Annals of Statistics
1994, Vol. 22, No. 2, 652–667

BAYESIAN ROBUSTNESS WITH MIXTURE CLASSES OF PRIORS[1]

BY SUDIP BOSE

George Washington University

Uncertainty in specification of the prior distribution is a common concern with Bayesian analysis. The robust Bayesian approach is to work with a class of prior distributions, which model uncertainty about the prior, instead of a single distribution. One is interested in the range of the posterior expectations of certain parametric functions as the prior varies over the class being considered—if this range is small, the analysis is robust to misspecification of the prior.

Relatively little research has dealt with robustness with respect to priors on several parameters, especially the problem of imposing shape and smoothness constraints on the priors in the class. To address this problem, we consider neighborhood classes of mixture priors. Results are presented for two kinds of "mixture classes," which yield different types of neighborhoods. The problem of finding suprema and infima of posterior expectations of parametric functions is seen to reduce to numerical maximization and minimization. In the applications we consider mixtures of uniform densities on variously shaped sets. This allows one to model symmetry and unimodality of different types in more than one dimension. Numerical examples are provided.

1. The robust Bayesian viewpoint. In a Bayesian analysis, one combines the likelihood and a prior to obtain a posterior distribution for the parameter(s) of interest and typically one is interested in the posterior expectation of one or more parametric functions. Possibly the most common criticism of a Bayesian analysis is that it supposedly requires one to quantify the available prior information as a probability measure on the parameter space, a process which would require probability judgments about uncountably many sets even for a real-valued parameter.

However, there has long been, at least since Good (1950), a robust Bayesian view, as it is now called [Berger (1984)]. Briefly put, this view assumes only that prior knowledge can be quantified in terms of a class Γ of prior distributions. A procedure is then said to be robust if its inferences are relatively insensitive to the variation of the prior distribution over the class Γ. Berger (1990) provides a thorough look at different approaches to the selection of Γ and techniques used in the analyses.

Depending on the class of priors, robust Bayesian analyses can be classified as "parametric" or "nonparametric." In parametric analyses, the prior is allowed to vary in a parametric class, usually consisting of priors that are conjugate to

Received February 1992; revised October 1993.

[1]This work was partially supported by NSF Grants DMS-87-02620 and DMS-89-23071.

AMS 1991 *subject classifications.* Primary 62A15; secondary 62F15.

Key words and phrases. Robust Bayes, classes of priors, density ratio, density bounded, mixture priors, mixtures of uniforms, shape constraints, smoothness constraints.

the likelihood. Restricting the prior to a parametric class often imposes fairly restrictive moment constraints and forces severe restrictions on the allowable prior tails. Tails of priors involve very small probabilities and are therefore very difficult to determine.

We feel that nonparametric classes of priors are a more accurate reflection of prior uncertainty, in that they usually place far fewer restrictions on the structure of the prior, particularly its moments. Various classes of nonparametric priors have been considered in the literature, ε-contamination classes, unimodal and symmetric priors with quantile constraints, density ratio and density bounded classes.

DeRobertis and Hartigan (1981) worked with the density ratio (DR) class of measures. For their class, a dominating measure λ exists, so it can be defined in terms of densities as

$$\Gamma = \left\{ \pi: \frac{L(\theta)}{U(\theta')} \le \frac{\pi(\theta)}{\pi(\theta')} \le \frac{U(\theta)}{L(\theta')} \text{ for almost all } \theta, \theta' \in \Theta \right\},$$

which can also be written as

$$\Gamma = \{\pi: L(\theta) \le \pi(\theta) \le U(\theta) \text{ for almost all } \theta \in \Theta\},$$

where L and U, satisfying $L \le U$, are specified nonnegative functions. In the above definitions, the densities were not normalized, that is, the corresponding measures were not necessarily probability measures.

Lavine (1991a, b) considers the density bounded (DB) class Γ (which is related to the density ratio class),

$$\Gamma = \left\{ \pi: L(\theta) \le \pi(\theta) \le U(\theta), \text{ for almost all } \theta \in \Theta, \int_\Theta \pi(\theta)\, d\lambda(\theta) = 1 \right\},$$

where $\int_\Theta L(\theta)\, d\lambda(\theta) < 1 < \int_\Theta U(\theta)\, d\lambda(\theta)$ and $\lambda(\theta)$ is a sigma-finite (dominating) measure. He considers the effects of uncertainty about the likelihood as well as the prior.

1.1. *Neighborhood classes with shape and smoothness constraints.*

Shape constraints in one or more dimensions. Classes of priors that incorporate shape constraints like symmetry and unimodality have been considered in the literature because such shape constraints may well reflect one's prior beliefs. It is not clear how one may impose such shape constraints on the DR and DB classes alluded to above. For that matter, what does one mean by symmetry and unimodality in several dimensions? We will see in later sections how one can impose various forms of unimodality and symmetry by considering mixture priors. We will obtain classes of priors that share the appealing features of the DR or DB classes and also impose shape constraints like unimodality and symmetry. The progression is a natural one. The DR and DB classes are neighborhood classes of priors—they are classes containing priors that are "close," in some sense, to one's base prior. Often, it is the case that one's base prior (e.g., a normal or Cauchy) is symmetric and/or unimodal in some sense. One can then look at a neighborhood class where the priors are "close" to the base prior with an additional degree of closeness in that they share some shape properties. At the same time one does not impose the moment constraints (often with severe associated constraints on tails) that are caused by considering the parametric classes.

Smoothness. It is seen with DR or DB classes that the priors that maximize or minimize posterior expectations have densities that jump back and forth between the functions L and U. In some cases such priors may not be reasonable in the sense of being realistic representations of prior beliefs. This could especially be the case when one is starting with a prior that has a smooth density, and one wants to look at a neighborhood of such a prior. The use of mixture classes provides us with a way to impose "smoothness" constraints that eliminate priors with densities that vary too rapidly.

