The Annals of Statistics
1989, Vol. 17, No. 2, 868–889

RANGES OF POSTERIOR MEASURES FOR PRIORS WITH UNIMODAL CONTAMINATIONS[1]

BY S. SIVAGANESAN AND JAMES O. BERGER

University of Cincinnati and Purdue University

We consider the problem of robustness or sensitivity of given Bayesian posterior criteria to specification of the prior distribution. Criteria considered include the posterior mean, variance and probability of a set (for credible regions and hypothesis testing). Uncertainty in an elicited prior, π_0, is modelled by an ε-contamination class $\Gamma = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q,\ q \in Q\}$, where ε reflects the amount of probabilistic uncertainty in π_0, and Q is a class of allowable contaminations. For $Q = \{$all unimodal distributions$\}$ and $Q = \{$all symmetric unimodal distributions$\}$, we determine the ranges of the various posterior criteria as π varies over Γ.

1. Introduction.

1.1. *The problem and motivation.* We observe X having density $f(x|\theta)$, and desire to perform a Bayesian analysis concerning the unknown real parameter θ. This requires specification of the prior distribution. Whether or not it is even conceptually possible to quantify exactly prior information in terms of a single distribution, time and other constraints introduce a degree of arbitrariness in the elicitation process. Thus, after an elicitation process which has led to a prior π_0, it is plausible that any prior "close" to π_0 would also be a reasonable representation of prior beliefs, and, that one should be "robust" with respect to such reasonable changes in π_0. [See Berger (1984, 1985) and Berger and Berliner (1986) for further motivation.] In this article we model "close" through the ε-contamination class

$$\Gamma = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q \colon q \in Q\}; \tag{1.1}$$

here ε determines the amount of probabilistic deviation from π_0 that is allowed, and Q is the class of allowed contaminations (see Section 1.3). In Section 1.2 we briefly indicate reasons for considering this class.

A natural goal of a robustness investigation is to find the range of the posterior quantity, $\rho(x, \pi)$, that is of interest, as π varies over Γ. Thus we will seek

$$\underline{\rho}(x, \pi) = \inf_{\pi \in \Gamma} \rho(x, \pi) \quad \text{and} \quad \bar{\rho}(x, \pi) = \sup_{\pi \in \Gamma} \rho(x, \pi). \tag{1.2}$$

Quantities that will be considered include the posterior mean, the posterior variance and the posterior probability of a set (allowing for credible sets or tests). If the range of the posterior quantity is small, then one can be assured of

---

Received February 1987; revised July 1988.
[1]Research supported by National Science Foundation Grants DMS-84-01996 and DMS-87-02620.
AMS 1980 *subject classifications*. Primary 62F15; secondary 62F35, 62C10.
Key words and phrases. Bayesian robustness, ε-contamination classes of priors, posterior mean, posterior variance, posterior probability.

robustness with respect to the elicitation process, providing Γ is large enough to reflect the possible uncertainty in π_0. If the range is large, one does not have robustness with respect to Γ, but our results provide indications as to which features of $\pi \in \Gamma$ are causing the nonrobustness, allowing for further elicitation or refinement of these features.

We make no attempt here to define what is a "small" or a "large" posterior range, that is, to define when one does or does not have posterior robustness. This is a problem-specific judgement. The idea, however, is simple: If the range is clearly so small that the differences between the various priors in Γ are irrelevant, then one can use π_0 with assurance, while if the range is not clearly small enough then further investigation is needed. It is not our purpose here to discuss how "further investigations" should be performed.

Bayesian robustness (or sensitivity) studies with respect to the prior have generally been carried out on an ad hoc basis (try a few different priors) because of the perceived technical difficulties in carrying out the minimization and maximization in (1.2) over realistically large classes of priors; DeRobertis and Hartigan (1981) is an important exception. We show that for the ε-contamination class the problem is often tractable and yields relatively simple answers. We hope that these techniques can lead to automatic checks for robustness with respect to the prior. Note, of course, that robustness with respect to the model is typically at least as important a concern. Thus our results provide at best one component of the overall study of robustness.

1.2. *History*. The "robust Bayesian" view alluded to above has been espoused (in various versions) by many statisticians; cf. Good (1983), Dempster (1975), Rubin (1977), Kadane and Chuang (1978), Hill (1980), Manski (1981), Wolfenson and Fine (1982), Berger (1984, 1985) (which contain general review and discussion) and Walley (1986). We discuss here only some of the articles directly related to our work.

Previous work on finding ranges of posterior measures has mainly considered classes of conjugate priors having parameters in certain ranges. Recent examples include Leamer (1978, 1982) and Polasek (1985), who call the endeavor "global sensitivity." While interesting, classes of conjugate priors are quite small and leave out many priors which are reasonable (such as priors with different tails than π_0), and against which it would be desirable to ensure robustness. [Similar comments can be made about classes based on moments, together with linear estimates, cf. Hartigan (1969) and Goldstein (1980).]

Several articles which do deal with large classes of priors are especially noteworthy. Huber (1973) determines the range of the posterior probability of a set for the class in (1.1) with $Q = \{\text{all distributions}\}$. DeRobertis and Hartigan (1981), in a breakthrough article, consider a class of priors specified by a type of upper and lower envelope on the prior density, and find ranges of general posterior quantities. DeRobertis (1978), Berliner and Goel (1986), Berger and O'Hagan (1988) and O'Hagan and Berger (1988) find the range of the posterior probability of certain sets over classes of priors with specified quantiles. West (1977) and Lambert and Duncan (1986) also have related analyses.

The main motivation for considering the ε-contamination class in (1.1) is that it easily lends itself to automatic checks for robustness with respect to the prior of standard Bayesian analysis. In other words, after specification of π_0 and the model and performance of a standard Bayesian analysis, one could automatically carry out a check of robustness with respect to π by, say, presenting the range of the desired posterior quantity as a function of ε in (1.1). (Q could be chosen in any of several automatic ways.) Note that other classes of priors require additional subjective specification; for instance, the DeRobertis and Hartigan (1981) class requires specification of "density bands" about π_0, and to ensure a rich enough range of tail behavior some thought must be given to the choice of the bands. Of course, additional thought is not a bad thing, but automatic checks have a certain pragmatic attractiveness.

The ε-contamination class of priors has also been utilized in other types of Bayesian robustness studies, including Schneeweiss (1964), Blum and Rosenblatt (1967), Bickel (1984), Marazzi (1985), Berger (1982) and Berger and Berliner (1986). This last article is primarily concerned with maximizing the marginal density, over π in Γ, and thus determining the "ML-II" prior. The mathematics used there is a simple version of that needed here. Also related is Edwards, Lindman and Savage (1963), Berger and Sellke (1987), Casella and Berger (1987), Berger and Delampady (1987) and Delampady (1986), which carry out the determination of the range of the posterior probability of a hypothesis when $\varepsilon = 1$ in (1.2) (i.e., when there is no specified subjective prior π_0). Because of the drastic differences that can arise in testing between Bayesian and classical measures, and because of the frequent lack of "objective" priors in such testing problems, they provide a particularly attractive domain for the application of robust Bayesian methodology.

1.3. *The choice of Q.* We alluded earlier to the choice $Q = \{\text{all distributions}\}$ made in Huber (1973). This choice is particularly easy to work with, and Sivaganesan (1988) extends Huber's results to deal also with the posterior mean and variance. The resulting class is attractive in that it certainly contains any prior "close" to π_0, so that if robustness obtains one is done.

Unfortunately, as pointed out in Berger and Berliner (1986), the range of the posterior quantity of interest will often be excessively large when $Q = \{\text{all distributions}\}$ is used, because this Q contains many unreasonable distributions (such as point masses which are far from π_0). Indeed, it is argued therein that more reasonable Q, when π_0 is unimodal, are the classes of all unimodal distributions (with the same mode as π_0) and the class of all symmetric unimodal distributions. These classes allow wide variation in the functional form and tails of $\pi \in \Gamma$, while retaining the overall shape features of π_0; this overall shape is often rather confidently known, so that it is not desirable to allow priors into Γ which have a very different shape. The ranges of posterior measures are substantially smaller for these classes, and a lack of robustness is thus much more likely to be indicative of a real problem. Section 2 deals with the symmetric unimodal class, and Section 3 with the unimodal class.

