SENSITIVITY OF POSTERIOR MEAN TO UNIMOBILITY
PRESERVING CONTAMINATIONS

S. Sivaganesan
Southern Methodist University
Dallas, Texas 75275

Department of Statistical Sciences
Southern Methodist University
Dallas, Texas 75275

SENSITIVITY OF POSTERIOR MEAN TO UNIMODALITY PRESERVING CONTAMINATIONS

S. Sivaganesan

ABSTRACT

The sensitivity or robustness of posterior mean to uncertainties in the specification of the prior distribution is considered. We model the uncertainty in an elicited prior π_0, which we assume unimodal, by means of a ε-contaminated class of priors $\Gamma = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q : q \in Q\}$ where ε is the amount of uncertainty in π_0 and Q is the set of all contaminations q which make the resulting prior $\pi = (1-\varepsilon)\pi_0 + \varepsilon q$ unimodal with the same mode as that of π_0. Then, we find the range of the posterior mean as prior varies over this class, and give an example involving normal distribution.

INTRODUCTION

In carrying out a Bayesian analysis about an unknown parameter θ, one is required to quantify the prior information in the form of a prior distribution. In practice, the available prior information, however valuable it may be, is typically vague. Due to this and the natural limitations on time and other resources it is virtually impossible to accurately quantify the prior information in the form of a single prior distribution. Thus, after arriving at a single prior, π_0, through a carefully carried out process of prior elicitation, one

would usually feel somewhat uncertain concerning the validity of π_0 in that any other prior "realistically close" to π_0 would also seem equally plausible. Therefore, only those statistical procedures which are insensitive or robust with respect to the changes in prior distribution within the class of all priors "realistically close" to π_0 would be desirable. This view, now known as the Robust Bayesian View, has been espoused by many authors at least since Good [9]. A detailed account of its implications along with a review is given in Berger [2]. Other related references are Dempster [8], Rubin [12], Kadane and Chuang [11], Hill [10] and Berger [3].

In adhering to the robust Bayesian view, two steps become fundamental to any investigation of Bayesian robustness. One is to determine the "right" class Γ which consists of the priors that remain plausible after the elicitation of π_0, and the other is to find the ranges of the relevant posterior criteria. Then, if these ranges are reasonably small, one can be satisfied that robustness occurs with respect to misspecification of prior distribution. Otherwise, one would conclude that robustness is not present with respect to Γ. For recommendations on how one may proceed in the latter case, see Berger [2].

An attractive way of specifying the class of priors close to π_0 is by ε-contamination class given by

$$\{\pi = (1-\varepsilon)\pi + \varepsilon q : \quad q \in Q\}$$

where ε, $(0 < \varepsilon < 1)$, is the amount of error or uncertainty deemed possible in π_0, and Q is a set of probability distributions representing all plausible contaminations of π_0. When θ is a real-valued parameter and π_0 is unimodal with mode θ_0, as will be assumed throughout this paper, an interesting choice for Q is one consisting of all contaminations $q(\cdot)$ for which the resulting prior $\pi = (1-\varepsilon)\pi_0 + \varepsilon q$ is also unimodal with the same mode θ_0. The resulting class of priors, denoted herein after by Γ is very

appealing, from subjective viewpoint, for two reasons. First, any prior which remains plausible after the elicitation process would naturally be close to π_0 and unimodal (with mode θ_0) and hence to be included in Γ. Second, most priors in Γ would typically be plausible after the elicitation of π_0. Thus, while consisting all plausible priors, this class also disallows most of the (apiriori) unrealistic priors.

In this paper, we model the uncertainty in π_0 by the class Γ, and find the range of the posterior mean (a commonly used Bayesian estimator) as the prior varies over Γ. Indeed, what class is appropriate in a given situation is almost entirely a subjective decision. However the class Γ can be a very realistic model for many situations.

Much of the early work related to this paper involved classes of conjugate priors and classes of priors specified by moments, both of which are not very sensible in view of the nature of the difficulties associated with prior elicitation. A discussion of this and other more sensible priors is given in Berger [2]. Recent work involving sensible classes include De Robertis and Hartigan [7], Berger and Berliner [4], Berger and O'Hagan [5], Berliner and Goel [6] and Sivaganesan [13]. Most closely related to this paper is that of Berger and Berliner [4], where they solve what is known as the ML-II problem for the class Γ considered in this paper.

2. Notations

We observe a random variable X which has density $f(x \mid \theta)$ where θ is unknown real valued parameter. For an observed x, we will denote the likelihood $f(x \mid \theta)$, a function of θ, by $f_X(\theta)$ and assume that $1/f_X(\theta)$ is convex. As one consequence, $f_X(\theta)$ is unimodal with unique mode, which we denote by $\hat{\theta}$. As indicated earlier, we also assume that the (base) prior π_0 is unimodal, with unique mode θ_0, and formally define the class Γ as

$$\Gamma = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q : \quad q \in Q\} \qquad (2.1)$$

where

    $Q = \{$probability distributions $q(\cdot)$ such that

        $\pi = (1-\varepsilon)\pi_0 + \varepsilon q$ is unimodal with (not

        necessarily unique) mode θ_0, and $\qquad (2.2)$

        $\pi(\theta_0) \leq \pi_0(\theta_0)\}$

3. Preliminaries

The posterior mean with respect to ε-contaminated prior $\pi = (1-\varepsilon)\pi_0 + \varepsilon q$ is given by