In summary, we are interested in modelling "closeness" to a base prior. The base prior is often unimodal and symmetric in some sense. It may also have a continuous or smooth density. We are interested in the behavior of posterior expectations of one or more parametric functions of interest as the prior varies over a class of priors that are "close" to the base prior and that share some of its shape and smoothness properties. Looking at neighborhood classes which do not impose shape constraints and smoothness constraints may well lead us to think incorrectly that robustness is lacking, in some cases. This may especially happen when one looks at parameters in several dimensions.

1.2. *A look at some of the literature.* We have already referred to Good (1950) and Berger (1984, 1990) in the context of the robust Bayesian viewpoint and the DR and DB classes which appear, respectively, in DeRobertis and Hartigan (1981) and Lavine (1991a, b).

A widely used class of priors is the ε-contamination class,

$$\{\pi = (1-\varepsilon)\pi_0 + \varepsilon q,\ q \in Q\},$$

where Q, the class of allowable contaminations, is a subset of P, the class of all probability measures on the parameter space Θ, and where π_0 is the starting prior or "base" prior. Huber (1973) presented results on the extremes of the posterior probabilities of sets for the case, $Q = P$. Berger (1984, 1985), Moreno and Cano (1991) and Sivaganesan and Berger (1989) consider other classes of contaminations and other posterior measures. In particular, symmetry and unimodality constraints are imposed. Wasserman (1989) provides a robust interpretation of likelihood regions in the case $Q = P$.

Mixture classes have been considered in several contexts. Some recent references are Dalal and Hall (1983) and Diaconis and Ylvisaker (1985), who consider continuous mixtures of conjugate priors and present results on approximating the prior and the posterior. Dicky and Chen (1985) discuss the elicitation of spherically symmetric and elliptically symmetric priors.

1.3. *A preview of the paper.* In Section 2, we introduce mixture classes and present results on finding the suprema and infima of posterior expectations for two kinds of mixture classes. They are classes of mixture distributions where the mixing density comes from a density ratio (DR) and a density bounded (DB) class, respectively.

Section 3 will deal with the commonly accepted concept of unimodality and symmetry in one dimension. There are several different possible definitions of symmetry and unimodality in many dimensions, and we shall consider a couple of them in Sections 4 and 5. We express unimodal and symmetric densities of various types as mixtures of uniforms over bounded sets. These mixtures of uniforms will be seen to satisfy certain "smoothness" conditions.

We also discuss ways of choosing the densities L and U that define the DR and DB classes of mixing densities. Several methods are discussed for the one-dimensional case in Section 3, and the ideas also apply to the multidimensional cases of Sections 4 and 5. As we point out in the next section, the formal mathematics of mixtures is equivalent to that of a hierarchical prior–hyperprior analysis. We do not emphasize that point of view because we consider mixtures of uniforms on bounded sets, and we do not propose such uniforms, by themselves, as priors to be considered seriously in a single prior Bayesian analysis.

2. Introduction to mixture classes. Let $\theta \in \Theta \subset R^n$ be an unknown parameter with likelihood $l(\theta)$. (We will assume, unless otherwise stated, that the data X is fixed.) Let $\{\pi_c(\theta)\}_{c \in A}$ be a family of prior densities for θ, where $A \subset R^m$. Define,

$$\mathcal{C} = \left\{\pi \colon \pi(\theta) = \int_A \pi_c(\theta)\, d\mu(c) \text{ for some } \mu \in \mathcal{D}\right\},$$

where $\mathcal{D}$ is some class of measures on A. This is formally equivalent to a hierarchical analysis with $\pi_c(\theta)$ as the prior, c as a hyperparameter and $\mathcal{D}$ as a class of hyperpriors μ.

Let $m(c) = \int l(\theta)\pi_c(\theta)\, d\theta$. Hence,

$$\int \phi(\theta) l(\theta) \pi_c(\theta)\, d\theta = m(c) E^{\pi_c}\left[\phi(\theta) \mid X\right].$$

Now, if $\pi \in \mathcal{C}$, there is a $\mu \in \mathcal{D}$ such that

$$\pi(\theta) = \int_A \pi_c(\theta)\, d\mu(c).$$

When $\pi \in \mathcal{C}$ and $\mu \in \mathcal{D}$ are as above we will say that π is *induced* by μ; μ will be referred to as a *mixing distribution* and its density, if it exists, will be referred to as a *mixing density*. For such a π, μ, pair, we have

$$E^{\pi}\left[\phi(\theta) \mid X\right] = \frac{\int_A m(c) E^{\pi_c}\left[\phi(\theta) \mid X\right] d\mu(c)}{\int_A m(c)\, d\mu(c)}.$$

2.1. *Density ratio classes of mixing distributions.* Let L and U be densities corresponding to sigma-finite measures μ_L and μ_U on A, and let $L(c) \le U(c)$ for all $c \in A$. Define

$$\mathcal{D}_1 = \{p: L(c) \le p(c) \le U(c)\ \forall c \in A\}$$

and

$$\mathcal{C}_1 = \left\{\pi: \pi(\theta) = \int_A \pi_c(\theta)\, d\mu(c) \text{ for some } \mu \text{ with density } p \in \mathcal{D}_1\right\}.$$

Such a class $\mathcal{C}_1$ will be referred to as a DR mixture class. As in Section 1, the measures in $\mathcal{D}_1$ are not normalized. The presence or absence of the normalizing constant does not affect the posterior expectation of a parametric function, that is, if, for a density π, we define

$$T(\pi) = \frac{\int \phi(\theta) l(\theta) \pi(\theta)\, d\theta}{\int l(\theta) \pi(\theta)\, d\theta},$$

then for $\alpha > 0$, $T(\alpha\pi) = T(\pi)$.

