

ON THE CONSTRUCTION OF BAYES MINIMAX ESTIMATORS¹

BY DOMINIQUE FOURDRINIER, WILLIAM E. STRAWDERMAN
AND MARTIN T. WELLS

Université de Rouen, Rutgers University and Cornell University

Bayes estimation of the mean of a multivariate normal distribution is considered under quadratic loss. We show that, when particular spherical priors are used, the superharmonicity of the square root of the marginal density provides a viable method for constructing (possibly proper) Bayes (and admissible) minimax estimators. Examples illustrate the theory; most notably it is shown that a multivariate Student- t prior yields a proper Bayes minimax estimate.

1. Introduction. When estimating the mean of a multivariate distribution, the two dominant approaches are the minimax approach and variants of the Bayes paradigm. The first has received the most extensive development while the second is most used in practice, due to its great flexibility. See [4] for a study of the interface between these two approaches. The problem of both methods is that neither necessarily leads to admissible estimators (hierarchical Bayes estimators are often only generalized Bayes estimators). Even if admissibility may provide nonreasonable estimators, it is a criterion which can be desirable when combining minimaxity and Bayesian properties. Indeed, since the sampling distribution is normal, under quadratic loss, the Bayes estimator is unique provided the Bayes risk is finite, so that the proper Bayes estimator is admissible (see [16]). In [5], Brown conjectured that, for estimating a multivariate normal mean using quadratic loss, a proper Bayes minimax estimator does not exist for four or less dimensions. This conjecture was proved by Strawderman [21], who also settled the conjecture for dimensions five or more that such estimators do indeed exist.

Stein [19] obtains minimaxity of a general estimator δ of the form $\delta(x) = x + \gamma(x)$ through the use of the unbiased estimator of the risk $k + 2 \operatorname{div} \gamma(x) + \|\gamma(x)\|^2$. Thus if for every $x \in \mathbb{R}^k$, $2 \operatorname{div} \gamma(x) + \|\gamma(x)\|^2 \leq 0$, then δ is minimax. As (formal) Bayes estimators are of the form $x + \nabla \operatorname{Log} m(x)$ (where ∇ denotes the gradient and m the marginal density), this condition becomes $\Delta \sqrt{m(x)} \leq 0$ (where Δ denotes the Laplacian). Stein gives examples for generalized Bayes estimates.

Stein's method of demonstrating minimaxity through the unbiased estimator of the risk is now the standard technique of proof. This supplants the method employed by Strawderman [21] and Faith [10], who used monotonicity

Received January 1997; revised June 1997.

¹Supported in part by NSF Grants DMS-96-25440 and DMS-97-04524.

AMS 1991 subject classifications. Primary 62C20, 62C15, 62C10; secondary 62A15.

Key words and phrases. Bayes estimate, minimaxity, multivariate normal mean, proper Bayes, quadratic loss, superharmonic functions.

and majorization conditions on the function r where

$$(1) \quad \gamma(x) = (k - 2)r\left(\frac{1}{2}\|x\|^2\right)x/\|x\|^2.$$

Alam [1] also follows from Stein's technique although in a less direct fashion.

In many cases the superharmonicity of \sqrt{m} is quite difficult to verify. The difficulty has led researchers to consider the superharmonicity of m rather than \sqrt{m} . This is reasonable as

$$(2) \quad \Delta\sqrt{m(x)} = \frac{1}{2\sqrt{m(x)}}\left(\Delta m(x) - \frac{1}{2}\frac{\|\nabla m(x)\|^2}{m(x)}\right) \quad \forall x \in \mathbb{R}^k.$$

Therefore, if $\Delta m(x) \leq 0$, then $\Delta\sqrt{m(x)} \leq 0$. A problem that occurs is that, as we will show below, if one uses a proper prior, the induced marginal cannot be superharmonic. Therefore if one insists on a superharmonic marginal the underlying prior is necessarily improper. See [4], [12] and [13] for applications of superharmonicity ideas to the construction of minimax estimates. Upon an examination of these papers, one can see that they deal with improper priors. If \sqrt{m} is superharmonic, this is not necessarily the case.

We will show that, for the class of variance mixtures of normal priors, Stein's result gives rise to a method for constructing admissible minimax estimators via solutions to the differential inequality $\Delta\sqrt{m} \leq 0$. It is important to work with the superharmonicity of the square root of the marginal rather than just the marginal itself because we wish to investigate the minimaxity of Bayes rules based on both proper and improper priors. This larger class contains many examples of priors that are routinely used in practice, such as the t -distribution. We are also able to construct new priors within this more general framework.

