d\theta$$

$$= C - 2\gamma + \frac{\gamma^2}{C} + Q_2(b) + \frac{\gamma^2}{D} + Q_3(b),$$

since

$$(2.11) \qquad Q_2(b) = \int (b'(\theta))^2 V(\theta) h(\theta) - \frac{\gamma^2}{C}$$

by (2.5) and $Q_3(b) = \int b^2(\theta) h(\theta) \, d\theta - \gamma^2/D$. The expression on the right of (2.10) is a quadratic form in γ. Let $\gamma_0 = (C^{-1} + D^{-1})^{-1}$ and complete the square on the right of (2.10) to get

$$B(g, \delta) \geq C - \gamma_0 + \gamma_0^{-1}(\gamma - \gamma_0)^2 + Q_2(b) + Q_3(b),$$

which is the same as (2.2). $\square$

REMARK 2.2. The assumption that csp(g) be compact with csp(g) $\subset \Theta^\circ$ is much stronger than necessary. It can be replaced by the weaker assumption that $D < \infty$ and the integration by parts in (2.5) is valid. [If $D < \infty$ and $C = \infty$, it is easy to show from (2.10) that $B(g, \delta) = \infty$.]

Corollary 2.3 is an immediate consequence of (2.2) since Q_i, $i = 1, 2, 3$, are all nonnegative quadratic forms in b and b'.

COROLLARY 2.3 [Borokov and Sakhanienko (1980)]. *Under the assumptions of Theorem 2.1,*

$$(2.12) \qquad B(g) \geq \frac{C^2}{C + D} = C - \left(\frac{1}{C} + \frac{1}{D} \right)^{-1}.$$

REMARK 2.4. The important inequality (2.12) is due to Borovkov and Sakhanienko (1980), and was established by them under somewhat weaker conditions than required in Corollary 2.3 or 2.7. Their method of proof is somewhat different than that used above and does not involve a result like Theorem 2.1. It is not at all clear to us whether one can use their methods to yield the somewhat improved inequalities contained in Theorems 2.7, 2.9 and 2.10.

An inequality related to (2.12) is developed in Bobrovsky, Mayer-Wolf and Zakai (1987) building from an earlier result of Van Trees (1968). That inequality agrees with (2.12) when $I(\theta)$ is constant. It is weaker in a number of other common examples, although there also exist examples where it is stronger than (2.12) or (2.16).

REMARK 2.5. The right side of (2.12) depends on V, h since $C = C(V, h)$, $D = D(V, h)$. Let V_1, h_1 be any two absolutely continuous functions with compact support. Suppose $0 \leq h_1 \leq h$ and $0 \leq V_1 h_1 \leq V h$. Then

$$(2.12') \qquad B(g) \geq \frac{C^2(V_1, h_1)}{C(V_1, h_1) + D(V_1, h_1)},$$

since $B(g, \delta) \geq \int (1 + b')^2 V h + \int b^2 h \geq \int (1 + b')^2 V_1 h_1 + \int b^2 h_1$ and the derivation in Theorem 2.1 can be legitimately applied to the right side of this

inequality. Similar remarks apply to inequalities stated later in this paper. The following formal statement is one consequence of this remark. It extends Corollary 2.3 to cases where $\mathrm{csp}(g)$ is not compact in $\Theta°$.

COROLLARY 2.6. *Assume that $V(\cdot)$ and $h(\cdot)$ are absolutely continuous on $\Theta° = (a, b)$ [with $(-a), b \leq \infty$] and that Assumption 1 is valid for all $\theta \in \Theta°$. Assume $D < \infty$. Let $a < c < b$.*
  If $b = \infty$ ($a = -\infty$, resp.), assume

$$(2.13a) \qquad \lim_{i \to \infty} i^{-2} \int_c^i V^2(\theta) h(\theta)\, d\theta = 0 \qquad \left[\lim_{i \to \infty} i^{-2} \int_{-i}^c \ , resp \right].$$

If $b < \infty$ ($a > -\infty$, resp.), assume

$$\lim_{i \to \infty} (\ln i)^{-2} \int_c^{b-i^{-1}} (b - \theta)^{-2} V^2(\theta) h(\theta)\, d\theta = 0$$

$$(2.13b)$$

$$\left[\lim_{i \to \infty} (\ln i)^{-2} \int_{a+i^{-1}}^c (\theta - a)^{-2} V^2(\theta) h(\theta)\, d\theta = 0, resp. \right].$$

Then (2.12) is valid.

