Ann. Inst. Statist. Math.
Vol. 55, No. 4, 803–816 (2003)

ON BAYES AND UNBIASED ESTIMATORS OF LOSS

DOMINIQUE FOURDRINIER[1] AND WILLIAM E. STRAWDERMAN[2*]

[1] *Laboratoire de Mathématiques R. Salem, UMR CNRS 6085, Université de Rouen, Mathématiques Site Colbert, 76821 Mont-Saint-Aignan, France*
[2] *Department of Statistics, Rutgers University, 561 Hill Center, Busch Campus, Piscataway, NJ 08854-8019, U.S.A.*

(Received July 18, 2000; revised December 24, 2002)

Abstract. We consider estimation of loss for generalized Bayes or pseudo-Bayes estimators of a multivariate normal mean vector, θ. In 3 and higher dimensions, the MLE X is UMVUE and minimax but is inadmissible. It is dominated by the James-Stein estimator and by many others. Johnstone (1988, On inadmissibility of some unbiased estimates of loss, *Statistical Decision Theory and Related Topics, IV* (eds. S. S. Gupta and J. O. Berger), Vol. 1, 361–379, Springer, New York) considered the estimation of loss for the usual estimator X and the James-Stein estimator. He found improvements over the Stein unbiased estimator of risk. In this paper, for a generalized Bayes point estimator of θ, we compare generalized Bayes estimators to unbiased estimators of loss. We find, somewhat surprisingly, that the unbiased estimator often dominates the corresponding generalized Bayes estimator of loss for priors which give minimax estimators in the original point estimation problem. In particular, we give a class of priors for which the generalized Bayes estimator of θ is admissible and minimax but for which the unbiased estimator of loss dominates the generalized Bayes estimator of loss. We also give a general inadmissibility result for a generalized Bayes estimator of loss.

Key words and phrases: Loss estimation, shrinkage estimation, Bayes estimation, unbiased estimation, superharmonicity.

1. Introduction

Suppose we observe x from a distribution P_θ where $\theta \in \mathbb{R}^p$. A basic statistical problem is to estimate θ by $\varphi(x)$ under a loss function $L(\theta, \varphi(x))$. The corresponding risk function is given by $R(\theta, \varphi) = E_\theta[L(\theta, \varphi(X))]$ (where E_θ denotes the expectation with respect to P_θ) and serves as a classical decision theoretic basis for evaluation of φ. However it is often of interest to assess the loss $L(\theta, \varphi(x))$ itself and a growing literature has recently developed on this subject (see, for instance, Johnstone (1988) for a rationale). A common approach to this assessment is to consider estimation of the loss $L(\theta, \varphi(x))$ by an estimator δ, called the loss estimator. For evaluation of this new type of estimator, another loss is required and it has become standard to use the squared error

$$L^*(\theta, \varphi(x), \delta(x)) = (\delta(x) - L(\theta, \varphi(x)))^2. \tag{1.1}$$

More precisely this evaluation will be done through the new risk function

$$\mathcal{R}(\theta, \varphi, \delta) = E_\theta[L^*(\theta, \varphi(X), \delta(X))] = E_\theta[(\delta(X) - L(\theta, \varphi(X)))^2]. \tag{1.2}$$

*Research supported by NSF Grant DMS-97-04524.

See Johnstone (1988), Lu and Berger (1989) and Fourdrinier and Wells (1995) for more details on this approach and Rukhin (1988) for an alternate view.

In this paper, we consider estimation of loss for generalized or (pseudo) Bayes estimators of a multivariate normal mean vector. Specifically let X be distributed as $\mathcal{N}_p(\theta, I)$ in dimension p and let the loss function for estimating θ be $L(\theta, \varphi) = \|\theta - \varphi\|^2$. For a given generalized prior π, we denote the generalized marginal by $m(x)$ and the generalized Bayes estimator of θ by

$$\varphi_m(x) = x + \frac{\nabla m(x)}{m(x)}. \tag{1.3}$$

Through Stein's lemma

$$E_\theta[(X - \theta) \cdot g(X)] = E_\theta[\operatorname{div} g(X)] \tag{1.4}$$

for any appropriate function g. The unbiased estimator of risk of φ_m is

$$\delta_u(x) = p + 2\frac{\Delta m(x)}{m(x)} - \frac{\|\nabla m(x)\|^2}{m^2(x)} \tag{1.5}$$

while the posterior risk of φ_m is

$$\delta_m(x) = p + \frac{\Delta m(x)}{m(x)} - \frac{\|\nabla m(x)\|^2}{m^2(x)}. \tag{1.6}$$

Here the symbols ∇ and Δ denote the gradient and the Laplacian for real valued functions and div denotes the divergence for vector valued functions. See Stein (1981) for development and details of the above.

Our results for generalized Bayes estimators will depend on $\pi(\theta)$ only through the marginal $m(x)$. Hence, in fact, they hold not only for generalized Bayes estimators but also for so called pseudo-Bayes estimators. In this paper, we define a pseudo-Bayes estimator of θ to be one of the form (1.3) for some function m which may or may not be a generalized marginal corresponding to some generalized prior π. Similarly, we define a pseudo-Bayes estimator of loss to be one of the form (1.6). See Bock (1988) for a discussion of pseudo-Bayes estimators.

In Section 2, we give an expression for the risk of a general estimator $\delta(X)$ of the loss of an estimator $\varphi(X)$ of θ and for the risk difference between two competing loss estimators. We give a specialization of this second result for δ_m and δ_u.

In Section 3, a primary interest is in comparing δ_m and δ_u for superharmonic marginals. In this case, it is well known and follows immediately from (1.5) that the generalized Bayes estimator φ_m of θ in (1.3) dominates the usual unbiased estimator, X, under the loss $\|\delta - \theta\|^2$. It is of interest to see if this domination of the generalized Bayes estimator over the unbiased estimator persists in the problem of estimation of the loss of φ_m. The somewhat surprising result is that this is not necessarily the case as we will demonstrate. In particular, we give a class of priors for which the generalized Bayes estimator of θ is admissible and minimax but for which the unbiased estimator of loss dominates the generalized Bayes estimator of loss.