The following theorem shows how to determine the supremum and infimum of the posterior expectation of real-valued ϕ for the prior lying in a DR mixture class. Let $\phi: \Theta \to R$ be a measurable, parametric function. For each real t, define

$$A_t = \left\{c \in A: E^{\pi_c}\left[\phi(\theta) \mid X\right] > t\right\}$$

and $B_t = A - A_t$.

THEOREM 2.1. *Let $\{\pi_c(\theta)\}_{c \in A}$, L, U, $\mathcal{D}_1$ and $\mathcal{C}_1$ be as above. Let ϕ: $\Theta \to R$ be the measurable parametric function of interest, and let A_t and B_t be as defined above. Then*

$$\sup_{\pi \in \mathcal{C}_1} E^{\pi}\left[\phi(\theta) \mid X\right] = \sup_{t \in R} M(t),$$
$$\inf_{\pi \in \mathcal{C}_1} E^{\pi}\left[\phi(\theta) \mid X\right] = \inf_{t \in R} N(t),$$

where

$$M(t) = \frac{\int_{A_t} m(c) E^{\pi_c}\left[\phi(\theta) \mid X\right] U(c)\, dc + \int_{B_t} m(c) E^{\pi_c}\left[\phi(\theta) \mid X\right] L(c)\, dc}{\int_{A_t} m(c) U(c)\, dc + \int_{B_t} m(c) L(c)\, dc}$$

and $N(t)$ is defined by reversing L and U in the definition of $M(t)$ above.

PROOF. We now prove the result for the supremum. For each real t, let ψ_t' be the density in $\mathcal{C}_1$ induced by the mixing distribution μ_t' that has density $p_t' \in \mathcal{D}_1$, defined below:

$$p_t'(c) = \begin{cases} U(c), & c \in A_t, \\ L(c), & c \in B_t. \end{cases}$$

Then

$$M(t) = E^{\psi_t'}\left[\phi(\theta)\,|\,X\right],$$

which implies that

$$\sup_{\pi \in \mathcal{C}_1} E^{\pi}\left[\phi(\theta)\,|\,X\right] \geq \sup_{t \in R} M(t).$$

To prove the reverse inequality, we proceed through steps (i) and (ii).

(i) Let $\pi \in \mathcal{C}_1$, with $E^{\pi}[\phi(\theta)\,|\,X] = s$, induced by a mixing distribution with density $p \in \mathcal{D}_1$. Let $\pi' = \psi_s'$ be the density which is induced by the mixing density p_s'. Then, writing p' for p_s',

$$p'(c) \geq p(c), \qquad c \in A_s$$
$$p'(c) \leq p(c), \qquad c \in B_s.$$

(ii) (We will now show that $E^{\pi'}[\phi(\theta)\,|\,X] \geq E^{\pi}[\phi(\theta)\,|\,X]$.) For $c \in A_s$, define $v(c) = p'(c) - p(c)$ and, for $c \in B_s$, define $w(c) = p(c) - p'(c)$. From (i), v and w are nonnegative on their domains.

Then

$$E^{\pi'}\left[\phi(\theta)\,|\,X\right]$$
$$= \frac{\int_{A_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]p'(c)\,dc + \int_{B_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]p'(c)\,dc}{\int_{A_s} m(c)p'(c)\,dc + \int_{B_s} m(c)p'(c)\,dc}$$
$$= \frac{\int_{A_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]\big(p(c)+v(c)\big)\,dc + \int_{B_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]\big(p(c)-w(c)\big)\,dc}{\int_{A_s} m(c)\big(p(c)+v(c)\big)\,dc + \int_{B_s} m(c)\big(p(c)-w(c)\big)\,dc}$$
$$\geq s.$$

To prove the last inequality, we split the numerator and denominator into four terms each. The first two terms in the numerator are

$$\int_{A_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]p(c)\,dc \quad \text{and} \quad \int_{B_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]p(c)\,dc$$

and the next two are

$$\int_{A_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]v(c)\,dc \quad \text{and} \quad -\int_{B_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]w(c)\,dc;$$

the corresponding terms in the denominator are the same except that the integrands do not contain $E^{\pi_c}[\phi(\theta)\,|\,X]$. Thus the first two terms in the numerator and denominator are, respectively, the numerator and denominator of $E^{\pi}[\phi(\theta)\,|\,X]$, which equals s, by assumption, and for the next two terms in the numerator and denominator we use the following inequalities:

$$\int_{A_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]v(c)\,dc \geq s\int_{A_s} m(c)v(c)\,dc;$$
$$\int_{B_s} m(c)E^{\pi_c}\left[\phi(\theta)\,|\,X\right]w(c)\,dc \leq s\int_{B_s} m(c)w(c)\,dc.$$

[These two inequalities are immediate consequences of the definitions of A_s and B_s and of the fact that $m(c)v(c)$ and $m(c)w(c)$ are nonnegative functions.]

This completes the proof in the supremum case.

The proof of the result for the infimum follows by replacing $\phi(\theta)$ by $-\phi(\theta)$ in the above proof for the supremum. □

REMARK. It can be shown that if $s = \sup_{t \in R} M(t)$, then $M(t) > t$ if $t < s$, $M(s) = s$ and $M(t) < t$ if $t > s$. This allows us to use a regula falsa method to find s as the solution to $M(t) - t = 0$, and at each stage of the search we have a lower and upper bound on s, one of which is improved at each step.

Similar inequalities hold for $N(t)$, that is, if $u = \inf_{t \in R} N(t)$, then $N(t) > t$ if $t < u$, $N(u) = u$ and $N(t) < t$ if $t > u$.