PROOF. Consider the case where $a = -\infty$, $b = \infty$. Let $k_i(\theta) = [(i - |\theta|)^+/i]$ and let $h_i(\theta) = k_i^2(\theta) h(\theta)$. Then $B(g) \geq C_i^2/(C_i + D_i)$, where $C_i = C(V, h_i)$, $D_i = D(V, h_i)$, by Remark 2.5. $C_i \to C$, since $h_i \leq h$ and $h_i \to h$. Also, direct calculation and the Cauchy–Schwarz inequality yield

$$(2.14) \qquad \left| \int_{-\infty}^\infty \frac{[(Vh)'(\theta)]^2}{h(\theta)}\, d\theta - \int_{-\infty}^\infty \frac{[(Vh_i)'(\theta)]^2}{h_i(\theta)}\, d\theta \right|$$

$$\leq \int_{-\infty}^\infty \frac{[(Vh)'(\theta)]^2}{h(\theta)} \left(1 - k_i^2(\theta)\right) d\theta + 2D^{1/2} m_i + m_i^2,$$

where

$$m_i^2 = \int_{-i}^i V^2(\theta) \left[\frac{k'_i(\theta)}{k_i(\theta)} \right]^2 k_i^2(\theta) h(\theta)\, d\theta = \frac{4}{i^2} \int_{-i}^i V^2(\theta) h(\theta)\, d\theta.$$

Note that $m_i \to 0$ by (2.13a), so that $D_i \to D$ by (2.14) and dominated convergence. Hence $C_i^2/(C_i + D_i) \to C^2/(C + D)$ and the corollary is proved in this case.
  The proof for $a = -\infty$, $b < \infty$ is similar except that for $\theta > c$ the definition of k_i should be replaced by $k_i(\theta) = (\ln(i(b - c)))^{-1} \ln^+(i(b - \theta))$. Entirely analogous arguments apply to the other possibilities for a and b. □

The inequality in Corollary 2.3 can generally be improved by a more careful treatment of the terms Q_i in Theorem 2.12. Here is our principal result in this direction.

THEOREM 2.7. *Make the assumptions in Theorem 2.1. Assume also that* (Vh) *is unimodal with maximum at* $\theta_0 \in \Theta^\circ$. *Let*

$$u(\theta) = \frac{\theta}{C} + \frac{(Vh)'(\theta)}{Dh(\theta)},$$

$$v(\theta) = -\frac{(Vh)'(\theta)}{\theta - \theta_0} \geq 0,$$

$$w(\theta) = \frac{v(\theta)h(\theta)}{v(\theta) + h(\theta)}$$

and

(2.15) $$E^{-1} = \int (u(\theta) - \bar{u})^2 w(\theta)\, d\theta,$$

where

$$\bar{u} = \frac{\int u(\theta) w(\theta)\, d\theta}{\int w(\theta)\, d\theta}.$$

Then

(2.16) $$B(g) \geq C - \left(\frac{1}{C} + \frac{1}{D} + \frac{1}{E}\right)^{-1}.$$

PROOF. Let $\beta(\theta) = b(\theta)/\gamma + \theta/C$ so that $Q_2 = \gamma^2 \int (\beta')^2 Vh$ and $Q_3 = \gamma^2 \int (\beta - u)^2 h$. Now note the inequality

(2.17) $$\int (\beta(\theta) - \beta(\theta_0))^2 v(\theta)\, d\theta \leq \int (\beta'(\theta))^2 V(\theta) h(\theta)\, d\theta,$$

which follows upon applying the Cauchy–Schwarz inequality to $\beta(\theta) - \beta(\theta_0) = \int_{\theta_0}^{\theta} \beta'(t)\, dt$ and then interchanging the order of integration. Let $\beta_0 = \beta(\theta_0)$. Then

$$Q_2 + Q_3 \geq \gamma^2 \int \left\{(\beta - \beta_0)^2 v + (\beta - \beta_0 + \beta_0 - u)^2 h\right\}$$

$$= \gamma^2 \int \left\{(\beta - \beta_0)^2 (v + h) + 2(\beta - \beta_0)(\beta_0 - u)h + (\beta_0 - u)^2 h\right\}.$$

Choose $\beta - \beta_0$ to minimize the integrand on the right and get

$$Q_2 + Q_3 \geq \gamma^2 \int (u - \beta_0)^2 w \geq \gamma^2 \int (u - \bar{u})^2 w = \gamma^2 E^{-1}.$$

Thus,

(2.18)
$$Q_1 + Q_2 + Q_3 \geq \left(\frac{1}{C} + \frac{1}{D}\right)\left(\gamma - \frac{CD}{C + D}\right)^2 + \gamma^2 E^{-1}$$

$$\geq -\left(\frac{1}{C} + \frac{1}{D} + \frac{1}{E}\right)^{-1} + \left(\frac{1}{C} + \frac{1}{D}\right)^{-1},$$

where the second inequality results from minimizing the quadratic in γ. Combined with Theorem 2.1, this is the desired result. $\square$

REMARK 2.8. The preceding result can be extended to cases where Vh is not unimodal, as follows. Let $0 < h_1 \le h$, where Vh_1 is unimodal and h_1 is absolutely continuous. Substitute h_1 in the definition of v; call the modified value v_1. Then let $w_1 = v_1 h / (v_1 + h)$ and define E^{-1} as in (2.15) but using w_1 in place of w. The conclusion (2.16) of the theorem is then valid with this modified value of E since (2.17) is valid with v_1 in place of v. (Remark 2.5 is of course also relevant to Theorem 2.7 and a result analogous to Corollary 2.6 can be proved.)