1.4. *Formulas and notation.* We will be working only with the observed likelihood function, $f(x|\theta)$, considered as a function θ, and, to emphasize that it is a function of θ, we will write it $f_x(\theta)$. We also assume that the base prior π_0 is unimodal with mode at θ_0 and density (with respect to Lebesgue measure) $\pi_0(\theta)$, and that the contamination q has density $q(\theta)$ with respect to Lebesgue measure; thus any $\pi \in \Gamma$ has a density of the form

$$\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta).$$

Using the notation $m(x|\pi)$ for the marginal distribution of X with respect to the prior π, namely

$$m(x|\pi) = \int f_x(\theta)\pi(\theta)\, d\theta,$$

and assuming all quantities in question exist, we get by simple computation

$$m(x|\pi) = (1 - \varepsilon)m(x|\pi_0) + \varepsilon m(x|q). \tag{1.3}$$

Also, the posterior density of θ with respect to π is

$$\pi(\theta|x) = \lambda(x)\pi_0(\theta|x) + (1 - \lambda(x))q(\theta|x), \tag{1.4}$$

where $\pi_0(\theta|x)$ and $q(\theta|x)$ are the posterior densities with respect to π_0 and q, respectively, and $\lambda(x) \in [0,1]$ is given by

$$\lambda(x) = \frac{(1 - \varepsilon)m(x|\pi_0)}{m(x|\pi)}. \tag{1.5}$$

[Note that $\lambda(x)$ could be thought of as the posterior probability that π_0 is the true prior, if a priori it was believed that π_0 or q were true with probabilities $(1 - \varepsilon)$ and ε, respectively.] Furthermore, the posterior mean δ^π and posterior variance V^π with respect to π can be written (when they exist) as

$$\delta^\pi(x) = \lambda(x)\delta^{\pi_0}(x) + (1 - \lambda(x))\delta^q(x) \tag{1.6}$$

and

$$\begin{aligned} V^\pi(x) &= \lambda(x)V^{\pi_0}(x) + (1 - \lambda(x))V^q(x) \\ &\quad + \lambda(x)(1 - \lambda(x))(\delta^{\pi_0}(x) - \delta^q(x))^2. \end{aligned} \tag{1.7}$$

Finally, if C is a measurable subset of the parameter space Θ, then the posterior probability of C with respect to π is given by

$$P^{\pi(\theta|x)}(\theta \in C) = \lambda(x)P^{\pi_0(\theta|x)}(\theta \in C) + (1 - \lambda(x))P^{q(\theta|x)}(\theta \in C). \tag{1.8}$$

In most of what follows Θ will be the whole real line R, and $C_0(R)$ will denote the set of all continuous real-valued functions vanishing at infinity. Cases where Θ is a subset of R can be similarly handled. As a last note, $U(a, b)$ will be used to denote the uniform distribution on the interval (a, b), and φ and Φ will stand for the standard normal density and c.d.f., respectively.

2. Symmetric unimodal contaminations.

2.1. *Introduction.* In view of the prior beliefs, it may often be natural to require that the contaminations be unimodal and symmetric. This would be particularly desirable when the base prior π_0 is also symmetric and unimodal. Thus, we define

$$Q = \{\text{all symmetric unimodal distributions with the same mode, } \theta_0, \text{ as that of } \pi_0\}, \tag{2.1}$$

and consider the class Γ given in (1.1). In Section 2.3 we find the ranges of posterior quantities such as the mean and probability of a set; in Section 2.4 we consider the range of the posterior variance; and in Section 2.5 we consider hypothesis testing. Applications to the normal distribution are given.

2.2. *Preliminaries.* The following lemma forms the basis of dealing with (2.1), and will be repeatedly used. The proof is standard (being based on representing a symmetric unimodal density as a mixture of symmetric uniforms), and will be omitted.

LEMMA 2.2.1. *For $q \in Q$ as in* (2.1) *and any g such that $\int |g(\theta)| f_x(\theta) q(\theta)\, d\theta < \infty$,*

$$\int g(\theta) f_x(\theta) q(\theta)\, d\theta = \int_0^\infty H^g(z)\, dF(z),$$

where F is some distribution function (the mixing distribution which yields q), and

$$H^g(z) = \begin{cases} \dfrac{1}{2z} \displaystyle\int_{\theta_0 - z}^{\theta_0 + z} g(\theta) f_x(\theta)\, d\theta & \text{if } z \neq 0, \\ g(\theta_0) f_x(\theta_0) & \text{if } z = 0. \end{cases} \tag{2.2}$$

The dependence of these quantities on x is suppressed for notational simplicity. Examples of g that will be considered include $g_0(\theta) \equiv 1$, $g_1(\theta) = \theta$, $g_2(\theta) = (\theta - \mu_0)^2$, $g(\theta) = I_C(\theta)$; here μ_0 is a constant and $I_C(\theta)$ is the indicator function on the set C. We always assume that $\int |g(\theta)| f_x(\theta) \pi(\theta)\, d\theta < \infty$. The $H^g(z)$ corresponding to these functions will be denoted by $H_0(z)$, $H_1(z)$, $H_2(z)$ and $H_C(z)$, for simplicity. Note that $H_0(z) = m(x|q)$.

2.3. *Range of the posterior expected value of $g(\theta)$.* Let Γ be as in (1.1) and $\Gamma_1 \subset \Gamma$ be given by

$$\Gamma_1 = \{\pi = (1 - \varepsilon)\pi_0 + \varepsilon q \colon q \text{ is } U(\theta_0 - z, \theta_0 + z) \text{ for some } z > 0\}. \tag{2.3}$$

In order to find the range of the posterior expected value of $g(\theta)$ over the large class Γ; it is in fact sufficient to do the maximization and minimization over the much smaller (and simpler) class Γ_1 (as shown in the following theorem), thus reducing the problem to that of finding the extrema of a function of one variable.

THEOREM 2.3.1. *Let $\rho^\pi(x)$ denote the posterior expected value of $g(\theta)$ with respect to the prior π. Then, for Γ and Γ_1 as above,*

$$\sup_{\pi\in\Gamma}\rho^\pi(x) = \sup_{\pi\in\Gamma_1}\rho^\pi(x) = \sup_z \frac{a_0 + H^g(z)}{a + H_0(z)}$$

and

$$\inf_{\pi\in\Gamma}\rho^\pi(x) = \inf_{\pi\in\Gamma_1}\rho^\pi(x) = \inf_z \frac{a_0 + H^g(z)}{a + H_0(z)},$$

where $a = (1-\varepsilon)m(x|\pi_0)/\varepsilon$, $a_0 = a\rho^{\pi_0}(x)$ and $H_0(z)$, $H^g(z)$ are as in the previous section.

PROOF. For $\pi \in \Gamma$, $\rho^\pi(x)$ can be written as

$$\rho^\pi(x) = \frac{a_0 + \int H^g(z)\,dF(z)}{a + \int H_0(z)\,dF(z)} \tag{2.4}$$

for some distribution function $F(\cdot)$. Now the result follows by an application of Lemma A.1 of the Appendix. □

EXAMPLE 2.3.1 (Range of the posterior mean). For $g(\theta) = \theta$, and hence $H^g(z) = H_1(z)$, Theorem 2.3.1 gives the range of the posterior mean. As an example, let $X|\theta \sim N(\theta, \sigma^2)$ and $\pi_0(\theta)$ be $N(\mu, \tau^2)$. Then

$$a = \left(\frac{1-\varepsilon}{\varepsilon}\right)\frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{(\sigma^2+\tau^2)}\right\}$$

and

$$a_0 = a\rho^{\pi_0}(x) = a\left(\frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x\right).$$

Furthermore,

$$H_0(z) = \frac{1}{2z}\int_{\mu-z}^{\mu+z}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\theta-x)^2}{2\sigma^2}\right)d\theta$$

and

$$H_1(z) = \frac{\sigma}{2z}\left(\frac{e^{-(\mu-z-x)^2/2\sigma^2}}{\sqrt{2\pi}} - \frac{e^{-(\mu+z-x)^2/2\sigma^2}}{\sqrt{2\pi}}\right) + xH_0(z).$$

Letting

$$t = \varphi\left(\frac{\mu+z-x}{\sigma}\right) + \varphi\left(\frac{\mu-z-x}{\sigma}\right),\quad u = \varphi\left(\frac{\mu+z-x}{\sigma}\right) - \varphi\left(\frac{\mu-z-x}{\sigma}\right)$$

and

$$v = \Phi\left(\frac{\mu+z-x}{\sigma}\right) - \Phi\left(\frac{\mu-z-x}{\sigma}\right),$$