2.2. *Density bounded classes of mixing distributions.* Let L and U, with $L(c) \leq U(c)$ for all $c \in A$, be densities of sigma-finite measures μ_L and μ_U, respectively, which satisfy

$$\mu_L(A) \leq 1 \leq \mu_U(A);$$

define

$$\mathcal{D}_2 = \left\{ p\colon L(c) \leq p(c) \leq U(c)\ \forall c \in A \text{ and } \int_A p(c)\, dc = 1 \right\}$$

and

$$\mathcal{C}_2 = \left\{ \pi\colon \pi(\theta) = \int_A \pi_c(\theta) d\mu(c) \text{ for some } \mu \text{ with density } p \in \mathcal{D}_2 \right\}.$$

The following argument is similar in spirit to that of Lavine (1991a, b). Once again, let $\phi\colon \Theta \to R$ and define

$$psup = \sup_{\pi \in \mathcal{C}_2} E^{\pi}\left[\phi(\theta) \,|\, X\right],$$
$$pinf = \inf_{\pi \in \mathcal{C}_2} E^{\pi}\left[\phi(\theta) \,|\, X\right].$$

We show how to compute *psup*. We can find *pinf* in a similar fashion. We assume throughout the rest of this subsection that we can find real numbers M_1 and M_2 satisfying $-\infty < M_1 \leq psup \leq M_2 < \infty$. The algorithm for finding *psup* to within any desired accuracy is based on being able to test, for any $s \in [M_1, M_2]$, whether *psup* is greater than s. One finds a $\psi_s \in \mathcal{C}_2$ such that

$$s < psup \quad \Rightarrow \quad E^{\psi_s}\left[\phi(\theta) \,|\, X\right] \geq s, \tag{1}$$

$$s > psup \quad \Rightarrow \quad E^{\psi_s}\left[\phi(\theta) \,|\, X\right] < s. \tag{2}$$

Once we know how to find such ψ_s's, we can conduct a simple bisection search and bracket *psup* to within desired "tolerance."

We now consider, given $s \in [M_1, M_2]$, how to construct a ψ_s satisfying (1) and (2). That condition (2) holds is obvious. Also, (1) is obvious for $s \leq pinf$. Therefore all that remains is to ensure that

$$s \in (pinf, psup) \Rightarrow E^{\psi_s}\big[\phi(\theta)\,|\,X\big] \geq s.$$

Theorem 2.2. deals with the existence of the ψ_s's.

THEOREM 2.2. *Let $s \in (pinf, psup)$, and let $h(c) = (E^{\pi_c}[\phi(\theta)\,|\,X] - s)m(c)$. Define, for each real z, the sets*

$$A_z = \{c \in A\colon h(c) > z\},$$
$$B_z = \{c \in A\colon h(c) = z\},$$
$$C_z = \{c \in A\colon h(c) < z\}.$$

Then the following hold:

(i) *There exists $p_s \in \mathcal{D}_2$, the density of a probability measure μ_s, satisfying $p_s(c) = U(c)$ for $c \in A_z$ and $p_s(c) = L(c)$ for $c \in C_z$, for some z.*

(ii) *If $\psi_s \in \mathcal{C}_2$ is the density induced by μ_s, that is, $\psi_s(\theta) = \int_A \pi_c(\theta)\,d\mu_s(c)$,*

$$E^{\psi_s}\big[\phi(\theta)\,|\,X\big] \geq s.$$

See the Appendix for the proof of Theorem 2.2.

3. Mixture distributions for a real parameter. Consider the case of a single real parameter θ. Without loss of generality, we shall take the origin as the mode and point of symmetry of unimodal symmetric prior densities. Densities that are unimodal and symmetric about 0 shall be referred to as *US*0 densities. *US*0 densities in one dimension can be expressed as mixtures of uniforms on intervals symmetric about 0. So, one has $A = [0, \infty] \subset R$ and, for $c \in A$,

$$\pi_c(\theta) = \frac{1}{2c} I_{(-c,\,c)}(\theta).$$

Suppose that one has a *US*0 base prior π_0, and one wants to robustify the single prior analysis using π_0 by considering a class of priors "close" to π_0. Let π_0 be induced by the mixing distribution μ in the sense of Section 2. Let μ have density f w.r.t. Lebesgue measure. Then

$$\pi_0(\theta) = \int_0^\infty \frac{1}{2t} I_{(-t,\,t)}(\theta) f(t)\,dt$$
$$= \int_{|\theta|}^\infty \frac{1}{2t} f(t)\,dt$$

and

$$f(t) = -2t\left[\frac{d}{d\theta}\pi_0(\theta)\right]_{\theta = t}$$

For $\theta \in [0, \infty)$, the density is decreasing, and the negative of the derivative is the rate of fall; hence, by the above representation, it follows that with a DR or DB class, the rate of fall of the mixed density is in between the rates of fall of the mixed densities induced by L and U, respectively. Thus these mixed densities are "smooth"—in particular they are continuous. The fact that priors in DR or DB classes can "hop" between L and U and that the priors that maximize and minimize posterior expectations do precisely that is, to many, an undesirable feature of DR and DB classes. With mixture classes one avoids this difficulty. Simply requiring continuity as an additional constraint for priors in a DR or DB class is not enough. If discontinuity is unappealing, prior with arbitrarily large rates of fall hardly seem more appealing. One wants to have some sort of "smoothness" requirement.

We now present suggestions for the choice of L and U for DR class and the DB class of mixing distributions. In the DR case, the statements about the densities are for unnormalized densities, whereas, for the DB case, they refer to normalized (i.e., probability) densities.