The alternate result below does not rely in any way on unimodality of (Vh). In the examples presented in Section 3, (2.16) is slightly better than (2.20) and (2.21), but there are other examples where the latter does better.

THEOREM 2.9. *Make the assumptions in Theorem 2.1. Let $q(\theta) \ge 0$ be absolutely continuous. Let $u(\theta)$ be as in Theorem 2.7 and let*

$$A_1 = \int u(\theta)(qVh)'(\theta)\, d\theta,$$

$$A_2 = \int q^2(\theta)V(\theta)h(\theta)\, d\theta,$$

(2.19)

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

$$E_1 = \frac{A_2 + A_3}{A_1^2}.$$

Assume $E_1 < \infty$. Then

(2.20) $$B(g) \ge C - \left(\frac{1}{C} + \frac{1}{D} + \frac{1}{E_1}\right)^{-1}.$$

REMARK. In several examples the choice

(2.21) $$q(\theta) = u'(\theta)V(\theta)h(\theta)$$

seems to yield satisfactory results in (2.20).

PROOF. The Cauchy–Schwarz inequality, integration by parts and minimization of a quadratic yields $Q_2 + Q_3 \ge \gamma^2 E_1^{-1}$. Proceeding then as in the proof of Theorem 2.7 completes the argument. $\square$

The preceding results all began by writing $e(\theta) = \theta + b(\theta)$ in the proof of Theorem 2.1. It is possible to fix $\alpha(\theta)$ and begin by writing $e(\theta) = \theta + \alpha(\theta) + a(\theta)$. This results in a formally different version of Theorem 2.1 and in

different inequalities analogous to those in Corollary 2.3 and Theorem 2.7. Theorem 2.10 expresses the analog of Corollary 2.3. Its proof is similar in style to that of Corollary 2.3 and is thus omitted.

THEOREM 2.10. *Make the assumptions in Theorem 2.1. Let $\alpha(\cdot)$ be an absolutely continuous function. Define*

$$r(\theta) = \alpha(\theta)h(\theta) - ((1 + r'(\theta))V(\theta)h(\theta))'.$$

Assume that $\int_{\Theta} r(\theta)\,d\theta = 0$, and define

$$R(\theta) = \int_{-\infty}^{\theta} r(t)\,dt.$$

Then

$$(2.22) \quad B(g) \geq C(\alpha) - \left[\left[\int \frac{R^2(\theta)}{V(\theta)h(\theta)}\,d\theta\right]^{-1} + \left[\int \frac{r^2(\theta)}{h(\theta)}\,d\theta\right]^{-1}\right]^{-1},$$

where

$$(2.23) \quad C(\alpha) = \int (1 + \alpha'(\theta))^2 V(\theta)h(\theta)\,d\theta + \int \alpha^2(\theta)h(\theta)\,d\theta.$$

Note that (2.22) is identical to (2.12) when $\alpha(\theta) = 0$ so that $r = (Vh)'$ and $R = Vh$.

REMARK. There evidently exist multivariate extensions of the above results. One such extension (not the best possible) is presented in Theorem 2.1′ and Corollary 2.2′ of Brown (1986); see also Shemyakin (1987).

3. Examples. The examples in this section display some feasible applications of the inequalities in Corollary 2.3 and Theorem 2.7. They have been chosen in part to display the degree of numerical accuracy of these inequalities. Some of the examples are also related to the minimax analyses discussed in Brown and Low (1991).

EXAMPLE 3.1. Let $m(\theta) \equiv 1$. Suppose $\{p_\theta\}$ is an exponential family with expectation parameter θ and g is a conjugate prior. Then (and only then), subject to mild regularity conditions, (2.12) is actually an equality. This follows since the Bayes procedure in this case is linear. This linearity is sufficient (and necessary) for equality at all θ in the information inequality. [See Lehmann (1983), page 123 and references cited therein.] Furthermore in this case both b and $(Vh)'/h$ are linear, and, as it is easy to check, it then follows that $Q_2 = Q_3 = 0$ so that (2.12) is an equality. If $m(\theta) \neq 1$, then equality holds in (2.12) if and only if h is proportional to a conjugate prior density. [The regularity conditions needed for the direct assertion are those in Corollary 2.6. They are usually (always?) satisfied if the Bayes risk is finite. Concerning the converse assertion, see Wijsman (1973), Joshi (1976) and Müller-Funk, Puckelsheim and Witting (1989)].

