Bayes minimax estimators of the mean of a scale mixture of multivariate normal distributions

Dominique Fourdriniera,∗, Othmane Kortbib, William E. Strawdermanc,1

aUniversité de Rouen, UMR CNRS 6085, BP 12 76801 Saint Étienne du Rouvray, France
bUniversité de Médéa, Algeria
cDepartment of Statistics, Rutgers University, New Brunswick, NJ 08903, USA

Received 7 February 2006
Available online 28 August 2006

Abstract

Bayes estimation of the mean of a variance mixture of multivariate normal distributions is considered under sum of squared errors loss. We find broad class of priors (also in the variance mixture of normal class) which result in proper and generalized Bayes minimax estimators. This paper extends the results of Strawderman [Minimax estimation of location parameters for certain spherically symmetric distribution, J. Multivariate Anal. 4 (1974) 255–264] in a manner similar to that of Maruyama [Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distribution, J. Multivariate Anal. 21 (2003) 69–78] but somewhat more in the spirit of Fourdrinier et al. [On the construction of bayes minimax estimators, Ann. Statist. 26 (1998) 660–671] for the normal case, in the sense that we construct classes of priors giving rise to minimaxity. A feature of this paper is that in certain cases we are able to construct proper Bayes minimax estimators satisfying the properties and bounds in Strawderman [Minimax estimation of location parameters for certain spherically symmetric distribution, J. Multivariate Anal. 4 (1974) 255–264]. We also give some insight into why Strawderman’s results do or do not seem to apply in certain cases. In cases where it does not apply, we give minimax estimators based on Berger’s [Minimax estimation of location vectors for a wide class of densities, Ann. Statist. 3 (1975) 1318–1328] results. A main condition for minimaxity is that the mixing distributions of the sampling distribution and the prior distribution satisfy a monotone likelihood ratio property with respect to a scale parameter.

AMS 1991 subject classification: primary 62C20; 62C15; 62C10

Keywords: Minimax estimators; Variance mixture of multivariate normal distributions; Quadratic loss; Bayes estimators

∗ Corresponding author. Fax: +33 232 103 794.
E-mail address: Dominique.Fourdrinier@univ-rouen.fr (D. Fourdrinier).

1 The support of NSA Grant 03G-112 is gratefully acknowledged.

© 2006 Elsevier Inc. All rights reserved.

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.06.012
1. Introduction

In this paper we study Bayes minimax estimation of the mean vector of a variance mixture of a multivariate normal distributions under sum of squared errors loss in dimension three and greater. It has been known since Stein [11], in the normal case, and Brown [4], generally, that the best equivariant and minimax estimator is inadmissible for \( p \geq 3 \). Explicit improvements in the normal case were given by James and Stein [8] and by many other authors thereafter. Explicit improvements for other subclasses of distributions were given by several authors starting with Strawderman [15] for the case of variance mixtures of normals and Berger [1] for certain general classes of spherically symmetric distributions. See also, for example [2,3]. All of these early papers (except [15]) in the non-normal setting did not consider (generalized) Bayes minimax estimators.

Strawderman [14] gave proper Bayes minimax estimators for the normal case for \( p \geq 5 \), while Fourdrinier et al. [7] constructed broad classes of proper and generalized Bayes estimators in this case as well.

Strawderman [15] gave generalized Bayes minimax estimators for certain variance mixtures of normals. Recently, Maruyama [9] extended these results and gave classes of proper and generalized Bayes estimators for the same subclass as Strawderman [15]. In particular these results are for the case where the mixing distribution has monotone likelihood ratio when considered as a scale family.

In this paper we restrict ourselves to the class of mixing distributions with monotone likelihood ratio. We give broad classes of priors (including Maruyama’s), somewhat in the spirit of Fourdrinier et al. [7], that lead to proper and generalized Bayes minimax estimators.

An interesting property of the class of priors, also shared by Maruyama’s, is that they too are a variance mixture of normals with monotone likelihood ratio.

The methods of proof in this paper are also generally similar to those of Maruyama with one exception. Both papers consider estimators of the form \( \delta(X) = (1 - r(\|X\|_2^2)/\|X\|_2^2)X \) where \( r(w) \) is non-decreasing, non-negative, and bounded. The proofs of monotonicity and of determination of the bound are quite similar. To determine minimaxity, one of two results is used in each paper; (a) the result of Strawderman [15] which also requires \( r(w)/w \) to be monotone non-increasing but gives a larger upper bound for \( r(w) \) or (b) a result of Berger [1] which requires the existence of more moments but does not require that \( r(w)/w \) be non-increasing.

Our methods and results differ from those of Maruyama in that we are able to give conditions that guarantee monotonicity of \( r(w)/w \) in certain general cases. In these cases it is possible to use the larger bound on \( r(w) \) in [15] and to find proper Bayes minimax estimators.

Another main difference in the two papers is that the class of priors considered by Maruyama is specific. In our setting, this class is characterized by a mixing density on the variance of the form \( h(t) = t^b(1 + t)^{a-2-b} \) with \( b \geq 0 \). Our class of mixing distributions, which contains this class, is given by \( h(t) \) such that \( h(t) \leq Kn^{-z} \) for \( 0 < t < t_0 \) and \( z < 1 \), \( h(0) < \infty \) and \( \lim_{t \to \infty} h(t)/t^\beta = c > 0 \).

Section 2 states the problem and develops the form of the Bayes estimators. Section 3 contains the main results. Section 4 gives examples illustrating the theory. Two basic examples, a gamma mixture and an inverse gamma mixture (the Student \( t \) case) indicate when our method does or does not lead to a proper Bayes minimax estimator with \( r(w)/w \) non-increasing. In Section 5, we give some concluding remarks. Finally, an Appendix gives certain of the proofs.
2. General expression of Bayes estimators

Let \( X \) be a random vector in \( \mathbb{R}^p \) (\( p \geq 3 \)) distributed as a variance mixture of multivariate normal distributions with mean vector \( \theta \). Thus we assume that the density function of \( X \) is of the form

\[
 f(x) = \int_0^\infty \frac{1}{(2\pi v)^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x - \theta\|^2}{v} \right) dG(v), \tag{2.1}
\]

where \( G \) is the distribution of a known non-negative random variable \( V \). More precisely, most of this paper is devoted to the case where \( G \) has a density \( g \) with respect to the Lebesgue measure on \( \mathbb{R}_+ \) (\( g \) is said the mixing density). The goal of this paper is to give sufficient conditions for Bayes estimators of \( \theta \) to be minimax, under the usual quadratic loss function \( L(\theta, \delta) = \|\theta - \delta\|^2 \).

For a prior probability measure \( \pi \) the marginal distribution of \( X \) has a density \( m \) with respect to the Lebesgue measure in \( \mathbb{R}^p \) given by

\[
 m(x) = \int_{\mathbb{R}^p} \int_0^\infty \frac{1}{(2\pi v)^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x - \theta\|^2}{v} \right) dG(v) d\pi(\theta). \tag{2.2}
\]

Upon an application of Fubini’s theorem for positive functions, for any \( x \in \mathbb{R}^p \), we have

\[
 m(x) = \int_0^\infty K(x, v) dG(v), \tag{2.3}
\]

where

\[
 K(x, v) = \int_{\mathbb{R}^p} \frac{1}{(2\pi v)^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x - \theta\|^2}{v} \right) d\pi(\theta). \tag{2.4}
\]

The Bayes estimator \( \delta_\pi = \delta_\pi(X) \), which is defined as the minimizer of the Bayes risk, is given for any \( x \in \mathbb{R}^p \) in the case of quadratic loss by \( \delta_\pi(x) = E[\theta|x] \), where this last expectation is considered with respect of the posterior distribution given \( x \). After classical calculations, we obtain

\[
 \delta_\pi(x) = x + \gamma(x) \tag{2.5}
\]

with

\[
 \gamma(x) = \frac{\nabla}{m(x)} \int_0^\infty vK(x, v) dG(v), \tag{2.6}
\]

where the symbol \( \nabla = (\partial/\partial x_1, \ldots, \partial/\partial x_p) \) denotes the gradient. Note that, in the special case where \( G \) is Dirac measure, we find again the normal case where \( \gamma(x) = \nabla \log(m(x)) \) considered in Fourdrinier et al. [7].