CASE 1. One can take, for the DR class,

$$L(t) = f(t),$$
$$U(t) = Kf(t), \qquad K > 1.$$

The larger value of K, the larger the class.

For DB class, assume that f is a probability density, that is, μ is a probability measure on $[0, \infty)$, and take $K_1 < 1$ and $K_2 > 1$,

$$L(t) = K_1 f(t),$$
$$U(t) = K_2 f(t).$$

The smaller the value of K_1 and/or the larger the value of K_2, the larger the class.

This is probably the simplest choice of L and U, requiring only the input K or K_1 and K_2, and is a reasonable way to robustify a Bayesian analysis using the prior π_0. However, this choice means that one does not allow too much variation in tail behavior from that of π_0.

CASE 2. One can also take, for the DR class,

$$L(t) = 0,$$
$$U(t) = f(t).$$

This means that there is no lower bound on the rate of fall, so one can have prior densities that are flat on intervals. Essentially, this is the class of (unnormalized) densities that are smaller than π_0 and whose rate of fall (for positive θ) is smaller than the rate of fall of π_0.

With this particular choice of L and U, one gets a large class of densities. The class of Case 2 is larger than the limit of the class of Case 1 as $K \to \infty$.

A similar DB class would have, for some $K > 1$,

$$L(t) = 0,$$
$$U(t) = Kf(t).$$

As above, with such a class, there would be no lower bound on the rate of fall, so this is the class of densities that are smaller than $K\pi_0(\theta)$ and whose rate of fall (for positive θ) is smaller than that of π_0.

CASE 3. Now suppose that one does not have a base prior π_0, but one wants a class of priors that contains thin-tailed and thick-tailed priors. In particular, one has a thin-tailed prior π_1 and a thick-tailed prior π_2, and one wants a class of priors that contains both of them and densities with tail thicknesses "in between" the tail thicknesses of π_1 and π_2.

Let f_1 and f_2 generate π_1 and π_2, respectively. One can take

$$L(t) = \min\big(f_1(t), f_2(t)\big),$$
$$U(t) = \max\big(f_1(t), f_2(t)\big).$$

If π_1 and π_2 are probability densities, then so are f_1 and f_2, hence the above applies to both the DR and the DB case.

CASE 4. In Case 3 we considered the situation where one wants to include a particular thin-tailed prior π_1 and a particular thick-tailed prior π_2. In such a situation one may only want the "tail" of the density bounded by the tails of π_1 and π_2, that is, one wants the mixing density to be between f_1 and f_2 only for t "large," say, $t > t_0$. It may also happen that in a neighborhood of 0 we may feel that the density closely resembles a density π_0, generated by f.

Then we can take

$$L(t) = \begin{cases} f(t), & 0 \le t \le t_0, \\ \min\big(f_1(t), f_2(t)\big), & t > t_0, \end{cases}$$

$$U(t) = \begin{cases} f(t), & 0 \le t \le t_0, \\ \max\big(f_1(t), f_2(t)\big), & t > t_0. \end{cases}$$

We may want to allow some more variation near 0. In the DR case, one may take $U(t)$ to be $Kf(t)$ for $0 \le t \le t_0$, for some $K > 1$. In the DB case, one can take $L = K_1 f$ and $U = K_2 f$ for $K_1 < 1 < K_2$.

EXAMPLE 1. We look at the posterior mean, that is, $\phi(\theta) = \theta$, and assume that the likelihood is normal, with unit variance:

$$l(\theta \,|\, X) \propto \exp\big[-\tfrac{1}{2}(\theta - X)^2\big].$$

Suppose that one is interested in studying robustness of the analysis using the standard normal prior $\pi_0(\theta) = (1/\sqrt{2\pi})\exp[-\theta^2/2]$. The prior π_0 is a $US0$

prior. A possible neighborhood class of $US0$ priors is a class of mixtures of uniforms on intervals symmetric about the origin, with a DR class of mixing densities. For such mixtures, $\pi_0(\theta)$ is generated by the mixing density $f(t) = (\sqrt{2}/\sqrt{\pi})t^2 \exp[-t^2/2]$. For the bounds on the (DR class of) mixing distributions, we use functions $L(t)$ and $U(t)$ that generate, respectively, π_0 and a multiple of π_0. This is the type of choice mentioned in Case 1 of this section. Thus

$$L(t) = \left(\frac{\sqrt{2}}{\sqrt{\pi}}\right)t^2 \exp\left[\frac{-t^2}{2}\right] \quad \text{and} \quad U(t) = K\left(\frac{\sqrt{2}}{\sqrt{\pi}}\right)t^2 \exp\left[\frac{-t^2}{2}\right]$$

[one can, of course, drop the common constant $(\sqrt{2}/\sqrt{\pi})$ in the expressions for L and U above]. We did not specify the value of K in the above. Naturally, the larger the value of K, the larger the neighborhood class. In Table 1, we present the range of the posterior mean for several values of K.

The intervals are presented in Table 1 for $X = 1.0, 2.0, 3.0, 4.0$ and for $K = 1.25, 1.50, 1.75, 2.00$. With the standard normal prior, the posterior expectation of $\phi(\theta) = \theta$ is, of course, just $X/2.0$, which has value 0.5, 1.0, 1.5 and 2.0 for the four rows of the table.

As K increases, the intervals get wider, since one is looking at larger classes of priors. It is also seen that the intervals get wider as X increases, that is, as the location of the likelihood is further and further away from the location of the prior. It is seen that by the time one gets to $X = 3.0$, the width has stabilized somewhat; the intervals for $X = 4.0$ are not much wider than those for $X = 3.0$.