In Section 4, we give general conditions under which δ_m is inadmissible and explicit improved estimators. The results fall into two classes; in the first class, the improved estimators shrink toward the origin while, in the second class, they expand away from 0.

In Section 5, we give some concluding remarks and an example which indicates that perhaps the loss function $(\delta(x) - \|\varphi(x) - \theta\|^2)^2$ may be inappropriate.

Finally, Appendix contains several lemmas which follow from Stein's lemma and the proof of the main result of Section 2.

2. Computations and comparisons of risk

In this section, we give an expression for the risk of a general estimator δ of the loss of an estimator φ of θ. We also give an expression for the difference in risk between two competing estimators of the loss of the estimator φ. We specialize this result to the comparison of a general estimator δ to δ_u and δ_m. In particular, we will focus on comparison of δ_u and δ_m.

Proofs are based on the unbiased estimation of risk technique of Stein (1981) (see also Johnstone (1988)). This method requires conditions of weak differentiability and finiteness of expectation which we make throughout without explicit mention beyond this paragraph. Thus, for example, we require that g and λ be twice weakly differentiable and that $E_\theta[\|g(X)\|^4]$ and $E_\theta[\lambda^2(X)]$ be finite.

Our first result gives an expression for the risk of a general estimator of loss.

THEOREM 2.1. *Let $\varphi(x) = x + g(x)$ and $\delta(x) = p + \lambda(x)$ and let the loss function be given by* (1.1). *Then*

$$
\begin{aligned}
\mathcal{R}(\theta, \varphi, \delta) = {} & 2p + E_\theta[(\|g(X)\|^2 + 2\operatorname{div} g(X) - \lambda(X))^2 \\
& + 4g(X) \cdot \nabla(\|g(X)\|^2 + 2\operatorname{div} g(X) - \lambda(X)) \\
& + 2\Delta(\|g(X)\|^2 + 2\operatorname{div} g(X) - \lambda(X)) \\
& + 4(\|g(X)\|^2 + 2\operatorname{div} g(X) + \operatorname{tr}(J_g^2(X)))]
\end{aligned}
\tag{2.1}
$$

where $J_g(X)$ and tr *denote respectively the Jacobian of g and the* trace.

PROOF. See Appendix. □

Recall that $\|g(X)\|^2 + 2\operatorname{div} g(X)$ is the unbiased estimator of risk difference between $X + g(X)$ and X as estimators of θ.

We now give an expression for the risk difference between two estimators of loss.

COROLLARY 2.1. *Let $\varphi(x) = x + g(x)$, $\delta(x) = p + \lambda(x)$ and $\delta^\gamma(x) = \delta(x) + \gamma(x)$. Then the risk difference $\mathcal{R}(\theta, \varphi, \delta^\gamma) - \mathcal{R}(\theta, \varphi, \delta)$ is given by*

$$
\begin{aligned}
& E_\theta[\gamma^2(X) - 2\gamma(X)(\|g(X)\|^2 + 2\operatorname{div} g(X) - \lambda(X)) \\
& \qquad - 4g(X) \cdot \nabla\gamma(X) - 2\Delta\gamma(X)].
\end{aligned}
\tag{2.2}
$$

PROOF. This is a direct application of (2.1). □

A simple situation where δ^γ dominates δ is when φ is a minimax estimator ($\|g(x)\|^2 + 2\operatorname{div} g(x) < 0$) and $E_\theta[\lambda(X)] > c > 0$, for all θ. This domination follows immediately from Corollary 2.1 if $\gamma(X) \equiv k$ with $-2c \leq k < 0$.

We now specialize to the case where the estimator φ of θ is a pseudo-Bayes estimator with respect to a pseudo-marginal m, that is, $\varphi(x) = \varphi_m(x) = x + \frac{\nabla m(x)}{m(x)}$. In this case, the two most natural loss estimators are the unbiased estimator δ_u given by (1.5) and the pseudo-Bayes estimator given by (1.6). We now compare these two estimators with a general estimator and with each other.

COROLLARY 2.2. *Let $\varphi_m(x) = x + \frac{\nabla m(x)}{m(x)}$, let δ_u and δ_m be given by (1.5) and (1.6) and let γ be a function from $\mathbb{R}^p$ into $\mathbb{R}$. Then we have*

$$\mathcal{R}(\theta, \varphi_m, \delta_m + \gamma) - \mathcal{R}(\theta, \varphi_m, \delta_m) = E_\theta\left[\gamma^2(X) - 2\frac{\Delta(m(X)\gamma(X))}{m(X)}\right], \tag{2.3}$$

$$\begin{aligned}\mathcal{R}(\theta, \varphi_m, \delta_u + \gamma) &- \mathcal{R}(\theta, \varphi_m, \delta_u) \\ &= E_\theta\left[\gamma^2(X) - 2\Delta\gamma(X) - 4\nabla\gamma(X)\cdot\frac{\nabla m(X)}{m(X)}\right]\end{aligned} \tag{2.4}$$

and

$$\mathcal{R}(\theta, \varphi_m, \delta_u) - \mathcal{R}(\theta, \varphi_m, \delta_m) = E_\theta\left[\left(\frac{\Delta m(X)}{m(X)}\right)^2 - 2\frac{\Delta^{(2)}m(X)}{m(X)}\right] \tag{2.5}$$

where $\Delta^{(2)}m = \Delta(\Delta m)$ is the bi-Laplacian of m.