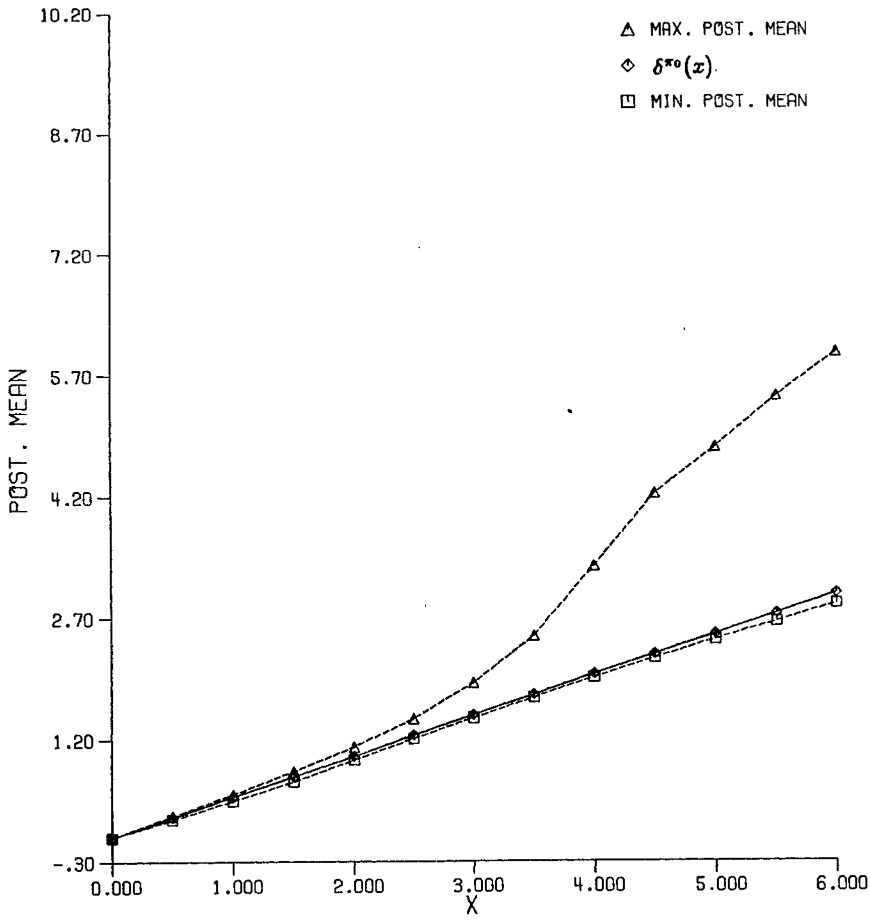

FIG. 1. *Graphs of range of the posterior mean (symmetric unimodal contaminations) and* $\delta^{\pi_0}(x)$ *against* x.

the values of z which maximize and minimize $(a_0 + H_1(z))/(a + H_0(z))$ are given by the solutions of the equation (obtained by differentiating the above)

$$z = \frac{(vx - \sigma u)(t + 2\sigma a) - v(2a_0\sigma + t\mu)}{2[auz + t(a\mu - a_0) - vu/2]}.$$

This equation may be iteratively solved for z by taking a number larger than $\delta^{\pi_0}(x)$ as the initial value of z when maximizing, and a number smaller than $\delta^{\pi_0}(x)$ as the initial value of z when minimizing.

As a specific example, let $\mu = 0$, $\sigma^2 = 1$, $\tau^2 = 1$ and $\varepsilon = 0.1$. In Figure 1, the ranges of $\delta^{\pi}(x)$ and the values of $\delta^{\pi_0}(x)$, for various values of x, are displayed. It can be seen from Figure 1 that the range of $\delta^{\pi}(x)$ is fairly small for small values of $x(|x| \leq 3)$ but is larger for large values of x. (Recall that our viewpoint here is a posterior viewpoint; we are imagining use of the methodology to find the possible range of the posterior criteria *after* the data is at hand. We present the range here as a function of x only because of several technical points we wish to make in Section 4.)

EXAMPLE 2.3.2 (Range of the posterior probability of a set). When constructing a credible set C for an unknown parameter θ it is of interest to find the range of the posterior probability of C as π varies over Γ. This can be done using Theorem 2.3.1 with $g(\theta) = I_C(\theta)$, $H^g(z) = H_C(z)$ and $a_0 = aP^{\pi_0(\theta|x)}(\theta \in C)$.

Defining $g_x(\theta) = f_x(\theta)I_C(\theta)$, $A(z) = g_x(\theta_0 + z) + g_x(\theta_0 - z)$ and $B(z) = f_x(\theta_0 + z) + f_x(\theta_0 - z)$, the values of z, at which the inf and sup of the probability of C are obtained, can be shown to be among the positive solutions [of which there are at most three when C is an interval and $B(z)$ has at most three local extrema] of the equation (which may be solved iteratively by carefully choosing the initial values)

$$2z(aA(z) - a_0B(z)) = (2a + B(z))\int_{\theta_0 - z}^{\theta_0 + z} g_x(\theta)\,d\theta - (2a_0 + A(z))\int_{\theta_0 - z}^{\theta_0 + z} f_x(\theta)\,d\theta.$$

In rare cases, the extrema may occur at boundary points of C.

As a specific example, let $X|\theta \sim N(\theta, 1)$, $\pi_0(\theta)$ be $N(0, 2)$ and $\varepsilon = 0.1$. When $x = 0.5$, the 95% HPD credible set is $C_0 = (-1.27, 1.93)$. The range of the posterior probability of C_0 is given by

$$\inf_{\pi \in \Gamma} P^{\pi(\theta|0.5)}(\theta \in C_0) = 0.945, \qquad \sup_{\pi \in \Gamma} P^{\pi(\theta|0.5)}(\theta \in C_0) = 0.958,$$

and these are attained when the contaminations are, respectively, the $U(-2.98, 2.98)$ distribution, and a point mass at 0. When $x = 4.0$, the 95% HPD credible set is $C_0 = (1.07, 4.27)$. The range of the posterior probability of C_0 is $(0.830, 0.965)$, the extreme values of which are attained when the contaminations are, respectively, $U(-6.1, 6.1)$ and $U(-4.3, 4.3)$.

2.4. *Range of the posterior variance for fixed posterior mean.* It is typically necessary, in estimation problems, to report also an accuracy measure; here we consider the posterior variance. Since the posterior mean is of primary interest, and since there will be a different range of the posterior variance for each (fixed) value of the posterior mean, it is natural to seek the range of the posterior variance corresponding to each possible value of the posterior mean. Thus, for $\mu_0 \in [\inf_\pi \delta^\pi(x), \sup_\pi \delta^\pi(x)]$, if we let

$$\Gamma_0 = \{\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta): q \in Q \text{ and } \delta^\pi(x) = \mu_0\}, \tag{2.5}$$

we want to find $\sup_{\pi \in \Gamma_0} V^\pi(x)$ and $\inf_{\pi \in \Gamma_0} V^\pi(x)$.

It is shown below that the posterior variance is maximized or minimized, subject to fixed posterior mean, when the contamination is a mixture of two symmetric (about θ_0) uniform distributions. Let

$$\Gamma_{02} = \{\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \in \Gamma_0: q \in Q_2\},$$

where

$$Q_2 = \{q = \alpha U(\theta_0 - z, \theta_0 + z) + (1 - \alpha)U(\theta_0 - z^*, \theta_0 + z^*): 0 \le \alpha \le 1 \text{ and } z, z^* \ge 0\}.$$

Note that one of the three quantities α, z, z^* is determined by the constraint that π have posterior mean μ_0. Hence maximizations over Γ_{02} are effectively two-dimensional maximizations. Note also that z^* (say) could be infinite, so that q would then be a single uniform, possibly with total mass less than 1.

THEOREM 2.4.1.

$$\sup_{\pi \in \Gamma_0} V^\pi(x) = \sup_{\pi \in \Gamma_{02}} V^\pi(x),$$

$$\inf_{\pi \in \Gamma_0} V^\pi(x) = \inf_{\pi \in \Gamma_{02}} V^\pi(x).$$

PROOF. Using (1.6) and (2.4) with $g(\theta) = \theta$, we have, for $\pi \in \Gamma_0$,

$$\int [H_1(z) - \mu_0 H_0(z) - a(\mu_0 - \delta^{\pi_0})] \, dF(z) = 0, \tag{2.6}$$

where F is the mixing distribution function corresponding to q. Furthermore, for $\pi_0 \in \Gamma_0$, we have

$$\begin{aligned} V^\pi(x) &= \frac{\int (\theta - \mu_0)^2 \pi(\theta) f_x(\theta) \, d\theta}{\int \pi(\theta) f_x(\theta) \, d\theta} \\ &= \frac{a V^{\pi_0}(x) + \int H_2(z) \, dF(z)}{a + \int H_0(z) \, dF(z)} \overset{\text{def}}{=} V(F), \end{aligned} \tag{2.7}$$

where $H_2(z)$ is as defined in Section 2.2. Thus, defining

$$\Psi_0 = \{\text{all probability distributions } F \text{ satisfying (2.6)}\},$$

we have

$$\sup_{\pi \in \Gamma_0} V^\pi(x) = \sup_{F \in \Psi_0} V(F). \tag{2.8}$$

Now, letting Ψ_{02} consist of those $F \in \Psi_0$ which are two-point distributions, we have; from a result of Mulholland and Rogers (1958), page 178, that Ψ_0 is the convex hull of Ψ_{02}. Note that Ψ_{02} can be expressed in the form of Φ_I, in Lemma A.1, with I defined by

$$I = \{\tau = (\tau_1, \tau_2) \in R^2: \text{support}(F) = \{\tau_1, \tau_2\} \text{ for some } F \in \Psi_{02}\}.$$