$$\delta^\pi(x) = \lambda(x)\delta^{\pi_0}(x) + (1-\lambda(x))\delta^q(x) \qquad (3.1)$$

where $\delta^{\pi_0}(x)$, $\delta^q(x)$ are the posterior means with respect to the priors π_0 and q respectively, and

$$\lambda(x) = \frac{(1-\varepsilon)\ m(x \mid \pi_0)}{(1-\varepsilon)m(x \mid \pi_0) + \varepsilon\ m(x \mid q)} \qquad (3.2)$$

here $m(x \mid \pi_0)$, $m(x \mid q)$ are the marginal distributions of X with respect to π_0 and q. Now, we introduce the notation (used as a generic notation) $\tilde{\pi}$ to denote any prior, in Γ, which is equivalent to $(1-\varepsilon)\pi_0$ everywhere except in some interval where it is a constant. That is, $\tilde{\pi} \in \Gamma$ is of the form

$$\tilde{\pi}(\theta) = \begin{cases} K & \theta \in B \\ (1-\varepsilon)\pi_0(\theta) & \theta \notin B \end{cases} \qquad (3.3)$$

for some interval B and an appropriate constant K. The value of the constant K and the length of the interval B are determined by the conditions that $\tilde{\pi}$ is unimodal with mode θ_0, and

$$\int_B (K - (1-\varepsilon)\pi_0(\theta))d\theta = \varepsilon \quad . \qquad (3.4)$$

We now give the specific form of $\tilde{\pi}$ for different invervals B. When $B \subset (\theta_0, \infty)$, writing $B = (\tilde{\theta}, w(\tilde{\theta}))$ for some $\tilde{\theta} > \theta_0$, the corresponding $\tilde{\pi}$ is

$$\tilde{\pi}(\theta) = \begin{cases} (1-\varepsilon)\pi_0(\tilde{\theta}) & \text{if } \theta \in [\tilde{\theta}, \ w(\tilde{\theta})] \\ (1-\varepsilon)\pi_0(\theta) & \text{otherwise} \end{cases}$$

where $w(\theta) > \theta$ is defined implicitly by

$$(1-\varepsilon)\pi_0(\theta) \ (w(\theta) - \theta) - (1-\varepsilon) \int_{\theta}^{w(\theta)} \pi_0(t)dt = \varepsilon.$$

When θ_0 is the left end-point of the interval B, $\tilde{\pi}$ is of the form

$$\tilde{\pi}(\theta) = \begin{cases} (1-\rho)\pi_0(\theta_0) & \text{if } \theta \in [\theta_0, \ \nu(\rho)] \\ (1-\varepsilon)\pi_0(\theta) & \text{otherwise} \end{cases}$$

for some $0 \leq \rho \leq \varepsilon$, where $\nu(\rho) > \theta_0$ is defined implicitly by

$$(1-\rho)\pi_0(\theta_0) \ (\nu(\rho)-\theta_0) - (1-\varepsilon) \int_{\theta_0}^{\nu(\rho)} \pi_0(\xi)d\xi = \varepsilon.$$

When θ_0 is an interior point of B, $\tilde{\pi}$ is of the form

$$\tilde{\pi}(\theta) = \begin{cases} (1-\varepsilon)\pi_0(\theta_0) & \theta \in [\theta', \ \theta''] \\ (1-\varepsilon)\pi_0(\theta) & \text{otherwise} \end{cases}$$

where $\theta' < \theta''$ are solutions to the equation

$$(1-\varepsilon)\pi_0(\theta_0) \ (\theta''-\theta') - (1-\varepsilon) \int_{\theta'}^{\theta''} \pi_0(\theta)d\theta = \varepsilon.$$

The forms of $\tilde{\pi}$ when $B \subset [-\infty, \ \theta_0]$ can be similarly defined by using the condition (3.3). Now, let $\bar{\Gamma} \subset \Gamma$ be given by

$$\bar{\Gamma} = \{\pi = (1-\varepsilon) \ \pi_0 + \varepsilon q \in \Gamma: \pi \text{ is of the form } \tilde{\pi}\}.$$

4. Range of the Posterior Mean

4.1 Statement of the Result and Example

In the following theorem, we show that in order to find the sup and
inf of $\delta^\pi(x)$ (as given in (3.1)) as π varies over Γ, (as given in
(2.1)), it is sufficient to do the maximization and minimization
over the sub-class $\tilde{\Gamma}$ of Γ. The latter can be done by means of
a simple numerical calculation due to the simplicity of the form of
$\tilde{\pi}$ (as given in (3.3)).

THEOREM Let Γ, $\tilde{\Gamma}$ be as in (2.1) and (2.2). If $1/f_x(\theta)$ is
convex, then

$$\sup_{\pi \in \Gamma} \delta^\pi(x) \quad = \quad \sup_{\pi \in \tilde{\Gamma}} \delta^\pi(x) \tag{4.1}$$

and

$$\inf_{\pi \in \Gamma} \delta^\pi(x) \quad = \quad \inf_{\pi \in \tilde{\Gamma}} \delta^\pi(x) . \tag{4.2}$$

The proof of the theorem is given in section 4.2.