The goal of this paper is not to check whether a Bayes estimator is minimax; it is to consider the inverse problem, that is, the construction of minimax Bayes estimators. Section 2 states the problem, gives the necessary set-up and discusses the superharmonicity of \sqrt{m} . In Section 3, examples illustrate the method of construction for spherical priors that are variance mixtures of normal distributions. In Section 4, we present some examples; in particular in Section 4.3 it will be shown that the multivariate Student- t prior yields a minimax Bayes rule. In the final section, we give some concluding remarks.

2. Marginals and superharmonicity.

2.1. *The model.* Let x be a random vector in \mathbb{R}^k distributed as a multivariate normal distribution, $\mathcal{N}_k(\theta, I_k)$, with mean vector θ and identity covariance matrix I_k . The central problem of this paper is that of constructing spherical Bayes minimax estimators of θ under the quadratic loss function $L(\theta, \delta) = \|\delta - \theta\|^2$. Before giving the main results, first recall the elements that lead to proper Bayes minimax estimates. The conditions that specify risk finiteness are critical and must be dealt with carefully.

Assume that θ is distributed according to a prior probability measure ν . Then the marginal distribution of x has density with respect to the Lebesgue

measure in \mathbb{R}^k given by

$$m(x) = \frac{1}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2}\|x - \theta\|^2\right) d\nu(\theta).$$

The Bayes estimate $\delta_\nu(x)$, which is defined as the minimizer of the Bayes risk (see [3], page 17) is given by $\delta_\nu(x) = x + \nabla \text{Log } m(x)$, where the symbol ∇ denotes the gradient. Thus the Bayes estimator δ_ν is of the form $\delta_\nu(x) = x + \gamma(x)$. Stein exhibits, under suitable integrability conditions, that an unbiased estimator of the risk of δ_ν equals $k+2 \text{div } \gamma(x) + \|\gamma(x)\|^2 = k+4\Delta\sqrt{m(x)}/\sqrt{m(x)}$ where the divergence and the Laplacian are denoted by div and Δ , respectively. Thus the risk of δ_ν equals $E_\theta[\|\delta_\nu - \theta\|^2] = k + 4E_\theta\Delta\sqrt{m(x)}/\sqrt{m(x)}$, where E_θ is the expectation with respect to $\mathcal{N}_k(\theta, I_k)$. Therefore a sufficient condition for δ_ν to dominate the usual minimax estimate $\delta_0(x) = x$ is that the square root of the marginal density is superharmonic; in this case δ_ν will be minimax as well.

Let us specify the conditions for which the above results are valid. As the risk of δ_0 is finite (namely, equal to k), a straightforward application of Schwarz's inequality shows that the risk of $\delta_\nu = \delta_0 + \gamma$ is finite if and only if $E_\theta[\|\gamma\|^2] < \infty$. In that case, Stein's identity states that $E_\theta[(X - \theta)\gamma] = E_\theta[\text{div } \gamma]$ and each expectation exists.

For $\gamma = \nabla \text{Log } m$, the finiteness condition means that

$$(3) \quad E_\theta[\|\nabla \text{Log } m\|^2] = E_\theta\left[\frac{\|\nabla m\|^2}{m^2}\right] < +\infty$$

and this, in turn, implies that

$$E_\theta[\Delta \text{Log } m] = E_\theta\left[\frac{\Delta m}{m} - \frac{\|\nabla m\|^2}{m^2}\right] < \infty.$$

2.2. The superharmonicity of $m^{1/2}$. In the following we give some comments about the various ways that \sqrt{m} can be superharmonic when a proper prior is used. There is a deep connection between properness and superharmonicity; the fact that the prior is proper implies that the marginal cannot be superharmonic. Indeed [19] shows that the posterior risk $E^x[\|\theta - \delta_\nu\|^2]$, where E^x denotes the expectation with respect to the a posteriori distribution given x , is greater than or equal to the unbiased estimator $\rho(x)$ of the risk of δ_ν , as soon as m is superharmonic since

$$E^x[\|\theta - \delta_\nu\|^2] = \rho(x) - \frac{\Delta m(x)}{m(x)}.$$

Hence, it follows that

$$E_\theta[E^x[\|\theta - \delta_\nu\|^2]] \geq E_\theta[\|\theta - \delta_\nu\|^2].$$

Integrating this inequality with respect to ν (and denoting by E^ν the corresponding expectation), the right-hand side $E^\nu[E_\theta[\|\theta - \delta_\nu\|^2]]$ is the usual Bayes risk. However, denoting by E^m the expectation with respect to the marginal

density, since ν is proper, we are guaranteed of the existence of the posterior (cf. [8], page 79-III). Hence we have

$$(4) \quad E^\nu[E_\theta[E^x[\|\theta - \delta_\nu\|^2]]] = E^m[E^x[\|\theta - \delta_\nu\|^2]],$$

which is also equal to the Bayes risk. So the previous inequality is actually an equality when ν is a probability measure, which excludes the fact that $\Delta m \leq 0$ with strict inequality on a set of nonzero measure (that is, as m is a smooth function, we have $\Delta m = 0$ everywhere). Note that, as m is also a positive function, the harmonicity condition $\Delta m = 0$ implies necessarily that m is constant (see [9]). As a consequence, the superharmonicity condition on the marginal contradicts the properness of the prior. Note, however, in the case where ν is improper, the formal posterior, as defined in (2.3) of [15], is not necessarily a regular conditional probability. In this case equality (4) does not hold.

