LOWER BOUNDS ON BAYES RISKS FOR ESTIMATING A NORMAL VARIANCE: WITH APPLICATIONS

by

Brani Vidakovic and Anirban DasGupta
Duke University Purdue University

Technical Report #92-38

Department of Statistics
Purdue University

August 1992

LOWER BOUNDS ON BAYES RISKS FOR ESTIMATING A NORMAL VARIANCE: WITH APPLICATIONS

by

Brani Vidakovic and Anirban DasGupta
Duke University Purdue University

Abstract

Brown and Gajek (1990) gave useful lower bounds on Bayes risks, which improve on earlier bounds by various authors. Many of these use the information inequality. For estimating a normal variance using the invariant quadratic loss and any arbitrary prior on the reciprocal of the variance that is a mixture of Gamma distributions, we obtain lower bounds on Bayes risks that are different from Borovkov-Sakhanienko bounds. The main tool is convexity of appropriate functionals as opposed to the information inequality. The bounds are then applied to many specific examples, including the multi-Bayesian setup (Zidek and his coauthors). Subsequent use of moment theory and geometry gives a number of new results on efficiency of estimates which are linear in the sufficient statistic. These results complement earlier results of Donoho, Liu and MacGibbon (1990) and Vidakovic and DasGupta (1992) for the location case.

General Goal

1. Introduction

1.1 General Goal. The purpose of this article is to address a number of issues of current interest in the context of estimating the variance of a normal distribution. The literature on estimation of variance and scale parameters is rich. Maatta and Casella (1990) is a good source for general exposition and other references. Much of the previous work has emphasized risk behavior and admissibility. Our intent is different. The principal goal of this article is to derive an expression for the Bayes rule and the corresponding Bayes risk in terms of the marginal density and its derivatives. These are then used, with the aid of

This research was supported in part by NSF Grant DMS-8923071 at Purdue University.
AMS 1980 *subject classifications.* Primary 62C20; Secondary 62F10.
Key words and phrases. Bayes risk, convex hull, efficiency, gamma, linear, lower bound, mixture, variance.

further mathematical tools, to establish easily calculable lower bounds on the Bayes risk, which in turn lead to an array of subsequent results on efficiency of linear rules (linear in a sufficient statistic). Each of these topics received its due attention for the case of estimating a normal mean: formulas for Bayes rules are available, among others, in Brown (1971), Brown and Hwang (1982), Haff (1991); expressions for the Bayes risk can be found in Brown (1971); lower bounds on Bayes risks were derived in Borovkov and Sakhanienko (1980) and Brown and Gajek (1990); efficiency of linear rules is discussed in Donoho, Liu and MacGibbon (1990) and Vidakovic and DasGupta (1992). The primary intent of this article is to now establish some corresponding results for estimating a variance. At this point, we would like to emphasize that in contrast to the existing literature on estimation of the mean, all results in this article deal with estimating a single unknown variance.

1.2. Model and Outline. Let $X_1, \ldots, X_n$ be n iid observations from the $N(0, \sigma^2)$ distribution. We consider estimation of σ^2 under the natural invariant quadratic loss $L(\sigma^2, a) = \left(\frac{a}{\sigma^2} - 1\right)^2$. For the rest of the article, θ will denote the precision parameter $\frac{1}{\sigma^2}$, and T will stand for the sufficient statistic $\sum_{i=1}^{n} X_i^2$.

In section 2, we derive formulas for the Bayes rule and the Bayes risk with respect to a general prior on θ in terms of the marginal density of T and its derivatives. The formula for Bayes risks involves an appropriate integral operator. We also prove in section 2 that this operator is convex. Section 3 specializes to prior densities on θ which are mixtures of Gamma densities. The convexity result from the previous section is utilized to establish a lower bound on the Bayes risk for an arbitrary prior density (on θ) that is a mixture of Gammas. A technical surprise in the obtained lower bound is that it only depends on the marginal mixing distribution of the Gamma shape parameter, with the scale parameter getting wiped out in a sequence of cancellations. The Borovkov-Sakhanienko method has proved very useful in the past for obtaining Bayes risk lower bounds. However, in this case, it leads to a disagreeable clutter of many integrals and our methods seem more rewarding. In section 4, we discuss at some length the richness of the family of Gamma mixtures. The lower bound derived in section 3 is explicitly evaluated for a number of examples. Section 5 addresses the issue of estimating σ^2 by estimates linear in T. We consider several natural families of priors and by a combination of some moment theory and convex programming methods, we derive upper bounds on the loss of efficiency due to the use of linear Γ-minimax rules as opposed to the usually incalculable exact Γ-minimax rule. In particular, in two important cases we obtain explicit sample sizes n such that the loss of efficiency is smaller than a prespecified (small) number. Section 6 contains some brief concluding remarks.