Recall that the quadratic risk of any estimator \( \delta(X) = X + \gamma(X) \), that is

\[
 R(\theta, \delta) = E_\theta[\|\theta - \delta\|^2]
\]

(where \( E_\theta \) denotes the expectation with respect to (2.1)), is finite as soon as the risk of \( X \) is finite (that is, if \( E[V] < \infty \) if and only if \( E[\|\gamma(X)\|^2] < \infty \) (this can be verified by an application of Schwarz’s inequality).
As the sampling density is a variance mixture of normal distributions of type (2.1), a natural choice for a prior distribution $\pi$ is to assume that it has a density with respect to the Lebesgue measure of the form

$$
\theta \mapsto \int_0^\infty \frac{1}{(2\pi t)^{p/2}} \exp \left( -\frac{1}{2} \frac{\|\theta\|^2}{t} \right) h(t) \, dt,
$$

(2.7)

where $h$ is a function from $\mathbb{R}_+$ into $\mathbb{R}_+$ such that this integral exists ($h$ is the mixing function).

In this context, the expression of $K(x, v)$ in (2.4) can be calculated through an application of Fubini’s theorem for positive functions. Indeed, for any $x \in \mathbb{R}^p$ and any $v \geq 0$, we have

$$
K(x, v) = \int_{\mathbb{R}^p} \frac{1}{(2\pi v)^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x - \theta\|^2}{v} \right) \int_0^\infty \frac{1}{(2\pi t)^{p/2}} \exp \left( -\frac{1}{2} \frac{\|\theta\|^2}{t} \right) h(t) \, dt \, d\theta
$$

$$
= \int_0^\infty \frac{1}{(2\pi (v + t))^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x\|^2}{v + t} \right) h(t) \, dt.
$$

(2.8)

Then the marginal density $m(x)$ in (2.3) becomes

$$
m(x) = \int_0^\infty \int_0^\infty \frac{1}{(2\pi (v + t))^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x\|^2}{v + t} \right) h(t) \, dt \, dG(v).
$$

(2.9)

Through the Lebesgue dominated convergence theorem we obtain, according to (2.6),

$$
\gamma(x) = \frac{1}{m(x)} \int_0^\infty v \nabla K(x, v) \, dG(v).
$$

(2.10)

Substituting $K(x, v)$ by its value given in (2.8), it follows that

$$
\gamma(x) = -\frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi (v + t))^{p/2}} \frac{v}{v + t} \exp \left( -\frac{1}{2} \frac{\|x\|^2}{v + t} \right) h(t) \, dt \, dG(v)}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi (v + t))^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x\|^2}{v + t} \right) h(t) \, dt \, dG(v)} x.
$$

(2.11)

Finally, the Bayes estimator resulting from a variance mixture of normal distributions prior is of the form

$$
\delta_h(x) = x - \frac{r(\|x\|^2)}{\|x\|^2} x,
$$

(2.12)

where

$$
r(w) = \frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi (v + t))^{p/2}} \frac{v}{v + t} \exp \left( -\frac{1}{2} \frac{w}{v + t} \right) h(t) \, dt \, dG(v)}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi (v + t))^{p/2}} \exp \left( -\frac{1}{2} \frac{w}{v + t} \right) h(t) \, dt \, dG(v)}.
$$

(2.13)
It is worth noting that the Bayes estimator has always finite risk. Indeed, for any \( x \in \mathbb{R}^p \), we have
\[
\|\hat{\gamma}(x)\|^2 = \|x\|^2 \left[ \frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \frac{v}{v+t} \exp \left( -\frac{1}{2} \frac{\|x\|^2}{v+t} \right) h(t) dt dG(v)}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp \left( -\frac{1}{2} \frac{\|x\|^2}{v+t} \right) h(t) dt dG(v)} \right]^2 \leq \|x\|^2
\]
using \( 0 < \frac{v}{v+t} \leq 1 \). Thus, it follows that
\[
E_\theta[\|\hat{\gamma}(X)\|^2] \leq E_\theta[\|X\|^2] \leq pE[V] + \|\theta\|^2 < \infty
\]
since we have supposed that \( E[V] < \infty \).

To prove the minimaxity of the Bayes estimator \( \hat{\delta}_h \) in (2.12), we cannot rely on an unbiased estimator of its risk, in contrast to Stein [12] in the normal case. Strawderman [15] and Berger [1], to prove minimaxity of an estimator of type (2.12) for variance mixture of normal distributions, use the fact the function \( r \) is bounded from above and non-decreasing. They also need the property that the function \( w \mapsto -\frac{r(w)}{w} \) is non-increasing. However, it is worth noting that Berger [1], proposing the class of densities \( x \mapsto f(\|x-\theta\|^2) \) such that \( c = \inf_{t \geq 0} \int_t^\infty f(u) du/f(t) > 0 \), does not use this last monotonicity condition. Furthermore it is interesting to notice that this class of densities contains some variance mixtures of normal distributions. Indeed, it is easy to show that a density of form (2.1) belongs to this class with \( c = 2E[V^{1-p/2}]/E[V^{-p/2}] \). This fact was used by Maruyama [9]. We recall below an adaptation of the results of Strawderman and Berger.

**Theorem 2.1.** Let \( f \) be a density with respect to Lebesgue measure of type (2.1) with \( p \geq 3 \). Let \( \hat{\delta}_h \) be an estimator of form (2.12) where \( r \) is non-decreasing. Then \( \hat{\delta}_h \) is a minimax estimator of \( \theta \) under quadratic loss if,

- either (a) \( 0 \leq r \leq c(p-2) \) with \( c = 2E[V^{1-p/2}]/E[V^{-p/2}] \),
- or (b) \( 0 \leq r \leq c^*(p-2) \) with \( c^* = 2/E[V^{-1}] \) and \( r(w)/w \) is non-increasing.

It is easy to see that \( c^* > c \).

### 3. Minimax Bayes estimators

Throughout this section we assumed that the mixing distribution \( G \) has a density \( g \). We will see that, for a sampling distribution of form (2.1) and for a prior of form (2.7), the main condition for obtaining minimaxity of the corresponding Bayes estimator is that both the mixing distribution \( g \) and the mixing (possibly improper) density \( h \) have monotone non-decreasing likelihood ratio when considered as a scale parameter family. It was noticed by Maruyama [9] that (expressed in terms of the function \( h \)) this property is, respectively, equivalent to
\[
h(s_1 t_1) h(s_2 t_2) \leq h(s_1 t_2) h(s_2 t_1) \quad (3.1)
\]
for any \( s_1 \leq s_2 \) and \( t_1 \leq t_2 \) and to the fact that the function
\[
t \mapsto t \frac{h'(t)}{h(t)} \quad (3.2)
\]
is non-increasing (if \( h \) is absolutely continuous). Actually, in the following lemma, this monotone likelihood property implies the monotonicity of the function \( r \) in (2.13) under a mild growth condition on \( h \).