By searching for priors among those for which the square root of the marginal density is superharmonic, we provide a richer class of minimax procedures including the proper Bayes minimax ones. A convenient subclass of proper spherical priors will be the class of variance mixtures of normal distributions (recall that Strawderman's prior is one of them). In the next section we construct Bayes minimax priors for this class.

3. The construction. Recall that a spherical (generalized) density with generating function g is a variance mixture of normal distributions if there exists a measurable positive function h on \mathbb{R}_+ such that

$$g(t) = \int_0^\infty \frac{1}{(2\pi v)^{k/2}} \exp\left(-\frac{t}{2v}\right) h(v) dv \quad \forall t \in \mathbb{R}_+$$

provided that this last integral exists. It is clear that a sufficient condition for this is that $g(0) < \infty$. The density is proper provided $\int_0^\infty h(v) dv$ is finite.

If $g(\|\theta\|^2)$ is taken as the (generalized) prior, the marginal density m can be seen as the corresponding mixture of marginal multivariate normal densities m_ν with the same mixing function h . Upon application of Fubini's theorem for positive functions, for any $x \in \mathbb{R}^k$,

$$(5) \quad \begin{aligned} m(x) &= \frac{1}{(2\pi)^{k/2}} \int_{\mathbb{R}^k} \exp\left(-\frac{1}{2}\|x - \theta\|^2\right) \\ &\quad \times \left[\int_0^\infty \frac{1}{(2\pi v)^{k/2}} \exp\left\{-\frac{\|\theta\|^2}{2v}\right\} h(v) dv \right] d\theta \\ &= \int_0^\infty m_\nu(x) h(v) dv, \end{aligned}$$

where

$$m_\nu(x) = \frac{1}{(2\pi(\nu + 1))^{k/2}} \exp\left(-\frac{\|x\|^2}{2(\nu + 1)}\right).$$

Lebesgue’s dominated convergence theorem ensures that differentiating under the integral sign is valid. Hence we obtain

$$(6) \quad \nabla m(x) = \int_0^\infty \nabla m_v(x)h(v) dv$$

and

$$(7) \quad \Delta m(x) = \int_0^\infty \Delta m_v(x)h(v) dv.$$

It is worth noting that these mixture results allow us to prove easily that any (generalized or proper) Bayes estimator resulting from a variance mixture of normal distributions has finite risk since (3) is satisfied. Indeed, for any $x \in \mathbb{R}^k$, we have

$$(8) \quad \frac{\|\nabla m(x)\|^2}{m^2(x)} = \|x\|^2 \left[\frac{\int_0^\infty (1/(v+1)^{k/2+1}) \exp(-\|x\|^2/2(v+1))h(v) dv}{\int_0^\infty (1/(v+1)^{k/2}) \exp(-\|x\|^2/2(v+1))h(v) dv} \right]^2 \leq \|x\|^2.$$

Thus, it follows that $E_\theta[\|\nabla m\|^2/m^2] \leq k + \|\theta\|^2 < \infty$ and, therefore, the necessary and sufficient risk finiteness condition is satisfied.

Before stating our main result, we add a comment about the dimension k of the sample space. It is well known that Stein’s phenomom only occurs when $k \geq 3$. However, dealing with proper priors, recall that [20] proves that, in three and four dimensions, spherically symmetric proper Bayes minimax estimators do not exist. So in the following, we assume that $k \geq 5$ if we wish only to use proper priors, while only $k \geq 3$ if we are willing to use improper priors.

Theorem 1 is the main result of this section.

THEOREM 1. *Let h be a positive function such that the function $v \rightarrow -(v+1)h'(v)/h(v)$ can be decomposed as $l_1(v) + l_2(v)$ where $l_1 \leq A$ and is nondecreasing while $0 < l_2 \leq B$ with $\frac{1}{2}A + B \leq (k-2)/4$. Assume also that $\lim_{v \rightarrow \infty} h(v)/(v+1)^{k/2} = 0$. Then the Bayes estimator corresponding to the mixing density h is minimax. Furthermore, if h is integrable, the resulting proper Bayes estimator is minimax.*

PROOF OF THEOREM 1. Since the Laplacian of \sqrt{m} is as in (2), the superharmonicity condition on \sqrt{m} is equivalent to

$$(9) \quad \frac{\Delta m(x)}{\|\nabla m(x)\|} - \frac{1}{2} \frac{\|\nabla m(x)\|}{m(x)} \leq 0 \quad \forall x \in \mathbb{R}^k.$$