When $\tilde{\pi} \in \tilde{\Gamma}$, $\delta^{\tilde{\pi}}(x)$ can be written, using (3.1) and
(3.3), as

$$\frac{a\delta^{\pi_0}(x) \; + \; \int_B t \, f_x(t) \, [k - (1-\varepsilon)\,\pi_0(t)]dt}{a \qquad + \; \int_B f_x(t) \, [k - (1-\varepsilon)\,\pi_0(t)]dt} \tag{4.3}$$

where $a = (1-\varepsilon)m(x \mid \pi_0)/m(x \mid \tilde{\pi})$. For example, when the interval B
is of the form $[\theta, w(\theta)]$ for some $\theta > \theta_0$, $\delta^{\tilde{\pi}}(x)$ can be written as

$$\frac{a\delta^{\pi_0}(x) \; + \; (1-\varepsilon) \int_\theta^{w(\theta)} t \, f_x(t) \, [\pi_0(\theta) - \pi_0(t)]dt}{a \; + \; (1-\varepsilon) \int_\theta^{w(\theta)} f_x(t) \, [\pi_0(\theta) - \pi_0(t)]dt} . \tag{4.4}$$

Similar expressions for $\delta^{\tilde{\pi}}(x)$ can be written for the
other cases of B.

EXAMPLE

Suppose $X \mid \theta \sim N(\theta, \sigma^2)$, $\pi_0 \equiv N(\theta_0, \tau^2)$ and σ, τ are known. Then

$$a = \frac{1-\epsilon}{\epsilon} \; \frac{1}{\sqrt{\{2\pi(\sigma^2+\tau^2)\}}} \; \exp\left\{ -\frac{1}{2} \frac{(x-\theta_0)^2}{(\sigma^2+\tau^2)} \right\} ,$$

and

$$\delta^{\pi_0}(x) = \frac{\sigma^2}{\sigma^2+\tau^2} \mu + \frac{\tau^2}{\sigma^2+\tau^2} x .$$

The range of $\delta^{\pi}(x)$ as π varies over Γ can be found by maximizing
(and minimizing) $\delta^{\tilde{\pi}}(x)$ over the sub-class $\tilde{\Gamma}$.
This can be done by maximizing (and minimizing) $\delta^{\tilde{\pi}}(x)$
in (4.3) over all possible B. Now, finding $\sup \delta^{\pi}(x)$ and
$\inf \delta^{\pi}(x)$ can be numerically carried out in view of the fact that
$\delta^{\tilde{\pi}}(x)$ can be treated as a function of one variable,
as is clear from (4.4). The ranges of $\delta^{\pi}(x)$ for the specific case
where $\theta_0 = 0$, $\sigma = \tau = 1$ and $\epsilon = 0.1$ were computed for various
values of x. These are displayed in Figure 1.

5. Proof of the Theorem

For convenience, we let $\hat{\theta} > \theta_0$ and prove (4.1) of the theorem.
The proof of (4.2) is similar.

For $\theta > \theta_0$, we now define

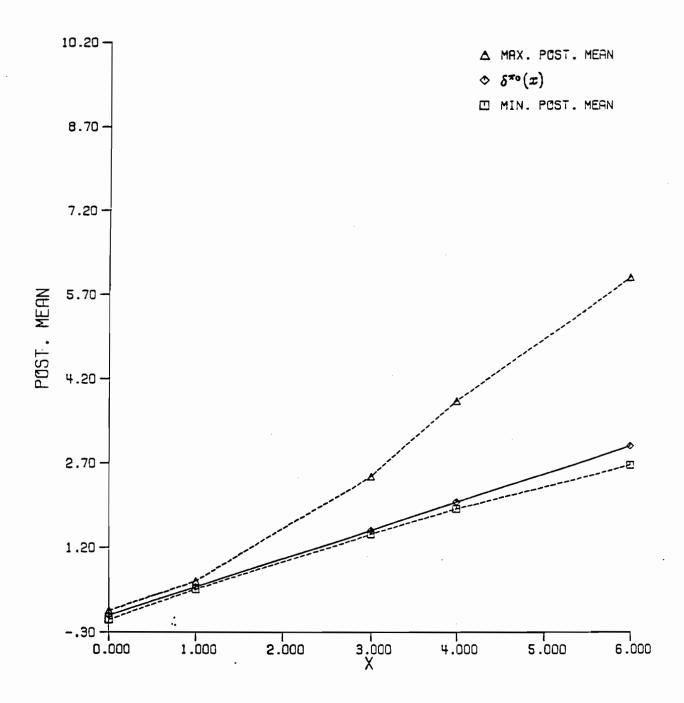$$W(\theta) = f_x(w(\theta)) \, (w(\theta)-\theta) - \int_{\theta}^{w(\theta)} f_x(t)dt . \tag{5.1}$$

Figure 1: Graphs of the Range of Posterior Mean and $\delta^{\pi 0}(x)$ against x.

It is shown in Lemma 4.1 of Berger and Berliner (1986) that, if $W(\theta_0) \geq 0$ ($V(\varepsilon)$ in their paper is the same as $W(\theta_0)$ here) then $W(\theta) = 0$ has a unique solution θ^*, $\theta_0 \leq \theta^* < \hat{\theta}$. It is then clear that $w(\theta^*) > \hat{\theta}$. In the following we consider the case $W(\theta_0) \geq 0$. The other case can be treated similarly. Before we begin the proof of (4.1), we prove some lemmas.