EXAMPLE 3.2. Note that the bounds in Section 2 rely on X only through the Fisher information of $\{p_\theta\}$. Thus the same bound applies to all families with the same information function. In particular, suppose p_θ is a location family, $p_\theta(x) = p(x - \theta)$, $m(\theta) \equiv 1$ and $g = N(\mu, \sigma^2)$ is normal with mean μ and variance σ^2. Then

$$(3.1) \qquad B(g) \geq \frac{C^2}{C + D} = \frac{\sigma^2}{i\sigma^2 + 1},$$

where $i = \int((p')^2/p)$ is the (constant) information for $\{p_\theta\}$. From Example 3.1, this bound is sharp if and only if $\{p_\theta\}$ is a normal location family.

Similarly, suppose $p_\theta(x) = \theta^{-1}p(x/\theta)$ is a scale family and g is an inverse gamma density with index λ and scale β (i.e., θ^{-1} has the indicated gamma density). The information function is $I(\theta) = i/\theta^2$. In this case it is natural to choose $m(\theta) = i/\theta^2$ to normalize the loss. Then

$$(3.2) \qquad B(g) \geq \frac{1}{i + \lambda + 1}.$$

This bound is sharp if and only if p_θ is a gamma density with index α and scale θ since then $h = mg$ is proportional to a conjugate prior density. See Gajek (1988) for further results related to (3.1) and (3.2).

EXAMPLE 3.3. Now suppose $X \sim N(\theta, 1)$, $m \equiv 1$ and

$$(3.3) \qquad g(\theta) = \frac{1}{L}\cos^2\frac{\pi}{2L}\theta, \qquad |\theta| \leq L.$$

Then Corollary 2.3 yields

$$(3.4) \qquad B(g) \geq \frac{1}{1 + \pi^2/L^2},$$

as noted in Borovkov and Sakhanienko (1980). See also Bickel (1981). Since this is not a conjugate prior this bound can be improved by use of Theorem 2.7. Direct manipulation yields

$$(3.5) \qquad E^{-1} = 4\frac{L^2}{\pi}\int_0^{\pi/2}(2t - \tan t)^2\frac{\cos^2 t}{\pi^2 + 2L^2 t \cot t}\,dt.$$

This expression can easily be numerically evaluated.

Table 1 gives values of the bound in (3.4), of the improved bound

$$(3.6) \qquad B(g) \geq 1 - \left(1 + \frac{L^2}{\pi^2} + E^{-1}\right)^{-1}$$

obtained from (2.16) of Theorem 2.7 and of the actual Bayes risk as evaluated by a rather lengthy, two-stage numerical integration. From this table one can see that here Theorem 2.7 provides a moderate improvement to Corollary 2.3. Note, as earlier, that these bounds on $B(g)$ apply to any family with constant

TABLE 1
Values of the bounds in Example 3.3

L	Bound(3.4)	Bound(3.6)	Bayes risk $B(g)$
0.5	0.02470	0.03145	0.03164
1	0.09198	0.1135	0.1156
2	0.2884	0.3305	0.3427
3	0.4770	0.5177	0.5375
5	0.7170	0.7399	0.7578
10	0.9102	0.9150	0.9214
20	0.9759	0.9765	0.9779

information. [For constant information $I(\theta) \equiv c$, the bounds should of course be evaluated at Lc and be multiplied by c^{-1}.]

EXAMPLE 3.4. Let X have a noncentral χ^2 distribution with p degrees of freedom and noncentrality parameter θ. Let $m \equiv 1$. Suppose θ is distributed under g as σ^2 times a central χ_p^2 variable. This prior is plausible for such a situation (as would also be any other gamma distribution) and has been specially chosen for the discussion here because its Bayes risk can be explicitly evaluated. Its value is

$$(3.7) \qquad B(g) = \frac{4p\sigma^4}{\sigma^2 + 2}\left(1 + \frac{\sigma^2}{2(\sigma^2 + 1)^2}\right),$$

corresponding to a Bayes estimator $\delta(x) = \sigma^4(1 + \sigma^2)^{-2}x + p\sigma^2(1 + \sigma^2)^{-1}$. Since the distributions of X are not an exponential family, the bounds of Section 2, which are calculated next, do not attain this value.

The information in X has no convenient closed form expression. However a useful bound is given by

$$(3.8) \qquad I(\theta) \le \frac{1}{4\theta}.$$

This can be verified as follows: For $p = 1$,

$$\begin{aligned} I(\theta) &= E\left(-\frac{\partial^2}{\partial\theta^2}\ln p_\theta(X)\right) \\ (3.9) \qquad &= \frac{1}{4\theta} - \frac{e^{-\theta/2}}{2\theta}\int_0^\infty \left(e^{x\sqrt{\theta}} + e^{-x\sqrt{\theta}}\right)^{-1}\sqrt{\frac{2}{\pi}}\, x^2 e^{-x^2/2}\, dx \\ &\le \frac{1}{4\theta}. \end{aligned}$$

The information is the same for larger p since for $p > 1$ *both* the variable and the parameter are the sum of p independent χ_1^2 variables and parameters, respectively.