For any $x \in \mathbb{R}^k$, applying (6) and (7), the two following inequalities are successively equivalent to (9):

$$\frac{\int_0^\infty (1/(v+1)^{k/2+1})((\|x\|^2/(v+1)) - k) \exp(-\|x\|^2/2(v+1))h(v) dv}{\|x\| \int_0^\infty (1/(v+1)^{k/2+1}) \exp(-\|x\|^2/2(v+1))h(v) dv} - \frac{1}{2} \frac{\|x\| \int_0^\infty (1/(v+1)^{k/2+1}) \exp(-\|x\|^2/2(v+1))h(v) dv}{\int_0^\infty (1/(v+1)^{k/2}) \exp(-\|x\|^2/2(v+1))h(v) dv} \leq 0$$

and

$$\frac{\int_0^\infty (1/(v+1)^{k/2+2}) \exp(-\|x\|^2/2(v+1))h(v) dv}{\int_0^\infty (1/(v+1)^{k/2+1}) \exp(-\|x\|^2/2(v+1))h(v) dv} - \frac{1}{2} \frac{\int_0^\infty (1/(v+1)^{k/2+1}) \exp(-\|x\|^2/2(v+1))h(v) dv}{\int_0^\infty (1/(v+1)^{k/2}) \exp(-\|x\|^2/2(v+1))h(v) dv} \leq \frac{k}{\|x\|^2}.$$

By changing of variable $t = 1/(v+1)$, (9) is also equivalent to

$$(10) \quad \frac{\int_0^1 t^{k/2} \exp(-(\|x\|^2/2)t)h((1-t)/t) dt}{\int_0^1 t^{k/2-1} \exp(-(\|x\|^2/2)t)h((1-t)/t) dt} - \frac{1}{2} \frac{\int_0^1 t^{k/2-1} \exp(-(\|x\|^2/2)t)h((1-t)/t) dt}{\int_0^1 t^{k/2-2} \exp(-(\|x\|^2/2)t)h((1-t)/t) dt} \leq \frac{k}{\|x\|^2}.$$

With this change of variables, the condition at infinity becomes

$$(11) \quad \lim_{t \rightarrow 0} t^{k/2} h\left(\frac{1-t}{t}\right) = 0.$$

To show that condition (10) is satisfied, first integrate by parts in both numerators on the left-hand side of (10), letting $s = \|x\|^2/2$. Then, for example, the first integral can be written as

$$(12) \quad \int_0^1 t^{k/2} \exp(-st)h\left(\frac{1-t}{t}\right) dt = -\frac{1}{s} \left[t^{k/2} h\left(\frac{1-t}{t}\right) \exp(-st) \right]_0^1 + \frac{k}{2s} \int_0^1 t^{k/2-1} \exp(-ts)h\left(\frac{1-t}{t}\right) dt - \frac{1}{s} \int_0^1 t^{k/2-2} \exp(-ts)h'\left(\frac{1-t}{t}\right) dt.$$

The second integral has the same form but with k replaced by $k-2$.

Condition (10) then becomes [on collecting like terms and noting that the integral in the middle expression of (12) is equal to the denominator and cancelling s]

$$(13) \quad \frac{\int_0^1 t^{k/2-2} \exp(-ts) \left(\frac{h'((1-t)/t)}{h((1-t)/t)}\right) h((1-t)/t) dt}{\int_0^1 t^{k/2-1} \exp(-ts) h((1-t)/t) dt} + \frac{1}{2} \frac{\int_0^1 t^{((k-2)/2)-2} \exp(-ts) \left(\frac{h'((1-t)/t)}{h((1-t)/t)}\right) h((1-t)/t) dt}{\int_0^1 t^{((k-2)/2)-1} \exp(-ts) h((1-t)/t) dt} + \left[\frac{-\lim_{t \rightarrow 1} t^{k/2} \exp(-ts) h((1-t)/t)}{\int_0^1 t^{k/2-1} \exp(-ts) h((1-t)/t) dt} + \frac{1}{2} \frac{\lim_{t \rightarrow 1} t^{(k-2)/2} \exp(-ts) h((1-t)/t)}{\int_0^1 t^{(k-2)/2-1} \exp(-ts) h((1-t)/t) dt} \right]$$

$$\begin{aligned}
& + \left[\frac{\lim_{t \rightarrow 0} t^{k/2} \exp(-ts) h((1-t)/t)}{\int_0^1 t^{k/2} \exp(-ts) h((1-t)/t) dt} \right. \\
& \quad \left. - \frac{1 \lim_{t \rightarrow 0} t^{(k-2)/2} \exp(-ts) h((1-t)/t)}{2 \int_0^1 t^{(k-2)/2-1} \exp(-ts) h((1-t)/t) dt} \right] \\
& \leq \frac{k}{2} - \frac{k}{2} + \frac{1}{2} \frac{(k-2)}{2} = \frac{k-2}{4}.
\end{aligned}$$

Now note that the two terms in brackets in (13) are both nonpositive. The first bracketed term is nonpositive because the limits in the numerators are equal, while the denominator of the second term is larger than that of the first. The second bracketed expression is nonpositive because the first term vanishes by (11) and the second term is nonpositive.