**Lemma 3.1.** If both \( h \) and \( g \) have monotone increasing likelihood ratio when considered as a scale parameter family and if \( h(0) < \infty \), \( h(t) \) is \( o(t^{p/2-1}) \) for \( t \) in a neighborhood of infinity and is absolutely continuous, then the function

\[
r(w) = w \frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \frac{v}{v+t} \exp \left( -\frac{1}{2} \frac{w}{v+t} \right) h(t) dt g(v) dv}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp \left( -\frac{1}{2} \frac{w}{v+t} \right) h(t) dt g(v) dv}
\]

is non-decreasing.

The proof of Lemma 3.1 is postponed to the Appendix.

The following lemma (whose proof is also given in the Appendix) gives conditions on \( h \) and \( g \) such that \( \lim_{w \to \infty} r(w) \) can be determined. Under the additional conditions of Lemma 3.1, this limit is the upper bound of \( r \).

**Lemma 3.2.** Assume that the mixing density \( g \) of the sampling distribution in (2.1) is such that

\[
E[V] = \int_0^\infty v g(v) dv < \infty \quad \text{and} \quad E[V^{-1}] = \int_0^\infty v^{-1} g(v) dv < \infty.
\]

Assume also that the mixing density \( h \) of the (possibly improper) prior distribution in (2.7) satisfies

\[
\lim_{t \to -\infty} \frac{h(t)}{t^\beta} = c
\]

for some \( \beta < p/2 - 1 \) and some \( c > 0 \), and assume that

\[
E[V^{1-\beta}] = \int_0^\infty v^{1-\beta} g(v) dv < \infty.
\]

Then

\[
\lim_{w \to \infty} r(w) = (p - 2 - 2\beta)E[V]
\]

provided there exist \( K > 0, t_0 > 0 \) and \( \alpha < 1 \) such that

\[
h(t) \leq K t^{-\alpha} \quad \text{for} \ 0 < t < t_0.
\]

Note that Condition (3.5) implies that \( h(t) \leq K t^\beta \) for \( t_0 \leq t < \infty \) (it is clear that we can use the same \( K \) as in (3.7)).

Combining Theorem 2.1.a, Lemmas 3.1 and 3.2 gives immediately our first domination result.

**Theorem 3.1.** Assume that the mixing density \( g \) of the sampling distribution in (2.1) is such that

\[
E[V] = \int_0^\infty v g(v) dv < \infty \quad \text{and} \quad E[V^{-p/2}] = \int_0^\infty v^{-p/2} g(v) dv < \infty.
\]

is non-increasing (if \( h \) is absolutely continuous). Actually, in the following lemma, this monotone likelihood property implies the monotonicity of the function \( r \) in (2.13) under a mild growth condition on \( h \).
Assume also that the mixing density \( h \) of the (possibly improper) prior distribution in (2.7) is absolutely continuous and satisfies Condition (3.5), for some \( \beta < p/2 - 1 \) and for \( 0 < c < \infty \). Assume finally that \( h \) and \( g \) have monotone increasing likelihood ratio when considered as a scale parameter family.

Then, if Condition (3.7) is satisfied for some \( K > 0, t_0 > 0 \) and \( \alpha < 1 \), the (generalized or proper) Bayes estimator \( \delta_0 \) with respect to the prior distribution corresponding to the mixing distribution \( h \) is minimax provided that \( \beta \) satisfies

\[
-(p - 2) \left[ \frac{E[V^{-\frac{\beta}{2} + \frac{1}{2}}]}{E[V]E[V^{-\frac{\beta}{2}}]} - \frac{1}{2} \right] \leq \beta.
\]

(3.9)

It is easy to see that Condition (3.8) implies Condition (3.6) since \( \beta < -1 \) corresponds to the proper priors. Note also, since \( E[V^{-\frac{\beta}{2} + \frac{1}{2}}]/(E[V]E[V^{-\frac{\beta}{2}}]) \leq 1 \), we have

\[
-(p - 2) \left[ \frac{E[V^{-\frac{\beta}{2} + \frac{1}{2}}]}{E[V]E[V^{-\frac{\beta}{2}}]} - \frac{1}{2} \right] \geq - \frac{p - 2}{2}.
\]

Hence, in the case of a proper prior, it is necessary that \(-(p - 2)/2 < -1\) which is equivalent to \( p > 4 \). This is related to the known fact in the normal case that \( p > 4 \) is required for the existence of a proper Bayes minimax estimator (see [13]).

Comment (Admissibility): For priors with mixing distribution \( h \) satisfying (3.5) and (3.7) an argument as in Maruyama [9] using Brown [5] and a Tauberian theorem suggests that the resulting generalized Bayes estimator is admissible if \( \beta \leq 0 \). A referee called to our attention that, recently, Maruyama and Takemura [10] have verified this under additional conditions which imply, in our setting, that \( E_0[\|X\|^3] < \infty \).

As mentioned in Section 2, the bound \( c^*(p - 2) \) in part (b) of Theorem 2.1 for minimaxity is greater than or equal to the bound \( c(p - 2) \) in part (a). However, this increase in the bound is obtained at the cost of the assumption of monotonicity of \( r(w)/w \). In general, it does not appear that the assumptions of Lemma 3.1 are sufficient to guarantee this monotonicity. The reason relies in the fact that both densities \( g \) and \( h \) have a monotone likelihood ratio property in the same direction. This will become clearer in the discussion below and in the examples in Section 4.

Here is one way to deal with the monotonicity of \( r(w)/w \). Note first that, according to (3.3), we have

\[
\frac{r(w)}{w} = E_w \left[ \frac{V}{V + T} \right],
\]

where \( E_w \) is the expectation with respect to the density

\[
f(t, v|w) \propto \left( \frac{1}{2\pi(v + t)} \right)^{p/2} \exp \left( -\frac{1}{2} \frac{w}{v + t} \right) h(t)g(v).
\]

Through the change of variable \( \lambda = v/(v + t) \) and \( z = v + t \), we have

\[
f(\lambda, z|w) \propto z^{-p/2 + 1} \exp \left( -\frac{1}{2} \frac{w}{z} \right) h(1 - \lambda)g(\lambda z),
\]
so that the density of \( \lambda \) given \( z \) and \( w \) does not depend on \( w \) since

\[
f(\lambda|z, w) = \frac{f(\lambda, z|w)}{\int_0^1 f(\lambda, z|w) d\lambda} = c(z)h(z(1 - \lambda))g(\lambda z) = f(\lambda|z).
\] (3.10)

Further the family of densities given by

\[
f(z|w) = \int_0^1 f(\lambda, z|w) d\lambda = K(z)z^{-p/2+1} \exp\left(-\frac{1}{2} w z\right),
\]

for some constant \( K(z) \), has monotone increasing likelihood ratio property with respect to the parameter \( w \) since, for \( w_1 < w_2 \), we have

\[
\frac{f(z|w_2)}{f(z|w_1)} = \exp\left(-\frac{1}{2} z (w_2 - w_1)\right).
\]

Hence the monotonicity of \( r(w)/w \) follows provided the function \( E[\lambda|z, w] = E[\lambda|z] \) is non-increasing in \( z \). The monotone decreasing likelihood ratio property for the density \( f(\lambda|z) \) in (3.10) is a natural sufficient condition to insure this monotonicity. Note, however, that \( f(\lambda|z) \propto h(z(1 - \lambda))g(z\lambda) \) where \( h(z(1 - \lambda)) \) and \( g(z\lambda) \) have monotone likelihood ratio in opposite direction according to Lemma 3.1. As we will see in Section 4, there are examples where \( f(\lambda|z) \) has decreasing monotone likelihood ratio in \( z \) (Example 4.2) and examples where \( f(\lambda|z) \) does not (Example 4.1).