LEMMA 1

Let θ_* be such that $f_X(\theta_*) = f_X(w(\theta^*))$, $\bar{\theta} \in (\theta^*, \theta_*)$ and $a > \bar{\theta}$ be such that $f_X(a) = f_X(w(\bar{\theta}))$. Suppose that $b \in (\bar{\theta}, a)$, and function $g_1(\theta)$ satisfies

    1. $g_1(\theta)$ is non-decreasing in $(b, w(\bar{\theta}))$

    2. $\int_b^{w(\bar{\theta})} g_1(\theta) d\theta = 0$.

Then,

$$\int_b^{w(\theta)} f_X(\theta) g_1(\theta) d\theta \leq \alpha g_1(b)$$

where $\alpha = \int_b^{w(\bar{\theta})} f_X(\theta) d\theta - (w(\bar{\theta}) - b) f(w(\bar{\theta}))$.

PROOF

This proof follows the same lines as the proof of Lemma 4.2 of Berger and Berliner [4]. Since $g_1(\theta)$ is non-decreasing in $(b, w(\bar{\theta}))$,

$$g_1(\theta) = K_1 - \int_b^{\theta} h(\xi) d\xi \text{ for } \theta \in (b, w(\bar{\theta})).$$

where $h(\xi)$ is some non-negative function. Then,

$$K_1 = \frac{1}{w(\tilde{\theta})-b} \int_b^{w(\tilde{\theta})} (w(\tilde{\theta})-\xi)h(\xi)d\xi$$

and

$$\int_b^{w(\tilde{\theta})} g_1(\theta)f_x(\theta)d\theta = K_1 \int_b^{w(\tilde{\theta})} f_x(\theta)d\theta - \int_b^{w(\tilde{\theta})} h(\xi) \int_\xi^{w(\tilde{\theta})} f_x(\theta)d\theta d\xi.$$

The condition (4.12) of the Lemma 4.2 of Berger and Berliner [4], namely

$$\int_\xi^{w(\tilde{\theta})} f_x(\theta)d\theta \geq (w(\tilde{\theta}) - \xi)f_x(w(\tilde{\theta})),$$

holds for $b \leq \xi \leq w(\tilde{\theta})$. Hence,

$$\int_b^{w(\tilde{\theta})} g_1(\theta)f_x(\theta)d\theta \leq K_1 \int_b^{w(\tilde{\theta})} f_x(\theta)d\theta$$

$$- f(w(\tilde{\theta})) \int_b^{w(\tilde{\theta})} (w(\tilde{\theta}) - \xi)h(\xi)d\xi$$

$$= K_1 \int_b^{w(\tilde{\theta})} f_x(\theta)d\theta - K_1(w(\tilde{\theta}) - b)f_x(w(\tilde{\theta}))$$

$$= K_1\alpha$$

where $\alpha = \int_b^{w(\tilde{\theta})} f_x(\theta)d\theta) - (w(\tilde{\theta}) - b)f_x w(\tilde{\theta})) > 0$. This proves the lemma since $K_1 = g_1(b)$.

LEMMA 2

Suppose that $\bar\theta$, a and b are as in Lemma 1. Let the function
g(θ) be such that

    1. g(θ) is non-increasing in $(b,w,(\bar\theta))$

    2. g(θ) $\geq$ 0 in $(b,w(\bar\theta))^c$

    3. $\int g(\theta)d\theta = 0$.

Then

$$\int f_X(\theta)g(\theta)d\theta < \alpha g(b).$$

PROOF:

We write

$$\int f_X(\theta)g(\theta)d\theta = \int_b^{w(\bar\theta)} f_X(\theta)g(\theta)d\theta + \int_{[b,w(\bar\theta)]^c} f_X(\theta)g(\theta)d\theta.$$

Let

$$\int_b^{w(\bar\theta)} g(\theta)d\theta = K = - \int_{[b,w(\bar\theta)]^c} g(\theta)d\theta$$

Now, $\int_{[b,w(\bar\theta)]^c} f_X(\theta)g(\theta)d\theta < (-K)f_X(w(\bar\theta))$. Furthermore,

$$\int_b^{w(\bar\theta)} f_X(\theta)g(\theta)d\theta = \int_b^{w(\bar\theta)} f_X(\theta)\left(g(\theta) - \frac{K}{(w(\bar\theta)-b)}\right)d\theta$$

$$+ \frac{K}{(w(\bar\theta)-b)} \int_b^{w(\bar\theta)} f_X(\theta)d\theta$$

$$= \int_b^{w(\bar\theta)} f_X(\theta)g_1(\theta)d\theta + \frac{K}{(w(\bar\theta)-b)} \int_b^{w(\bar\theta)} f_X(\theta)d\theta$$

$$\leq \alpha g_1(b) + \frac{K}{(w(\tilde{\theta})-b)} \int_b^{w(\tilde{\theta})} f_X(\theta)d\theta,$$

where $g_1(\theta) = g(\theta) - \dfrac{K}{(w(\tilde{\theta})-b)}$ and the last step in the above follows

from Lemma 1. Hence,

$$\int f_X(\theta)g(\theta)d\theta < -Kf_X(w(\tilde{\theta})) + \frac{K}{(w(\tilde{\theta})-b)} \int_b^{w(\tilde{\theta})} f(\theta)d\theta$$
$$+ \alpha g_1(b)$$