Also note that, holding s fixed and defining

$$g_k(t) = \frac{\exp(-ts) t^{k/2-1} h((1-t)/t)}{\int_0^1 \exp(-ts) t^{k/2-1} h((1-t)/t) dt},$$

(13) implies that a sufficient condition for minimaxity of the Bayes estimator corresponding to the mixing density $h(\cdot)$ is

$$(14) \quad E_k(f) - \frac{1}{2} E_{k-2}(f) \leq \frac{k-2}{4},$$

where E_k denotes expectation with respect to the density $g_k(t)$ and where $f(t) = -h'((1-t)/t)/(th((1-t)/t))$.

Let $l_i(v) = f_i(1/(v+1))$ ($i = 1, 2$). Then $f(t) = f_1(t) + f_2(t)$ where $f_1 \leq A$ and is nonincreasing while $0 < f_2 \leq B$. Since $g_k(t)$ is a family with monotone (increasing) likelihood ratio in k and f_1 is nonincreasing and bounded by A , $E_k(f_1) - \frac{1}{2} E_{k-2}(f_1) \leq \frac{1}{2} E_{k-2}(f_1) \leq \frac{1}{2} A$. Furthermore, as $0 < f_2 \leq B$, $E_k(f_2) - \frac{1}{2} E_{k-2}(f_2) \leq B$. Hence the left-hand side of (14) is bounded above by $\frac{1}{2} A + B$ which in turn is bounded above by $(k-2)/4$ by the hypothesis of the theorem. Therefore (14) holds so that the Bayes estimator is minimax. If h is integrable, the proper Bayes estimator is minimax. This completes the proof of the theorem. \square

A corollary which is very useful in constructing Bayes minimax estimators follows easily from the theorem and will be applied in the examples in the next section.

COROLLARY 1. *Let ψ be a continuous function which can be decomposed as $\psi = \psi_1 + \psi_2$, where $\psi_1 \leq C$ and is nondecreasing, while $0 < \psi_2 \leq D$ for $\frac{1}{2}C + D \leq 0$. Define*

$$(15) \quad h(v) = \exp \left[-\frac{1}{2} \int_{v_0}^v \frac{2\psi(u) + k - 2}{u + 1} du \right] \quad \forall v > 0,$$

where v_0 is a nonnegative real constant.

Assume also that

$$(16) \quad \lim_{v \rightarrow \infty} \frac{h(v)}{(v+1)^{k/2}} = 0.$$

Furthermore, if h is integrable, the resulting proper Bayes estimator is minimax.

PROOF. Note that $-((v+1)h'(v)/h(v)) = \psi_1(v) + \psi_2(v) + (k-2)/2$. Let $l_1(v) = \psi_1(v) + (k-2)/2$ and $l_2(v) = \psi_2(v)$. With $A = (k-2)/2 + C$ and $B = D$ the result follows from Theorem 1. \square

COMMENT 1. In applications of Theorem 1 it often suffices to take $l_2 = 0$. Such an identification is equivalent to a monotone $r(\cdot)$ function in the representation of the shrinkage function in (1), which in turn corresponds to choosing $\psi_2 = 0$ in Corollary 1. The inclusion of a nonzero ψ_2 term leads to the enrichment of the class of priors which yield Bayes minimax estimators and consequently which have shrinkage functions with a nonmonotone $r(\cdot)$. The multivariate Student t example below illustrates this point. In such cases, the usual proof of minimaxity based on the unbiased estimate of risk and monotonicity does not necessarily apply.

COMMENT 2. Bayesians do not necessarily use proper priors. However, in such cases, it is still of interest to verify the admissibility of the Bayes minimax estimator. Brown [5] showed that if $\int_1^\infty dr/r^{k-1}m^*(r) = \infty$, where $m^*(\|x\|) = m(x)$, then the corresponding generalized Bayes estimator is admissible. In the variance mixture of normals prior case we are considering, (5) and a change of variable yield

$$(17) \quad m(r) = \mathcal{L} \left[t^{k/2-2} h \left(\frac{1-t}{t} \right) \mathbf{1}_{(0,1)}(t) \right] (r^2/2),$$

where $\mathcal{L}[g](u)$ is the Laplace transform of g evaluated at u and $\mathbf{1}_{(0,1)}$ is the indicator function of the interval $(0, 1)$. A Tauberian theorem in [22] gives a nice technique for relating the tail behavior of a function and its Laplace transform. We will use this result to relate the tail behavior of the mixing density to that of the marginal, then use this rate to establish Brown's admissibility condition.