PROOF. Applying Corollary 2.1 with $g(x) = \frac{\nabla m(x)}{m(x)}$ and $\lambda(x) = \frac{\Delta m(x)}{m(x)} - \frac{\|\nabla m(x)\|^2}{m^2(x)}$ gives that the left hand side of (2.3) is equal to

$$E_\theta\left[\gamma^2(X) - 2\left(\gamma(X)\frac{\Delta m(X)}{m(X)} + \Delta\gamma(X) + 2\nabla\gamma(X)\cdot\frac{\nabla m(X)}{m(X)}\right)\right]$$

which reduces to the right hand side of (2.3) since the expression in parentheses is equal to $\frac{\Delta(m(X)\gamma(X))}{m(X)}$.

Similarly, still with $g(x) = \frac{\nabla m(x)}{m(x)}$ but with $\lambda(x) = \frac{2\Delta m(x)}{m(x)} - \frac{\|\nabla m(x)\|^2}{m^2(x)}$, the term $\|g(X)\|^2 + 2\operatorname{div} g(X) - \lambda(X)$ in (2.2) vanishes and (2.4) follows directly.

Finally (2.5) follows from (2.3) since $\delta_u(x) = \delta_m(x) + \gamma(x)$ with $\gamma(x) = \frac{\Delta m(x)}{m(x)}$. □

Comment 1. Our main interest in this paper is to compare the estimators δ_m and δ_u and more generally to find procedures which dominate δ_m. Sections 3 and 4 respectively are devoted to these questions. It is worth noting at this stage that domination of δ_m (via the unbiased estimation of risk technique) requires the existence of a function $\gamma(X)$ so that $\Delta(m(X)\gamma(X)) > 0$ and $E_\theta[\gamma^2(X)] < \infty$ by (2.3). Similarly, domination of δ_m by δ_u requires by (2.5) that the bi-Laplacian of m, $\Delta^{(2)}m$, is positive.

Comment 2. Johnstone (1988) applies his proposition 4.1 (equivalent to (2.4)) to the improvement of $\delta_u(X)$ for the two estimators of θ, $\varphi_m(X) = X$ (the "usual" estimator) and $\varphi_m(X) = (1 - \frac{p-2}{\|X\|^2})X$ the James-Stein estimator). Note, however, that $\varphi_m(X) = X$ (a generalized Bayes estimator) corresponds to $m(x) \equiv 1$ and $\varphi_m(X) = (1 - \frac{p-2}{\|X\|^2})X$ (a pseudo-Bayes estimator) corresponds to $m(x) = \|x\|^{-(p-2)}$. Since both of these "marginals" are harmonic (i.e. $\Delta m(x) \equiv 0$) we have $\delta_m(X) = \delta_u(X)$ and hence

(2.3) and (2.4) coincide. Therefore Johnstone's results may be seen as improving either on $\delta_m(X)$ or $\delta_u(X)$.

Comment 3. Much of our focus is on pseudo-marginals $m(x)$ for which $\Delta m(x) < 0$ and hence the corresponding pseudo-Bayes estimator of θ is minimax. Such pseudo-marginals cannot be integrable. See Fourdrinier *et al.* (1998) for more discussion of this issue. Of course, if the marginal corresponds to a proper prior, the posterior risk is the unique Bayes estimator of loss and is therefore admissible.

We now turn our attention to generalized Bayes estimators and give a result which implies domination of δ_u over δ_m.

COROLLARY 2.3. *Suppose that the prior density $\pi(\theta)$ satisfies*

$$\left(\frac{\Delta\pi(\theta)}{\pi(\theta)}\right)^2 - 2\Delta^{(2)}\pi(\theta) \leq 0 \qquad \forall\theta \in \mathbb{R}^p. \tag{2.6}$$

Then δ_u dominates δ_m.

PROOF. Note that $m(x) = E_x[\pi(\theta)]$ where E_x denotes the expectation with respect to $\mathcal{N}_p(x, I)$. By interchange of differentiation and integration and Stein's lemma, we have

$$\frac{\partial}{\partial x_i} E_x[f(\theta)] = E_x[(\theta_i - x_i)f(\theta)] = E_x\left[\frac{\partial}{\partial \theta_i} f(\theta)\right]$$

for any weakly differentiable function f. Then it follows that

$$\nabla_x m(x) = E_x[\nabla_\theta \pi(\theta)],$$
$$\Delta_x m(x) = E_x[\Delta_\theta \pi(\theta)],$$

and

$$\Delta_x^{(2)} m(x) \equiv E_x[\Delta_\theta^{(2)} \pi(\theta)].$$

We now show that (2.6) implies that the right hand side of (2.5) is negative and hence gives the result. Let $E_{\theta/x}$ be the posterior expectation given x. The bracketed term in the right hand side of (2.5) is equal to

$$\begin{aligned}\left(\frac{\Delta m(x)}{m(x)}\right)^2 - 2\frac{\Delta^{(2)} m(x)}{m(x)} &= \left(\frac{E_x[\Delta\pi(\theta)]}{E_x[\pi(\theta)]}\right)^2 - 2\frac{E_x[\Delta^{(2)}\pi(\theta)]}{E_x[\pi(\theta)]} \\ &= \left(E_{\theta/x}\left[\frac{\Delta\pi(\theta)}{\pi(\theta)}\right]\right)^2 - 2E_{\theta/x}\left[\frac{\Delta^{(2)}\pi(\theta)}{\pi(\theta)}\right] \\ &\leq E_{\theta/x}\left[\left(\frac{\Delta\pi(\theta)}{\pi(\theta)}\right)^2 - 2\frac{\Delta^{(2)}\pi(\theta)}{\pi(\theta)}\right] \\ &\leq 0\end{aligned}$$

by (2.6). □

3. Comparison of δ_m and δ_u

In this section, for a pseudo-marginal m, we compare δ_m and δ_u as estimators of the loss of the pseudo-Bayes estimator φ_m of θ. In particular, we are interested in conditions of domination of δ_u over δ_m.