$$= \frac{K}{(w(\tilde{\theta})-b)}\alpha + \alpha\left(g(b) - \frac{K}{(w(\tilde{\theta})-b)}\right)$$

$$= \alpha g(b).$$

LEMMA 3

Suppose $1/f_X(\theta)$ is convex. Let M be the class of sub-probability measures defined by

$$M = \{\mu(\cdot): \int \mu(d\theta) = c_1 \int f_X(\theta)\mu(d\theta) = c_2\},$$

where $0 \leq c_1 \leq 1$ and $c_2 = c_1 f_X(\hat{\theta})$ for some $\hat{\theta} > \hat{\theta}$.
Suppose that $I \subset (\hat{\theta}, \infty)$ is an interval containing $\hat{\theta}$, and $\mu_1(\cdot)$, $\mu_2(\cdot) \in M$ are such that

$$\mu_1(I) = c_1 = \mu_2(I^c).$$

Then

$$\int \theta f_X(\theta)\mu_1(d\theta) \geq \int \theta f_X(\theta)\mu_2(d\theta).$$

PROOF:

Since discrete measures form a dense set, it is sufficient to prove the lemma for the case where $\mu_1(\cdot)$ and $\mu_2(\cdot)$ are discrete measures in M. Suppose, therefore, that $\mu_1(\cdot)$ and $\mu_2(\cdot)$ are of the form

$$\mu_1(\cdot) = \sum_{i=1}^{m} \beta_i I_{\{\theta^{(i)}\}}(\cdot) \qquad \mu_2(\cdot) = \sum_{i=1}^{n} \alpha_i I_{\{\theta_i\}}(\cdot),$$

where $\theta^{(i)} \epsilon I$ and $\theta_i \epsilon I^c$. Without loss of generality, we may assume that $\theta_i \geq \hat{\theta}$. For, otherwise, we could always replace any $\theta_i < \hat{\theta}$ by two appropriate point-masses in $[\hat{\theta}, \infty)$. We now use induction on n and then on m to prove

$$\sum_{i=1}^{n} \alpha_i \theta_i a_i < \sum_{i=1}^{m} \beta_i \theta^{(i)} b_i \qquad (5.2)$$

where $a_i = f_X(\theta_i)$ and $b_i = f_X(\theta^{(i)})$. When $n = m = 2$,

$$\alpha_1 + \alpha_2 = \beta_1 + \beta_2 \text{ and } \alpha_1 a_1 + \alpha_2 a_2 = \beta_1 b_1 + \beta_2 b_2.$$

It is possible to find $\alpha_1' > 0, \alpha_2' > 0$ such that

$$\alpha_1' + \alpha_2' = \beta_1$$

and

$$\alpha_1' a_1 + \alpha_2' a_2 = \beta_1 b_1 \qquad (5.3)$$

Similarly, there exists $\alpha_1'' > 0$, $\alpha_2'' > 0$ such that

$$\alpha_1'' + \alpha_2'' = \beta_2$$

and

$$\alpha_1'' a_1 + \alpha_2'' a_2 = \beta_2 b_2. \qquad (5.4)$$

Now,

$$\therefore \quad \alpha_1' \theta_1 a_1 + \alpha_2' \theta_2 a_2 < \beta_1 \theta^{(1)} b_1$$

and

$$\alpha_1'' \theta_1 a_1 + \alpha_2'' \theta_2 a_2 < \beta_2 \theta^{(2)} b_2$$

(since, by Lemma 4, $\int \theta f_x(\theta) \mu(d\theta)$ is maximized, subject to $\int \mu(d\theta)$ = c_1 and $\int f_x(\theta) \mu(d\theta) = c_2$, when $\mu(\cdot)$ is a point-mass at some $\theta > \hat{\theta}$). Hence,

$$(\alpha_1' + \alpha_1'') \theta_1 a_1 + (\alpha_2' + \alpha_2'') \theta_2 a_2 < \beta_1 \theta^{(1)} b_1 + \beta_2 \theta^{(2)} b_2 \qquad (5.5)$$

Now, from the equations (5.3) and (5.4) we have

$$(\alpha_1' + \alpha_1'') a_1 + (\alpha_2' + \alpha_2'') a_2 = \beta_1 b_1 + \beta_2 b_2$$

$$= (\beta_1 + \beta_2) \bar{b}, \text{ say.}$$

But,

$$\alpha_1 a_1 + \alpha_2 a_2 = \beta_1 b_1 + \beta_2 b_2 = (\beta_1 + \beta_2) \bar{b}.$$

Hence,

$$\alpha_1 = \alpha_1' + \alpha_2'',$$

$$\alpha_2 = \alpha_2' + \alpha_2'',$$

Thus, from (5.5)

$$\alpha_1 \theta_1 a_1 + \alpha_2 \theta_2 a_2 < \beta_1 \theta^{(1)} b_1 + \beta_2 \theta^{(2)} b_2 .$$