2. Bayes Rule and Bayes Risk Identity

In this section, we first derive formulas for the Bayes rule for σ^2 and the corresponding Bayes risk. The prior is general; no assumptions on the prior are made other than the existence of all relevant integrals. The following notation will be used:

$$\begin{aligned} f(t|\theta) &= \frac{e^{\frac{-\theta t}{2}} t^{\frac{n}{2}-1} \theta^{\frac{n}{2}}}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}},\ t, \theta > 0 \\ m(t) &= \frac{t^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} \int e^{-\frac{\theta t}{2}} \theta^{\frac{n}{2}} \pi(\theta) d\theta \\ m_0(t) &= \int e^{-\frac{\theta t}{2}} \theta^{\frac{n}{2}} \pi(\theta) d\theta \\ \psi(h) &= \int t^{\frac{n}{2}-1} \frac{(h'(t))^2}{h''(t)} dt. \end{aligned} \tag{2.1}$$

In the above, it is understood that all integrals are on $(0, \infty)$. Notice $f(t|\theta)$ is the density of the sufficient statistic T and $m(t)$ is the marginal density of T with respect to a general prior density π.

Theorem 2.1. Let $X_1, \ldots, X_n$ be iid $N(0, \sigma^2)$ and let $\theta = \frac{1}{\sigma^2}$ have the prior density $\pi(\theta)$. Then the Bayes rule for σ^2 under the loss $\left(\frac{a}{\sigma^2} - 1\right)^2$ equals

$$\delta_\pi(t) = -\frac{1}{2} \frac{m_0'(t)}{m_0''(t)} \tag{2.2}$$

and the corresponding Bayes risk equals

$$r(\pi) = 1 - \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} \psi(m_0) \tag{2.3}$$

Proof: The Bayes rule for σ^2 equals

$$\delta_\pi(t) = \frac{E(\theta|T=t)}{E(\theta^2|T=t)}.$$

It is transparent from the definition of $m_0(t)$ that it is infinitely differentiable and in fact,

$$m_0^{(k)}(t) = (-\frac{1}{2})^k \int e^{-\frac{\theta t}{2}} \theta^{\frac{n}{2}+k} \pi(\theta) d\theta \tag{2.4}$$

However, on direct computation,

$$E(\theta^k/t) = \frac{\int e^{-\frac{\theta t}{2}} \theta^{\frac{n}{2}+k} \pi(\theta) d\theta}{\int e^{-\frac{\theta t}{2}} \theta^{\frac{n}{2}} \pi(\theta) d\theta} \tag{2.5}$$

(2.5), in conjunction with (2.4), yields (2.2). To derive (2.3), notice that

$$\begin{aligned} r(\pi) &= E_\pi E_{T|\theta}[\theta \delta_\pi(T) - 1]^2 \\ &= E_m E_{\theta|t}[\theta \delta_\pi(t) - 1]^2 \\ &= E_m(1 - \frac{(m_0'(t))^2}{m_0(t) m_0''(t)}) \\ &= 1 - \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} \int t^{\frac{n}{2}-1} \frac{(m_0'(t))^2}{m_0''(t)} dt \\ &= 1 - \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} \psi(m_0); \end{aligned}$$

here, the first two equalities are trivial, the third uses (2.2) and some algebra, the next line uses the definition of $m(t)$ and the final inequality uses the definition of the operator ψ. This proves the Theorem.

Remark 1. Formulae (2.2) and (2.3) resemble the corresponding Brown-Stein identities for the mean, although they are different.

We will now prove that the operator ψ is convex on any domain of convex functions. This convexity result will be useful in the next section.

Theorem 2.2. Let $\mathcal{F}$ be any class of convex functions defined on $(0, \infty)$ such that $\psi(h)$ is defined for every h in $\mathcal{F}$. Then ψ is a convex operator on $\mathcal{F}$.

Proof: It is necessary to show that

$$\psi(\lambda g + (1-\lambda)h) \leq \lambda \psi(g) + (1-\lambda)\psi(h) \tag{2.6}$$

for $0 < \lambda < 1$ and g, h in $\mathcal{F}$.

Notice that since ψ is an integral operator, (2.6) will follow if one has

$$t^{\frac{n}{2}-1}\frac{(\lambda f'(t)+(1-\lambda)g'(t))^2}{\lambda f''(t)+(1-\lambda)g''(t)} \leq \lambda t^{\frac{n}{2}-1}\frac{(f'(t))^2}{f''(t)}+(1-\lambda)t^{\frac{n}{2}-1}\frac{(g'(t))^2}{g''(t)}, \tag{2.7}$$

pointwise in t.

Labelling $f'(t)=a$, $g'(t)=b$, $f''(t)=c$ and $g''(t)=d$, we therefore need to have

$$\frac{(\lambda a+(1-\lambda)b)^2}{\lambda c+(1-\lambda)d} \leq \lambda\frac{a^2}{c}+(1-\lambda)\frac{b^2}{d}. \tag{2.8}$$

However, the difference of the RHS and the LHS of (2.8) can be directly verified to be nonnegative if c, d are both nonnegative. This simply corresponds to the convexity of the members of $\mathcal{F}$, thereby establishing the Theorem.