A first example gives a class of pseudo-marginals corresponding to minimax estimators of θ for which δ_u dominates δ_m. It is then worth noting that the pseudo-Bayes estimator φ_m of θ dominates the unbiased estimator X while the pseudo-Bayes estimator δ_m of loss is dominated by the unbiased estimator of loss.

Example 1. Let $m(x)$ be a function proportional to $(\frac{1}{\|x\|^2/2+a})^b$ where a and b are nonnegative constants. It is convenient to set $y = \frac{\|x\|^2}{2}$ and $m(x) = f(y) = (\frac{1}{y+a})^b$. Then it is straightforward to express $\Delta m(x)$ and $\Delta^{(2)}m(x)$ as

$$\Delta m(x) = 2yf''(y) + pf'(y) \tag{3.1}$$

and

$$\Delta^{(2)}m(x) = p(p+2)f''(y) + 4(p+2)yf'''(y) + 4y^2 f^{(4)}(y). \tag{3.2}$$

Hence the unbiased estimator of the risk difference between δ_u and δ_m given by (2.5) is expressed in term of y as

$$\eta(y) = \left(2y\frac{f''(y)}{f(y)} + p\frac{f'(y)}{f(y)}\right)^2 - 2\left(p(p+2)\frac{f''(y)}{f(y)} + 4(p+2)y\frac{f'''(y)}{f(y)} + 4y^2\frac{f^{(4)}(y)}{f(y)}\right).$$

Noting that

$$\frac{f^{(k)}(y)}{f(y)} = (-1)^k b(b+1)\cdots(b+k-1)(y+a)^{-k} \tag{3.3}$$

straightforward calculations lead to

$$\eta(y) = -b(y+a)^{-2}(Az^2 - Bz + C)$$

where

$$\begin{aligned} A &= 4(b+1)(b^2+9b+12),\\ B &= 4(b+1)((p+4)b+4(p+2)),\\ C &= p((p+4)b+2(p+2)) \end{aligned}$$

and

$$z = y(y+a)^{-1}$$

for $(a,y) \neq (0,0)$. Note that $A > 0$, $B > 0$ and $C > 0$ and hence $\eta(0) < 0$ provided $b > 0$ and $a > 0$. It follows that one cannot show that δ_m dominates δ_u using this unbiased estimator of risk technique if $a > 0$.

The next lemma gives conditions which guarantee that $\eta(y) \leq 0$ for all y and hence δ_u dominates δ_m in the case where $a > 0$. We consider the case $a = 0$ in Lemma 3.2.

LEMMA 3.1. *Suppose $a > 0$. A sufficient condition for which $\eta(y) \le 0$ for all y is*

$$p \ge b + 1 + \left[(b+1)^2 + 8\frac{(b+1)(b+2)^2}{b^2+3b+4}\right]^{1/2}. \tag{3.4}$$

In particular this condition holds if $p \ge 2(b+3)$.

PROOF. A necessary and sufficient condition for $\eta(y) \le 0$ for all $y \ge 0$ is that $Az^2 - Bz + C \ge 0$ for all $z \in [0,1[$. Then it suffices that $B^2 - 4AC \le 0$. Straightforward calculations lead to

$$\frac{B^2-4AC}{32(b+1)} = -(b^2+3b+4)p^2 + 2(b+1)(b^2+3b+4)p + 8(b+1)(b+2)^2.$$

It is clear that this quadratic in p will be nonpositive as soon as p is at least as big as its only positive root which is given by the right hand side of (3.4).

Since

$$\frac{(b+1)(b+2)^2}{b^2+3b+4} = \frac{b^2+3b+2}{b^2+3b+4}(b+2) \le b+2 \le b+3$$

an upper bound of the right hand side of (3.4) is $2(b+3)$. □

Comment 4. Note that, by (3.1) and (3.3), the Laplacian of m is given by

$$\Delta m(x) = \frac{b}{(y+a)^{b+1}}\left(2(b+1)\frac{y}{y+a} - p\right) \tag{3.5}$$

and is nonpositive for $b \le \frac{p-2}{2}$ and, in particular, for $b \le \frac{p-6}{2}$ which is equivalent to the last condition of Lemma 3.1. Hence, under this condition, the James-Stein-like pseudo-Bayes estimator φ_m given by $\varphi_m(x) = (1 - \frac{2b}{\|x\|^2+2a})x$ is minimax and dominates the usual unbiased estimator X. The lemma shows however that the unbiased estimator of risk δ_u dominates the (pseudo) posterior risk of φ_m as an estimator of loss.

Example 2. Let $\pi(\theta) = (\frac{1}{\|\theta\|^2/2+a})^b$. It follows immediately from calculations in Example 1 that, if $p \ge 2(b+3)$ then (2.6) holds and hence δ_u dominates δ_m. Since π is integrable if and only if $b > \frac{p}{2}$ (for $a > 0$), the prior π is improper whenever this condition for domination of δ_u over δ_m holds. Of course, whenever π is proper, the Bayes estimator δ_m is admissible provided its Bayes risk is finite.

We now turn to the case where $a = 0$.

LEMMA 3.2. 1) *If $a = 0$, a necessary and sufficient condition for which $\eta(y) \le 0$ (resp. < 0) for any $y > 0$ is that $A - B + C \ge 0$ (resp. > 0).*

2) *If $p > 4$, then $A - B + C \ge 0$ (resp. > 0) for $b \ge \frac{p-2}{2}$ (resp. $> \frac{p-2}{2}$) and for $0 \le b \le b_0$ (resp. $< b_0$) where $b_0 < \frac{p-2}{2}$ is a positive root of the cubic equation in b, $A - B + C = 0$.*

PROOF. The first part of the lemma follows immediately upon using the alternative representation

$$\eta(y) = -b\left\{(A-B+C)\frac{1}{(y+a)^2} - (2A-B)\frac{a}{(y+a)^3} + A\frac{a^2}{(y+a)^4}\right\}.$$

A straightforward calculation shows that, for $p > 4$, the cubic equation (in b), $f(b) = A - B + C = 0$, has the following properties: a) $f(\frac{p-2}{2}) = 0$, b) $f(0) > 0$, c) $f'(\frac{p-2}{2}) > 0$ and d) $f'''(b) = 24 > 0$. This implies that there is one negative root and two positive roots b_0 and $\frac{p-2}{2}$ of which $\frac{p-2}{2}$ is the larger. Hence $f(b) \geq 0$ for $0 \leq b \leq b_0$ and for $b \geq \frac{p-2}{2}$. □

Comment 5. The above re-expression for $\eta(y)$ was kindly pointed out by a referee. The referee also raises the possibility that $E[\eta(y)] \leq 0$ for a range of positive values of a for values of $b \geq \frac{p-2}{2}$. This seems plausible but does not appear to follow directly from our method of proof.