Now invoke Remark 2.5 with $V_1 = 4\theta \geq V(\theta)$. Use of Corollary 2.3 then easily yields

$$(3.10) \qquad\qquad B(g) \geq \frac{4p\sigma^4}{\sigma^2 + 2} = \text{Bound}(2.12).$$

This bound cannot be improved by use of Theorem 2.7 with $V_1 = 4\theta$, since $E^{-1} = 0$. This is inevitable since $V_1 = 4\theta$ is the variance function for estimation of $\theta = 2\lambda$, where λ is a Poisson expectation and g is a conjugate prior for the Poisson family. Note that (3.9) is reasonably accurate since

$$\frac{B(g)}{\text{Bound}(2.12)} \to 1, \quad \text{as } \sigma^2 \to 0, \infty$$

and

$$\frac{B(g)}{\text{Bound}(2.12)} \leq \left[\frac{9}{8}\right],$$

where the value $\frac{9}{8}$ is attained at $\sigma^2 = 1$.

It is possible to use (3.9) to obtain better bounds on $I(\theta)$. For example, $I(\theta) \leq \min\{(4\theta)^{-1}, [2(\theta + p/2)]^{-1}\}$. For $p \geq 4$, this bound can be used via Remark 2.5 in Corollary 2.3 or in Theorem 2.7 but yields negligible improvement to (3.10). For $p \leq 3$, this bound cannot be used because the conditions of Corollary 2.6 are not satisfied. For example, for $p = 2$, use of this bound in (2.12) of Corollary 2.3 yields the illegitimate statement, $B(g) \geq 3.033$, whereas actually $B(g) = 3$.

EXAMPLE 3.5. Let X have a binomial (n, p) distribution. Then X is distributed according to an exponential family with natural parameter $\theta = \ln(p/(1 - p))$, the log-odds-ratio. Here, $I(\theta) = ne^\theta/(1 + e^\theta)^2 = np(1 - p)$. Consider the problem of estimating θ under normalized squared error loss, $L(\theta, a) = I(\theta)(a - \theta)^2$. [Note that estimation of θ under this loss is equivalent to estimation of the odds ratio, ρ, under the plausible loss function $L_1(\rho, b) = m(\rho)\ln^2(b/\rho)$ with $m(\rho) = n\rho/(1 + \rho)^2$.] Since θ is not a linear function of the expectation parameter, one should not expect the bounds of Section 2 to be sharp.

Consider the prior density $g(\theta) = 6e^{2\theta}(1 + e^\theta)^{-4}$. This corresponds to the Dirichlet $(1,1)$ prior, $6p(1 - p)$, for p; this choice is explained further in Brown and Low (1991). The bound of Corollary 2.3 yields

$$(3.11) \qquad\qquad B(g) \geq \frac{n}{n + 8} = \text{Bound}(2.12).$$

The bound in Theorem 2.7 can also be calculated, though it requires somewhat more algebraic manipulation as well as numerical integration of the resulting expression for E^{-1}. (One nice feature is that a symmetry argument immediately yields $\bar{u} = 0$.) Table 2 gives values of this bound [labelled Bound(2.16)] as well as Bound(2.12). It also gives, for comparison, the values of $B(g)$ calculated via numerical integration.

TABLE 2
Values of the bounds for Example 3.5

$n =$	1	3	5	10	25	100
Bound(2.12)	0.1111	0.2727	0.3846	0.5556	0.7576	0.9259
Bound(2.16)	0.1314	0.3053	0.4186	0.5829	0.7707	0.9279
$B(g)$	0.1358	0.3183	0.4360	0.6043	0.7893	0.9360

Note that (3.11) yields $B(g) \geq 1 - (8/n) + O(1/n^2)$. It is of interest to ask whether this bound is precise to order $1/n$; that is, whether $B(g) = 1 - 8/n + o(1/n)$? We will show in a forthcoming manuscript that this is not the case. The argument there involves an extension of the results of Section 2.

APPENDIX

The information inequality. This appendix is devoted to a statement and proof of the information inequality which does not impose regularity conditions on the statistic T. The results here extend those of Fabian and Hannan (1977). Alternate conditions, which can be used in place of those given here, are established in Simons and Woodroofe (1983); see also Müller-Funk, Puckelsheim and Witting (1989). For simplicity, only the one dimensional case is considered in detail. The multidimensional case is analogous and is summarized at the end of this appendix.

Let $\{p_\theta \colon \theta \in \Theta\}$ be a family of probability densities relative to some Borel measure ν. Θ is an open subset of $\mathbb{R}$. The form of the information inequality to be derived below is valid at $\theta_0 \in \Theta$ for any real valued statistic T having a finite variance at θ_0. Thus, concerning the statistic T assume only that

$$\text{(A.1)} \qquad \text{Var}_{\theta_0}(T) < \infty.$$

Let $e(\theta) = E_\theta(T)$, whenever the expectation on the right exists. Because of (A.1), this expectation exists at θ_0.