3. Lower Bound on Bayes Risks

Formula (2.3) for Bayes risk is general and should have broad applicability. However, for the purpose of this section, we specialize to mixture Gamma priors with densities of the form

$$\pi(\theta) = \int \frac{e^{-\frac{\theta}{\tau}}\theta^{\alpha-1}}{\tau^{\alpha}\Gamma(\alpha)}dG(\alpha,\tau). \tag{3.1}$$

We will later provide evidence that the mixture Gamma class is in fact quite rich. For now, we will only mention that restriction to priors of the form (3.1) enables one to derive a very neat lower bound on the Bayes risk. This is the content of the following result.

Theorem 3.1. For any prior density of the form (3.1),

$$r(\pi) \geq E_G\frac{1}{\frac{n}{2}+1+\alpha}.$$

Remark 2. Notice the rather curious absence of the scale parameter τ in this lower bound. The reward is the obvious calculational simplicity because one only has to work with the marginal distribution of α. We will cite examples later.

Proof: The crux of the argument is that the mixture form of the prior $\pi(\theta)$ translates into the mixture representation

$$m_0(t) = \int m_0(t|\alpha,\tau)dG(\alpha,\tau) \tag{3.2}$$

where

$$\begin{aligned} m_0(t|\alpha,\tau) &= \int e^{-\frac{\theta t}{2}}\theta^{\frac{n}{2}}\frac{e^{-\frac{\theta}{\tau}}\theta^{\alpha-1}}{\Gamma(\alpha)\tau^{\alpha}}d\theta \\ &= \frac{\Gamma(\alpha+\frac{n}{2})}{\Gamma(\alpha)\tau^{\alpha}(\frac{1}{\tau}+\frac{t}{2})^{\alpha+\frac{n}{2}}}. \end{aligned} \tag{3.3}$$

It is then a consequence of Theorem (2.2) that

$$\psi(m_0) \leq \int \psi(m_0(\cdot|\alpha,\tau))dG(\alpha,\tau), \tag{3.4}$$

where by definition,

$$\psi(m_0(\cdot|\alpha,\tau)) = \int t^{\frac{n}{2}-1}\frac{(m_0'(t|\alpha,\tau))^2}{m_0''(t|\alpha,\tau)}dt. \tag{3.5}$$

However, on using (3.3),

$$m_0'(t|\alpha,\tau) = -\frac{1}{2}\cdot\frac{\Gamma(\alpha+\frac{n}{2})}{\Gamma(\alpha)\tau^{\alpha}(\frac{1}{\tau}+\frac{t}{2})^{\alpha+\frac{n}{2}+1}}$$

and

$$m_0''(t|\alpha,\tau) = \frac{1}{4}\cdot\frac{\Gamma(\alpha+\frac{n}{2})(\alpha+\frac{n}{2}+1)}{\Gamma(\alpha)\tau^{\alpha}(\frac{1}{\tau}+\frac{t}{2})^{(\alpha+\frac{n}{2}+2)}}. \tag{3.6}$$

Substitution of (3.6) into (3.5) and elementary integration now gives

$$\psi(m_0(\cdot|\alpha,\tau)) = 2^{\frac{n}{2}}\Gamma(\frac{n}{2})(1-\frac{1}{\frac{n}{2}+1+\alpha}) \tag{3.7}$$

$$\Rightarrow \psi(m_0) \leq 2^{\frac{n}{2}}\Gamma(\frac{n}{2})E_G(1-\frac{1}{\frac{n}{2}+1+\alpha}), \tag{3.8}$$

by virtue of (3.4).

The Bayes risk identity (2.3) now establishes the stated lower bound on $r(\pi)$.

Remark 3. In contrast to the developments in Borovkov and Sakhanienko (1980) and Brown and Gajek (1990), our lower bound makes no use of the Information inequality.

Convexity of the operator ψ is the main ingredient. However, application of the Borovkov-Sakhanienko inequality does indeed provide another lower bound on Bayes risk for the problem we have. The following result is needed for reference.

Theorem 3.2 (Borovkov-Sakhanienko). Let $X \sim f(x|\theta)$, $\theta \sim \pi(\theta)$, and $L(\theta, a) = w(\theta)(\theta - a)^2$. Let $I(\theta)$ denote Fisher information for the family $\{f(x|\theta)\}$. Then under some additional conditions,

$$r(\pi) \geq \frac{C^2}{C + D}, \tag{3.9}$$

where

$$C = \int \frac{w(\theta)\pi(\theta)}{I(\theta)} d\theta \tag{3.10}$$

and

$$D = \int \frac{\left(\left(\frac{w(\theta)\pi(\theta)}{I(\theta)}\right)'\right)^2}{w(\theta)\pi(\theta)} d\theta. \tag{3.11}$$

Corollary. Let us apply the Borovkov-Sakhanienko bound to the exact problem we have and see what emerges. Recall the prior is a mixture Gamma as in (3.1). The following are straightforward:

$$\begin{aligned} I(\sigma^2) &= \frac{n}{2(\sigma^2)^2} \\ w(\sigma^2) &= \frac{1}{(\sigma^2)^2} \\ C &= \frac{2}{n} \\ D &\leq \frac{4}{n^2}(1 + E_G\alpha). \end{aligned} \tag{3.12}$$

The inequality involving the quantity D involves use of the Cauchy-Schwartz inequality; the exact details can be obtained from the authors.