Example 3. We continue with Example 1 when $a = 0$. In this case, Lemma 3.2 implies that $\eta(y) < 0$ for $0 \leq b < b_0$ and $b > \frac{p-2}{2}$ when $p > 4$. Further, the finiteness conditions stated prior to Theorem 2.1 require $E_\theta[\|X\|^{-4}] < \infty$ which holds when $p > 4$. Hence, under these conditions, δ_u dominates δ_m (and φ_m dominates X for $\frac{p-2}{2} \leq b < p-2$).

However note that δ_m dominates δ_u and also φ_m dominates X for $b_0 < b < \frac{p-2}{2}$. Also note that the risks of δ_m and δ_u coincide when $b = \frac{p-2}{2}$ and when $b = b_0$; actually δ_m and δ_u coincide when $b = \frac{p-2}{2}$ but not when $b = b_0$.

We now consider the prior distribution in Example 2 but for the case $a = 0$. This is perhaps the most striking of our examples since it gives a class of priors for which the generalized Bayes estimator of θ is minimax and admissible but for which the unbiased estimator of risk δ_u dominates the generalized Bayes estimator δ_m. Hence $\pi(\theta)$ provides an admissible and minimax estimator of θ but an inadmissible estimator of loss.

Example 4. Consider $\pi(\theta) = (\frac{1}{\|\theta\|^2})^b$. By the calculations in Lemma 3.2 and Example 3, the conditions of Corollary 2.3 are satisfied provided $b \geq \frac{p-2}{2}$. We also require that $b < p$ in order that π is locally integrable in a neighborhood of the origin (otherwise m will not exist). It follows that, if $\frac{p-2}{2} < b < p$, the generalized Bayes estimator δ_m is inadmissible and is dominated by the unbiased estimator of risk δ_u. We show below however that for $\frac{p-2}{2} \leq b \leq \frac{p-2}{2} + \frac{p-2}{2(p+1)}$ the generalized Bayes estimator of θ, φ_m, is admissible and minimax.

It is easily seen that π has a representation as a hierarchical prior where the first stage of the prior is the conjugate prior $\theta \mid v \sim \mathcal{N}(0, v^{-1}I)$ and where v has "density" proportional to v^{-a} for $a = \frac{p}{2} - b + 1$. A standard calculation shows that the generalized Bayes estimator of θ, φ_m, is given by

$$\varphi_m(X) = \left(1 - \frac{\int_0^1 u^{p/2-a+1}(1-u)^{a-2}e^{-u\|X\|^2/2}du}{\int_0^1 u^{p/2-a}(1-u)^{a-2}e^{-u\|X\|^2/2}du}\right) X.$$

By Theorems 2.3 and 2.4 of Maruyama (1998), φ_m is admissible and minimax provided $3 - \frac{p}{2} \leq a \leq 2$ and $a - 2 \geq -(a + \frac{p}{2} - 3)/p$ or equivalently $\frac{3(p+2)}{2(p+1)} \leq a \leq 2$. In terms of b $(= \frac{p}{2} - a + 1)$ this calculation becomes $\frac{p-2}{2} \leq b \leq \frac{p-2}{2} + \frac{p-2}{2(p+1)}$ as stated above.

Hence if $\frac{p-2}{2} < b \leq \frac{p-2}{2} + \frac{p-2}{2(p+1)}$, φ_m is admissible and minimax as an estimator of θ while δ_u dominates δ_m as an estimator of loss.

4. General inadmissibility results for δ_m

The results of this section are of the form: if $m(x) \geq h(x)$ for all x, where h is in some class, then δ_m is inadmissible as an estimator of loss and we give an explicit class of dominating estimators. The main result is given by the following theorem.

THEOREM 4.1. *Let $g(x)$ be a strictly positive function on $\mathbb{R}^p$ such that either $\Delta g(x) < 0$ or $\Delta g(x) > 0$ for all $x \in \mathbb{R}^p$ and such that $E_\theta[(\frac{\Delta g(X)}{g(X)})^2] < \infty$. Assume that $m(x) > K\frac{g^2(x)}{|\Delta g(x)|}$ for all x and for some $K > 0$ and that $K_0 = \inf_{x\in\mathbb{R}^p} m(x)\frac{|\Delta g(x)|}{g^2(x)}$ (which is also the supremum of such K).*

Then δ_m is inadmissible and a class of dominating estimators is given by

$$\delta_m(x) = \alpha \operatorname{sgn}(\Delta g(x))\frac{g(x)}{m(x)} \qquad \textit{for} \quad 0 < \alpha < 2K_0.$$

PROOF. We give the proof only for $\Delta g < 0$ the case $\Delta g > 0$ being similar. We apply the first part of Corollary 2.2 with $\gamma(x) = -\alpha\frac{g(x)}{m(x)}$. The sufficient domination condition becomes

$$2\alpha\Delta g(x) + \alpha^2\frac{g^2(x)}{m(x)} \leq 0. \tag{4.1}$$

By assumption on m and g, we have $\Delta g(x) \leq -K_0\frac{g^2(x)}{m(x)}$. So the left hand side of (4.1) is bounded above by

$$\alpha\frac{g^2(x)}{m(x)}(\alpha - 2K_0) \leq 0.$$

Also the integrability condition $E_\theta[\gamma^2(X)] < \infty$ is satisfied since $\gamma^2(x) = \alpha^2\frac{g^2(x)}{m^2(x)} \leq \frac{\alpha^2}{K_0^2}(\frac{\Delta g(x)}{g(x)})^2$. $\square$