Suppose the mixing density is such that $h(v) \sim c_1 v^\beta$, for some generic constant c_1 , as $v \rightarrow \infty$. Then $t^{k/2-2} h((1-t)/t) \mathbf{1}_{(0,1)}(t) \sim c_1 t^{k/2-\beta-2}$ as $t \rightarrow 0^+$. By Corollary 1b, page 182 of [22], $\mathcal{L}[t^{k/2-2} h((1-t)/t) \mathbf{1}_{(0,1)}(t)](s) \sim c_2 s^{\beta-k/2-1}$ as $s \rightarrow \infty$. Therefore by the representation in (17), $m^*(r) \sim c_3 r^{2\beta-k+2}$ as $r \rightarrow \infty$. Consequently, Brown's admissibility condition for generalized Bayes estimators is satisfied for the variance mixture of normals prior if the mixing density is such that $h(v) \sim c_1 v^\beta$ as $v \rightarrow \infty$ for $\beta \leq 0$. Note that the condition for propriety is $\beta < -1$.

4. Examples. In this section we will give several examples of the constructive method given by Corollary 1. It suffices to propose functions ψ_1 and ψ_2 such that the conditions of the corollary hold.

4.1. *Strawderman-type priors.* Suppose $\psi_1(v) = b$, for $b \leq 0$, and $\psi_2(v) = 0$. It follows from (15) that

$$(18) \quad \begin{aligned} h(v) &= \exp\left[-\frac{1}{2} \int_{v_0}^v \frac{2b+k-2}{u+1} du\right] \\ &= c_1(v+1)^{-(b+(k-2)/2)}, \end{aligned}$$

where c is the normalizing constant. As Condition (16) reduces to $b > 1 - k$, the corresponding Bayes estimator is minimax when $2 - k/2 < b \leq 0$. Note that h is integrable provided $b > 2 - k/2$ and, since $b \leq 0$, it follows that $k \geq 5$. The tail rate results in Comment 2 imply that, if $1 - k/2 \leq b \leq 0$, the minimax generalized Bayes rule is admissible.

It can be easily shown that Strawderman's prior corresponds to the mixing density $h(v) = (1-a)(1+v)^{a-2}$. Then the above conditions reduce to $3 - k/2 \leq a < 1$ for proper Bayes minimaxity. The corresponding conditions for admissible minimaxity are $3 - k/2 \leq a < 2$.

As an example of the generality of Corollary 1, this example could be extended through the choice of function ψ_2 such that $0 < \psi_2 \leq -b/2$ to give

$$h(v) = c_2(v+1)^{-(b+(k-2)/2)} \exp\left[-\frac{1}{2} \int_{v_0}^v \frac{\psi_2(u)}{u+1} du\right].$$

Any choice of nonmonotone ψ_2 leads to a nonmonotonic $r(\cdot)$ discussed in Comment 1.

4.2. *Shifted inverted gamma priors.* Suppose $\psi_1(v) = b - (a/(v+1))$, with $a > 0$ and $b \leq 0$ and $\psi_2(v) = 0$. In this case,

$$(19) \quad \begin{aligned} h(v) &= \exp\left[-\int_{v_0}^v \left[\frac{-a}{(u+1)^2} + \frac{b+(k-2)/2}{u+1}\right] du\right] \\ &= c \exp\left[\frac{-a}{v+1} - \left(b + \frac{k-2}{2}\right) \log(v+1)\right] \\ &= c \exp\left(-\frac{a}{v+1}\right) (v+1)^{1-(b+k/2)}. \end{aligned}$$

The integrability condition is $b + k/2 - 2 > 0$. Indeed the normalizing constant can be found by integrating (19); that is,

$$\int_0^\infty (v+1)^{1-(b+k/2)} \exp\left(-\frac{a}{v+1}\right) dv = a^{2-(b+k/2)} \gamma\left(b + \frac{k}{2} - 2, a\right),$$

where $\gamma(\cdot, \cdot)$ is the incomplete gamma function (see [14]). As Condition (16) reduces to $b + k - 1 > 0$, we obtain a proper Bayes minimax estimate for $2 - k/2 < b \leq 0$. The tail rate results in Comment 2 imply that, if $1 - k/2 \leq b \leq 0$, the minimax generalized Bayes rule is admissible.