(3.9) and (3.12) now imply

$$\begin{aligned} r(\pi) &\geq \frac{\frac{4}{n^2}}{\frac{3}{n} + \frac{4}{n^2}E_G(1+\alpha)} \\ &= \frac{1}{E_G(\frac{n}{2} + 1 + \alpha)}; \end{aligned} \tag{3.13}$$

clearly, this is weaker than the bound we present in Theorem 3.1. However, due to the use of a bound on D, (3.13) is not exactly the Borovkov-Sakhanienko found for the problem. The exact bound that uses the exact expression for D is an abhorrent medley of various integrals and does not seem to be useful without resorting to substantive numerical work.

4. The Mixture Gamma Class: An Appraisal

Since the Bayes risk lower bound in Theorem 3.1 requires the prior to be a mixture Gamma, some discussion of its inclusiveness seems necessary. We will restrict this discussion to a few examples. All Gamma distributions are evidently in this class. We will exhibit two more families included here and another natural one that is not.

Example 1. For a fixed shape parameter α, consider a Gamma scale mixture

$$\pi(\theta) = \frac{\theta^{\alpha-1}}{\Gamma(\alpha)} \int \frac{e^{-\frac{\theta}{\tau}}}{\tau^{\alpha}} dH(\tau) \tag{4.1}$$

This itself is a rich class. Even the case $\alpha = 1$ includes completely monotone densities (see Feller (1971)). A special subclass of Gamma scale mixtures is obtained by taking H to be an inverse Gamma with density

$$h(\tau|\beta,\gamma) = \frac{e^{-\frac{1}{\gamma\tau}}}{\tau^{\beta+1}\gamma^{\beta}\Gamma(\beta)}, \quad \beta,\gamma,\tau > 0. \tag{4.2}$$

This mixture results in the family of type 2 Beta densities

$$\pi(\theta|\beta,\gamma) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\theta^{\alpha-1}}{\gamma^{\beta}(\theta+\frac{1}{\gamma})^{\alpha+\beta}}. \tag{4.3}$$

Another important example is the Weibull distribution with cdf

$$F(\theta) = 1 - e^{-\frac{\theta^p}{\lambda}}, \ \theta > 0, \ 0 < p < 1$$

(see Jewel (1982)).

Example 2. Another distribution contained in the mixture Gamma class is the noncentral chi-square distribution with arbitrary noncentrality parameter λ and degrees of freedom k. Indeed, this is a Gamma mixture with τ fixed as 2 and $\alpha = \frac{k}{2} + i$ where $i \sim P_{0i}(\lambda)$.

We will come back to this example later. Notice that therefore the mixture Gamma class contains densities that are not necessarily unimodal. See Min'ko and Petunin (1988).

Example 3. Here is an example of a natural prior in the problem that cannot be a Gamma mixture. A Gamma distribution on σ^2 corresponds to an inverse Gamma distribution on $\theta = \frac{1}{\sigma^2}$ with density

$$\pi(\theta) = \frac{e^{-\frac{1}{\theta b}}}{\theta^{a+1} b^a \Gamma(a)}, \; a, b, \theta > 0. \tag{4.4}$$

The density in (4.4) has the property that $\lim_{\theta \to 0} \pi^{(k)}(\theta) = 0$ for every $k \geq 0$. Such a function cannot be a mixture of Gammas.

Remark 4. These examples seem to suggest that while the mixture Gamma class is quite rich, at the same time it fails to include some natural priors for the problem as well.

5. Applications

5.1. Linear Bayes Estimation and Sample Size Determination. A principal motivation for derivation of lower bounds on Bayes risks is that they enable one to verify how close to optimal a simpler and easily calculable rule is in comparison to the exact (and frequently messy) optimal rules. For instance, it is a completely trivial calculation to verify that for any prior π in our problem, the best linear Bayes estimate for σ^2 is $\frac{T}{n+2}$ with Bayes risk $\frac{2}{n+2}$. Theorem 3.1 therefore immediately gives that the efficiency of the intuitive and the simple estimate $\frac{T}{n+2}$ is at least as large as $\frac{n+2}{2} \cdot E_G(\frac{1}{\frac{n}{2}+1+\alpha})$. One anticipates that for any fixed G, this will be close to 1 for large n. That is indeed the case; a formal proof is transparent. Let us see a few concrete examples.

Example 4. Consider the Weibull distribution given in Example 1. This corresponds to a scale mixture of Gammas with the shape parameter α fixed at 1. Hence, for any Weibull distribution with $\lambda > 0$ and $0 < p < 1$, the efficiency of the rule $\frac{T}{n+2}$ is at least as large as $\frac{n+2}{n+4}$. Therefore, even for a sample of size 16, the linear Bayes rule has an efficiency of at least 90%.

Example 5. Suppose $\theta = \frac{1}{\sigma^2}$ has the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter λ. Then, Theorem 3.1 says that for this prior,

$$r(\pi) \geq E\frac{1}{\frac{n+k+2}{2} + i}, \tag{4.5}$$

where $i \sim P_{0i}(\lambda)$.