Example 5. Let $g_b(x) = (\frac{1}{\|x\|^2+a})^b$. By calculations essentially equivalent to those leading to (3.5), we have $\Delta g_b(x) < 0$ for $a \geq 0$ and $0 < 2(b+1) < p$. Also $\Delta g_b(x) > 0$ if $a = 0$ and $2(b+1) > p$. Furthermore

$$\frac{g_b^2(x)}{|\Delta g_b(x)|} = \frac{1}{2b\left|p - 2(b+1)\frac{\|x\|^2}{\|x\|^2+a}\right|}\frac{1}{(\|x\|^2+a)^{b-1}}.$$

a) Suppose that $0 < 2(b+1) < p$ and $a \geq 0$. Then

$$\frac{g_b^2(x)}{|\Delta g_b(x)|} \leq \frac{1}{2b(p-2(b+1))}\frac{1}{(\|x\|^2+a)^{b-1}}$$

and $E_\theta[(\frac{\Delta g_b(X)}{g_b(X)})^2] < \infty$ since it is proportional to $E_\theta[\frac{1}{(\|X\|^2+a)^2}]$ which is finite for $a > 0$ or for $a = 0$ and $p > 4$.

Suppose that $m(x)$ is greater than or equal to some multiple of $(\frac{1}{\|x\|^2+a})^{b-1}$ or equivalently

$$m(x) \geq \frac{k}{2b(p-2(b+1))}\left(\frac{1}{\|x\|^2+a}\right)^{b-1} \tag{4.2}$$

for some $k > 0$. The theorem implies that $\delta_m(X)$ is inadmissible and is dominated by

$$\delta_m(X) - \frac{\alpha}{m(X)(\|X\|^2+a)^b}$$

for

$$0 < \alpha < 4b(p-2(b+1)) \inf_{x \in \mathbb{R}^p} (m(x)(\|x\|^2+a)^{b-1}).$$

Alternatively, if $m(x) \geq \frac{k}{(\|x\|^2+a)^c}$ for $0 < c < \frac{p-4}{2}$, δ_m is inadmissible and the above gives an explicit improvement upon substituting $c-1$ for b. Note that the improved estimators shrink towards 0.

Suppose, for example, that $m(x) \equiv 1$. Then (4.2) is satisfied for $b \geq 1$. Here $\varphi_m(X) = X$ and $\delta_m(X) = p$. Choosing $b = 1$, an improved class of estimators is given by $p - \frac{\alpha}{\|X\|^2+a}$ for $0 < \alpha < 4(p-4)$. The case $a = 0$ is equivalent to Johnstone's result for this marginal.

b) Suppose that $2(b+1) > p > 4$ and $a = 0$. Then

$$\frac{g_b^2(x)}{|\Delta g_b(x)|} = \frac{1}{2b(2(b+1)-p)} \frac{1}{\|x\|^{2(b-1)}}.$$

A development similar to the above implies that, when $m(x)$ is greater than or equal to some multiple of $(\frac{1}{\|x\|^2})^{b-1}$, an improved estimator is

$$\delta_m(X) + \frac{\alpha}{m(X)\|X\|^{2b}}$$

for

$$0 < \alpha < 4b(2(b+1)-p) \inf_{x \in \mathbb{R}^p} (m(x)\|x\|^{2(b-1)}).$$

Note that, in this case, the correction term is positive and hence the estimator expands away from 0. Note also that this result only works for $a = 0$ and hence applies to pseudo-marginals which are unbounded in a neighbourhood of 0. Since all marginals corresponding to a generalized prior π are bounded, this result can never apply to generalized Bayes procedures but only to pseudo-Bayes procedures.

Suppose, for example, that $m(x) = (\frac{1}{\|x\|^2})^{(p-2)/2}$. Here $\varphi_m(X) = (1 - \frac{p-2}{\|x\|^2})X$ is the James-Stein estimator and $\delta_m(X) = p - \frac{(p-2)^2}{\|X\|^2}$. In particular, the above applies for $b - 1 = \frac{p-2}{2}$, that is, for $b = \frac{p}{2} > \frac{p-2}{2}$. An improved estimator is given by $\delta_m(X) + \frac{\gamma}{\|X\|^2}$ for $0 < \gamma < 4p$. This again agrees with Johnstone's result for James-Stein estimators.

5. Conclusion

In this paper, we have studied pseudo-Bayes and generalized Bayes estimators of loss. We have given conditions under which these estimators are inadmissible and have provided explicit improved estimators.

In a sense, the most interesting result of our investigation is that, in certain common situations where the usual estimator, X, of θ is dominated by a pseudo-Bayes or generalized Bayes estimator with respect to some "marginal" $m(x)$, the corresponding "Bayes" estimator of loss δ_m is dominated by the unbiased estimator of loss δ_u.

This phenomenon may reflect a deficiency in the loss function $(\delta - \|\varphi_m - \theta\|^2)^2$ of the loss estimation problem. The situation is somewhat analogous to estimation of a strictly convex function of θ, say for example $g(\theta) = \theta^2$, under quadratic loss. The following simple result concerning estimation of the square of a location parameter in $\mathbb{R}^1$ indicates a general problem with squared error in this setting.