4. Uniforms on rectangular setsorthant symmetry and coordinatewise unimodality. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_n) \in R^n$. Without loss of generality we take the origin as the point of symmetry as well as the mode. We consider mixtures of uniforms over rectangle sets in R^n, that is, sets of the form,

$$[-a_1, a_1] \times [-a_2, a_2] \times \cdots \times [-a_n, a_n],$$

for $a_i \geq 0, i = 1, 2, \ldots, n$. Letting $c = (a_1, a_2, \ldots, a_n)$, we denote the above rectangle set by $[-c, c]$. We then have

$$\pi_c(\theta) = \left[2^n \prod_{i=1}^{n} a_i\right]^{-1} I_{[-c,c]}(\theta)$$

TABLE 1
Range of posterior mean with mixtures

Values of X	Values of K: 1.25	1.50	1.75	2.00
1.0	(0.477, 0.523)	(0.458, 0.542)	(0.442, 0.558)	(0.429, 0.571)
2.0	(0.961, 1.039)	(0.929, 1.070)	(0.903, 1.096)	(0.879, 1.119)
3.0	(1.452, 1.547)	(1.414, 1.586)	(1.381, 1.618)	(1.352, 1.646)
4.0	(1.948, 2.052)	(1.905, 2.094)	(1.868, 2.131)	(1.837, 2.162)

and $c \in A = [0, \infty) \times [0, \infty) \times \cdots \times [0, \infty) \subset R^n$. Hence,

$$\pi(\theta) = \int_A \left[2^n \prod_{i=1}^{n} a_i \right]^{-1} I_{[-c,c]}(\theta) f(c)\, dc$$
$$= \int_{|\theta_n|}^{\infty} \cdots \int_{|\theta_1|}^{\infty} \left[2^n \prod_{i=1}^{n} a_i \right]^{-1} f(a_1, \dots, a_n)\, da_1 \cdots da_n$$

and

$$f(a_1, a_2, \dots, a_n) = (-2)^n \left(\prod_{i=1}^{n} a_i \right) \left[\frac{\partial^n}{\partial\theta_1\, \partial\theta_2 \cdots \partial\theta_n} \pi(\theta) \right]_{\theta = c}.$$

Such densities are symmetric in the sense that they depend on their arguments only through their absolute values, and unimodal in the sense that if $0 \le \theta_{11} \le \theta_{12}$, $0 \le \theta_{21} \le \theta_{22}, \dots, 0 \le \theta_{n1} \le \theta_{n2}$, then

$$\pi(\theta_{11}, \theta_{21}, \dots, \theta_{n1}) \ge \pi(\theta_{12}, \theta_{22}, \dots, \theta_{n2}),$$

and if we define π_L and π_U to be the mixture densities generated by L and U in the same way that π is generated by f above, then the rate of fall of π is bounded above and below by the rates of fall of π_L and π_U in the sense that if $0 \le \theta_{11} \le \theta_{12}$, $0 \le \theta_{21} \le \theta_{22}, \dots, 0 \le \theta_{n1} \le \theta_{n2}$, then

$$\begin{aligned} &\pi_L(\theta_{11}, \theta_{21}, \dots, \theta_{n1}) - \pi_L(\theta_{12}, \theta_{22}, \dots, \theta_{n2}) \\ &\quad \le \pi(\theta_{11}, \theta_{21}, \dots, \theta_{n1}) - \pi(\theta_{12}, \theta_{22}, \dots, \theta_{n2}) \\ &\quad \le \pi_U(\theta_{11}, \theta_{21}, \dots, \theta_{n1}) - \pi_U(\theta_{12}, \theta_{22}, \dots, \theta_{n2}). \end{aligned}$$

Actually, mixtures of uniforms on rectangular sets satisfy a stronger condition than the coordinatewise unimodality referred to above. Consider the rectangle M whose vertices consist of the 2^n points of the form $x = (t_1, t_2, \dots, t_n)$, where t_i is either θ_{i1} or θ_{i2}. Let $\mathrm{sgn}_M(x)$, the signum of the vertex, be +1 or −1, according as the number of i, $1 \le i \le n$, satisfying $t_i = \theta_{i2}$ is even or odd. The condition is

$$D(\pi) = \sum \mathrm{sgn}_M(x)\pi(x) \ge 0,$$

where the sum extends over all vertices x of M. In fact, all priors π in the class satisfy the condition $D(\pi_L) \le D(\pi) \le D(\pi_U)$.

The symmetry above is referred to as *orthant symmetry*, and the above form of unimodality is referred to as *block unimodality*.

The suggested ways of choosing L and U, as discussed in Cases 1–4, now apply.

EXAMPLE 2. We consider a situation with $\theta = (\theta_1, \theta_2)$ and a normal likelihood,

$$l(\theta_1, \theta_2 \,|\, X) \propto \exp\left[-\tfrac{1}{2} \left((\theta_1 - X_1)^2 + (\theta_2 - X_2)^2 \right) \right],$$

TABLE 2
Effect of shape and smoothness by taking mixtures

Values of K	DR class	DR mixture class	Width ratio
1.25	(0.2050, 0.2872)	(0.2390, 0.2486)	0.117
1.50	(0.1769, 0.3259)	(0.2354, 0.2526)	0.115
1.75	(0.1556, 0.3606)	(0.2321, 0.2561)	0.117
2.00	(0.1388, 0.3920)	(0.2294, 0.2591)	0.117

for $X_1 = 0.5, X_2 = 0.5$. The parametric function is $\phi(\theta) = I_{(0,1)\times(0,1)}(\theta_1, \theta_2)$ so that the posterior expectation of ϕ is the posterior probability that θ is in the rectangle $(0,1) \times (0,1)$. We determine the supremum and infimum of the range of the above posterior probability for two classes of priors. Both these classes will represent neighborhoods of the normal prior π_0 with zero means and the identity matrix as the dispersion matrix.