From (4.5), the efficiency of $\frac{T}{n+2}$ exceeds

$$\begin{aligned} e &\stackrel{\text{def}}{=} e^{-\lambda} \cdot \frac{n+2}{2} \cdot \sum_{i=0}^{\infty} \frac{\lambda^i}{(\frac{n+k+2}{2} + i)i!} \\ &= e^{-\lambda} \cdot \frac{n+2}{n+k+2} \cdot \Phi(\frac{n+k+2}{2}, \frac{n+k+4}{2}, \lambda) \end{aligned} \tag{4.6}$$

where $\Phi(\alpha, \gamma, \lambda)$ denotes the confluent hypergeometric function with parameters α and γ. Using the relation

$$\Phi(\alpha, \gamma, \lambda) = e^{\lambda}\Phi(\gamma - \alpha, \gamma, -\lambda)$$

whenever γ is not a nonpositive integer (see, e.g., Lebedev (1972), correcting misprint), we then have from (4.6) that

$$e = \frac{n+2}{n+k+2} \cdot \Phi(1, \frac{n+k+4}{2}, -\lambda). \tag{4.7}$$

The following table gives values of sample sizes sufficient for the efficiency of the optimal linear rule to exceed .8 and .9; the actual value of e is given in parentheses.

Table 1.

$k=1$		n	0.8	n	0.9
	λ =0.25	4	(0.811608)	11	(0.90034)
	λ =1	9	(0.800431)	24	(0.900591)
	λ =3	25	(0.802126)	60	(0.900767)
	λ =5	41	(0.801666)	96	(0.900576)
	λ =10	81	(0.801004)	186	(0.900340)
	λ =20	161	(0.800549)	366	(0.900184)
$k=2$					
	λ =0.25	8	(0.804476)	21	(0.903252)
	λ =1	14	(0.807456)	34	(0.902157)
	λ =3	29	(0.800967)	69	(0.900429)
	λ =5	45	(0.801110)	105	(0.900414)
	λ =10	85	(0.800830)	195	(0.900289)
	λ =20	165	(0.800500)	375	(0.900170)
$k=3$					
	λ =0.25	12	(0.802363)	30	(0.902087)
	λ =1	18	(0.804819)	43	(0.901391)
	λ =3	33	(0.800252)	78	(0.900214)
	λ =5	49	(0.800701)	114	(0.900293)
	λ =10	89	(0.800682)	204	(0.900245)
	λ =20	169	(0.800456)	384	(0.900157)
$k=5$					
	λ =0.25	20	(0.800989)	48	(0.901183)
	λ =1	26	(0.802493)	61	(0.900716)
	λ =3	42	(0.803096)	97	(0.900879)
	λ =5	57	(0.800162)	132	(0.900130)
	λ =10	97	(0.800447)	222	(0.900175)
	λ =20	177	(0.800376)	402	(0.900133)

5.2. Linear Γ-Minimax Estimation. The development in section 5.1 assumes a single prior. The standard Γ-minimax analysis assumes instead a family Γ of priors. This is quite common as is evidenced by a large literature on the subject. Berger (1985) contains a lucid discussion. The exact Γ-minimax rule can be and frequently is unpleasantly messy. So as in section 5.1, one can argue for the use of the linear Γ-minimax rule, the ultimate defense being that the efficiency of the linear rule measured as the ratio of the exact Γ-minimax and the linear Γ-minimax risk is very large. The negligible loss of efficiency coupled with the attractive user friendliness then makes a strong argument in favor of the linear rule. This approach was taken in Donoho, Liu and MacGibbon (1990) and later in Vidakovic and DasGupta (1992). Again, let us apply the bound of Theorem 3.1 to several natural examples. But first, a general result on the linear Γ-minimax risk is needed for the ensuing efficiency calculations.

Theorem 5.1. For a specified class of priors Γ, let

$$\mathcal{C} = \{(c_1, c_2): \quad \pi \in \Gamma\},$$

where $c_i = E_\pi(\theta^i)$.

Assume $\mathcal{C}$ is bounded. Then,

$$\inf_{\delta \in \mathcal{D}_L} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{(c_1,c_2)\in\mathcal{C}} \frac{c_2 - c_1^2}{c_1^2 + (\frac{n}{2}+1)(c_2 - c_1^2)},$$

where $\mathcal{D}_L$ denotes the class of all linear estimates $\delta(T) = aT + b$ and $r(\pi, \delta)$ denotes Bayes risk of δ with respect to π.