Hence, using Lemma A.1,

$$\sup_{F \in \Psi_0} V(F) = \sup_{F \in \Psi_{02}} V(F).$$

Now, using (2.8) and observing that

$$\sup_{F \in \Psi_{02}} V(F) = \sup_{\pi \in \Gamma_{02}} V^\pi(x)$$

completes the proof for the "sup." The proof for the "inf" is similar. □

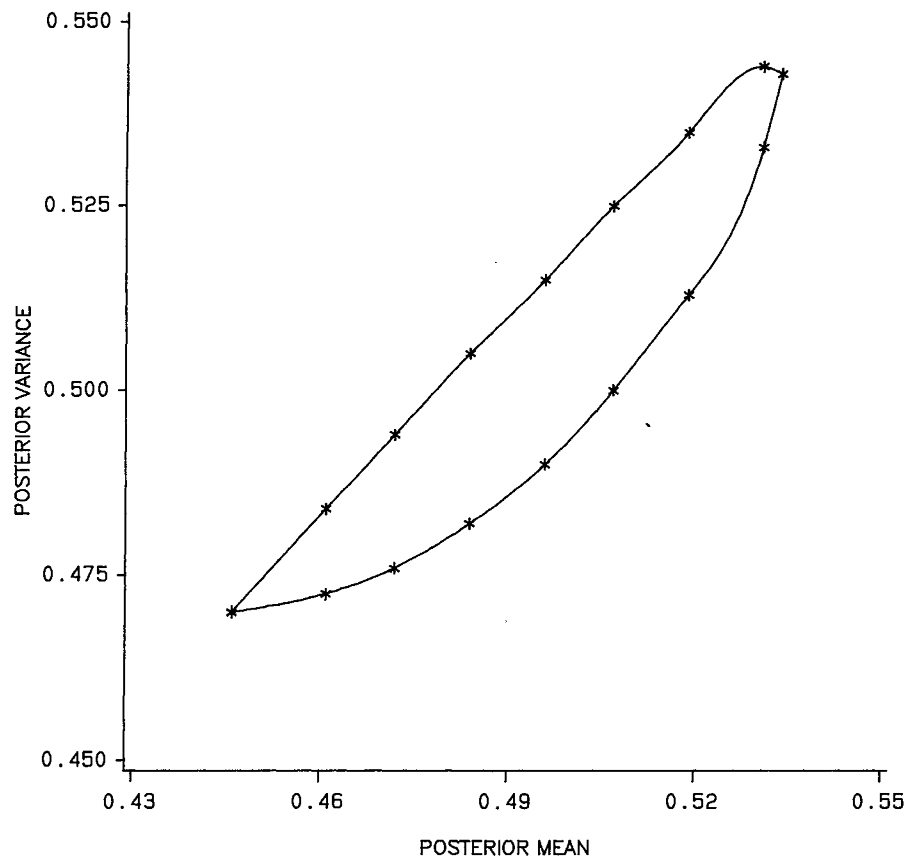

FIG. 2. *Range of posterior variance as posterior mean varies from* 0.45 *to* 0.53: $x = 1.0$.

EXAMPLE 2.4.1. Suppose $X|\theta \sim N(\theta, 1)$, $\pi_0(\theta)$ is $N(\theta, 1)$ and $\varepsilon = 0.1$. Then the range of the posterior variance [calculated by numerical optimization of $V^{\pi}(x)$ over Γ_{02}], when the posterior mean is fixed at various levels, is displayed in Figures 2 and 3 for $x = 1.0$ and $x = 4.0$, respectively.

A stronger characterization of the infimum of the posterior variance. We will show that the infimum of $V^{\pi}(x)$ over Γ_0 can be obtained by minimizing over the smaller class

$$\Gamma_{01} = \{\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \in \Gamma_0 \colon q \equiv \alpha U(\theta_0 - z, \theta_0 + z) \text{ for some } 0 \le \alpha \le 1 \text{ and } z \ge 0\}. \tag{2.9}$$

Here $\alpha U(\theta_0 - z, \theta_0 + z)$ denotes the subprobability distribution having mass α uniformly distributed over the interval $(\theta_0 - z, \theta_0 + z)$. Note that α will be determined by the constraint that $\delta^{\pi} = \mu_0$ (i.e., that $\pi \in \Gamma_0$), so that this will be only a one-dimensional minimization. It is interesting that one must consider the possibility of mass escaping to infinity (i.e., $\alpha < 1$). For use in the following

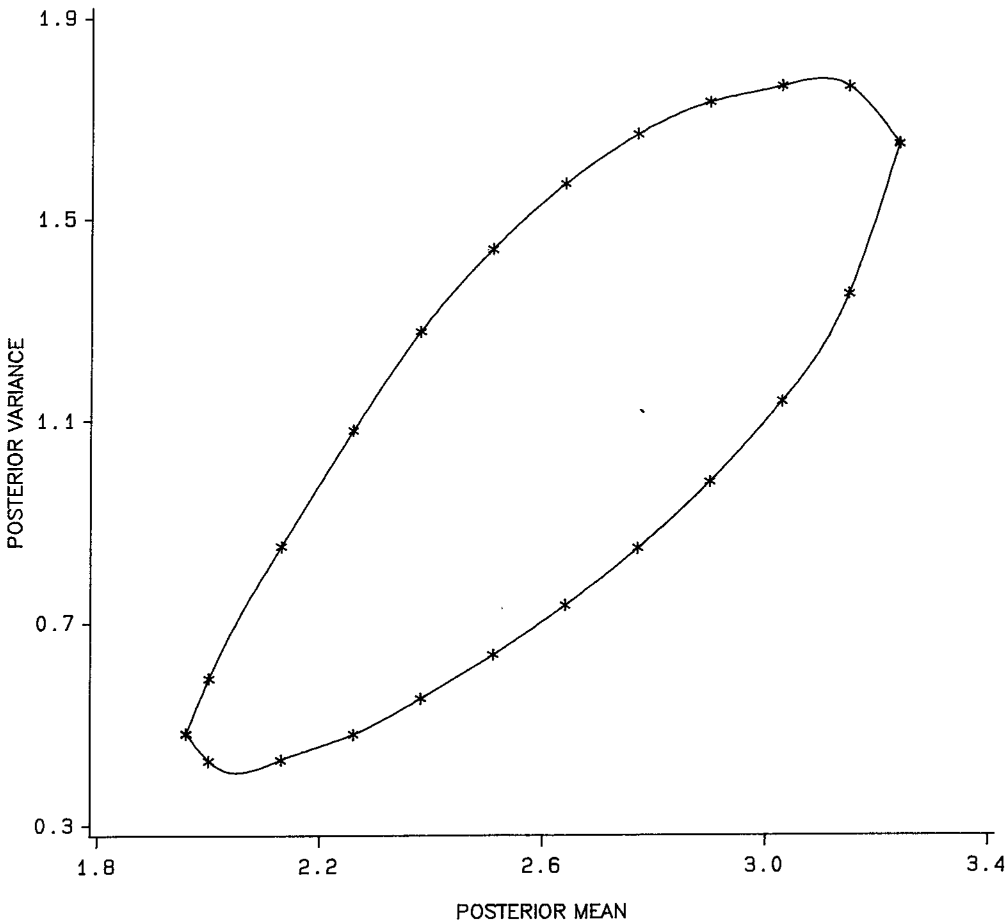

FIG. 3. *Range of posterior variance as posterior mean varies from* 1.96 *to* 3.24: $x = 4.0$.

theorem, define

$$\psi(z) = \frac{H_2(z) - H_0(z)\left[V^{\pi_0} + (\delta^{\pi_0})^2\right]}{H_0(z)\,\delta^{\pi_0} - H_1(z)}.$$

THEOREM 2.4.2. *Suppose that* $f_x(\theta)$ *is such that, if* a *and* b *are real constants, then* $[h(\theta_0 + z) + h(\theta_0 - z)]$ *has at most one local minimum at which the function is positive; here*

$$h(\theta) = (\theta^2 + a\theta + b)f_x(\theta).$$

Then

$$\inf_{\pi \in \Gamma_0} V^{\pi}(x) = \inf_{\pi \in \Gamma_{01}} V^{\pi}(x). \tag{2.10}$$

When $\delta^{\pi_0}(x) \neq \mu_0$, *the infimum is given by*

$$V^{\pi_0} + (\delta^{\pi_0})^2 - \mu^2 + (\delta^{\pi_0} - \mu)\psi(\hat{z}), \tag{2.11}$$

where $\hat{z}$ *is that value of* z *minimizing* $[sgn(\delta^{\pi_0} - \mu)]\psi(z)$ *over the region*

$$\left\{z: \frac{\mu H_0(z) - H_1(z)}{\mu^*} \geq a = \frac{(1-\varepsilon)m(x|\pi_0)}{\varepsilon}\right\}. \tag{2.12}$$

When $\delta^{\pi_0}(x) = \mu_0$, *the infimum is the smaller of* $V^{\pi_0}(x)$ *and*

$$\inf_z \left\{\frac{aV^{\pi_0} + H_2(z) - H_0(z)(\delta^{\pi_0})^2}{a + H_0(z)} : H_1(z) = \delta^{\pi_0}H_0(z)\right\}. \tag{2.13}$$

PROOF. Using the results and notation of the proof of Theorem 2.4.1, we want to minimize $V(F)$ over Ψ_{02}. For $F = \alpha I_{\{z_1\}} + (1-\alpha)I_{\{z_2\}}$ in Ψ_{02}, $V(F)$ and the condition (2.6), using a (1–1) transformation, can be written [writing $V = V(F)$]

$$V = \beta v(z_1) + (1-\beta)v(z_2) \quad \text{and} \quad M \equiv \beta m(z_1) + (1-\beta)m(z_2) = 0,$$

where $\beta = \alpha(a + H_0(z_1))/(a + \alpha H_0(z_1) + (1-\alpha)H_0(z_2))$, $v(z) = V^{\pi}(x)$ and $m(z) = \delta^{\pi}(x) - \mu_0$ for prior $\pi = (1-\varepsilon)\pi_0 + \varepsilon U(\theta_0 - z, \theta_0 + z)$.

