Ranges of Posterior Measures for Priors with Unimodal Contaminations*

by

S. Sivaganesan, University of Cincinnati
and
James Berger, Purdue University

Technical Report #86-41

Department of Statistics
Purdue University

January 1987
Revised September 1987

* Research supported by the National Science Foundation, Grants DMS–8401996 and DMS–8702620.

Ranges of Posterior Measures for Priors with Unimodal Contaminations*

by

S. Sivaganesan and James Berger
University of Cincinnati Purdue University

Abstract

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

* Research supported by the National Science Foundation, Grants DMS–8401996 and DMS–8702620.

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 exactly quantify 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 and 1985, and Berger and Berliner, 1986 for further motivation.) In this paper we model "close" through the ε-contamination class

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

here ε determines the amount of probabilistic deviation from π_0 that is allowed, and Q is the class of allowed contaminations (see subsection 1.3). In subsection 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

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

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 robustness with respect to the elicitation process. If the range is large, one does not have robustness with respect to Γ, but our results provide indications as to which features of $\pi \epsilon \Gamma$ 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, i.e. 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 adhoc 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; the technical problem appears to be a difficult variational problem. We show, however, that the problem is often tractable, and yields relatively simple answers. (The mathematical basis for solution of this type of variational problem is briefly sketched in the Appendix.) 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), Wolfenson and Fine (1982), Berger (1984, 1985) (which contain general review and discussion), and Walley (1986). We discuss here only some of the papers 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 papers which do deal with large classes of priors are especially noteworthy. Huber (1973) determines the range of the prior probability of a set for the class in (1.1) with $Q = \{$ all distributions $\}$. DeRobertis and Hartigan (1981), in a breakthrough paper, 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 (1987) and O'Hagan and Berger (1987) find the range of the posterior probability of certain sets over classes of priors with specified quantiles. West (1979) 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.) Automating robustness checks is probably necessary to have them actually used.

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 paper 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 $Q = \{$ all distributions $\}$ made in Huber (1973). This choice is particularly easy to work with, and Sivaganesan (1986b) 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 $Q = \{$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 $\pi \epsilon \Gamma$, 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 $\pi \epsilon \Gamma$ 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 $\lambda(x) \epsilon [0,1]$ 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 apriori it was believed that π_0 or q were true with probabilities $(1-\epsilon)$ 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

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

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

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

and consider the class Γ given in (1.1). In sections 2.3, 2.4, and 2.5 we find the ranges, as π varies over Γ, of the posterior mean, posterior variance, and posterior probability of a set, respectively. 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.

<u>Lemma</u> 2.2.1: For $q \epsilon Q$ as in (2.1)

$$m(x|q) = \int_0^\infty H(z)dF(z), \int \theta^i f_x(\theta)q(\theta)d\theta = \int_0^\infty H_i(z)dF(z),$$

and

$$\int 1_C(\theta)f_x(\theta)q(\theta)d\theta = \int_0^\infty H_C(z)dF(z),$$

where 1_C is the indicator function on the set C, F is some distribution function (the mixing distribution which yields q),

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

$$H_i(z) = \begin{cases} \frac{1}{2z}\int_{\theta_0-z}^{\theta_0+z} \theta^i f_x(\theta)d\theta & \text{if } z \neq 0 \\ \theta_0^i f_x(\theta_0) & \text{if } z = 0, \end{cases} \tag{2.3}$$

and

$$H_C(z) = \begin{cases} \frac{1}{2z}\int_{\theta_0-z}^{\theta_0+z} 1_C(\theta) f_x(\theta)d\theta & \text{if } z \neq 0 \\ f_x(\theta_0) & \text{if } z = 0 \text{ and } \theta_0 \epsilon C \\ 0 & \text{if } z = 0 \text{ and } \theta_0 \notin C. \end{cases} \tag{2.4}$$

The dependence of these quantities on x is suppressed for notational simplicity.

2.3 Range of the Posterior Mean

Let Γ be as in (1.1) and $\Gamma_1 \subset \Gamma$ be given by

$$\Gamma_1 = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q : q \text{ is } U(\theta_0 - z, \theta_0 + z) \text{ for some } z > 0\}. \tag{2.5}$$

In order to find the range of the posterior mean 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.