The first class will be a density ratio class with the lower bound being the same as π_0 and the upper bound a multiple of π_0. This class allows priors that are not unimodal or symmetric about the origin and also priors with discontinuous densities. Thus it does not impose shape or smoothness constraints. The bounds are

$$L(\theta) = (2\pi)^{-1} \exp\left[-0.5(\theta_1^2 + \theta_2^2)\right],$$
$$U(\theta) = K(2\pi)^{-1} \exp\left[-0.5(\theta_1^2 + \theta_2^2)\right],$$

for the values $K = 1.25, 1.5, 1.75, 2.0$. [The common factor $(2\pi)^{-1}$ can, of course, be dropped from both L and U without affecting the DR class.]

The second class of priors that we will consider is a neighborhood class with shape and smoothness constraints. We consider a class of mixtures of uniforms on rectangles symmetric about both axes, with the mixing distributions in a DR class. The bounds on the DR class (of mixing distributions) will be the functions L_1 and U_1 that generate, respectively, the functions L and U. The expressions are $L_1(a,b) = 4a^2b^2(2\pi)^{-1}\exp[-0.5(a^2+b^2)]$ and $U_1(a,b) = KL_1(a,b)$. Thus the choice of L_1 and U_1 is similar to that suggested in Case 1 of Section 3.

Notice that the second class is strictly contained in the first—it consists of the densities between L and U whose rates of fall in any direction are bounded by those of L and U and which are orthant symmetric and block unimodal. With the smaller class of priors one naturally gets intervals of less width. To observe the magnitude of the effect of imposing the additional shape and smoothness constraints, we calculate the ratio of widths, smaller divided by larger. For the single prior π_0, the posterior probability is 0.2438. All the intervals, therefore, contain the value 0.2438.

The interval for the mixture class is more than eight times narrower than the other. For the DR class, the upper bound is a little further from the single prior value than the lower bound, and the difference becomes more pronounced as K gets larger. On the other hand, for the mixture class, the intervals are almost symmetric around 0.2438, with the upper bound being very slightly further away from 0.2438. Once again, remember that the densities are unnormalized. Even though, nominally, only the upper density changes as K increases, one can as well imagine that the lower bound is being made smaller with the upper bound held fixed, or that the lower bound is being made smaller as well as the upper bound being made larger.

5. Mixtures of uniforms on spheres and ellipses. We will now consider classes of elliptically symmetric or spherically symmetric densities. Once again, the origin will be the point of symmetry as well as the mode.

Suppose that one has a base prior ψ_0, and one wants to robustify the single prior analysis using ψ_0 by considering a class of priors "close" to ψ_0. Suppose that ψ_0 is elliptically symmetric in θ, that is,

$$\psi_0(\theta) = \pi_0(\theta'\Sigma^{-1}\theta),$$

where $\pi_0\colon [0,\infty) \to [0,\infty)$ is a nonincreasing, absolutely continuous function and Σ is a positive definite matrix. Whenever ψ_0 is of the above form, the dispersion matrix is $c\Sigma$, for some $c > 0$. If $\Sigma = I$, the identity matrix, then the density is spherically symmetric.

With $\Sigma = I$ and a density that is radially decreasing away from 0, we represent it by

$$\pi(\theta) = \int_{\|\theta\|}^{\infty} \frac{1}{g_n\sigma^n}\, d\mu(\sigma),$$

where $\|\cdot\|$ denotes the Euclidean norm and $g_n = 2\pi^{n/2}/n\Gamma(n/2)$ is the volume of the n-dimensional unit sphere. Thus, π is a mixture of π_c's, where $\pi_c(\theta) = I_{S_c}(\theta)/g_n c^n$, S_c denoting the n-dimensional sphere of radius c, centered at the origin, and I_B being the indicator function of the set B. Elliptically symmetric densities are generated as mixtures (with respect to c) of densities (uniform on ellipses) $\pi_c^A(\theta) = I_{[0,c]}(\theta' A\theta)/g_n^A c^n$, where g_n^A denotes the volume of the n-dimensional ellipse $\{\theta' A\theta \le 1\}$ and A is a positive definite, $n \times n$ matrix. The mixing μ is a measure on subsets of $[0,\infty)$. Dickey and Chen (1985) discuss elicitation of spherically symmetric and elliptically symmetric priors. One can get a mixing distribution μ_0 that generates the base prior π_0. For the DR and DB classes, one has to choose L and U, but since they are densities on $[0,\infty)$ the ideas of Section 3 apply.

APPENDIX

PROOF OF THEOREM 2.2. We use the notation of Section 2, especially subsection 2.2.

Proof of (i). Define the functions $\underline{g}$ and $\bar{g}$, both from R to $[0, \infty)$ by

$$\underline{g}(y) = \mu_U(A_y) + \mu_L(B_y) + \mu_L(C_y)$$

and

$$\bar{g}(y) = \mu_U(A_y) + \mu_U(B_y) + \mu_L(C_y).$$

Then $\underline{g}$ and $\bar{g}$ are both nonincreasing functions. Take $z = \inf\{y: \underline{g}(y) < 1\}$. We shall now show $\underline{g}(z) \leq 1$ and $\bar{g}(z) \geq 1$.

From the definition of $\underline{g}(y)$, we have, for any $t > 0$,

$$\underline{g}\left(z + 1/t\right) = \underline{g}(z) - (\mu_U - \mu_L)\left(\left\{c: h(c) \in (z, z + 1/t]\right\}\right).$$

Hence $\lim_{t\to\infty} \underline{g}(z + 1/t) = \underline{g}(z)$ by continuity of measures. However, by the definition of z, $\underline{g}(z + 1/t) < 1$ for $t > 0$, so $\underline{g}(z) \leq 1$. Similarly, again for $t > 0$,

$$\bar{g}(z - 1/t) = \bar{g}\left(z\right) + (\mu_U - \mu_L)\left(\left\{c: h(c) \in \left[z - 1/t, z\right)\right\}\right).$$

Thus, $\lim_{t\to\infty} \bar{g}(z - 1/t) = \bar{g}(z)$ so that $\bar{g}(z) \geq 1$.