As Fabian and Hannan observe, what are needed are regularity conditions under which $e(\theta)$ exists on a neighborhood of θ_0, can be differentiated at θ_0 and under which there is a measurable function q such that

$$\text{(A.2)} \qquad e'(\theta_0) = E_{\theta_0}(T(X)q(X)).$$

Think of q as $q(x) = [(\partial/\partial\theta_0)\ln p_\theta(x)]_{\theta=\theta_0}$, for this is what it will be in all standard cases.

The only step requiring justification is (A.2). To dissect this step, define $\omega(dx) = p_{\theta_0}(x)\nu(dx)$. Then $T(\cdot) \in L_2(\omega_{\theta_0})$ because of (A.1). Assume

$$\text{(A.3)} \qquad p_{\theta_0}(x) = 0 \Rightarrow p_\theta(x) = 0 \quad \text{a.e. } (\nu),$$

for all θ in a neighborhood of θ_0. Then (A.2) can equivalently be written as

$$\text{(A.4)} \qquad \lim_{\Delta \to 0} \int T(x)\left[\Delta^{-1}\left[\frac{p_{\theta_0+\Delta}(x)}{p_{\theta_0}(x)} - 1\right] - q(x)\right]\omega(dx) = 0,$$

for every $T \in L_2(\omega)$. This can be expressed abstractly by saying that p_θ/p_{θ_0} is weakly differentiable in $L_2(\omega)$ at $\theta = \theta_0$, with weak derivative q. This abstractly expressed condition on $\{p_\theta\}$ is thus sufficient for validity of the information inequality for all T satisfying (A.1). It is also necessary. This equivalence is pointed out by Fabian and Hannan (1977).

This condition makes it plain that the information inequality requires regularity conditions only on $\{p_\theta\}$ and not on T [except for the trivial condition (A.1)]. However, explicit conditions are desirable which imply (A.4) and can be easily checked. The following discussion provides three separate sufficient conditions. They are progressively easier to check, but apply in successively less generality. The first sufficient condition is that

$$(A.5) \qquad \lim_{\Delta \to 0} \int \left[\Delta^{-1} \left[\frac{p_{\theta_o + \Delta}(x)}{p_{\theta_0}(x)} - 1 \right] - q(x) \right]^2 \omega(dx) = 0.$$

which says that p_θ/p_{θ_0} is strongly differentiable in $L_2(\omega)$ at $\theta = \theta_0$. That (A.5) implies (A.4) follows from the Cauchy–Schwarz inequality.

Now suppose on a neighborhood of θ_0, $p_\theta(x)$ is absolutely continuous in θ for a.e. $x(\nu)$. Let $p_\theta^*(x) = (\partial/\partial\theta) p_\theta(x)$ and $q(x) = p_{\theta_0}^*(x)/p_{\theta_0}(x)$. [Assume $p_{\theta_0}^*(x)$ exists a.e. (ν).] Writing $p_{\theta_0+\Delta}$ as the integral of its derivative, applying Cauchy–Schwarz and interchanging order of integration yields

$$\int \left[\Delta^{-1} \left[\frac{p_{\theta_0 + \Delta}(x)}{p_{\theta_0}(x)} - 1 \right] - q(x) \right]^2 \omega(dx)$$

$$= \int \left[\frac{\left[\Delta^{-1} \left[\int_{\theta_0}^{\theta_0 + \Delta} p_t^*(x) \, dt \right] - p_{\theta_0}^*(x) \right]^2}{p_{\theta_0}^2(x)} \right] \omega(dx)$$

$$= \int \left[\frac{\Delta^{-2} \left[\int_{\theta_0}^{\theta_0 + \Delta} \left(p_t^*(x) - p_{\theta_0}^*(x) \right) dt \right]^2}{p_{\theta_0}^2(x)} \right] \omega(dx)$$

$$\leq \int \left[\frac{\Delta^{-1} \int_{\theta_0}^{\theta_0 + \Delta} \left(p_t^*(x) - p_{\theta_0}^*(x) \right)^2 dt}{p_{\theta_0}^2(x)} \right] \omega(dx)$$

(by Cauchy–Schwarz)

$$= \Delta^{-1} \int_{\theta_0}^{\theta_0 + \Delta} \int \left[\frac{p_t^*(x) - p_{\theta_0}^*(x)}{p_{\theta_0}(x)} \right]^2 \omega(dx) \, dt$$

$$= \Delta^{-1} \int_{\theta_0}^{\theta_0 + \Delta} E_{\theta_0} \left[\left[\frac{p_t^*(X) - p_{\theta_0}^*(X)}{p_{\theta_0}(X)} \right]^2 \right] dt.$$

Consequently, (A.5) is satisfied if

$$(A.6) \qquad E_{\theta_0}\left[\left[\frac{p_t^*(X) - p_{\theta_0}^*(X)}{p_{\theta_0}(X)}\right]^2\right] \to 0 \quad \text{as } t \to \theta_0.$$