<u>Theorem</u> 2.3.1: For Γ and Γ_1 as above ,

$$\sup_{\pi\epsilon\Gamma} \delta^\pi(x) = \sup_{\pi\epsilon\Gamma_1} \delta^\pi(x) = \sup_z \frac{a_0 + H_1(z)}{a + H(z)}$$

and

$$\inf_{\pi\epsilon\Gamma} \delta^\pi(x) = \inf_{\pi\epsilon\Gamma_1} \delta^\pi(x) = \inf_z \frac{a_0 + H_1(z)}{a + H(z)},$$

where $a = (1-\varepsilon)m(x|\pi_0)/\varepsilon$, $a_0 = a\delta^{\pi_0}(x)$, and $H(z), H_1(z)$ are as in the previous subsection.

Proof: For $\pi\epsilon\Gamma$, $\delta^{\pi}(x)$ can be written as

$$\delta^{\pi}(x) = \frac{a_0 + \int H_1(z)dF(z)}{a + \int H(z)dF(z)}$$

for some distribution function $F(\cdot)$. Now the result follows by a direct application of Lemma A1 of the Appendix. ||

Example 2.3.1: 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

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

Furthermore,

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

and

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

Letting $t = \varphi(\frac{\mu+z-x}{\sigma}) + \varphi(\frac{\mu-z-x}{\sigma})$, $u = \varphi(\frac{\mu+z-x}{\sigma}) - \varphi(\frac{\mu-z-x}{\sigma})$ and $v = \Phi(\frac{\mu+z-x}{\sigma}) - \Phi(\frac{\mu-z-x}{\sigma})$, the values of z which maximize and minimize

$$\frac{a_0 + H_1(z)}{a + H(z)}$$

are given by the solutions of the equation (obtained by differentiating the above),

$$z = \frac{(vx - \sigma u)(t + 2\sigma a) - v(2a_0\sigma + t\mu)}{2\left[auz + t(a\mu - a_0) - vu/2\right]}. \qquad (2.6)$$

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 2.1, 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 2.1 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.)

2.4 Range of the Posterior Variance For Fixed Posterior Mean

It is typically necessary, in estimation problems, to also require 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 $\mu \in [\inf_{\pi} \delta^{\pi}(x), \sup_{\pi} \delta^{\pi}(x)]$, if we let

$$\Gamma_0 = \{\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta) : q \epsilon Q \text{ and } \delta^{\pi}(x) = \mu\}, \tag{2.7}$$

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

2.4.1 Determining the Supremum of the Posterior Variance

It is shown below, under some regularity conditions on the likelihood function, that the posterior variance is maximized, subject to fixed posterior mean, when the contamination is a mixture of two symmetric uniform (symmetric with respect to θ_0) distributions. Let

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

where

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

Note that one of the three quantities α, z, z^* is determined by the constraint that π have posterior mean μ. 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 one.

<u>THEOREM</u> 2.4.1: Suppose $f_x(\theta)$ is such that, if a and b are real constants, then $[h(\theta_0+\theta)+h(\theta_0-\theta)]$ and $[g(\theta_0+\theta)+g(\theta_0-\theta)]$ each have at most two positive local maxima; here

$$h(\theta) = (\theta - a)f_x(\theta) \text{ and } g(\theta) = (\theta^2 - a\theta - b)f_x(\theta). \tag{2.8}$$

Then

$$\sup_{\pi \epsilon \Gamma_0} V^\pi(x) = \sup_{\pi \epsilon \Gamma_{02}} V^\pi(x). \tag{2.9}$$

<u>PROOF:</u> Using equation (1.6) of section 1.4, for $\pi \in \Gamma_0$ we have

$$\mu = \lambda(x)\delta^{\pi_0}(x) + (1-\lambda(x))\delta^q(x). \tag{2.10}$$

First we consider the problem of maximizing $V^\pi(x)$ over a subset Γ_{0m} of Γ_0 given by

$$\Gamma_{0m} = \{\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \epsilon \Gamma_0 : m(x|q) = m\}, \tag{2.11}$$

where

$$m \in \Lambda(\mu) = \left\{ \int f_x(\theta)q(\theta)d\theta : q \epsilon Q \text{ satisfies } (2.10) \right\}. \tag{2.12}$$