The following theorem is useful in situations where decreasing monotone likelihood ratio property holds for \( f(\lambda|z) \).

**Theorem 3.2.** Under the conditions of Theorem 3.1, assume that \( f(\lambda|z) \) in (3.10) has decreasing monotone likelihood ratio in \( z \). Then minimaxity of \( \delta_h \) is satisfied as soon as

\[
-(p - 2) \left[ \frac{1}{E[V^{-1}]} - \frac{1}{2} \right] \leq \beta < \frac{p}{2} - 1.
\] (3.11)

**Proof.** The proof follows from Theorem 2.1.b, Lemmas 3.1 and 3.2 and the above discussion. \( \square \)

The minimaxity results of Theorem 3.1 and 3.2 rely on the monotone likelihood ratio property of the densities \( g \) and \( h \). This class of densities is essentially described in the following lemma.

**Lemma 3.3.** Let \( \varphi \) be a non-increasing function on \( \mathbb{R}_+ \) such that the indefinite integral of \( \varphi(u)/u \) exists. Then the function \( h \) defined by

\[
h(t) = C \exp\left\{ \int_t^\infty \frac{\varphi(u)}{u} du \right\}
\] (3.12)

with \( C > 0 \) and \( \alpha > 0 \) has monotone non-decreasing likelihood ratio when considered as a scale parameter family.
Proof. Let $0 < \sigma_1 < \sigma_2$. For any $t \in \mathbb{R}_+$, we have
\[
\frac{1/\sigma_2 h(t/\sigma_2)}{1/\sigma_1 h(t/\sigma_1)} = \frac{\sigma_1}{\sigma_2} \exp \left\{ \int_{t/\sigma_2}^{t/\sigma_1} \frac{-\varphi(u)}{u} \, du \right\}
\]
\[
= \frac{\sigma_1}{\sigma_2} \exp \left\{ \int_{1/\sigma_2}^{1/\sigma_1} \frac{-\varphi(tv)}{v} \, dv \right\}
\]
which is non-decreasing in $t$ by monotonicity of $\varphi$. □

Note that the fact that Lemma 3.3 gives virtually all smooth functions with monotone likelihood ratio follows since $\varphi(t) = th'(t)/h(t)$ and thus $h$ satisfies (3.2).

The monotone likelihood ratio property of $f(\lambda|z)$ needed in Theorem 3.2 involves the two densities $g$ and $h$. For densities of form (3.12), this link is made explicit in the following lemma.

**Lemma 3.4.** Let $\varphi$ (respectively $\varphi_0$) a derivable non-increasing function on $\mathbb{R}_+$ such that the indefinite integral of $\varphi(u)/u$ (respectively of $\varphi_0(u)/u$) exists. Then, for
\[
h(t) = C \exp \left\{ \int_{z}^{t} \frac{\varphi(u)}{u} \, du \right\}
\]
with $C > 0$ and $\alpha > 0$

and
\[
g(v) = C_0 \exp \left\{ \int_{z_0}^{v} \frac{\varphi_0(u)}{u} \, du \right\}
\]
with $C_0 > 0$ and $z_0 > 0$,

the density $f(\lambda|z)$ given in (3.10) has decreasing monotone likelihood ratio in $z$ if and only if the derivatives of $\varphi$ and $\varphi_0$ satisfy
\[
\varphi'_0(z\lambda) \leq \varphi'(z(1 - \lambda))
\]
for any $z \geq 0$ and any $\lambda \in ]0, 1[$.

**Proof.** Let $0 \leq z_1 < z_2$. According to (3.10), we have to show that the function
\[
\frac{h(z_2(1 - \lambda))g(z_2\lambda)}{h(z_1(1 - \lambda))g(z_1\lambda)}
\]
is non-increasing in $\lambda$ which reduces to
\[
\frac{d}{d\lambda} \left\{ \int_{z}^{z_2(1 - \lambda)} \frac{\varphi(u)}{u} \, du - \int_{z}^{z_1(1 - \lambda)} \frac{\varphi(u)}{u} \, du + \int_{z_0}^{z_2\lambda} \frac{\varphi_0(u)}{u} \, du - \int_{z_0}^{z_1\lambda} \frac{\varphi_0(u)}{u} \, du \right\} \leq 0
\]
that is to
\[
\frac{\varphi(z_1(1 - \lambda))}{1 - \lambda} - \frac{\varphi_0(z_1\lambda)}{\lambda} \leq \frac{\varphi(z_2(1 - \lambda))}{1 - \lambda} - \frac{\varphi_0(z_2\lambda)}{\lambda}.
\]
The latter inequality expresses that the function
\[ z \mapsto \frac{\varphi(z(1 - \lambda))}{1 - \lambda} - \frac{\varphi_0(z\lambda)}{\lambda} \]
is non-decreasing in \( z \), for any \( \lambda \in [0, 1] \), and hence that
\[ \varphi'(z(1 - \lambda)) - \varphi'_0(z\lambda) \geq 0 \]
for any \( z \geq 0 \) and any \( \lambda \in [0, 1] \). □

4. Examples

In this section, we consider sampling distributions where the mixing density \( g \) is a gamma or an inverse gamma density. We start with the latter which corresponds to a generalized Student \( t \).

Example 4.1. Inverse gamma: Let \( g(v) \propto \frac{1}{v^{a_0 + 1}} \exp(-b_0/v) \) with \( a_0 > 1 \) and \( b_0 > 0 \). Let also the mixing prior \( h \) be a (possibly) generalized inverse gamma, that is, \( h(t) \propto \frac{1}{t^{a + 1}} \exp(-b/t) \) with \( a > -p/2 \) and \( b > 0 \). Note that \( h \) and \( g \) have monotone increasing likelihood ratio when considered as a scale parameter family.

(1) Minimaxity using Theorem 3.1: It is clear that Conditions (3.8) and (3.5) are satisfied with \( a_0 > 1 \) and \( \beta = -(a + 1) \). It is also clear that Condition (3.7) holds for any \( z < 1 \). Finally a simple calculation shows that
\[ \frac{E[V^{-\frac{p}{2} + 1}]}{E[V]E[V^{-\frac{p}{2}}]} = \frac{a_0 - 1}{p/2 + (a_0 - 1)} \]
so that Condition (3.9) reduces to
\[ a \leq (p - 2) \left[ \frac{a_0 - 1}{p/2 + a_0 - 1} - \frac{1}{2} \right] - 1 \quad (4.1) \]
which guarantees the minimaxity of the (generalized or proper) Bayes estimator \( \delta_h \).

Note that the condition \( E[V^{1-\beta}] = E[V^{a+2}] < \infty \) is equivalent to \( a_0 - 2 > a \). Thus \( a_0 \neq a \) and the mixing densities \( g \) and \( h \) cannot be equal. This is consistent with the fact that, when the sampling and the prior distributions are identical, necessarily, the Bayes estimator is \( X/2 \).