Thus, (5.2) is true for $n = m = 2$. Suppose that (5.2) is true for $n = k$ and $m = 2$. Let

$$\sum_{i=1}^{k+1} \alpha_1 = \beta_1 + \beta_2,$$

$$\sum_{i=1}^{k+1} \alpha_i a_i = \beta_1 b_1 + \beta_2 b_2.$$

Choose θ_1, θ_2 such that $\max\{\theta_1, \theta_2\} < \min\{\theta^{(1)}, \theta^{(2)}\}$ (or $\min\{\theta_1, \theta_2\} > \max\{\theta^{(1)}, \theta^{(2)}\}$). If $\alpha_1 a_1 + \alpha_2 a_2 = (\alpha_1 + \alpha_2)\bar{a}_{12}$ and $f_x(\bar{\theta}_{12}) = \bar{a}_{12}$ $(\theta_1 < \theta_{12} < \theta_2)$, then

$$\alpha_1 \theta_1 a_1 + \alpha_2 \theta_2 a_2 < (\alpha_1 + \alpha_2)\bar{\theta}_{12}\bar{a}_{12}.$$

Now,

$$\sum_{i=1}^{k+1} \alpha_i \theta_i a_i < (\alpha_1 + \alpha_2)\bar{\theta}_{12}a_{12} + \sum_{i=3}^{k+1} \alpha_i \theta_i a_i \qquad (5.6)$$

Also,

$$(\alpha_1 + \alpha_2) + \alpha_3 + \ldots \alpha_{k+1} = \beta_1 + \beta_2$$

$$(\alpha_1 + \alpha_2)\bar{a}_{12} + \ldots \alpha_{k+1} a_{k+1} = \beta_1 b_1 + \beta_2 b_2.$$

Thus, by the assumption that (5.2) is true for $n=k$,

$$(\alpha_1 + \alpha_2)\bar{\theta}_{12}\bar{a}_{12} + \sum_{i=3}^{k+1} \alpha_i \theta_i a_i < \beta_1 \theta^{(1)} b_1 + \beta_2 \theta^{(2)} b_2.$$

Hence, by (5.6),(5.2) is true for $m=2, n=k+1$. So, (5.2) is true for $m = 2$, $n \geq 2$. Now, induction on m, similarly carried out, leads to the proof of (5.2) and hence the lemma.

LEMMA 4

If $1/f_x(\theta)$ is convex, then $\sup \int \theta f_x(\theta)\mu(d\theta)$ is attained, subject to the conditions $\int \mu(d\theta) = c_1$ and $\int f_x(\theta)\mu(d\theta) = c_2$, by a point-mass measure.

Proof:

Without loss of generality we may assume $c_1 = 1$. Let V be the convex hull generated by the set $U = \{(f_X(\theta), \theta f_X(\theta)): \theta \in R\}$. Now, the lemma is a simple consequence of the fact that all the boundary points of V are elements of U.

PROOF OF (4.1)

Let $0 < m < \sup_{\pi \ \Gamma} m(x \mid \pi)$ and let Γ_m be given by

$$\Gamma_m = \{\pi \in \Gamma \ : \ m(x \mid \pi) = m\}$$

Then, it is sufficient to prove that

$$\sup_{\Gamma_m} \delta^\pi(x) \ = \ \sup_{\Gamma_m \cap \bar{\Gamma}} \delta^\pi(x). \tag{5.7}$$

Now there exists $\bar{\theta} > \theta^*$ such that

$$\int f_X(\theta) \ \bar{\pi}(\theta) d\theta = m$$

where $\bar{\pi} \in \bar{\Gamma}$ is given by

$$\bar{\pi}(\theta) = \begin{cases} (1-c) \ \pi_0(\bar{\theta}) \ \text{if} \ \theta \in [\bar{\theta}, \ w(\bar{\theta})] \\ (1-\varepsilon) \ \pi_0(\theta) \ \text{otherwise} \ . \end{cases}$$

We now note that, to prove (5.7), it is sufficient to prove

$$\int \theta f_X(\theta)(\pi(\theta) - \bar{\pi}(\theta))d\theta \leq 0$$

for all $\pi \in \Gamma_m$.

Let $\pi \in \Gamma_m$ and define a function $g(\theta)$ by

$$g(\theta) = \pi(\theta) - \bar{\pi}(\theta).$$

Then, $g(\theta)$ is non-increasing in $(\bar{\theta}, w(\bar{\theta}))$, and

$$\int g(\theta)d\theta = 0,$$

$$\int f_X(\theta)g(\theta)d\theta = 0.$$

Let b be that number for which

$$g(\theta) > 0 \quad \text{for } \theta \ (\bar{\theta},b)$$

and

$$g(b) = 0.$$

Then $g(\theta) < 0$ in $I = (b,w(\bar{\theta}))$ and, when $b \geq \hat{\theta}$ we have, by letting

$$u_1(d\theta) = g(\theta)d\theta \text{ in } I^c$$

$$\mu_2(d\theta) = |g(\theta)| \, d\theta \text{ in } I$$

and using Lemma 3, that

$$\int \theta f_X(\theta)g(\theta)d\theta = \int \theta f_X(\theta)u_1(d\theta) - \int \theta f_X(\theta)\mu_2(d\theta) < 0.$$

Now, let $b < \hat{\theta}$ (and hence $\theta* < \bar{\theta} < \hat{\theta}$). First, we note that if $f_X(\bar{\theta}) < f_X(w(\bar{\theta}))$ then $b > a$ (where a is as in Lemma 1, that is $a < \hat{\theta}$ and $f_X(a) < f_X(w(\bar{\theta}))$); for otherwise, using Lemma 2,

$$\int f_X(\theta)g(\theta)d\theta < \alpha g(b) = 0.$$

The rest of the proof for the case $f_X(\bar{\theta}) > f_X(w(\bar{\theta}))$, $\bar{\theta} < \hat{\theta}$ follows almost the same line, so we only consider the case where $f_X(\bar{\theta}) < f_X(w(\bar{\theta}))$, $\bar{\theta} < \hat{\theta}$.