Now, using the Lagrange multiplier method, we can show that if z_1, z_2 minimize V subject to $M = 0$, then

$$\inf V = G(z_1) = G(z_2),$$

and z_1, z_2 are both local minima for $G(z)$,

$$G(z) = v(z) + \lambda m(z)$$

and λ is the Lagrange multiplier.

When $\inf V > G(\infty) = V^{\pi_0}(x) + \lambda(\delta^{\pi_0}(x) - \mu_0)$, one can show that, for $F \in \Psi_{02}$ with support$(F) = \{z_1, \infty\}$ or $\{z_2, \infty\}$, $V(F) < \inf V$ leading to a contradiction unless either of z_1 or z_2 is infinite. When $\inf V \leq G(\infty)$ it can be shown, under the condition on $h(\theta)$ stated in the theorem, that at most one of z_1 and z_2 can be finite unless they are equal. This proves (2.10). For more details of these steps, see Sivaganesan and Berger (1986).

For $q \equiv \alpha U(\theta_0 - z, \theta_0 + z)$ to yield $\pi = (1-\varepsilon)\pi_0 + \varepsilon q \in \Gamma_0$, it must be the case that (when $\delta^{\pi_0} \neq \mu_0$)

$$\alpha = \frac{a(\mu_0 - \delta^{\pi_0})}{H_1(z) - \mu_0 H_0(z)}.$$

Using this expression for α, together with (1.7), yields after simplification the expression for $V^{\pi}(x)$ given in (2.12). This expression is then minimized over the set of all allowable z, namely those for which $0 \leq \alpha \leq 1$; this set is equivalent to (2.13). When $\delta^{\pi_0} = \mu_0$ [equivalent to the condition $H_1(z) = \delta^{\pi_0}H_0(z)$ in (2.14)], α can be arbitrary. Then, however, V^{π} is a linear function of λ, and hence a monotonic function of α, so that only $\alpha = 0$ (yielding V^{π_0}) and $\alpha = 1$ [yielding the expression in (2.14)] need be considered. This completes the proof. □

2.5. *Testing of hypotheses.* To test the hypothesis H_0: $\theta \in \Theta_0$ vs. H_1: $\theta \in \Theta \setminus \Theta_0$, suppose it is desired to determine their posterior probabilities. Then, robustness can be investigated by determining the range of these probabilities as π ranges over the class of priors Γ. These ranges can be directly obtained as in Example 2.3.2 when Θ_0 and $\Theta \setminus \Theta_0$ both have positive Lebesgue measure.

Also quite interesting is the testing of point null hypotheses, because of the dramatic discrepancies between classical P-values and posterior probabilities [see Berger and Sellke (1987)]. Thus, suppose we want to test H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$. A typical prior distribution for this problem specifies a point mass, α, to be assigned to θ_0, and a continuous density, $g(\theta)$, to be assigned to $\{\theta \neq \theta_0\}$. We assume that $|\alpha - \alpha_0| \leq \delta$, and that g is of the form $g(\theta) = (1-\varepsilon)g_0(\theta) + \varepsilon q(\theta)$; thus α is specified as α_0, but could be in error by an amount δ, and g is within a certain ε-contamination class of the elicited g_0. Specifically, we consider

$$G = \{g(\theta) = (1-\varepsilon)g_0(\theta) + \varepsilon q(\theta) \colon q \text{ is symmetric unimodal with mode } \theta_1\};$$

here θ_1 is the (assumed unique) mode of g_0, not necessarily equal to θ_0. Let Γ_0 denote the implied class of priors π. The range of the posterior probabilities of H_0, as π varies over Γ_0, is given in the following theorem, the proof of which is straightforward and is omitted.

THEOREM 2.5.1.

$$\inf_{\pi \in \Gamma_0} P^{\pi(\theta|x)}(H_0) = \frac{f_x(\theta_0)}{f_x(\theta_0) + \left(\dfrac{1-(\alpha_0-\delta)}{\alpha_0-\delta}\right)\left((1-\varepsilon)m(x|g_0) + \varepsilon\hat{H}\right)}$$

and

$$\sup_{\pi \in \Gamma_0} P^{\pi(\theta|x)}(H_0) = \frac{f_x(\theta_0)}{f_x(\theta_0) + \left(\dfrac{1-(\alpha_0+\delta)}{\alpha_0+\delta}\right)(1-\varepsilon)m(x|g_0)},$$

where $\hat{H} = \sup_{z>0} H_0(z)$, *with* θ_0 *being replaced by* θ_1 *in the definition of* H_0.

EXAMPLE 2.5.1. Let $X|\theta \sim N(\theta, 1)$, $\theta_0 = 0$, $g_0(\theta)$ be $N(\theta_1, 1)$, $\alpha_0 = \frac{1}{2}$, $\delta = 0.1$ and $\varepsilon = 0.1$. Then the range of the posterior probability of H_0, for various values of θ_1 and x, is given in Table 1. When $x \leq 1.5$, the uncertainty in $P^{\pi(\theta|x)}(H_0)$ is

TABLE 1
Range of the posterior probability of H_0

x	$\theta_1 = 0$	$\theta_1 = 1.0$	$\theta_2 = 2.0$
0.5	(0.46, 0.69)	(0.46, 0.68)	(0.59, 0.78)
1.0	(0.42, 0.65)	(0.35, 0.59)	(0.42, 0.65)
1.5	(0.35, 0.57)	(0.24, 0.45)	(0.24, 0.45)
4.0	(0.004, 0.007)	(0.002, 0.007)	(0.0008, 0.002)

almost entirely due to the uncertainty in the prior probability of H_0. When $x = 4.0$, the uncertainty in g also contributes significantly to that of $P^{\pi(\theta|x)}(H_0)$.

3. Unimodal contaminations.

3.1. *Introduction.* When π_0 is not symmetric about its mode θ_0 (and even in some situations when it is), it may be desired to drop the symmetry assumption on q that was made in Section 2. Then, the class of prior distributions is given by (1.1), where

$$Q = \{\text{all distributions which are unimodal with the same mode, } \theta_0, \text{ as that of } \pi_0\}.$$

Here we present the analogs of the results in Section 2 for this class. Proofs are similar and hence are omitted.

3.2. *Preliminaries.*

LEMMA 3.2.1. *As in Lemma* 2.2.1, *for* $q \in Q$ *from above, and any* g *such that* $\int |g(\theta)| f_x(\theta) q(\theta)\, d\theta < \infty$,

$$\int g(\theta) f_x(\theta) q(\theta)\, d\theta = \int_{-\infty}^{\infty} H^g(z)\, dF(z),$$

where F *is some distribution function and*

$$H^g(z) = \begin{cases} \dfrac{1}{z}\displaystyle\int_{\theta_0}^{\theta_0+z} g(\theta) f_x(\theta)\, d\theta & \text{if } z \neq 0, \\ g(\theta_0) f_x(\theta_0) & \text{if } z = 0. \end{cases}$$

We will consider the same examples of g as in Section 2. For simplicity, we therefore adopt the same notation, viz., H_0, H_1, H_2 and H_C, for the corresponding H^g's as described in Section 2.2.

3.3. *Range of the posterior expected value of* $g(\theta)$. Let Γ be as in (1.1), and $\Gamma_1 \subseteq \Gamma$ be given by

$$\Gamma_1 = \{\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \colon q(\theta) \text{ is } U(\theta_0, \theta_0 + z) \text{ or } U(\theta_0 - z, \theta_0) \text{ for some } z > 0\}.$$

Then, Theorem 2.3.1 is also valid here, with H^g, H_0, Γ and Γ_1 as defined above.