4.3. *Generalized Student priors.* Suppose $\psi(v) = \alpha + \beta/v + \gamma/v^2$. We shall consider two cases. The conditions $\alpha \leq 0, \beta \leq 0$ and $\gamma < 0$ correspond to a monotonic $r(\cdot)$ in Comment 1. The second case we consider is $\alpha \leq 0, \beta > 0$ and $\gamma < 0$, which corresponds to a nonmonotonic $r(\cdot)$ in Comment 1. In each case we require specific conditions on α, β and γ so that the hypotheses of Corollary 1 are satisfied. In both cases

$$h(v) = \exp\left[-\int_{v_0}^v \frac{\alpha + \beta/u + \gamma/u^2 + (k-2)/2}{u+1} du\right] \\ = c(v+1)^{\beta-\alpha-\gamma-(k-2)/2} v^{\gamma-\beta} \exp\left(\frac{\gamma}{v}\right),$$

where c is the normalizing constant. The function h is integrable if $2 - k/2 < \alpha \leq 0$. As in the previous examples, the integrability condition implies (16). The admissibility condition of Comment 2 is $1 - k/2 \leq \alpha \leq 0$. Hence if $1 - k/2 \leq \alpha \leq 0$ and $\gamma \leq 0$ the corresponding minimax generalized Bayes estimate is admissible.

In the first case with $\alpha \leq 0, \beta \leq 0$ and $\gamma < 0$, we take $\psi = \psi_1$ and $\psi_2 = 0$. So the conditions $2 - k/2 < \alpha \leq 0$ and $\gamma < 0$ give rise to a proper Bayes minimax estimate.

When $\alpha \leq 0, \beta > 0$ and $\gamma < 0$ take $\psi_1(v) = \alpha + (\gamma/v)(1/v + \beta/\gamma)\mathbf{1}_{(0, -2\gamma/\beta)}(v)$, $\psi_2(v) = (\gamma/v)(1/v + \beta/\gamma)\mathbf{1}_{(-2\gamma/\beta, \infty)}(v)$, $C = \alpha$ and $D = \beta^2/4\gamma$. Then the minimaxity condition of Corollary 1 is $\alpha \leq \beta^2/2\gamma, \beta > 0$, and $\gamma < 0$.

It is of course possible to continue with functions ψ of the form $\psi(u) = \sum_{i=0}^n (\alpha_i/u^i)$. We prefer to pursue the connection with spherical multivariate Student- t priors with m degrees of freedom and a scale parameter equal to τ . This corresponds to $\psi(v) = \alpha + \beta/v + \gamma/v^2$ with $\alpha = (m - k + 4)/2, \beta = (m(1 - \tau) + 2)/2$ and $\gamma = -m\tau/2$. The general conditions on α, β and γ in case 1 reduce to $k - m - 4 \geq 0$ and $\tau \geq (m + 2)/m$. As mentioned above, this corresponds to a monotonic $r(\cdot)$ in Comment 1. This case is the example developed by [10], involving a monotonicity-type argument. However, if we use the second case of this example, we obtain a nonmonotone shrinkage function; that is when $\tau < (m + 2)/m$. The conditions for this case reduce to $k - m - 4 \geq 0, m(1 + \tau) + 2 > 0$ and $k \geq 2 + m\tau/2 + (m/2 + 2 + 2/m)/\tau$.

The case where the scale τ equals one is of particular interest, but cannot occur under the conditions of case 1; hence the $r(\cdot)$ function is necessarily nonmonotone. The conditions of case 2 when $\tau = 1$ reduce to $k \geq m + 4 + 2/m$. Specifically, if $m = 1$ (the spherical multivariate Cauchy) and $k \geq 7$, the corresponding Bayes estimate is minimax. Furthermore, if $m \geq 2$ and $k \geq m + 5$, we also get the desired minimaxity result. Numerical studies indicate that, for $k = 5$ and $m = 1$, we indeed need a scale parameter $\tau > 2$ for minimaxity, while, when $k = 6$, a scale of $\tau = 1$ will suffice.

It is important to note that these results for the t -distribution prior cannot be deduced using a classical monotonicity argument. In [2], [6] and [11], the Bayesian robustness of the Cauchy prior was examined; the minimaxity result serves as a nice complement. That is, one can carry out a simple Bayesian

analysis with a Cauchy prior and still maintain the desired frequentist conservatism of minimaxity.

5. Concluding remarks. Brown [5] conjectured a rate at which the tail of the prior must decay to attain minimaxity. He suggested that prior distributions on θ should have a tail that behaves as $\|\theta\|^{4-2k}$ in order to be minimax. We can use the Tauberian-type results in Comment 2 to study the tail behavior of the prior since $\pi(\theta) = \mathcal{L}[t^{k/2-2}h(1/t)](\|\theta\|^2)$. In this case it is easy to show that if $h(v) \sim cv^\beta$ as $v \rightarrow \infty$, then $\pi(\theta) \sim c\|\theta\|^{4\beta-2k+4}$ as $\|\theta\| \rightarrow \infty$. In the case of Strawderman's prior, it follows that the tail of $\pi(\theta)$ behaves as $\|\theta\|^{2a-k-2}$, so that if we choose $a = 3 - k/2$ the tail rate agrees with Brown's rate. Similarly, the tail of a multivariate- t prior in k dimensions with m degrees of freedom behaves as $\|\theta\|^{-m-k}$. This agrees with Brown's prediction for minimaxity when $m = k - 4$, in agreement with the result of Section 4.3. Hence it seems our results for the t -priors are sharp in the sense that, if $k - m \leq 4$, the resulting Bayes procedures will not be minimax. These rates are a polynomial, hence ruling out a variety of lighter to medium tailed priors such as the multivariate normal and multivariate double exponential.