Now, letting

$$\lambda(x) = \frac{(1-\varepsilon)m(x|\pi_0)}{(1-\varepsilon)m(x|\pi_0) + \varepsilon m} = \lambda,$$

and

$$\delta^q(x) = \frac{\mu - \lambda\delta_0^\pi}{(1-\lambda)} = \delta,$$

we have, for $\pi\epsilon\Gamma_{0m}$, using equation (1.7), that

$$V^{\pi}(x) = \lambda V^{\pi_0}(x) + \lambda(1-\lambda)(\delta^{\pi_0}(x)-\delta)^2 + \frac{(1-\lambda)}{m}\int \theta^2 f_x(\theta)q(\theta)d\theta.$$

Thus, finding $\sup_{\pi\epsilon\Gamma_{0m}} V^{\pi}(x)$ is equivalent to finding

$$\sup_{q\epsilon Q_m}\int \theta^2 f_x(\theta)q(\theta)d\theta,$$

where

$$Q_m = \left\{\, q(\cdot) : m(x|q) = m, \int \theta f_x(\theta)q(\theta)d\theta = m\delta \,\right\}.$$

This problem of maximization may be re-formulated using the symmetry and unimodality of $q(\cdot)$ (with mode θ_0) as that of finding

$$\sup_{\mu(\cdot)}\int H_2(z)\mu(dz),$$

where $\mu(\cdot)$ is subject to the following conditions:

$$\begin{aligned} \int H_1(z)\mu(dz) &= m\delta \\ \int H(z)\mu(dz) &= m \\ \int \mu(dz) &= 1, \end{aligned} \tag{2.13}$$

and H, H_1, and H_2 are given by equations (2.2) and (2.3). Clearly $(m, m\delta)\epsilon V$, the convex hull generated by the set $\{(H(z), H_1(z)) : \; z\epsilon R\}$. When $(m, m\delta)$ is not an interior point of V, Lemma A2 implies that $\mu(\cdot)$ satisfying (2.13) must be a point mass, proving the result. Thus, suppose that $(m, m\delta)$ is an interior point of V. Then, using Theorem A1 of the Appendix,

$$\begin{aligned} \sup_{\mu(\cdot):(\text{ 2.13})}\int H_2(z)\mu(dz) &= \hat{a}m\delta + \hat{b}m + \hat{c} \\ &= \inf\{\, am\delta + bm + c : (a,b,c) \;\in\; \Lambda \,\}, \end{aligned}$$

where

$$\Lambda = \left\{(a,b,c)\epsilon R^3 : H_2(z) \le aH_1(z) + bH(z) + c \quad \forall\, z \ge 0\right\}.$$

Furthermore, the maximizing $\mu_0(\cdot)$ has support given by

$$support\,(\mu_0(\cdot)) = \left\{\; z \geq 0 : H_2(z) = \hat{a}H_1(z) + \hat{b}H(z) + \hat{c} \;\right\}.$$

Now, since $(\hat{a}, \hat{b}, \hat{c}) \in \Lambda$ we have, by letting $g(\theta) = (\theta^2 - \hat{a}\theta - \hat{b})f_x(\theta)$, that

$$\int_{\theta_0 - z}^{\theta_0 + z} g(\theta)d\theta - 2\hat{c}z \leq 0 \quad \forall\, z > 0$$

and $\hat{c} \geq 0$. Thus, if $G(z)$ represents the function on the left hand side above, then

Number of elements in *support* $(\mu_0(\cdot))$ $\leq$ Number of zeroes of G(z) .

Now, $G(0) = 0$ and

$$\text{Number of zeroes of G(z)} \;\leq \frac{1}{2}\;\text{Number of zeroes of}\; \frac{d}{dz}G(z).$$

But,

$$\frac{d}{dz}G(z) = g(\theta_0 + z) + g(\theta_0 - z) - 2\hat{c}$$

and

$$\frac{d}{dz}G(z)(\text{at } z{=}0) = 2g(\theta_0) - 2\hat{c} \leq 0.$$

Now, suppose $\hat{c} > 0$. Then $\mu_0(\cdot)$ is a probability measure (see Theorem A1) and, if $g(\theta_0 + z) + g(\theta_0 - z)$ has at most two local maxima, then

$$\text{Number of zeroes of}\; \frac{d}{dz}G(z) \leq 4 \quad (4 + 1 \text{ if } g(\theta_0) - \hat{c} = 0).$$