EXAMPLE 3.3.1 (Range of the posterior mean). We can find the range of the posterior mean as in Example 2.3.1 with H_1 and H_0 given as in Section 3.2. As an example, let $X|\theta \sim N(\theta, \sigma^2)$ and $\pi_0(\theta)$ be $N(\mu, \tau^2)$. Then a, a_0 are as given in

Example 2.3.1 and

$$H_0(z) = \frac{1}{z}\int_{\mu}^{\mu+z} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\theta - x)^2}{2\sigma^2} \right) d\theta,$$

$$H_1(z) = \frac{\sigma}{z\sqrt{2\pi}} \left(\exp\left\{ -\frac{(\mu - x)^2}{2\sigma^2} \right\} - \exp\left\{ -\frac{(\mu + z - x)^2}{2\sigma^2} \right\} \right) + xH_0(z).$$

Letting

$$t = \varphi\left(\frac{\mu - x}{\sigma}\right) - \varphi\left(\frac{z + \mu - x}{\sigma}\right),\ u = \Phi\left(\frac{z + \mu - x}{\sigma}\right) - \Phi\left(\frac{\mu - x}{\sigma}\right)$$

and

$$v = \varphi\left(\frac{z + \mu - x}{\sigma}\right),$$

the values of z which maximize and minimize $(a_0 + H_1(z))/(a + H_0(z))$ are given by the solutions of the equation (obtained by differentiating the above)

$$z = \frac{(\sigma t + ux)(\sigma a + v) - q(\sigma a_0 + r\mu)}{v(za + a\mu + u - a_0)}.$$

As a specific example, let $\mu = 0$, $\sigma^2 = 1$, $\tau^2 = 1$ and $\varepsilon = 0.1$. In Figure 4 the ranges of $\delta^{\pi}(x)$ and the values of $\delta^{\pi_0}(x)$ for various values of x are displayed. Note that the ranges are larger than those in Figure 1, as would be expected.

EXAMPLE 3.3.2 (Range of the posterior probability of a set). The range of the posterior probability of a measurable set C, as π varies over Γ, can be obtained as in Example 2.3.2 with H_C and H_0 given as in Section 3.2. As an example, let $X|\theta \sim N(\theta, 1)$, $\pi_0(\theta)$ be $N(0, 2)$ and $\varepsilon = 0.1$. When $x = 0.5$, the 95% HPD credible set for θ is $C_0 \equiv (-1.27, 1.93)$, and the range of the posterior probability of C_0 is

$$\inf_{\pi\in\Gamma} P^{\pi(\theta|0.5)}(\theta \in C_0) = 0.94, \qquad \sup_{\pi\in\Gamma} P^{\pi(\theta|0.5)}(\theta \in C_0) = 0.96.$$

These are attained when the contaminations are, respectively, $U(0,3)$ and $U(0, 0.75)$. When $x = 4.0$, the 95% HPD credible set for θ is $C_0 \equiv (1.07, 4.27)$, and the range of the posterior probability of C_0 is

$$\inf_{\pi\in\Gamma} P^{\pi(\theta|4.0)}(\theta \in C_0) = 0.77, \qquad \sup_{\pi\in\Gamma} P^{\pi(\theta|4.0)}(\theta \in C_0) = 0.97.$$

These are attained when the respective contaminations are $U(0, 6.3)$ and $U(0, 0.97)$.

3.4. *Range of the posterior variance for fixed posterior mean.* Let Γ_0 and Γ_{02} be as in Section 2.4, but with Q as in (3.1) and

$$Q_2 = \{q = \alpha U_z + (1 - \alpha)U_{z^*}: 0 \le \alpha \le 1 \text{ and } z, z^* \ge 0\},$$

with U_y representing a uniform distribution of the form $U(\theta_0, \theta_0 + y)$ or $U(\theta_0 - y, \theta_0)$. Then, Theorem 2.4.1 is valid with Γ_0 and Γ_{02} as defined above.

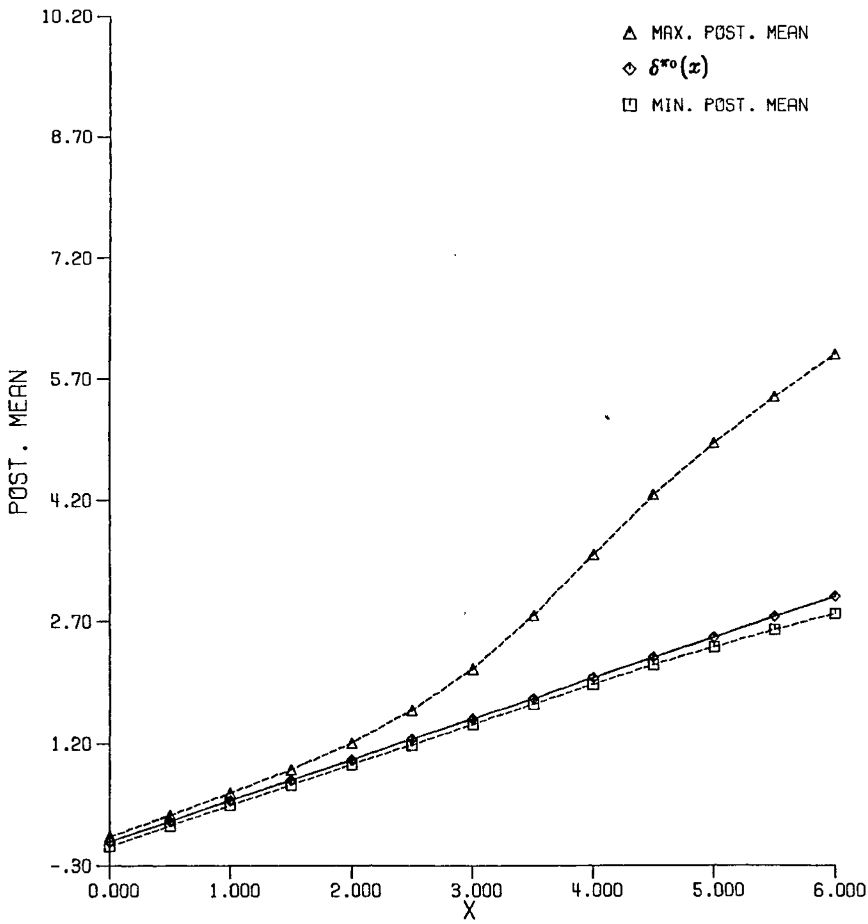

FIG. 4. *Graphs of range of the posterior mean (unimodal contaminations) and $\delta^{\pi_0}(x)$ against x.*

Furthermore, when $1/f_x(\theta)$ is convex in θ the infimum, as in Section 2.4, can be expressed as the minimum of a function of one variable over a specified range.

EXAMPLE 3.4.1. Let $X|\theta \sim N(\theta, 1)$, $\pi_0(\theta)$ be $N(0, 1)$ and $\varepsilon = 0.1$. Ranges of the posterior variance, when the posterior mean is held fixed at different values, are displayed in Figures 5 and 6 for $x = 1.0$ and $x = 4.0$, respectively.

3.5. *Testing hypotheses.* The discussion in Section 2.5 applies here, though now we will constrain g to be in

$$G = \{g(\theta) = (1 - \varepsilon)g_0(\theta) + \varepsilon q(\theta) \colon q \text{ is unimodal with mode } \theta_1\}.$$

Then Theorem 2.5.1 is valid here, with $\hat{H}$ replaced by $\sup_z H_0(z)$.

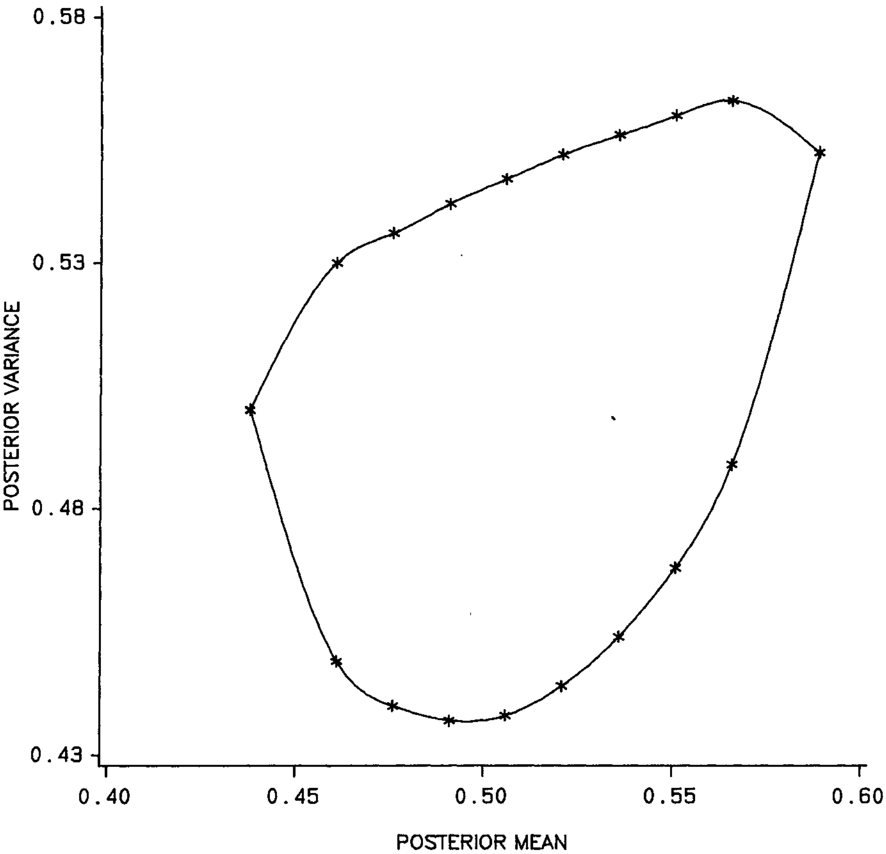

FIG. 5. *Range of posterior variance as posterior mean varies from* 0.44 *to* 0.59: $x = 1.0$.