We know that $\underline{g}(z) \leq 1 \leq \bar{g}(z)$.

(a) If $\underline{g}(z) = 1$, then define $p_s(c) = U(c)$, for $c \in A_z$, and equal to $L(c)$, for other c.

(b) If (a) does not hold but instead $\bar{g}(z) = 1$, then define $p_s(c) = U(c)$, for $c \in A_z \cup B_z$, and equal to $L(c)$, for other c.

(c) If neither (a) nor (b) holds, that is, if $\underline{g}(z) < 1 < \bar{g}(z)$, then define p_s by

$$p_s(c) = \begin{cases} U(c), & c \in A_z, \\ L(c), & c \in C_z, \\ \alpha U(c) + (1 - \alpha)L(c), & c \in B_z, \end{cases}$$

where $\alpha = (1 - \underline{g}(z))/(\bar{g}(z) - \underline{g}(z))$.

This completes the proof of (i) of Theorem 2.2.

Proof of (ii). Given $s < psup$, let p_s and μ_s be as above, satisfying (i). Further, since $s < psup$, there is a $\psi \in \mathcal{C}_2$ such that $E^{\psi}[\phi(\theta)\,|\,X] \geq s$. Also, suppose that ψ is induced by a mixing distribution μ which has density $p \in \mathcal{D}_2$. Define S and T as

$$S = \left\{c \in A: p_s(c) > p(c)\right\},$$
$$T = \left\{c \in A: p_s(c) < p(c)\right\}.$$

Then, $S \subset A \setminus C_z$ and $T \subset A \setminus A_z$. Now,

$$\begin{aligned}
&\int_A h(c)\, d\mu_s(c) - \int_A h(c)\, d\mu(c) \\
&\quad = \int_A h(c)\left(p_s(c) - p(c)\right) dc \\
&\quad = \int_S h(c)\left(p_s(c) - p(c)\right) dc - \int_T h(c)\left(p(c) - p_s(c)\right) dc \\
&\quad \geq 0
\end{aligned}$$

since, for $c \in S$, $h(c) \geq z$; for $c \in T$, $h(c) \leq z$; and $\int_S (p_s(c) - p(c))\, dc = \int_T (p(c) - p_s(c))\, dc$ and $p_s(c) - p(c) \geq 0$, for $c \in S$, and $p(c) - p_s(c) \geq 0$, for $c \in T$.

Further,

$$E^{\psi}\left[\phi(\theta)\,|\,X\right] \geq s \quad \Rightarrow \quad \int_A m(c) E^{\pi_c}\left[\phi(\theta)\,|\,X\right] d\mu(c) \geq s \int_A m(c)\, d\mu(c).$$

Therefore, $\int_A h(c)\, d\mu(c) \geq 0$, and so by what we have shown above, $\int_A h(c)\, d\mu_s(c) \geq 0$; but then this, in turn, means that $\int_A m(c) E^{\pi_c}[\phi(\theta)\,|\,X]\, d\mu_s(c) \geq s \int_A m(c)\, d\mu_s(c)$. Since μ_s is a mixing distribution that induces ψ_s, it follows that $E^{\psi_s}[\phi(\theta)\,|\,X] \geq s$, which completes the proof of (ii). □

Acknowledgments. This article forms part of the author's Ph.D. dissertation at Purdue University. The author expresses his deep appreciation for the invaluable guidance and support provided by Professor James Berger. The author is also grateful to Eric Slud for his useful and interesting comments when the author was visiting the Department of Mathematics at the University of Maryland, College Park. The author would also like to express his appreciation for the helpful remarks of the Editor, an Associate Editor and a referee.

REFERENCES

BERGER, J. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness of Bayesian Analysis* (J. Kadane, ed.) 63–144. North-Holland, Amsterdam.

BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

BERGER, J. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference* **25** 303–328.

DALAL, S. R. and HALL, W. J. (1983). Approximating priors by mixtures of natural conjugate priors. *J. Roy. Statist. Soc. Ser. B* **45** 278–286.

DEROBERTIS, L. and HARTIGAN, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9** 235–244.

DIACONIS, P. and YLVISAKER, D. (1985). Quantifying prior opinion. In *Bayesian Statistics 2. Proceedings of the Second Valencia Meeting* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 133–156. North-Holland, Amsterdam.

DICKEY, J. M. and CHEN, C.-H. (1985). Direct subjective-probability modelling using ellipsoidal distributions. In *Bayesian Statistics 2. Proceedings of the Second Valencia Meeting* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 157–182. North-Holland, Amsterdam.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. Griffin, London.

HUBER, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Inst. Internat. Statist.* **45** 181–191.

LAVINE, M. (1991a). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.* **86** 396–399.

LAVINE, M. (1991b). An approach to robust Bayesian analyses for multidimensional parameter spaces. *J. Amer. Statist. Assoc.* **86** 400–403.

MORENO, E. and CANO, J. A. (1991). Robust Bayesian analysis with ε-contaminations partially known. *J. Roy. Statist. Soc. Ser. B* **53** 143–156.

SIVAGANESAN, S. and BERGER, J. O. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17** 868–889.

WASSERMAN, L. A. (1989). Robust Bayesian interpretation of likelihood regions. *Ann. Statist.* **17** 1387–1393.

DEPARTMENT OF STATISTICS
GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC 20052