Since $\int f_X(\theta)g(\theta)d\theta = 0$, there exists $\bar{b}\epsilon(b', w(\tilde{\theta}))$ such that

$$\int_b^{w(\tilde{\theta})} f_X(\theta)g(\theta)d\theta = f_X(\bar{b}) \int_b^{w(\tilde{\theta})} g(\theta)d\theta$$

where $b' > \hat{\theta}$ is such that $f_X(b') = f_X(b)$. Also,

$$\int_{-\infty}^b f_X(\theta)g(\theta)d\theta = f_X(\bar{a}) \int_{-\infty}^b g(\theta)d\theta \quad \text{for some } \bar{a}\epsilon(a,b).$$

Let $\bar{a}'\epsilon(b',w(\tilde{\theta}))$ such that $f_X(\bar{a}) = f_X(\bar{a}')$. Then $\bar{a}' < \bar{b}$. Now, there exists $t_0\epsilon(\hat{\theta},\bar{a}')$ such that

$$\int_b^{\bar{a}'} f_X(\theta)g(\theta)d\theta = f_X(t_0) \int_b^{\bar{a}'} g(\theta)d\theta.$$

Letting

$$h(t) = \frac{\int_b^t f_X(\theta)g(\theta)d\theta}{\int_b^t g(\theta)d\theta}, \quad \theta < t < w(\tilde{\theta}),$$

we have

1. $h(t)$ is continuous;

2. $h(\bar{a}') = f_X(t_0) \quad t_0\epsilon(\hat{\theta},\bar{a}')$;

3. $h(w(\tilde{\theta})) = f_X(\bar{b}) \quad \bar{b}\epsilon(b',w(\tilde{\theta}))$,

where $f_X(t_0) < f_X(\bar{b})$, $t_0 < \bar{a}' < \bar{b}$. Thus there exists $b_0\epsilon(\bar{a}',w(\tilde{\theta}))$ such that

$$h(b_0) = f_X(\bar{a}')$$

or, equivalently,

$$\int_b^{b_0} f_X(\theta)g(\theta)d\theta = f_X(\bar{a}) \int_b^{b_0} g(\theta)d\theta.$$

Suppose that

$$\int_{-\infty}^{b_0} g(\theta)d\theta < 0 \quad,$$

and let

$$g_1(\theta) = \begin{cases} 0 & \text{if } -\infty < \theta < b \\ (1-\lambda)g(\theta) & \text{if } b < \theta < b_0 \\ g(\theta) & \text{otherwise} \end{cases}$$

where $0 < \lambda < 1$ is given by

$$\lambda \left| \int_b^{b_0} g(\theta)d\theta \right| = \int_{-\infty}^{b} g(\theta)d\theta.$$

Then, $\int g_1(\theta)d\theta = \int g(\theta)d\theta = 0$. Furthermore,

$$\int_{-\infty}^{b_0} f_X(\theta)g(\theta)d\theta = f_X(\bar{a}) \int_{-\infty}^{b_0} g(\theta)d\theta$$

$$= \int_{-\infty}^{b} f_X(\theta)g(\theta)d\theta + \int_{b}^{b_0} f_X(\theta)g(\theta)d\theta$$

$$= f_X(\bar{a}) \int_{-\infty}^{b} g(\theta)d\theta + f_X(\bar{a}) \int_{b}^{b_0} g(\theta)d\theta$$

$$= f_X(\bar{a}) \int_{-\infty}^{b_0} g(\theta)d\theta = (1-\lambda) \int_{b}^{b_0} f_X(\theta)g(\theta)d\theta$$

$$= \int_{-\infty}^{b_0} f_X(\theta)g_1(\theta)d\theta \quad .$$

Thus,

$$\int f_X(\theta)g(\theta)d\theta = \int f_X(\theta)g_1(\theta)d\theta.$$

But

$$\int f_X(\theta)g_1(\theta)d\theta = \int_b^{w(\tilde{\theta})} f_X(\theta)g_1(\theta)d\theta + \int_{w(\tilde{\theta})}^{\infty} f_X(\theta)g_1(\theta)d\theta < 0,$$

since $f_X(\theta) < f_X(\theta')$ when $\theta \in (w(\tilde{\theta}),\infty)$ and $\theta' \in (b,w(\tilde{\theta}))$, which is a contradiction. Hence,

$$\int_{-\infty}^{b_0} g(\theta)d\theta > 0.$$

Now, let

$$g_1(\theta) = \begin{cases} (1-\lambda)g(\theta) & \text{if } -\infty < \theta < b \\ 0 & \text{if } b < \theta < b_0 \\ g(\theta) & \text{otherwise} \end{cases}$$

where $0 < \lambda < 1$ is given by

$$\left| \int_b^{b_0} g(\theta)d\theta \right| = \lambda \int_{-\infty}^{b} g(\theta)d\theta.$$