Proof: By direct calculation, for any given prior π, the Bayes risk of $\delta(T) = aT + b$ equals

$$r(\pi, \delta) = \underset{\sim}{\mu}'\Sigma\underset{\sim}{\mu} - 2\underset{\sim}{\mu}'\underset{\sim}{\beta} + 1, \tag{5.1}$$

where

$$\underset{\sim}{\mu} = (a\ b)', \Sigma = \begin{pmatrix} n(n+2) & nc_1 \\ nc_1 & c_2 \end{pmatrix}, \ \underset{\sim}{\beta} = (n\ c_1)'.$$

It is now clear that the particular problem at hand can be viewed as a S-game and because of the assumption made on the set $\mathcal{C}$,

$$\begin{aligned} &\inf_{\delta\in\mathcal{D}_L} \sup_{\pi\in\Gamma} r(\pi,\delta) \\ &= \sup_{\pi\in\Gamma} \inf_{\delta\in\mathcal{D}_L} r(\pi,\delta) \\ &= \sup_{c_1,c_2} \inf_{\underset{\sim}{\mu}} (\underset{\sim}{\mu}'\Sigma\underset{\sim}{\mu} - 2\underset{\sim}{\mu}'\underset{\sim}{\beta} + 1) \\ &= \sup_{c_1,c_2} (1 - \underset{\sim}{\beta}'\Sigma^{-1}\underset{\sim}{\beta}) \\ &= \sup_{c_1,c_2} \frac{c_2 - c_1^2}{c_1^2 + (\frac{n}{2}+1)(c_2 - c_1^2)}, \end{aligned}$$

establishing the Theorem.

An important application of this general Theorem is to the case when we again consider mixture Gamma priors. Let us make it precise.

Theorem 5.2. Let S be a specified set in $\mathsf{R}^+ \otimes \mathsf{R}^+$. Let $T = \{(\alpha\tau, \alpha\tau^2 + \alpha^2\tau^2) : (\alpha, \tau) \in S\}$. Consider the family of priors Γ defined as

$$\Gamma = \{\pi : \ \pi \text{ is of the form (3.1) with } G \text{ an arbitrary probability measure on } S\}.$$

Then

$$\inf_{\delta G \mathcal{D}_L} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{(u,v) \in \mathcal{H}(T)} \frac{v - u^2}{u^2 + (\frac{n}{2} + 1)(v - u^2)}, \tag{5.2}$$

where $\mathcal{H}(T)$ denotes the convex hull of T.

Remark 5. A verbal description of the statement of the above Theorem is perhaps useful. Suppose we elicit a Gamma prior for $\theta = \frac{1}{\sigma^2}$ with parameters α and τ. Due to the subjective nature of our elicitation, we can only specify (α, τ) up to a set S. We then take Γ as all possible mixtures of these Gamma priors. The Theorem above then gives a recipe for evaluating the linear Γ-minimax risk by forming the convex hull of the set of first two moments of θ if θ had exactly one of these Gamma distributions.

Proof: Define

$$\begin{aligned} u &= u(\alpha, \tau) = \alpha\tau \\ v &= v(\alpha, \tau) = \alpha\tau^2 + \alpha^2\tau^2. \end{aligned} \tag{5.3}$$

Since the priors are mixture Gamma,

$$\begin{aligned} c_1 &= E_\pi(\theta) = E_G(u) \\ c_2 &= E_\pi(\theta^2) = E_G(v) \end{aligned} \tag{5.4}$$

Since G is an arbitrary probability measure on S, the set $\mathcal{C}$ of all pairs (c_1, c_2) as π varies in Γ is simply the convex hull of the set of pairs (u, v) as (α, τ) varies in S. This establishes the assertion made in the Theorem.

Let us see some explicit illustrations of Theorem 5.2.

Example 6. Linear Estimation in Multi-Bayesian Setup. While working with a family of priors Γ, the questions of multi-Bayesian decision naturally arise. Let us suppose that N ($N \geq 2$) Bayesians having the loss function $(\frac{a}{\sigma^2} - 1)^2$ are each eliciting a Gamma prior on θ say with parameters (α_i, τ_i), $i = 1, \ldots, N$. We define a family of priors Γ as

$$\Gamma = \{\pi | \pi(\theta) = \sum_{i=1}^{N} \lambda_i \Gamma(\theta | \alpha_i, \tau_i),\ \lambda_i \geq 0,\ \sum_{i=1}^{N} \lambda_i = 1\},$$

i.e. as the class of all linear combinations of the elicited Gammas. This is a standard *Linear Opinion Pool* in combining probabilistic evidence. See Zidek (1988). In terms of the preceding discussion the set S is the discrete set of N pairs (α_i, τ_i).

The transformation (5.3) transforms the points (α_i, τ_i), $i = 1, N$ to a convex polygon in the (u, v) plane. This is the set $\mathcal{H}(T)$. To determine the least favorable prior, we equivalently solve a convex programming problem

$$\max \frac{v - u^2}{u^2 + (\frac{n}{2} + 1)(v - u^2)} \text{ over } (u, v) \text{ in } \mathcal{H}(T).$$

We elaborate the case $N = 2$ in more detail. Let us suppose that the pairs (α_1, τ_1) and (α_2, τ_2) were elicited. Then $\mathcal{H}(T)$ is a line segment $C_1 C_2$, where $C_i = (\alpha_i \tau_i, \alpha_i \tau_i^2 + \alpha_i^2 \tau_i^2) = (u_i, v_i)$.

It is easy to see that the level curves of (5.2) are parabolae

$$c_2 = \frac{1 - \mu n/2}{1 - \mu n/2 - \mu} c_1^2, \text{ with } 0 \leq \mu \leq \frac{2}{n+2}. \tag{5.5}$$

A parabola from the class (5.5) touches the segment $C_1 C_2$ at one of the endpoints, or at a point C_0 from the interior of the segment $C_1 C_2$ (Figure 1).