Note also that properness of \( h \) requires \( a > 0 \) so that
\[ 0 < a \leq (p - 2) \left[ \frac{a_0 - 1}{p/2 + (a_0 - 1)} - \frac{1}{2} \right] - 1 \]
and this double inequality can hold if and only if \( p \geq 5 \) and
\[ a_0 > 1 + \frac{p}{2} \frac{p}{p - 4} \]

In particular, if the sampling distribution is a \( p \)-variate Student \( t \) with \( n_0 \) degrees of freedom, the mixing density \( g \) is the inverse gamma \( (n_0/2, n_0/2) \), that is \( g(v) \propto v^{-\frac{n_0 + 2}{2}} \exp(-n_0/2v) \). Similarly, if the prior is Student \( t \) distribution with \( n \) degrees of freedom, we have \( h(t) \propto t^{-\frac{n + 2}{2}} \exp(-n/2t) \). This corresponds to \( a_0 = n_0/2, b_0 = n_0/2, a = n/2 \) and \( b = n/2 \).
The condition \( a_0 > 1 \) corresponds to \( n_0 > 2 \) and the condition \( a > -p/2 \) is of course satisfied since it is a “true” Student \( t \) distribution. The condition for minimaxity (4.1) becomes
\[
n \leq \frac{2(n_0 - 2)}{p + n_0 - 2} (p - 2) - p. \tag{4.2}
\]
Note also that the above remark \( a_0 - 2 > a \) corresponds to \( n_0 > n + 4 \).

The properness condition imposes that the left-hand side of (4.2) is positive which implies
\[
n_0 > 2 + \frac{p^2}{p - 4}.
\]
It is worth noting that, for large \( n_0 \), Condition (4.2) becomes approximatively \( n \leq p - 4 \) which corresponds to the condition for proper Bayes minimaxity under a normal sampling distribution given in [7].

(2) Minimaxity using Theorem 3.2: Coming back to the general case of inverse gamma densities, it is clear that Theorem 3.2 does not apply with the choice of a prior mixing density in the same class as the sampling mixing density. Indeed it is easy to show that, for \( z_1 < z_2 \),
\[
\frac{f(\lambda|z_2)}{f(\lambda|z_1)} = \psi(z_1, z_2) \exp\left(\frac{1}{z_1} - \frac{1}{z_2}\right) \left(\frac{b}{1 - \frac{b}{\lambda}} + \frac{b_0}{\lambda}\right)
\]
(for some function \( \psi \)) which is non-monotone in \( \lambda \).

In fact note that \( g \) (respectively \( h \)) corresponds, in Lemma 3.3, to the choice of \( \varphi_0(v) = -(a_0 + 1) + b_0/v \) (respectively \( \varphi(t) = -(a + 1) + b/t \)) and it is easy to check that Condition (3.13) in Lemma 3.4 is not satisfied. To find a density \( h \) corresponding to the choice of the density \( g \) in order that Lemma 3.4 applies reduces to exhibit a function \( \varphi \) such that
\[
0 \leq -\varphi'(z(1 - \lambda)) \leq \frac{b_0}{z^2 \lambda^2} \tag{4.3}
\]
for any \( z \geq 0 \) and any \( \lambda \in [0, 1] \). For fixed \( u \geq 0 \), Condition (4.3) implies that
\[
0 \leq -\varphi'(u) \leq \frac{b_0}{(z - u)^2}
\]
for any \( z \geq 0 \) and hence, when \( z \) goes to infinity, it follows that \( -\varphi'(u) = 0 \). Thus the function \( \varphi \) is constant, that is, \( \varphi(u) = \gamma \) and necessarily, according to (3.12), we have that \( h(t) = Ct^\gamma \) which is improper for any \( \gamma \). Now Conditions (3.5) and (3.7) impose that \( \gamma = \beta \geq -\alpha > -1 \).

Minimaxity of \( \hat{\beta}_h \) will follow from (3.11), that is, from the existence of a non-empty interval of the form
\[
\left[-(p - 2)\left(\frac{a_0 - 1}{a_0} - \frac{1}{2}\right), \frac{p}{2} - 1\right] \tag{4.4}
\]
for the value of \( \beta \). Using the fact that \( \beta > -1 \) the following cases arise.

When \( p = 3 \) or 4, it is easy to check that \( -(p - 2)((a_0 - 1)/a_0 - 1/2) > -1 \) since \( 1/a_0 > (p - 4)/(p - 2) \). Thus the range of values of \( \beta \) is exactly the interval in (4.4). This is still the case, when \( p \geq 5 \), as soon as \( a_0 < 2(p - 2)/(p - 4) \). However, when \( p \geq 5 \) and \( a_0 \geq 2(p - 2)/(p - 4) \), the range of value of \( \beta \) reduces to \((-1, p/2 - 1) \).

Hence, for the case of an inverse gamma mixing distribution, \( g \), there exists a mixing distribution, \( h \), of the form \( h(t) \propto t^\beta \) which results in a minimax (improper) Bayes estimator. Note that
this result implies that, in the Student case mentioned above, minimaxity of \( \delta_h \) is guaranteed for any \( n_0 \geq 3 \).

**Example 4.2. Gamma:** Let \( g(v) \propto v^{a_0-1} \exp(-b_0v) \) with \( a_0 > 0 \) and \( b_0 > 0 \). Clearly \( g \) has increasing monotone likelihood ratio property as a scale parameter family.

1) **Minimaxity using Theorem 3.1:** The choice of a function \( h \) also proportional to a gamma density, that is, \( h(t) \propto t^{a-1} \exp(-bt) \) with \( b > 0 \) seems natural but, for any \( \beta \in \mathbb{R} \), \( c = \lim_{t \to \infty} h(t)/t^\beta = 0 \) and Theorem 3.1 does not apply since Condition (3.5) needs \( c > 0 \).

Now it is worth noting that the choice of an inverse gamma type density for \( h \) allows to use Theorem 3.1. Indeed, for \( h(t) \propto 1/t^{a+1} \exp(-b/t) \) with \( b > 0 \) and \( a > -p/2 \), satisfies Condition (3.5) with \( \beta = -(a+1) \). It is also clear that, for \( a_0 > p/2 \) Condition (3.8) is satisfied. Finally a simple calculation shows that

\[
\frac{E[V^{-\frac{p+1}{2}}]}{E[V]E[V^{-\frac{p}{2}}]} = \frac{a_0 - p/2}{a_0}
\]

so that Condition (3.9) reduces to

\[
p - 2\left[1 - \frac{p}{a_0}\right] \geq a + 1.
\]

(4.5)

Then it follows from Theorem 3.1 that the Bayes estimator \( \delta_h \) in (2.12) is minimax provided that (4.5) is satisfied. Note that, properness of the estimator \( \delta_h \) needs that the right-hand side of (4.5) be greater than 1, which requires that \( a_0 > p(2-p)/(p-4) \), provided \( p \geq 5 \).

2) **Minimaxity using Theorem 3.2:** Note that the densities \( h \) and \( g \) correspond, respectively, to the functions \( \varphi(t) = -(a+1) + b/t \) and \( \varphi_0(v) = a_0 - 1 - b_0v \) in Lemma 3.4. Obviously Condition (3.13) is not satisfied since it reduces to

\[
-b_0 \leq \frac{-b}{(z(1 - \lambda))^2}
\]

for any \( z \) and any \( \lambda \). Hence, we cannot apply Theorem 3.2 for this choice of \( h \).

The choice of \( \varphi(t) = v - \mu t/\sigma \), with \( v \in \mathbb{R} \), \( \mu \geq 0 \) and \( \sigma > 0 \), leads to the density \( h(t) \propto (t/\sigma)^\gamma (1 + t/\sigma)^{-\mu} \) according to (3.12). Thus Condition (3.13) expresses that, for any \( z \geq 0 \) and any \( \lambda \in ]0, 1[ \),

\[
-b_0 \leq \frac{-\mu/\sigma}{(1+z(1-\lambda))/\sigma^2}
\]

and is satisfied as soon as \( b_0 \geq \mu/\sigma \).