Thus,

$$\text{Number of zeroes of}\; G(z) \leq 2$$

and hence, support (μ_0) has at most two points. By means of a similar argument it can be shown that μ_0 is a one point (sub)probability measure when $\hat{c} = 0$. Thus we have established that

$$\sup_{\Gamma_{0m}} V^{\pi}(x) = \sup_{\Gamma_{0m} \cap \Gamma_{02}} V^{\pi}(x).$$

It follows directly that

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

2.4.2 Determining 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

$$\Gamma_{01} = \{\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \epsilon \Gamma_0 : \quad q \equiv \alpha U(\theta_0 - z, \theta_0 + z) \text{ for some } 0 \le \alpha \le 1 \text{ and } z \ge 0\}. \tag{2.14}$$

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 $\delta^{\pi} = \mu$ (i.e., that $\pi \epsilon \Gamma_0$), 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 theorem, define

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

where H, H_1, and H_2 are as in (2.2) and (2.3).

<u>Theorem</u> 2.4.2: Suppose that $[h(\theta_0+\theta) - h(\theta_0-\theta)]$ has at most two positive local maxima and that $[g(\theta_0+z) + g(\theta_0-z)]$ has at most one positive local minimum, where h and g are defined in (2.8). Then

$$\inf_{\pi \epsilon \Gamma_0} V^{\pi}(x) = \inf_{\pi \epsilon \Gamma_{01}} V^{\pi}(x). \tag{2.15}$$

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

$$V^{\pi_0} + (\delta^{\pi_0})^2 - \mu^2 + (\delta^{\pi_0} - \mu)\psi(\hat{z}), \tag{2.16}$$

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

$$\{z : \frac{\mu H(z) - H_1(z)}{\mu^*} \ge a = \frac{(1-\varepsilon)m(x|\pi_0)}{\varepsilon}\}. \tag{2.17}$$

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

$$\inf_z\{\frac{aV^{\pi_0} + H_2(z) - H(z)(\delta^{\pi_0})^2}{a + H(z)} : \quad H_1(z) = \delta^{\pi_0} H(z)\}. \tag{2.18}$$

<u>Proof</u>: The proof of (2.15) is similar to that of Theorem 2.4.1, and will be omitted. For $q \equiv \alpha U(\theta_0 - z, \theta_0 + z)$ to yield $\pi = (1-\varepsilon)\pi_0 + \varepsilon q \epsilon \Gamma_0$, it must be the case that (when $\delta^{\pi_0} \neq \mu$)

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

Using this expression for α, together with (1.7), yields after simplification the expression for $V^{\pi}(x)$ given in (2.16). 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 (2.17).

When $\delta^{\pi_0} = \mu$ (equivalent to the condition $H_1(z) = \delta^{\pi_0} H(z)$ in (2.18)), 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 (2.18)) need be considered. This completes the proof. ||

<u>Example</u> 2.4.1: (Normal Distribution)

Suppose $X|\theta \sim N(\theta, \sigma^2)$ where σ is known. Then it can be shown that $\frac{d}{dz}[g(\theta_0 + z) + g(\theta_0 - z)]$ has at most three positive zeroes, which would mean that $g(\theta_0 + z) + g(\theta_0 - z)$ has at most two positive local maxima and at most one positive local minimum. Similarly, the other regularity conditions can be verified; thus, when $X|\theta \sim N(\theta, \sigma^2)$ and $\pi_0(\cdot)$ is any unimodal distribution with mode θ_0, the results of Theorems 2.4.1 and 2.4.2 hold.

As a specific example, let $X|\theta \sim N(\theta, 1)$, $\pi_0(\theta)$ be $N(0,1)$ and $\varepsilon = 0.1$. Then the range of the posterior variance, when the posterior mean is fixed at various levels, is displayed in Figure 2.2 for certain values of x.

2.5 Posterior Probability of a Set

2.5.1 Credible 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. For $\pi \in \Gamma$, (1.8),(2.2), and (2.4) yield

$$P^{\pi(\theta|x)}(\theta \epsilon C) = \frac{k_1 + \int H_C(z)dF(z)}{k + \int H(z)dF(z)},$$

where $k = \left(\frac{(1-\varepsilon)}{\varepsilon}\right) m(x|\pi_0)$ and $k_1 = kP^{\pi_0(\theta|x)}(\theta \epsilon C)$. Determination of the range of the posterior probability of C is considerably simplified by the following theorem.