Now suppose on a neighborhood of θ_0, $p_\theta^*(x)$ is also absolutely continuous for a.e. $x(\nu)$. Let $p_\theta^{**}(x) = (\partial/\partial\theta) p_\theta^*(x) = (\partial^2/\partial\theta^2) p_\theta(x)$. Then a similar sequence of steps yields

$$E_{\theta_0}\left[\left[\frac{p_t^*(X) - p_{\theta_0}^*(X)}{p_{\theta_0}(X)}\right]^2\right] = E_{\theta_0}\left[\left[\frac{\int_{\theta_0}^{\theta_0+t} p_w^{**}(X)\, dw}{p_{\theta_0}(X)}\right]^2\right]$$

$$\leq t E_{\theta_0}\left[\int_{\theta_0}^{\theta_0+t}\left[\frac{p_w^{**}(X)}{p_{\theta_0}(X)}\right]^2 dw\right]$$

$$= t\int_{\theta_0}^{\theta_0+t} E_{\theta_0}\left[\left[\frac{p_w^{**}(X)}{p_{\theta_0}(X)}\right]^2\right] dw.$$

Consequently (A.6) is satisfied if for some $B < \infty$,

$$(A.7) \qquad E_{\theta_0}\left[\left[\frac{p_\theta^{**}(X)}{p_{\theta_0}(X)}\right]^2\right] < B < \infty,$$

for every θ in some neighborhood of θ_0. To summarize:

THEOREM A.1. *Suppose (A.3) and at least one of (A.4) to (A.7) are satisfied at θ_0. Then the information inequality is satisfied at θ_0 for any statistic T for which $\mathrm{Var}_{\theta_0}(T) < \infty$.*

REMARKS. (i) It is implicit in the theorem that (A.1), (A.3) and any of the conditions (A.4)–(A.7) imply that $e(\theta)$ exists on a neighborhood of θ_0 and is differentiable at θ_0.

(ii) Condition (A.6) cannot be replaced by

$$E_{\theta_0}\left[\left[\frac{p_t^*(X)}{p_t(X)} - \frac{p_{\theta_0}^*(X)}{p_{\theta_0}(X)}\right]^2\right] \to 0 \quad \text{as } t \to \theta_0$$

for this latter condition does not imply (A.6) nor (A.4). For an example where the implication fails let $|\theta| < \frac{1}{8}$ and let $p_\theta(x) = \theta^2/x^{2/3}$ for $0 < x < |\theta|^3$, $= 1$ for $|\theta|^3 \leq x < \frac{1}{2}$, $= 1 - 4|\theta|^3/3$ for $\frac{1}{2} \leq x < 1$ relative to Lebesgue measure on $0 < x < 1$. Then the above condition holds and $p_\theta(x)$ is absolutely continuous in θ for every $x \in (0, 1)$. However at $\theta_0 = 0$, (A.6) does not hold and (A.4) fails when, for example, $T(x) = x^{-1/3}$.

(iii) If (A.3) fails, the information inequality may still be valid under a condition slightly stronger than (A.1) but still suitable for occasional applica-

tions. Replace (A.1) by the assumption that $E_\theta(T^2) < B < \infty$ for some B and all θ in a neighborhood of θ_0. Let S_{θ_0} denote the support of p_{θ_0}. Assume

$$\Delta^{-1} \int_{S_{\theta_0}^c} p_{\theta_0 + \Delta}(x)\nu(dx) \to 0 \quad \text{as } \Delta \to 0.$$

Then any one of the conditions (A.4)–(A.7) implies validity of the information inequality.

Multivariate case. In this case Θ is an open subset of $\mathbb{R}^k$. The real valued function $q(x)$ is replaced by a $(k \times 1)$ vector valued function $q(x)$. Think of $q(x)_i$ as

$$E_{\theta_0}\left[\left[\frac{\partial \ln p_\theta(x)}{\partial \theta_i}\right]_{\theta_0}\right], \qquad \text{`}$$

for this is what it will be in standard cases. The information matrix at θ_0 is defined by

$$I(\theta_0) = E_{\theta_0}(q(X)q^\tau(X)),$$

where q^τ denotes the transpose of q. Let T be a $(l \times 1)$ vector valued statistic and let $e(\theta) = E_\theta(T)$, when it exists. When the partial derivatives of $e(\cdot)$ exist, define $\nabla^\tau e(\theta_0)$ to be the $(l \times k)$ matrix with $(\nabla^\tau e(\theta))_{ij} = (\partial e_i(\theta))/\partial\theta_j$. The multivariate information inequality asserts that

$$(\text{A.8}) \qquad \text{Var}_{\theta_0}(T) \ge (\nabla^\tau e(\theta_0))I^-(\theta_0)(\nabla^\tau e(\theta_0))^\tau.$$

Here $\text{Var}_{\theta_0}(T)$ denotes the $(l \times l)$ variance-covariance matrix of T and the inequality symbol means that the difference of the two sides of (A.8) is positive semi-definite. I^- denotes the symmetric generalized inverse of I.