In this case Theorem 3.2 does apply with \( \beta = v - \mu \) and \( v > -1 \). Condition (3.11) becomes

\[
-(p-2)\left[\frac{1}{2} - \frac{1}{a_0}\right] \leq v - \mu < \frac{p}{2} - 1.
\]

(4.6)

This interval of values of \( \beta = v - \mu \) is non-void for all \( a_0 > 1 \). Hence it is always possible to find pairs \( (\mu, v) \) with \( \mu > 0 \) and \( v > -1 \) which satisfy (4.6). It remains for any such choice to choose a scale parameter, \( \sigma \), for the mixing density \( h \) such that \( b_0 \geq \mu/\sigma \).
Propriety of $h$ requires in addition that $\beta = \nu - \mu < -1$. This condition in turn implies that $-(p - 2)[1/2 - 1/a_0] < -1$, since $a_0 > 1$, $p$ must be at least 5, and $a_0 > 2(p - 2)/(p - 4)$. Under this condition it is always possible to apply Theorem 3.2 to obtain (many) proper Bayes minimax estimators corresponding to priors of the form $h(t) = t^\mu (1 + t/\sigma)^{-\mu}$.

It should be noted that when $\sigma = 1$, this class of priors corresponds to the class of Maruyama. It is also interesting to note (and this is how we initially found the class) that this class arises as an inverse gamma mixture of gamma densities.

The class of priors $h$, to which Theorem 3.2 applies, is, however, much broader, as is the class of mixing densities $g$. This is the subject of our final example.

**Example 4.3. Mixtures with bounded $\psi$:** Suppose $g$ is any mixing density such that

$$\psi_0(v) = \frac{d}{dv} \left[ v \frac{d}{dv} g(v)/g(v) \right] = \frac{d}{dv} [\varphi_0(v)]$$

satisfies $-\infty < \psi_0(v) \leq -b_0 < 0$. The gamma class of Example 4.2 is such a density. The inverse gamma class of Example 4.1 is not.

Let $h_1(t)$ be any mixing density satisfying $\lim_{t \to \infty} h_1(t)/t^\beta = c$ and such that, $0 \geq \psi_1(t) = \frac{d}{dt} \varphi_1(t)] > -b_1$. Then, if $h_\sigma(t) = \frac{1}{\sigma} h_1(t/\sigma)$, an easy calculation gives $\psi_\sigma(t) = \frac{d}{dt} \varphi_\sigma(t) = \frac{1}{\sigma} \psi_1(t/\sigma)$. Hence $\psi_\sigma(t) = \frac{1}{\sigma} \psi_1(t/\sigma) > -b_1/\sigma \geq -b_0$ as soon as $\sigma \geq b_1/b_0$.

The mixing density $h_1(t) = t^\psi (1 + t)^{-\mu}$ of Example 4.2 is such a density. It is interesting that scaling the prior mixing density does not affect the value $\beta$, but it does affect the lower bound of the function $\psi_\sigma(t)$.

It is easy to construct other such examples of $h_1(t)$. One such example is

$$h_1(t) = t^\psi (1 + t)^{-\mu_1} (2 + t)^{-\mu_2} \quad \text{with } \psi > -1, \mu_1 + \mu_2 > 0$$

for which

$$0 > \psi_1(t) = -\mu_1 \left( \frac{1}{1 + t} \right)^{-2} - \mu_2 \left( \frac{1}{2 + t} \right)^{-2} \geq -\mu_1 - \frac{\mu_2}{4}$$

and for which $\beta = \psi - \mu_1 - \mu_2$.

In such cases Theorem 3.2 applies. It results in proper Bayes estimators whenever $p \geq 5$, $E[V]E[V^{-1}] > 2(p - 2)/(p - 4)$ so that the left-hand side of (3.11) is less than $-1$. In this case, choosing

$$-(p - 2) \left[ \frac{1}{E[V^{-1}]E[V]} - \frac{1}{2} \right] \leq \beta < -1,$$

and a scale parameter $\sigma$ such that $-b_1/\sigma > -b_0$, the resulting procedure is proper Bayes and minimax. Choosing the same scale parameter and $-1 \leq \beta < p/2 - 1$ gives a generalized (non-proper) Bayes minimax estimator. As noted above $\beta \leq 0$ corresponds to admissible estimator.

5. Conclusions

We have studied Bayes minimax estimation for the case of variance mixture of normal distributions in dimension 3 and higher. We have assumed throughout that the prior distribution is
also a variance mixture of normals and that both sampling and prior mixing distributions have monotone non-decreasing likelihood ratio when considered as a scale parameter family. We, and Maruyama [9], use the monotone likelihood ratio property to establish monotonicity of the function $r(w)$.

Our minimaxity results (and Maruyama’s) rely on the results of Strawderman [15] when $r(w)/w$ is non-increasing, and on a result of Berger [1] when $r(w)/w$ cannot be shown to be non-increasing. In either case $r(w)$ is required to be non-decreasing, non-negative and bounded by a constant. This constant is always larger for the Strawderman case than for the Berger case.

The class of mixing priors in Maruyama, in our setting (his parameterization is slightly different than ours), is given by $h(t) \propto t^b (1 + t)^{a-2-b}$ with $b > -1$. These priors are proper for $a < 1$ and improper for $a \geq 1$.

Our class is a generalization of Maruyama’s in that we assume $h(t) \leq K t^{-\alpha}$ for $0 < t < t_0$ and $\alpha < 1$, $h(0) < \infty$ and $\lim_{t \to \infty} h(t)/t^\beta = c > 0$. These priors are proper for $\beta < -1$. Each paper establishes that $\lim_{w \to \infty} r(w) = (p - 2 - 2\beta)E[V]$ (however, Maruyama’s $V$ is our $V^{-1}$).

Both papers use this bound, the monotonicity of $r$ and the Strawderman or Berger result to obtain minimaxity of the resulting proper or generalized Bayes estimator. With (our $\beta$) equal to (Maruyama’s $a - 2$), the conditions for minimaxity in the two papers agree.

A point of departure in the present paper is that we study in some detail, conditions under which $r(w)/w$ is non-increasing. In fact we give broad classes of examples where $r(w)/w$ is non-increasing and where the resulting proper Bayes estimator is minimax. This development depends on the fact that $h(t) = C \exp \left( \int_{u_0}^{t} \frac{\varphi(u)}{u} du \right)$ (respectively $g(v) = C_0 \exp \left( \int_{v_0}^{v} \frac{\varphi_0(u)}{u} du \right)$) with $\varphi(u)$ (respectively $\varphi_0(u)$) non-increasing, then $h$ (respectively $g$) has non-increasing monotone likelihood ratio as a scale family. Using these expressions we show that $r(w)/w$ is non-increasing whenever

$$\sup \varphi_0'(u) \leq \inf \varphi'(u). \quad (5.1)$$

In particular, if the function $g$ is such that $\varphi_0'(v) \leq -b_0 < 0$, we construct functions $h$ such that (5.1) holds. One way to do this is to consider a scaled version of Maruyama’s class of priors and to choose the scale parameter sufficiently large.

This method leads to proper Bayes minimax estimators (with $r(w)/w$ decreasing) in the case where $g(v)$ is a gamma distribution (with exponential tails) and it fails to do so in the case where $g(v)$ is an inverse gamma distribution (with polynomial tails).

As in Maruyama, in the inverse gamma case, we give generalized Bayes minimax estimators with decreasing $r(w)/w$. Both Maruyama and we give proper Bayes minimax estimators based on the Berger result.

**Acknowledgments**

The authors are grateful to two referees for their careful reading and for their comments which improved the manuscript.