LEMMA 5.1. *Suppose $X \in \mathbb{R}^1 \sim f((X-\theta)^2)$. Consider estimation of θ^2 under loss $(\delta - \theta^2)^2$. Then the generalized Bayes estimator δ_π of θ^2 with respect to the uniform prior $\pi(\theta) \equiv 1$ is inadmissible for any $f(\cdot)$ such that $E_\theta[X^4] < \infty$ and is dominated by the unbiased estimator $\delta_u = X^2 - E_0[X^2]$.*

PROOF. The generalized Bayes estimator of θ^2 is given by

$$\delta_\pi(X) = \frac{\int \theta^2 f((X-\theta)^2) d\theta}{\int f((X-\theta)^2) d\theta} = X^2 + E_0[X^2].$$

Since this estimator has constant bias $2E_0[X^2]$, it is dominated by the unbiased estimator $X^2 - E_0[X^2]$ (the risk difference is $(E_0[X^2])^2$). □

The phenomenon of inadmissibility of the generalized Bayes estimator with respect to the uniform distribution extends easily to the case of a strictly convex (or concave) function $g(\theta)$ with a positive lower bound on $|g''(\theta)|$. The uniform prior in one dimension is virtually universally accepted as the correct non-informative or reference prior for θ. It seems that the loss function in this basic problem as well as in the loss estimation problem is at the heart of the somewhat paradoxical nature of the result.

Another interesting aspect of the loss estimation problem studied in this paper is the domination of the unbiased (and pseudo-Bayes) estimator of loss of the James-Stein estimator by an estimator that adjusts away from 0 i.e. expands δ_u. We have indicated that this phenomenon seems to be connected with the singularity of the pseudo-marginal at $\theta = 0$.

It seems that the unbiased estimation of risk technique cannot provide improved estimators that always "expand away from 0" if the marginal is non-singular at $\theta = 0$ (e.g. $(\|X\|^2 + a)^{-b}$ for $a > 0$). In fact, Fourdrinier and Wells (1995), in a slightly different context (general spherically symmetric distributions in the presence of a residual vector U of dimension k), have shown that shrinkage towards 0 improves on the unbiased estimator of loss of the James-Stein estimator $(1 - \frac{(p-2)\|U\|^2}{(k+2)\|X\|^2})X$.

Acknowledgements

The authors wish to thank a referee for suggesting the development in Lemma 3.2 and Example 3. We also thank him for his very careful reading of the paper.

Appendix

We first recall Stein's lemma and derive several useful expressions which follow upon repeated applications of it.

LEMMA A.1. (Stein (1981)) *For any weakly differentiable function g from $\mathbb{R}^p$ into $\mathbb{R}^p$, we have*

$$E_\theta[(X-\theta)\cdot g(X)] = E_\theta[\operatorname{div} g(X)] \tag{A.1}$$

provided these expectations exist.

LEMMA A.2. (Stein (1981) and see also Johnstone (1988)) *For any twice weakly differentiable function g from $\mathbb{R}^p$ into $\mathbb{R}^1$, we have*

$$E_\theta[\|X-\theta\|^2 g(X)] = E_\theta[\Delta g(X) + p g(X)] \tag{A.2}$$

provided these expectations exist.

PROOF. It follows from (A.1) applied to $(X-\theta)g(X)$. □

LEMMA A.3. *For any twice weakly differentiable function g and h from $\mathbb{R}^p$ into $\mathbb{R}^p$, we have*

$$\begin{aligned} E_\theta&[(X-\theta)\cdot g(X)\times(X-\theta)\cdot h(X)] \\ &= E_\theta[g(X)\cdot h(X) + \operatorname{div} g(X)\operatorname{div} h(X) + g(X)\cdot\nabla(\operatorname{div} h(X)) \\ &\qquad + h(X)\cdot\nabla(\operatorname{div} g(X)) + \operatorname{tr}(J_g(X)J_h(X))] \end{aligned} \tag{A.3}$$

provided these expectations exist. Here $J_g(X)$ and tr *denote respectively the Jacobian matrix of g and the* trace.

PROOF. We have

$$\begin{aligned} E_\theta&[(X-\theta)\cdot g(X)\times(X-\theta)\cdot h(X)] \\ &= E_\theta[\operatorname{div}([(X-\theta)\cdot h(X)]g(X))] \\ &= E_\theta[(X-\theta)\cdot h(X)\operatorname{div} g(X) + \nabla[(X-\theta)\cdot h(X)]\cdot g(X)] \\ &= E_\theta[\operatorname{div} h(X)\operatorname{div} g(X) + \nabla(\operatorname{div} g(X))\cdot h(X) \\ &\qquad + h(X)\cdot g(X) + (J_h^t(X)(X-\theta))\cdot g(X)] \\ &= E_\theta[\operatorname{div} h(X)\operatorname{div} g(X) + \nabla(\operatorname{div} g(X))\cdot h(X) \\ &\qquad + h(X)\cdot g(X) + (X-\theta)\cdot(J_h(X)g(X))] \\ &= E_\theta[h(X)\cdot g(X) + \operatorname{div} h(X)\operatorname{div} g(X) + \nabla(\operatorname{div} g(X))\cdot h(X) + \operatorname{div}(J_h(X)g(X))]. \end{aligned}$$

As we have

$$\operatorname{div}(J_h(X)g(X)) = \nabla(\operatorname{div} h(X))\cdot g(X) + \operatorname{tr}(J_h(X)J_g(X))$$

this is the desired result. □

COROLLARY A.1. *For any twice weakly differentiable function g from $\mathbb{R}^p$ into $\mathbb{R}^p$, we have*

$$\text{(A.4)}\qquad \begin{aligned} &E_\theta[((X-\theta)\cdot g(X))^2] \\ &\quad = E_\theta[\|g(X)\|^2 + 2g(X)\cdot\nabla(\operatorname{div} g(X)) + (\operatorname{div} g(X))^2 + \operatorname{tr}(J_g^2(X))] \end{aligned}$$

provided these expectations exists. Furthermore, if g is three times weakly differentiable, then

$$\text{(A.5)}\qquad E_\theta[\|X-\theta\|^2(X-\theta)\cdot g(X)] = E_\theta[(p+2)\operatorname{div} g(X) + \Delta(\operatorname{div} g(X))].$$

PROOF. Formula (A.4) follows from Lemma A.3 for $h = f$ while formula (A.5) follows with $h(X) = X - \theta$ and upon an additional application of Lemma A.1. □

We now use these lemmas to prove Theorem 2.1.