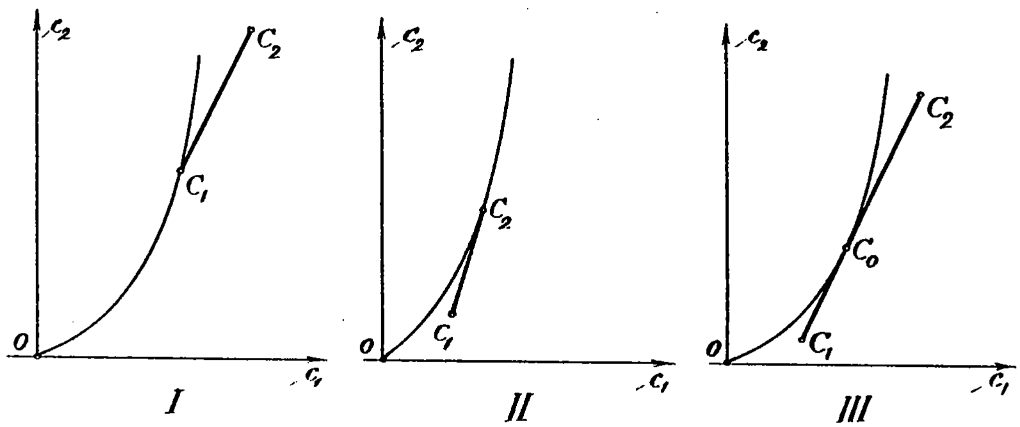


Figure 1: Three cases in two Bayesians setup

The first two cases when the touching point is an endpoint are trivial. In these cases the corresponding least favorable prior is a single Gamma. The nontrivial case III has the following solution:

The touching point C_0 has the coordinates

$$u_0 = \frac{2\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}}{v_2 - v_1}, \qquad v_0 = \frac{\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}}{u_2 - u_1} \tag{5.6}$$

assuming $u_1 < u_2$ and $v_1 < v_2$.

The least favorable prior is a linear combination

$$\lambda \Gamma(\theta|\alpha_1, \tau_1) + (1-\lambda)\Gamma(\theta|\alpha_2, \tau_2),$$

where

$$\lambda = \frac{u_2 - u_0}{u_2 - u_1} \tag{5.7}$$

The linear Γ-minimax risk equals

$$\frac{v_0 - u_0^2}{v_0 + \frac{n}{2}(v_0 - u_0^2)} \tag{5.8}$$

Proof: The proof of this is straightforward.

Remark 6: This is the solution when u_0, v_0 as defined above satisfy $u_1 < u_0 < u_2$, $v_1 < v_0 < v_2$. Otherwise, the required maximum is obtained at one of the endpoints C_i.

The case $N > 2$ is identical. It is only necessary to locate the side of the polygon touched by a parabola from the family (5.5). This can easily be done on a computer, using the solution for the case $N = 2$.

Numerical Example: Let the elicited priors be $\Gamma(\theta|1,1)$ and $\Gamma(\theta|2,2)$. The segment C_1C_2 is a part of the line $v = \frac{22}{3}u - \frac{16}{3}$, between the points (1,2) and (4,24). The point C_0 is $(\frac{16}{11}, \frac{16}{3})$; that gives the least favorable prior as

$$\pi(\theta) = \frac{28}{33}\Gamma(\theta|1,1) + \frac{5}{33}\Gamma(\theta|2,2).$$

It is interesting that the point C_0, and therefore the least favorable prior, does not depend on the sample size n. The linear Γ-minimax risk equals

$$\frac{3.21763}{(\frac{n}{2}+1)3.21763 + 2.11570} \tag{5.9}$$

Finally, Theorem 3.1 gives as a lower bound on the exact Γ-minimax risk the quantities $\max_i \frac{1}{\frac{n}{2}+1+\alpha_i}$. In conjunction with (5.9), this then results in the following numbers.

Table 2: Lower bounds on efficiency of the linear Γ-minimax rule

n	lower bound
10	.8082
20	.8884
100	.9741

Example 7. Let us consider the priors $\pi(\theta)$ obtained by mixing all $\Gamma(\theta|\alpha,\tau)$ Gamma densities that have a mode equal to k, $k>0$. Assume $k_1 \le k \le k_2$ and $\alpha \ge 1+\varepsilon$ for some specified ε. The assumption of bounding the mode is perfectly natural. The assumption on α is a technical necessity. Then the set S of pairs (α,τ) is defined by

$$k_1 \le (\alpha-1)\tau \le k_2$$
$$\alpha \ge 1+\varepsilon.$$

The corresponding image $\mathcal{H}(T)$ is the area between the four parabolae

$$c_2 = \frac{2+\varepsilon}{1+\varepsilon}c_1^2,\ c_2 = c_1^2,\ c_2 = 2c_1(c_1-\frac{k_1}{2}),\ \text{and}\ c_2 = 2c_1(c_1-\frac{k_2}{2}). \tag{5.10}$$