EXAMPLE 3.5.1. As a specific example, let $X|\theta \sim N(\theta, 1)$, $\theta_0 = 0$, $g_0(\theta)$ be $N(\theta_1, 1)$, $\alpha_0 = \frac{1}{2}$, $\delta = 0.1$ and $\varepsilon = 0.1$. Then the range of the posterior probability of H_0, for various values of θ_1 and x, is given in Table 2.

4. Discussion. As is clear from the figures given earlier, the degree of robustness present, in any given situation, can depend heavily on the observed value of x. The large ranges (of the posterior quantities) that we observed for large x were due to our choice of π_0 as normal; the resulting Γ contained priors with tails ranging from normal to uniform, and robustness is generally lacking when there is such a wide variety of tails and the likelihood function is located in the tail of the prior. Had we chosen π_0 to be, say, Cauchy the ranges for large x would have been much smaller (and indeed go to 0 as $|x| \to \infty$); note that the tail can, in some sense, never get sharper than that of $(1 - \varepsilon)\pi_0(\theta)$. Whether or not one can rule out exponential tails, however, is a subjective decision, although a large difference between x and θ_0 does indicate that sharp tails for the overall

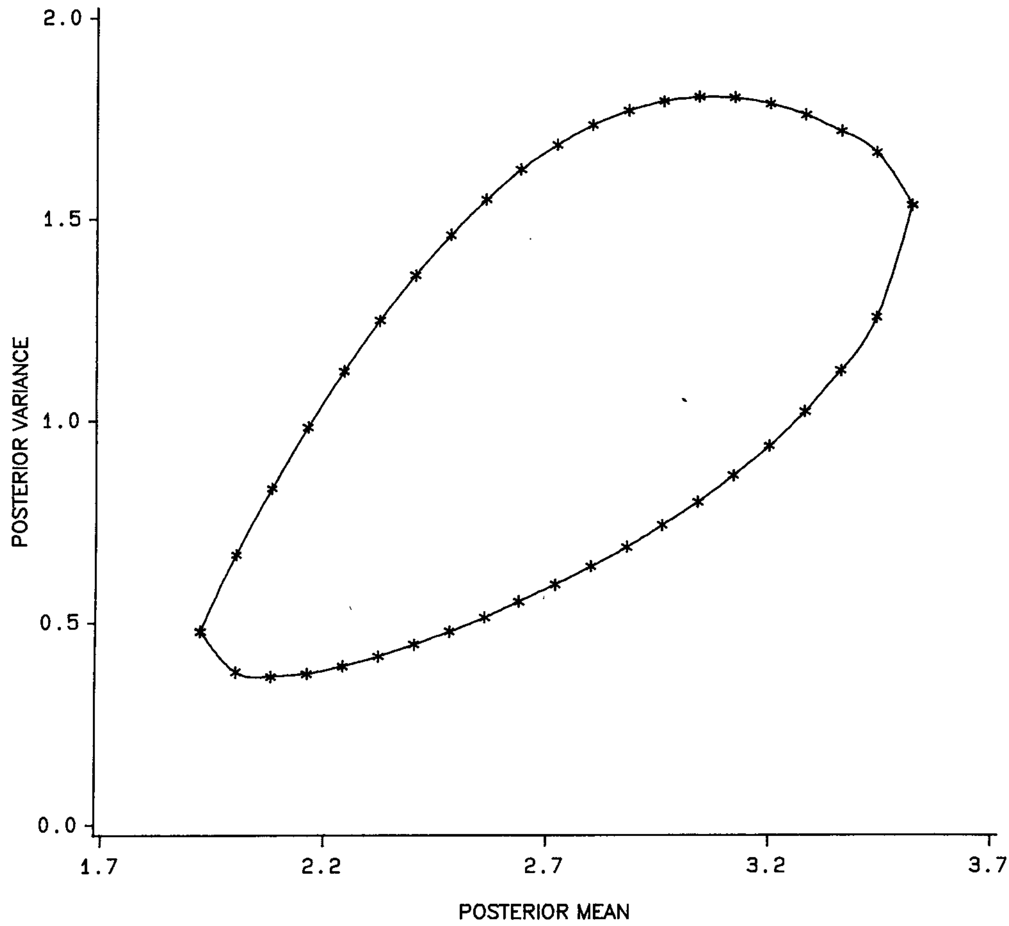

FIG. 6. *Range of posterior variance as posterior mean varies from* 1.92 *to* 3.52: $x = 4.0$.

TABLE 2
Posterior probabilities of H_0

x	$\theta_1 = 0$	$\theta_1 = 1.0$	$\theta_1 = 2.0$
0.5	(0.46, 0.69)	(0.46, 0.69)	(0.57, 0.79)
1.0	(0.41, 0.65)	(0.35, 0.59)	(0.41, 0.65)
1.5	(0.35, 0.57)	(0.24, 0.45)	(0.24, 0.45)
4.0	(0.004, 0.04)	(0.002, 0.007)	(0.0007, 0.002)

prior may not be very appropriate [see Berger (1985)], although the large difference could, of course, be due to an "outlying" x.

As mentioned in Section 1.3, an ε-contamination class can be "too big," in the sense of containing unreasonable priors which artificially inflate the ranges of the posterior criteria. We mentioned that choosing $Q = \{\text{all distributions}\}$ is generally "too big." Further evidence of this comes from noting that, if $X_1, \dots, X_n$ are i.i.d. $N(\theta, 1)$ and $Q = \{\text{all distributions}\}$ is used, then $|\sup_{\pi \in \Gamma} \delta^{\pi}(x) - \bar{x}|$ converges to 0 as $n \to \infty$ at the rate of $\sqrt{\ln(n)/n}$. For the classes we consider in this article, the rate of convergence can be shown to be the correct rate $\sqrt{1/n}$ [see Sivaganesan (1986)], correct in the sense that this is the usual rate at which single posterior distributions converge.

A reasonable alternative to the choices considered here is the choice $Q = \{$all distributions such that the resulting $\pi = (1 - \varepsilon)\pi_0 + \varepsilon q$ is unimodal$\}$. This was considered for the ML-II problem in Berger and Berliner (1986), and for the posterior mean in Sivaganesan (1987). Besides being substantially more difficult to work with, there is some indication in Berger and Berliner (1986) that this class might also be too big. The classes we have considered seem to strike a reasonable compromise between the desire to have Γ include all reasonable priors, and the problems of having a too-large Γ. Whether one uses either of these classes is, of course, dependent on believing that either symmetry and unimodality, or just unimodality, are reasonable.

Comparison of the numerical results in Sections 2 and 3 indicate that the effect of the symmetry assumption is relatively modest. The largest effect was in the situation of Examples 2.3.2 and 3.3.2, for $x = 4.0$; the nominal 95% credible set had posterior probability ranging between 0.830 and 0.965 with the symmetry assumption, and ranging between 0.77 and 0.97 without. The difference is not great. In contrast, the unimodality assumption can have a dramatic effect; compare the ranges in this article with those in Berger and Berliner (1986), where $Q = \{\text{all priors}\}$ was considered.

A slight modification of the Section 3 unimodal class, that might be appealing, is to impose the additional constraint that θ_0 be the median of q; this prevents all the contaminating mass, ε, from being concentrated on one side. Analysis for this class can be done similarly.

A second possible modification arises from observing that the key fact driving much of the mathematics of the article is that unimodal densities are mixtures of uniforms, and that the relevant extreme points for calculating infimums and supremums are just one or two point mixtures of uniforms. This suggests the possibility of replacing Q by the class of mixtures of a family of distributions other than uniform, for example, the class of normal or maybe Cauchy distributions. Indeed this will typically work, the only real change being the replacement of (2.2) or (3.2) by the corresponding mixtures over the new base parametric family.

Alternative mixture classes might be useful in a variety of situations. First, if say one feels quite certain that π has a smooth bell shape, then only allowing smooth bell-shaped contaminations, as would result for example from letting Q

be all mixtures of symmetric (about θ_0) normal distributions, can be reasonable and will reduce the size of the ranges of posterior quantities. A second possibility would be to consider a mixture class of, say, bimodal distributions, if one desires to allow the possibility of departures from unimodality. The basic point is that a great deal of flexibility is possible, without complicating most of the mathematics.