The general regularity conditions needed to establish (A.8) are (A.3),

$$(\text{A.1}') \qquad\qquad\qquad E_{\theta_0}(\|T\|^2) < \infty$$

and

$$(\text{A.4}') \qquad \lim_{\Delta \to 0} \int T(x)\left[\Delta^{-1}\left[\frac{p_{\theta_0+\Delta v}(x)}{p_{\theta_0}(x)} - 1\right] - q^\tau(x)v\right]\omega(dx) = 0,$$

for every $T \in L_2(\omega)$, $v \in \mathbb{R}^k$.

Condition (A.5'), which implies (A.4'), is related to (A.5) as (A.4') is to (A.4); that is,

$$(\text{A.5}') \qquad \lim_{\Delta \to 0} \int \left[\Delta^{-1}\left[\frac{p_{\theta_0+\Delta v}(x)}{p_{\theta_0}(x)} - 1\right] - q^\tau(x)v\right]^2 \omega(dx) = 0$$

for every $v \in \mathbb{R}^k$. Next, assume that $\{p_\theta\}$ is absolutely continuous on a neighborhood of θ_0 a.e. (ν), in the sense that

$$p_\theta(x) - p_{\theta_0}(x) = \int_0^1 (\theta - \theta_0)^\tau p^*_{\theta_0 + t(\theta - \theta_0)}(x)\, dt,$$

for every θ in this neighborhood. Here $p^* = \nabla_\theta p$ denotes the $(k \times 1)$ vector with coordinates $\partial p_\theta / \partial \theta_i$ and $q(x) = \nabla_\theta(\ln p)$. The condition (A.6) can conveniently be replaced by

$$(A.6') \qquad E_{\theta_0}\left[\left\|\frac{p_\theta^*(X) - p_{\theta_0}^*(X)}{p_{\theta_0}(X)}\right\|^2\right] \to 0 \quad \text{as } \theta \to \theta_0.$$

A convenient way to rewrite (A.7) is to assume $p_\theta^*(x)$ is absolutely continuous near θ_0 and

$$(A.7') \qquad E_{\theta_0}\left[\left[\frac{\partial^2 \theta}{\partial \theta_i \, \partial \theta_j}\ln p_\theta(X)\right]^2\right] < B < \infty \qquad \forall \, 1 \le i, j \le k,$$

for every θ in a neighborhood of θ_0. To summarize:

THEOREM A.1′. *Suppose* (A.3) *and at least one of* (A.4′) *to* (A.7′) *are satisfied at* θ_0. *Then the information inequality* (A.8) *is satisfied at* θ_0 *for any statistic* T *for which* $\mathrm{Var}_{\theta_0}(T)$ *exists.*

REFERENCES

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

BOBROVSKY, B. Z., MAYER-WOLF, E. and ZAKAI, M. (1987). Some classes of global Cramér–Rao inequalities. *Ann. Statist.* **15** 1421–1438.

BOROVKOV, A. A. and SAKHANIENKO, A. U. (1980). On estimates of the expected quadratic risk. *Probab. Math. Statist.* **1** 185–195. (In Russian.)

BROWN, L. D. (1986). Information inequalities for the Bayes risk. Preprint, Statist. Center, Cornell Univ.

BROWN, L. D. (1988). An information inequality for truncated quadratic loss. Preprint, Statist. Center, Cornell Univ.

BROWN, L. D. and LOW, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann. Statist.* **19** (1).

CRAMÉR, H. (1946). A contribution to the theory of statistical estimation. *Skand. Aktuarietidskr.* **29** 85–94.

FABIAN, V. and HANNAN, J. (1977). On the Cramér–Rao inequality. *Ann. Statist.* **5** 197–205.

GAJEK, L. (1988). On the minimax value in the scale model with truncated data. *Ann. Statist.* **16** 669–677.

JOSHI, V. M. (1976). On the attainment of the Cramér–Rao lower bound. *Ann. Statist.* **4** 998–1002.

LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

MÜLLER-FUNK, U., PUCKELSHEIM, F. and WITTING, H. (1989). On the attainment of the Cramér-Rao bound in L_r-differentiable families of distributions. *Ann. Statist.* **17** 1742–1748.

SHEMYAKIN, A. (1987). On bounds for the Bayes risk under violated regularity conditions. *Theory Probab. Appl.* **32** 201.

SIMONS, G. and WOODROOFE, M. (1983). The Cramér-Rao inequality holds almost everywhere. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 69–93. Academic, New York.

VAN TREES, H. (1968). *Detection, Estimation, and Modulation Theory* **1**. Wiley, New York.

WIJSMAN, R. A. (1973). On the attainment of the Cramér-Rao lower bound. *Ann. Statist.* **1** 538–542.

MATHEMATICS DEPARTMENT
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

POLITECHNIKA KODZKA
INSTYTUT MATEMATYKI
AL. POLITECHNIKI 11
90–924 LODZ
POLAND