Then

$$\int f_X(\theta)g(\theta)d\theta = \int f_X(\theta)g_1(\theta)d\theta \quad , \quad \int g_1(\theta)d\theta = 0,$$

and

$$\lambda \int_{-\infty}^{b} \theta f_X(\theta) g(\theta) d\theta \;\leq\; \lambda b \int_{-\infty}^{b} f_X(\theta) g(\theta) d\theta$$

$$= \lambda b f_X(\bar{a}) \int_{-\infty}^{b} g(\theta) d\theta$$

$$= b f_X(\bar{a}) \int_{b}^{b_0} \Big| g(\theta) \Big| d\theta$$

$$= b \int_{b}^{b_0} f_X(\theta) \Big| g(\theta) \Big| d\theta$$

$$< \int_{b}^{b_0} \theta f_X(\theta) \Big| g(\theta) \Big| d\theta.$$

Thus,

$$\int_{-\infty}^{b_0} \theta f_X(\theta) g(\theta) d\theta \;<\; (1-\lambda) \int_{-\infty}^{b} \theta f_X(\theta) g(\theta) d\theta.$$

Hence,

$$\int \theta f_X(\theta) g(\theta) d\theta \;<\; \int \theta f_X(\theta) g_1(\theta) d\theta.$$

Further, $g_1(\theta) < 0$ only in $[b_0, w(\tilde{\theta})] \subset (\hat{\theta}, \infty)$. Hence,
letting $\mu_1(\cdot) = \Big| g_1(\cdot) \Big|$ on $[b_0, w(\tilde{\theta})]$ and $\mu_2(\cdot) = g_1(\cdot)$ on
$[b_0, w(\tilde{\theta})]^c$ and using Lemma 3, we get

$$\int \theta f_X(\theta) g_1(\theta) d\theta < 0$$

and hence

$$\int \theta f_X(\theta) g(\theta) d\theta < 0,$$

proving the theorem.

and

$$\lambda \int_{-\infty}^{b} \theta f_X(\theta)g(\theta)d\theta \;\leq\; \lambda b \int_{-\infty}^{b} f_X(\theta)g(\theta)d\theta$$

$$= \;\lambda b f_X(\overline{a}) \int_{-\infty}^{b} g(\theta)d\theta$$

$$= \;b f_X(\overline{a}) \int_{b}^{b_0} \Big| g(\theta)\Big| \, d\theta$$

$$= \;b \int_{b}^{b_0} f_X(\theta)\Big| g(\theta)\Big| \, d\theta$$

$$< \;\int_{b}^{b_0} \theta f_X(\theta)\Big| g(\theta)\Big| \, d\theta .$$

Thus,

$$\int_{-\infty}^{b_0} \theta f_X(\theta)g(\theta)d\theta \;<\; (1-\lambda) \int_{-\infty}^{b} \theta f_X(\theta)g(\theta)d\theta .$$

Hence,

$$\int \theta f_X(\theta)g(\theta)d\theta \;<\; \int \theta f_X(\theta)g_1(\theta)d\theta .$$

Further, $g_1(\theta) < 0$ only in $[b_0, w(\tilde{\theta})] \subset (\hat{\theta}, \infty)$. Hence, letting $\mu_1(\cdot) = \big| g_1(\cdot)\big|$ on $[b_0, w(\tilde{\theta})]$ and $\mu_2(\cdot) = g_1(\cdot)$ on $[b_0, w(\tilde{\theta})]^c$ and using Lemma 3, we get

$$\int \theta f_X(\theta)g_1(\theta)d\theta \;<\; 0$$

and hence

$$\int \theta f_X(\theta)g(\theta)d\theta \;<\; 0 ,$$

proving the theorem.

[2] Berger, J. The robust Bayesian viewpoint (with   discussion). In Robustness of Bayesian Analysis (J. Kadane, Ed.). North Holland, Amsterdam (1984).

[3] Berger, J. Statistical Decision Theory and Bayesian Analysis Springer-Verl, New York (1985).

[4] Berger, J. and Berliner, L. M. Robust Bayes and empirical Bayes analysis with -contaminated priors. Ann. Statist. 14, 461-486 (1986).

[5] Berger, J. and O'Hagan, A. Range of posterior probabilities for the class of unimodal priors with specified quantiles. Technical Report, Department of Statistics, Purdue University (1986).

[6] Berliner, L. M. and Goel, P. Ranges of posterior probabilities for the class of priors with specified quantiles. Technical Report, Department of Statistics, Ohio State University (1986).

[7] DeRobertis, L. and Hartigan, J. A. Bayesian inference using intervals of measures. Ann. Statist. 1, 235-244 (1981).

[8] Dempster, A.P. A subjective look at robustness. Bull. Int. Statist. Inst. 46, 349-374 (1975).

[9] Good, I.J. Probability and the weighing of Evidence. Griffin, London (1950).

[10] Hill, B. Robust analysis of the random model and weighted least squares regression. In Evaluation of Econometric Models. Academic Press, New York (1980).

[11] Kadane, J.B. and Chuang, D.T. Stable decision problems. Ann. Statist. 6, 1095-1110 (1978).

[12] Rubin, H. Robust Bayesian estimation. In Statistical Decision Theory and Related Topics II, (S.S. Gupta and D. Moore, Eds.). Academic Press, New York (1977).

[13] Sivaganesan, S. Robust Bayesian Analysis with ϵ-contaminated Classes. Ph.D. Thesis, Purdue University, West Lafayette (1986).

[14] Sivaganesan, S. and Berger, J. O. Range of posterior measures for priors with unimodal contaminations. Technical Report, Department of Statistics, Purdue University.