PROOF OF THEOREM 2.1. The risk $\mathcal{R}(\theta, X+g(X), p+\lambda(X))$ of the loss estimator $p+\lambda(X)$ of the loss $\|X+g(X)-\theta\|^2$ is given by

$$E_\theta[(p+\lambda(X) - \|X+g(X)-\theta\|^2)^2] = E_\theta[A(X)+B(X)+C(X)]$$

where

$$\begin{aligned} A(X) &= E_\theta[(p+\lambda(X) - \|g(X)\|^2)^2], \\ B(X) &= E_\theta[(\|X-\theta\|^2 + 2(X-\theta)\cdot g(X))^2] \end{aligned}$$

and

$$C(X) = -2E_\theta[(p+\lambda(X) - \|g(X)\|^2)(\|X-\theta\|^2 + 2(X-\theta)\cdot g(X))].$$

Expanding $B(X)$ and using Corollary A.1 give

$$\begin{aligned} B(X) &= p(p+2) + 4E_\theta[((X-\theta)\cdot g(X))^2] + 4E_\theta[\|X-\theta\|^2(X-\theta)\cdot g(X)] \\ &= p(p+2) + 4E_\theta[\|g(X)\|^2 + 2g(X)\cdot\nabla(\operatorname{div} g(X)) + (\operatorname{div} g(X))^2 + \operatorname{tr}(J_g^2(X))] \\ &\quad + 4E_\theta[(p+2)\operatorname{div} g(X) + \Delta(\operatorname{div} g(X)]. \end{aligned}$$

Also, using Lemmas A.1 and A.2, $C(X)$ becomes

$$\begin{aligned} C(X) &= -2E_\theta[\Delta(p+\lambda(X) - \|g(X)\|^2) + p(p+\lambda(X) - \|g(X)\|^2)] \\ &\quad - 4E_\theta[\operatorname{div}((p+\lambda(X) - \|g(X)\|^2)g(X))] \\ &= -2E_\theta[\Delta(\lambda(X) - \|g(X)\|^2) + p(p+\lambda(X) - \|g(X)\|^2)] \\ &\quad - 4E_\theta[(p+\lambda(X) - \|g(X)\|^2)\operatorname{div} g(X) + \nabla(\lambda(X) - \|g(X)\|^2)\cdot g(X)]. \end{aligned}$$

Expanding $A(X)$ and gathering all the terms give

$$\begin{aligned} &\mathcal{R}(\theta, X+g(X), p+\lambda(X)) \\ &\quad = 2p + E_\theta[\lambda^2(X) + \|g(X)\|^4 + 4\|g(X)\|^2 \\ &\qquad\qquad - 2\lambda(X)\|g(X)\|^2 + 8g(X)\cdot\nabla(\operatorname{div} g(X)) + 4(\operatorname{div} g(X))^2 \\ &\qquad\qquad + 4\operatorname{tr}(J_g^2(X)) + 8\operatorname{div} g(X) + 4\Delta(\operatorname{div} g(X)) \end{aligned}$$

$$
\begin{aligned}
&\quad - 2\Delta(\lambda(X)) + 2\Delta(\|g(X)\|^2) - 4\lambda(X) \operatorname{div} g(X) \\
&\quad + 4\|g(X)\|^2 \operatorname{div} g(X) - 4\nabla(\lambda(X)) \cdot g(X) + 4\nabla(\|g(X)\|^2) \cdot g(X)] \\
&= 2p + E_\theta[(\|g(X)\|^2 + 2 \operatorname{div} g(X) - \lambda(X))^2 \\
&\quad + 4g(X) \cdot \nabla(\|g(X)\|^2 + 2 \operatorname{div} g(X) - \lambda(X)) \\
&\quad + 2\Delta(\|g(X)\|^2 + 2 \operatorname{div} g(X) - \lambda(X)) \\
&\quad + 4(\|g(X)\|^2 \\
&\quad + 2 \operatorname{div} g(X) + \operatorname{tr}(J_g^2(X)))]. \qquad \square
\end{aligned}
$$

References

Bock, M. E. (1988). Shrinkage estimators: pseudo-Bayes rules for normal mean vectors, *Statistical Decision Theory and Related Topics, IV* (eds. S. S. Gupta and J. O. Berger), Vol. 1, 281–297, Springer, New York.

Fourdrinier, D. and Wells, M. T. (1995). Estimation of a loss function for spherically symmetric distributions in the general linear model, *Ann. Statist.*, **23**(2), 571–592.

Fourdrinier, D., Strawderman, W. E. and Wells, M. T. (1998). On the construction of Bayes minimax estimators, *Ann. Statist.*, **26**(2), 660–671.

Johnstone, J. (1988). On inadmissibility of some unbiased estimates of loss, *Statistical Decision Theory and Related Topics, IV* (eds. S. S. Gupta and J. O. Berger), Vol. 1, 361–379, Springer, New York.

Lu, K. L. and Berger, J. O. (1989). Estimation of normal means: Frequentist estimation of loss, *Ann. Statist.*, **17**, 890–906.

Maruyama, Y. (1998). A unified and broadened class of admissible minimax estimators of a multivariate normal mean, *J. Multivariate Anal.*, **64**, 196–205.

Rukhin, A. L. (1988). Estimated loss and admissible loss estimators, *Statistical Decision Theory and Related Topics, IV* (eds. S. S. Gupta and J. O. Berger), Vol. 1, 409–418, Springer, New York.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution, *Ann. Statist.*, **9**, 1135–1151.