A final possible modification that should be mentioned is that of allowing variation in the prior mode. It is certainly reasonable to assume that there is some uncertainty in this mode. We did not explicitly incorporate this uncertainty because: (i) we feel that the central part of the prior is easier to elicit than the tail of the prior, and were hence mainly concerned with including contaminations that allowed very general tail behavior; (ii) small variation in the prior mode will typically not change the Bayesian answers much; and (iii) allowing variation in the mode can easily be done utilizing our results, with the appending of an optimization over the mode. (In a sense, parametric optimizations are easy; the purpose of the article is partly to show how optimizations over large classes of distributions can often be reduced to parametric optimizations.)

When robustness fails to obtain for a given ε, π_0 and Q, one must reconsider these subjective inputs. [This is, of course, a somewhat controversial statement; see Berger (1984), Section 2.4, and Berger (1985), Sections 3.7 and 4.12, for discussion.] In particular, further refinement of ε, π_0 or Q may lead to robustness. Note that this refinement need not be simply a reduction in ε or Q. One might well change π_0 itself. (Indeed, the old π_0 might not even be an element of the refined Γ, such as when the data strongly supports a subset of Γ which does not contain π_0.) In any case, knowledge of the priors in Γ, at which the extremes occur, can be invaluable in suggesting where to concentrate such efforts at refinement.

APPENDIX

LEMMA A.1. *Let Ψ be the convex hull of a set Φ_I of probability measures on the real line given by*

$$\Phi_I = \{\mu_\tau : \tau \in I\},$$

where $I \subseteq R^k$ is some index set. Suppose that f and g are both real-valued functions on R such that $\int |f(x)|\, dF(x) < \infty$ for any $F \in \Psi$, and $B + g(x) > 0$ for some constant B. Then

$$\sup_{F \in \Psi} \frac{A + \int f(x)\, dF(x)}{B + \int g(x)\, dF(x)} = \sup_{\tau \in I} \frac{A + \int f(x) \mu_\tau(dx)}{B + \int g(x) \mu_\tau(dx)}.$$

The same result holds with "sup" *replaced by* "inf."

PROOF. Note that

$$A + \int f(x)\, dF(x) = \int (A + f(x)) \int_I \mu_\tau(dx) \gamma(d\tau) \tag{A.1}$$

for some probability measure γ on I.

Using Fubini's theorem, we can write (A.1) as

$$A + \int f(x)\,dF(x) = \int_I \left(\int (A + f(x))\mu_\tau(dx) \right) \gamma(d\tau)$$
$$= \int_I \left[\left(\frac{\int (A + f(x))\mu_\tau(dx)}{\int (B + g(x))\mu_\tau(dx)} \right) \int (B + g(x))\mu_\tau(dx) \right] \gamma(d\tau)$$
$$\leq \left(\sup_{\tau \in I} \frac{\int (A + f(x))\mu_\tau(dx)}{\int (B + g(x)\mu_\tau(dx)} \right) \left(B + \int g(x)\,dF(x) \right).$$

Thus,

$$\sup_{F \in \Psi} \frac{A + \int f(x)\,dF(x)}{B + \int g(x)\,dF(x)} \leq \sup_{\tau \in I} \frac{A + \int f(x)\mu_\tau(dx)}{B + \int g(x)\mu_\tau(dx)}.$$

But, since $\Psi \supseteq \Phi_I$, equality must obtain. The proof for the "inf" is similar. □

Acknowledgments. We are grateful to Prem Goel, Herman Rubin, Stephen Samuels and Bill Studden for helpful comments, and to the Editor, Associate Editor and three referees for very valuable suggestions. In particular, the Editor and a referee graciously provided substantially more elegant versions of most of the theorems and proofs.

REFERENCES

BERGER, J. (1982). Bayesian robustness and the Stein effect. *J. Amer. Statist. Assoc.* **77** 358–368.

BERGER, J. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness of Bayesian Analysis* (J. Kadane, ed.) 63–124. North-Holland, Amsterdam.

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

BERGER, J. (1987). Robust Bayesian analysis: Sensitivity to the prior. Technical Report 87-10, Dept. Statistics, Purdue Univ.

BERGER, J. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

BERGER, J. and DELAMPADY, M. (1987). Testing precise hypotheses (with discussion). *Statist. Sci.* **2** 317–352.

BERGER, J. and O'HAGAN, A. (1988). Range of posterior probabilities for the class of unimodal priors with specified quantiles. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Oxford Univ. Press, New York.

BERGER, J. and SELLKE, T. (1987). Testing a point null hypothesis: The irreconcilability of *p*-values and evidence (with discussion). *J. Amer. Statist. Assoc.* **82** 112–139.

BERLINER, L. M. and GOEL, P. (1986). Incorporating partial prior information: Ranges of posterior probabilities. Technical Report 357, Dept. Statistics, Ohio State Univ.

BICKEL, P. (1984). Parametric robustness: Small biases can be worthwhile. *Ann. Statist.* **12** 864–879.

BLUM, J. R. and ROSENBLATT, J. (1967). On partial a priori information in statistical inference. *Ann. Math. Statist.* **38** 1671–1678.

CASELLA, G. and BERGER, R. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J. Amer. Statist. Assoc.* **82** 106–111.

DELAMPADY, M. (1986). Testing a precise hypothesis: Interpreting *p*-values from a robust Bayesian viewpoint. Ph.D. dissertation, Purdue Univ.

DEMPSTER, A. P. (1975). A subjectivist look at robustness. *Bull. Internat. Statist. Inst.* **46** 349–374.

DEROBERTIS, L. (1978). The use of partial prior knowledge in Bayesian inference. Ph.D. dissertation, Yale Univ.

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

EDWARDS, W., LINDMAN, H. and SAVAGE, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Rev.* **70** 193–242.

GOLDSTEIN, M. (1980). The linear Bayes regression estimator under weak prior assumptions. *Biometrika* **67** 621–628.

GOOD, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. Univ. Minnesota Press, Minneapolis.

HARTIGAN, J. A. (1969). Linear Bayesian methods. *J. Roy. Statist. Soc. Ser. B* **31** 446–454.

HILL, B. (1980). Robust analysis of the random model and weighted least squares regression. In *Evaluation of Econometric Models* (J. Kmenta and J. B. Ramsey, eds.) 197–217. Academic, New York.

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

KADANE, J. B. and CHUANG, D. T. (1978). Stable decision problems. *Ann. Statist.* **6** 1095–1110.

KUDŌ, H. (1967). On partial prior information and the property of parametric sufficiency. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 251–265. Univ. California Press.

LAMBERT, D. and DUNCAN, G. (1986). Single-parameter inference based on partial prior information. *Canad. J. Statist.* **14** 297–305.

LEAMER, E. E. (1978). *Specification Searches*. Wiley, New York.

LEAMER, E. E. (1982). Sets of posterior means with bounded variance prior. *Econometrica* **50** 725–736.

MANSKI, C. F. (1981). Learning and decision making when subjective probabilities have subjective domains. *Ann. Statist.* **9** 59–65.

MARAZZI, A. (1985). On constrained minimization of the Bayes risk for the linear model. *Statist. Decisions* **3** 277–296.

MULHOLLAND, H. P. and ROGERS, C. A. (1958). Representation theorems for distribution functions. *Proc. London Math. Soc.* (3) **8** 177–223.

O'HAGAN, A. and BERGER, J. (1988). Ranges of posterior probabilities for quasi-unimodal priors with specified quantiles. *J. Amer. Statist. Assoc.* **83** 503–508.

POLASEK, W. (1985). Sensitivity analysis for general and hierarchical linear regression models. In *Bayesian Inference and Decision Techniques with Applications* (P. K. Goel and A. Zellner, eds.). North-Holland, Amsterdam.

RUBIN, H. (1977). Robust Bayesian estimation. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. Moore, eds.) 351–356. Academic, New York.

SCHNEEWEISS, H. (1964). Eine Entscheidungsregel für den Fall partiell bekannter Wahrscheinlichkeiten. *Unternehmensforschung* **8** (2) 86–95.

SIVAGANESAN, S. (1986). Robust Bayesian analysis with ε-contaminated classes. Ph.D. dissertation, Purdue Univ.

SIVAGANESAN, S. (1987). Sensitivity of posterior mean to unimodality preserving contamination. Technical Report, Dept. Statistics, Southern Methodist Univ.

SIVAGANESAN, S. (1988). Range of posterior measures for priors with arbitrary contaminations. *Comm. Statist. A—Theory Methods* **17** 1591–1612.

SIVAGANESAN, S. and BERGER, J. (1986). Range of posterior measures for priors with unimodal contaminations. Technical Report 86-41, Dept. Statistics, Purdue Univ.

WALLEY, P. (1986). Rationality and vagueness. In preparation.

WEST, S. A. (1977). Upper and lower probability inferences pertaining to Poisson processes. *J. Amer. Statist. Assoc.* **72** 448–453.

WOLFENSON, M. and FINE, T. L. (1982). Bayes-like decision making with upper and lower probabilities. *J. Amer. Statist. Assoc.* **77** 80–88.

DEPARTMENT OF MATHEMATICAL SCIENCES
MAIL CODE 25
UNIVERSITY OF CINCINNATI
CINCINNATI, OHIO 45221

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907