The points $C_1=(k_1,k_1^2)$, $C_2=(k_2\frac{1+\varepsilon}{\varepsilon},\frac{k_2^2(1+\varepsilon)(2+\varepsilon)}{\varepsilon^2})$, $C_3=(k_1\frac{1+\varepsilon}{\varepsilon},\frac{k_1^2(1+\varepsilon)(2+\varepsilon)}{\varepsilon^2})$, and $C_4=(k_2,k_2^2)$ are the points (with positive coordinates) at which the four parabolae intersect. Since for all $k_1,k_2,k_1<k_2$ and $\varepsilon>0$ the inequality

$$\frac{\frac{k_2^2(1+\varepsilon)(2+\varepsilon)}{\varepsilon^2}-k_1^2}{k_2\frac{1+\varepsilon}{\varepsilon}-k_1}k_1\frac{1+\varepsilon}{\varepsilon} > \frac{k_1^2(1+\varepsilon)(2+\varepsilon)}{\varepsilon^2}$$

holds, we conclude that the point C_3 is below the line C_1C_2. This means that the least favorable distribution is determined by the points C_1C_2. The distribution corresponding to C_1 is a point mass at k_1 and the least favorable prior is of the form

$$\lambda\delta\{k_1\}+(1-\lambda)\Gamma(\theta|1+\varepsilon,\frac{k_2}{\varepsilon}). \tag{5.11}$$

Notice the very interesting fact that the solution is determined by just two priors from the infinite family, a phenomenon that has been called 'dictatorship of two Bayesians'.

6. Conclusion. The results in this paper complement earlier similar results in the location case. Our results indicate that a variety of technical tools are probably usable in deriving Bayes risk lower bounds: the information inequality and convexity being two of them. We are particularly pleased by the simplistic charm of the lower bound for mixture Gamma priors. The resulting linear estimation problems lead to nice geometric questions. The case of many unknown variances was not addressed here. We hope to do it elsewhere.

References

Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Borovkov, A. and Sakhanienko, A. (1980). On estimates of the expected quadratic risk. *Prob. and Math. Stat.*, **1**, 185–195.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems, *Ann. Math. Stat.* **42**, 855–903.

Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof, In Statistical Decision Theory and Related Topics, III, Ed. S. S. Gupta and J. Berger, Academic Press, New York.

Brown, L. D. and Gajek, L. (1990). Information inequalities for the Bayes risk, *Ann. Stat.*, **18**, 1578–1594.

Donoho, D., Liu, R. and MacGibbon, B. (1990). Minimax risk over hyperrectangles, *Ann. Stat.*, **18**, 1416–1437.

Feller, W. (1971). An Introduction to Probability Theory with Applications, John Wiley, New York.

Haff, L. R. (1991). The variational form of certain Bayes estimators, *Ann. Stat.* **19**, 1163–1190.

Jewel, N. (1982). Mixtures of exponential distributions, *Ann. Stat.*, **10**, 479–484.

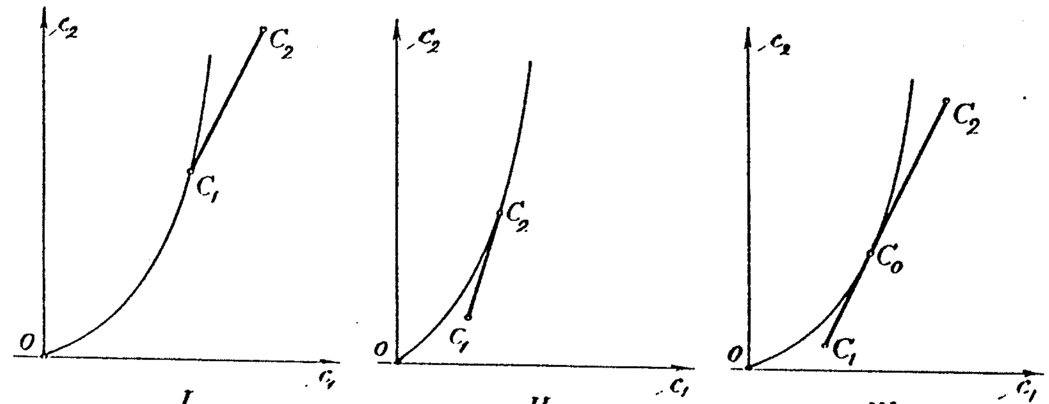
Lebedev, N. N. (1972). Special functions and their applications, Dover publications, New York.

Maatta, J. M. and Casella, G. (1990). Developments in decision-theoretic variances estimation, *Stat. Science* **5**, 90–120.

Min'ko, A. A. and Petunin, Yu. I. (1988). Unimodality of a noncentral chi-square distribution, *Theor. Prob. and Math. Stat.* **36**, 91–96.

Vidakovic, B. and DasGupta, A. (1992). Efficiency of linear rules for estimating bounded normal means, Tech. Report #92-23, Dept. of Statistics, Purdue University, submitted.

Zidek, J. V. (1988). Group decision analysis and its application to combining opinions, *Jour. Stat. Planning and Inf.*, **20**, 307–326.



I II III