The tail rates needed to attain minimaxity shed some light on the area of "what if" asymptotics, that is, asymptotics which show how posterior densities behave as the discrepancy between the likelihood and prior information tends to infinity. This theory classifies distributions into three classes based on asymptotic tail behavior. Dawid [7] and O'Hagan [17] showed that, when θ is a location parameter, assuming a Student- t prior density and normal sampling density causes the posterior tail of θ to behave like the tail of a normal density. Sansó and Pericchi [18] examined behavior for a normal likelihood and Laplace prior, finding that the posterior mean tends to $x - c$ where c is some constant, and thus the prior exerts *bounded influence*. Angers and Berger [2] examined the behavior for a Cauchy prior. Chance and Wells [6] give some results for general location families.

For the class of variance mixtures of normal distribution priors, we have seen that superharmonicity of the square root of the marginal density gives rise to a differential inequality that can be solved to yield admissible Bayes minimax rules. We present some new priors that should serve as useful default priors for Bayesians who wish to use estimates that have good frequentist properties. Our results should also be comforting to Bayesians who routinely use t -priors for robustness consideration.

REFERENCES

- [1] ALAM, K. (1973). A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **1** 517–525.
- [2] ANGERS, J. F. and BERGER, J. O. (1991). Robust hierarchical Bayes estimation of exchangeable means. *Canad. J. Statist.* **19** 39–56.
- [3] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.
- [4] BERGER, J. O. and ROBERT, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.* **18** 617–651.

- [5] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.
- [6] CHANCE, B. L. and WELLS, M. T. (1994). Characterizing hierarchical model behavior. In *Proceedings of the 26th Symposium on the Interface. Computationally Intensive Statistical Methods* (J. Sall and A. Lehman, eds.) 231–239.
- [7] DAWID, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60** 664–667.
- [8] DELLACHERIE, C. and MEYER, P.A. (1978). *Probabilities and Potential*. North-Holland, Amsterdam.
- [9] DOOB, J. L. (1984). *Classical Potential Theory and Its Probabilistic Counterpart*. Springer, New York.
- [10] FAITH, R. E. (1978). Minimax Bayes point estimators of a multivariate normal mean. *J. Multivariate Analysis* **8** 372–379.
- [11] FOURDRINIER, D. and WELLS, M. T. (1996). Spherically symmetric Bayes estimators for a linear subspace of a normal law. In *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 569–579. Oxford Univ. Press.
- [12] GEORGE, E. I. (1986). Combining minimax shrinkage estimators. *J. Amer. Statist. Assoc.* **81** 437–445.
- [13] GEORGE, E. I. (1986). Minimax multiple shrinkage estimators. *Ann. Statist.* **14** 188–205.
- [14] GRADSHTEYN, I. and RYZHIK, I. (1994). *Tables of Integrals, Series and Products*, 5th ed. Academic Press, New York.
- [15] HEATH, D. and SUDDERTH, W. (1989). Coherent inference from improper priors and from finitely additive priors. *Ann. Statist.* **17** 907–919.
- [16] LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.
- [17] O'HAGAN, A. (1988). Modeling with heavy tails. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 345–359. Oxford Univ. Press.
- [18] SANSÓ, B. and PERICCHI, L. R. (1992). Near ignorance classes of log-concave priors for the location model. *Test* **1** 39–46.
- [19] STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.
- [20] STRAWDERMAN, W. E. (1970). On the existence of proper Bayes minimax estimators of the mean of a multivariate normal distribution. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 51–55. Univ. California Press, Berkeley.
- [21] STRAWDERMAN, W. E. (1971). Proper bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.
- [22] WIDDER, D. V. (1946). *The Laplace Transform*. Princeton Univ. Press.

D. FOURDRINIER
 UNIVERSITÉ DE ROUEN
 URA CNRS 1378
 76821 MONT SAINT AIGNAN CEDEX
 FRANCE

W. E. STRAWDERMAN
 DEPARTMENT OF STATISTICS
 RUTGERS UNIVERSITY
 NEW BRUNSWICK, NEW JERSEY 08903

M. T. WELLS
 DEPARTMENT OF SOCIAL STATISTICS
 CORNELL UNIVERSITY
 358 IVES HALL
 ITHACA, NEW YORK 14851-0952
 E-MAIL: mtwl@cornell.edu