**Appendix A.**

We first recall here the FKG inequality established by Fortuin et al. [6] which will be used in the proof of Lemma 3.1.
**Lemma A1** (FKG inequality). Let $\xi$ denote a probability density function with respect to a $\sigma$-finite measure $\nu$ on $\mathbb{R}^n$. For any two points $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$, define
\[
y \wedge z = (y_1 \wedge z_1, \ldots, y_n \wedge z_n),
y \vee z = (y_1 \vee z_1, \ldots, y_n \vee z_n)
\]
(where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$) and suppose that $\xi$ satisfies
\[
\xi(y) \xi(z) \leq \xi(y \wedge z) \xi(y \vee z).
\]
If the functions $f(y)$ and $g(y)$ are non-decreasing in each argument and if $f$, $g$ and $fg$ are integrable with respect to $\nu$, then
\[
\int_{\mathbb{R}^n} f(y) g(y) \nu(y) \geq \int_{\mathbb{R}^n} f(y) \xi(y) \nu(y) \int_{\mathbb{R}^n} g(y) \xi(y) \nu(y).
\]

**Proof of Lemma 3.1.** Through the change of the variables $u = \frac{1}{v}$ and $t = \frac{1}{v} - \frac{\lambda}{u}$ in integral (2.13), the function $r$ characterizing the Bayes estimator can be expressed, for any $w \geq 0$, as
\[
r(w) = w \frac{\phi_1(w)}{\phi_0(w)},
\]
where, for $i = 0, 1$,
\[
\phi_i(w) = \int_0^1 \int_0^\infty \lambda^{\frac{p}{2} - 2 + i} u^{\frac{p}{2} - 3} \exp \left( -\frac{1}{2} \lambda u w \right) h \left( \frac{1}{u} \frac{1 - \lambda}{\lambda} \right) g \left( \frac{1}{u} \right) du d\lambda.
\]
Then, to study the monotonicity of $r$, differentiating (A.1) with respect to $w$, we have
\[
r'(w) = \frac{\phi_1(w)}{\phi_0(w)} - \frac{1}{2} w \frac{A_1(w) \phi_0(w) - A_0(w) \phi_1(w)}{\phi_0^2(w)},
\]
where, for $i = 0, 1$,
\[
A_i(w) = \int_0^\infty \eta_i(u) u^{\frac{p}{2} - 2} g \left( \frac{1}{u} \right) du
\]
with
\[
\eta_i(u) = \int_0^1 \lambda^{\frac{p}{2} - 1 + i} \exp \left( -\frac{1}{2} \lambda u w \right) h \left( \frac{1}{u} \frac{1 - \lambda}{\lambda} \right) d\lambda.
\]
Now integrating by parts in (A.5) and applying $h(0) < \infty$ and the growth condition $h(t) = o(t^{p/2-1})$ for $t$ in a neighborhood of infinity to the bracketed term yield
\[
\eta_i(u) = \frac{2}{uw} \left\{ -h(0) \exp \left( -\frac{1}{2} \lambda u w \right)
+ \left( \frac{p}{2} - 1 + i \right) \int_0^1 \lambda^{\frac{p}{2} - 2 + i} \exp \left( -\frac{1}{2} \lambda u w \right) h \left( \frac{1}{u} \frac{1 - \lambda}{\lambda} \right) d\lambda
- \frac{1}{u} \int_0^1 \lambda^{\frac{p}{2} - 3 + i} \exp \left( -\frac{1}{2} \lambda u w \right) h' \left( \frac{1}{u} \frac{1 - \lambda}{\lambda} \right) d\lambda \right\}.
\]
Hence (A.4) and (A.6) give, for \( i = 0, 1 \),

\[
A_i(w) = \frac{2}{w} \left[ \left( \frac{p}{2} - 1 + i \right) \phi_i(w) - B_i(w) - C(w) \right],
\tag{A.7}
\]

where

\[
B_i(w) = \int_0^1 \int_0^\infty \lambda^{\frac{p}{2}-3+i} u^{\frac{p}{2}-4} \exp \left( -\frac{1}{2} \lambda u w \right) h' \left( \frac{1}{u} \right) g \left( \frac{1}{u} \right) du d\lambda
\tag{A.8}
\]

and

\[
C(w) = h(0) \int_0^\infty u^{\frac{p}{2}-3} \exp \left( -\frac{1}{2} u w \right) g \left( \frac{1}{u} \right) du.
\tag{A.9}
\]

Finally, combining (A.3) and (A.7) gives, after some algebra, the following expression:

\[
r'(w) = \frac{1}{\phi_0^2(w)} \left[ \phi_0(w) B_1(w) - \phi_1(w) B_0(w) + C(w) (\phi_0(w) - \phi_1(w)) \right].
\tag{A.10}
\]

As we want to prove that \( r'(w) \geq 0 \), it follows from (A.10) that it suffices that

\[
\frac{B_1(w)}{\phi_0(w)} \geq \frac{\phi_1(w) B_0(w)}{\phi_0(w)}
\tag{A.11}
\]

since, according to (A.2) and to (A.9), \( \phi_0(w) \geq \phi_1(w) \) and \( C(w) \geq 0 \), respectively. This can be proved through a new expression of \( \phi_i(w) \) and \( B_i(w) \). Using the change of variable \( u = \frac{1 - \lambda}{\lambda} t \) in (A.2) and (A.8) and setting

\[
G_w(t, \lambda) = t^{\frac{p}{2}-3} (1 - \lambda)^{\frac{p}{2}-2} \exp \left( -\frac{1}{2} (1 - \lambda) t w \right) h \left( \frac{1}{t} \right) g \left( \frac{1}{t} \right) \left( \frac{\lambda}{1 - \lambda} \right)
\tag{A.12}
\]

it can be shown that, for \( i = 0, 1 \),

\[
\phi_i(w) = \int_0^\infty \int_0^1 \lambda^i G_w(t, \lambda) d\lambda dt
\]

and

\[
B_i(w) = \int_0^\infty \int_0^1 \frac{\lambda^i}{1 - \lambda t} \frac{h' \left( \frac{1}{t} \right)}{h \left( \frac{1}{t} \right)} G_w(t, \lambda) d\lambda dt.
\]

Thus inequality (A.11) can be interpreted as follows: with respect to the density \( \xi_w : (\lambda, t) \mapsto G_w(t, \lambda)/\phi_0(w) \), its left-hand side appears as the expectation of the product of the functions \( \varphi : (\lambda, t) \mapsto \frac{1}{1 - \lambda t} \frac{h' \left( \frac{1}{t} \right)}{h \left( \frac{1}{t} \right)} \) and \( \psi : (\lambda, t) \mapsto \lambda \) while its right-hand side is the product of the respective expectations of \( \varphi \) and \( \psi \). Note that both functions \( \varphi(\lambda, t) \) and \( \psi(\lambda, t) \) are non-decreasing in both arguments \( \lambda \) and \( t \) (the monotonicity in \( t \) of \( \varphi(\lambda, t) \) coming from the monotone likelihood ratio assumption of \( h \) as a scale parameter family mentioned above).
Then inequality (A.11) will follow from the FKG inequality recalled in Lemma A1. Indeed, fix \( w \geq 0 \), \( \lambda_1 \leq \lambda_2 \) and \( t_1 \leq t_2 \). First it is clear that
\[
\exp \left( -\frac{1}{2} (1 - \lambda_1) t_2 w \right) \exp \left( -\frac{1}{2} (1 - \lambda_2) t_1 w \right) \\
\leq \exp \left( -\frac{1}{2} (1 - \lambda_1) t_1 w \right) \exp \left( -\frac{1}{2} (1 - \lambda_2) t_2 w \right).
\] (A.13)

Now, by monotone likelihood ratio property of \( g \) considered as a scale parameter family and since the functions \( \lambda \mapsto \lambda/(1 - \lambda) \) and \( t \mapsto 1/t \) are, respectively, increasing in \( \lambda \) and decreasing in \( t \), we have according to (3.1)
\[
g \left( \frac{1}{t_2} \frac{\lambda_1}{1 - \lambda_1} \right) g \left( \frac{1}{t_1} \frac{\lambda_2}{1 - \lambda_2} \right) \\
\leq g \left( \frac{1}{t_1} \frac{\lambda_1}{1 - \lambda_1} \right) g \left( \frac{1}{t_2} \frac{\lambda_2}{1 - \lambda_2} \right).
\] (A.14)

Therefore, according to (A.12), it follows from (A.13) and (A.14) that
\[
\xi_w(t_1, \lambda_2) \xi_w(t_2, \lambda_1) \leq \xi_w(t_1, \lambda_1) \xi_w(t_2, \lambda_2).
\]

Finally Lemma A1 gives inequality (A.11) which proves that the function \( r \) is non-decreasing. \( \square \)

**Proof of Lemma 3.2.** We give the proof of the case \( \beta < 0 \), the case \( \beta \geq 0 \) being similar and somewhat simpler. Dividing numerator and denominator of the right-hand side of (3.3) by \( w^\beta \) and using the change of variable \( t = v(w - s)/s \), we obtain
\[
r(w) = \frac{\int_0^\infty \int_0^\infty M_1(w, s, v) \, ds \, dv}{\int_0^\infty \int_0^\infty M_0(w, s, v) \, ds \, dv}
\] (A.15)

where, for \( i = 0, 1 \),
\[
M_i(w, s, v) = w^{-\beta} 1_{[0, w]}(s) s^{\frac{p}{2} - 2 + i} \exp \left( -\frac{s}{2v} \right) h \left( v \frac{w - s}{s} \right) v^{-p/2 + 1} g(v).
\] (A.16)

We now bound the integral of \( M_i \) to apply the Lebesgue dominated convergence theorem. Before, for simplicity, let
\[
\lambda_0(v) = \lambda_0 = \frac{v/t_0}{1 + v/t_0},
\]
\[
K_0(w, v) = \int_0^{\lambda_0 w} w^{-\beta} s^{\frac{p}{2} - 2 + i} \exp \left( -\frac{s}{2v} \right) h \left( v \frac{w - s}{s} \right) \, ds
\]
and
\[
K_1(w, v) = \int_{\lambda_0 w}^w w^{-\beta} s^{\frac{p}{2} - 2 + i} \exp \left( -\frac{s}{2v} \right) h \left( v \frac{w - s}{s} \right) \, ds
\]
so that
\[
\int_0^\infty \int_0^\infty M_i(w, s, v) \, ds \, dv = \int_0^\infty \left[ K_0(w, v) + K_1(w, v) \right] v^{-p/2 + 1} g(v) \, dv.
\]
Note now that, for $0 < s < w$, under Condition (3.7), we have

\[ h \left( \frac{w - s}{s} \right) \leq K \left( \frac{w - s}{s} \right)^{-\alpha} \]

if $0 < \lambda_0 < s/w < 1$. Therefore

\[
K_1(w, v) \leq K \int_{0}^{w} w^{-\beta} s^{p/2 - 2 + i} \exp \left( -\frac{s}{2v} \right) \left( \frac{w - s}{s} \right)^{-\alpha} ds
\]

\[
= K w^{-\beta + p/2 - 1 + i} v^{-\alpha} \int_{0}^{\lambda_0} \lambda^{p/2 - 2 + \alpha + i} (1 - \lambda)^{-\alpha} \exp \left( -\frac{\lambda w}{2v} \right) d\lambda
\]

(A.17)

through the change of variable $s = \lambda w$. Using $w^B e^{-\lambda w} \leq B^B e^B A^{-B}$, it follows that, for some constant $C$,

\[
K_1(w, v) \leq KC v^{-\alpha + p/2 - 2 + i} \int_{0}^{1} \lambda^{\beta + \alpha - 1} (1 - \lambda)^{-\alpha} d\lambda.
\]

since $\beta < 0$ and $\alpha - 1 < 0$. Finally, after some algebra, we have, for some constant $\mu$,

\[
\int_{0}^{\infty} K_1(w, v) v^{-p/2 + 1} g(v) dv \leq \mu \int_{0}^{\infty} \left( 1 + \frac{v}{t_0} \right)^{-\beta - \alpha} v^{i - 1} g(v) dv
\]

\[
\leq \mu 2^{1 - \beta - \alpha} \int_{0}^{\infty} \left[ 1 + \left( \frac{v}{t_0} \right)^{-1 - \beta - \alpha} \right] v^{i - 1} g(v) dv,
\]

(A.18)

where the right-hand side of (A.18) is finite since $\int_{0}^{\infty} v^{i - \beta - \alpha} g(v) dv < \infty$ and $\int_{0}^{\infty} v^{i - 1} g(v) dv < \infty$ by assumption.

We now consider the case $0 < s/w < \lambda_0 < 1$. We have (see comment after Lemma 3.2)

\[ h \left( \frac{w - s}{s} \right) \leq K \left( \frac{w - s}{s} \right)^{\beta} \]

which implies that

\[
K_0(w, v) \leq K \int_{0}^{\lambda_0 w} w^{-\beta} s^{p/2 - 2 + i} \exp \left( -\frac{s}{2v} \right) \left( \frac{w - s}{s} \right)^{\beta} ds
\]

\[
\leq K v^{\beta} \int_{0}^{\lambda_0 w} s^{p/2 - 2 - \beta + i} \exp \left( -\frac{s}{2v} \right) (1 - s/w)^{\beta} ds
\]

\[
\leq K v^{\beta} \left( \frac{1}{1 + v/t_0} \right)^{\beta} \int_{0}^{\lambda_0 w} s^{p/2 - 2 - \beta + i} \exp \left( -\frac{s}{2v} \right) ds
\]

since $\beta < 0$. Hence

\[
K_0(w, v) \leq K v^{\beta} \left( \frac{1}{1 + v/t_0} \right)^{\beta} \Gamma(p/2 - 1 - \beta + i)(2v)^{p/2 - 1 - \beta + i}
\]
and then, for some constant $K'$, we have
\[
\int_{0}^{\infty} K_0(w, v) v^{-p/2 + 1} g(v) \, dv \leq K' \int_{0}^{\infty} (1 + v/t_0)^{-\beta} v^i g(v) \, dv \\
\leq K' 2^{-\beta} \int_{0}^{\infty} \left[ 1 + (v/t_0)^{-\beta} \right] v^i g(v) \, dv, \quad (A.19)
\]
where the right-hand side of (A.19) is finite since \( \int_{0}^{\infty} v^{i-i} g(v) \, dv < \infty \).

The finiteness of the integrals in (A.18) and (A.19) allows to apply the Lebesgue dominated convergence theorem. Then, according to (A.16) and (3.5)
\[
\lim_{w \to \infty} \int_{0}^{\infty} \int_{0}^{\infty} M_i(w, s, v) \, ds \, dv = \int_{0}^{\infty} \int_{0}^{\infty} \lim_{w \to \infty} M_i(w, s, v) \, ds \, dv \\
= c \int_{0}^{\infty} \int_{0}^{\infty} s^{\beta - 2 + i} \exp \left( -\frac{s}{2v} \right) v^{\beta - p/2 + 1} g(v) \, ds \, dv \\
= c 2^{p/2 - \beta - 1 + i} \Gamma(p/2 - \beta - 1 + i) E[V^i].
\]
Finally (A.15) gives
\[
\lim_{w \to \infty} r(w) = 2 \left( \frac{p}{2} - \beta - 1 \right) E[V]
\]
which is the desired result. □

References