<u>Theorem</u> 2.5.1: For any measurable subset C of Θ,

$$\sup_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \sup_{z \geq 0} \frac{k_1 + H_C(z)}{k + H(z)}$$

and

$$\inf_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \inf_{z \geq 0} \frac{k_1 + H_C(z)}{k + H(z)}.$$

<u>Proof</u>: Proof follows directly from Lemma A1 of the Appendix. ||

Defining $g_x(\theta) = f_x(\theta)I_C(\theta)$, $A(z) = g_x(z) + g_x(-z)$ and $B(z) = f_x(z) + f_x(-z)$, the values of z, at which the inf and sup of the probabilities are obtained, are given by the positive solutions of the equation (which may be solved iteratively by carefully choosing the initial values)

$$2z(kA(z) - k_1B(z)) = (2a + B(z)) \int_{-z}^{z} g_x(\theta)d\theta - (2k_1 + A(z)) \int_{-z}^{z} f_x(\theta)d\theta.$$

In choosing the initial values for iteration when C is an interval, it is helpful to note that the sup is attained at a value in the interior of C or, to the right of C or left of C according as whether $x > \theta_0$ or $x < \theta_0$, and the inf is always attained at an exterior point of z (mostly to the right of C when $x > \theta_0$ and left of C when $x < \theta_0$).

<u>Example</u> 2.5.1 (Normal Distribution) 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

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

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 (.830,.965), the extreme values of which are attained when the contaminations are, respectively, $U(-6.1, 6.1)$ and $U(-4.3, 4.3)$.

2.5.2 Testing of Hypotheses

To test the hypothesis $H_0 : \theta \epsilon \Theta_0$ versus $H_1 : \theta \epsilon \Theta \backslash \Theta_0$, 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 from Theorem 2.5.1 when Θ_0 and $\Theta \backslash \Theta_0$ 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 $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. 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

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

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.

<u>Theorem</u> 2.5.2:

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

and

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

where

$$\hat{H} = \sup_{z\geq 0}\left(\frac{1}{2z}\int_{\theta_1-z}^{\theta_1+z} f_x(\theta)d\theta\right).$$

<u>Example</u> 2.5.2: 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 2.5.1. When $x \leq 1.5$, the uncertainty in $P^{\pi(\theta|x)}(H_0)$ is 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 $P^{\pi(\theta|x)}(H_0)$.

Table 2.5.1 Range of the Posterior Probability of H_0

x	$\theta_1 = 0$	$\theta_1 = 1.0$	θ_2=2.0
0.5	(.46,.69)	(.46,.68)	(.59,.78)
1.0	(.42,.65)	(.35,.59)	(.42,.65)
1.5	(.35,.57)	(.24,.45)	(.24,.45)
4.0	(.004,.007)	(.002,.007)	(.0008,.002)

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

$$Q = \{ \text{ all distributions which are unimodal with the same mode }, \theta_0, \text{ as that of } \pi_0 \}. \tag{3.1}$$

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

3.2 Preliminaries

LEMMA3.2.1: For $q \in Q$ from (3.1),

$$m(x|q) = \int_{-\infty}^{\infty} H(z) dF(z),$$

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

and

$$\int 1_C(\theta) f_x(\theta) q(\theta) d\theta = \int_{-\infty}^{\infty} H_C(z) dF(z),$$

where 1_C is the indicator function on the set C, F is some distribution function,

$$H(z) = \begin{cases} \frac{1}{z} \int_{\theta_0}^{\theta_0+z} f_x(\theta) d\theta & \text{if } z \neq 0 \\ f_x(\theta_0) & \text{if } z = 0, \end{cases} \tag{3.2}$$

$$H_i(z) = \begin{cases} \frac{1}{z}\int_{\theta_0}^{\theta_0+z} \theta^i f_x(\theta)d\theta & \text{if } z \neq 0 \\ \theta_0^i f_x(\theta_0) & \text{if } z = 0, \end{cases} \tag{3.3}$$

and

$$H_C(z) = \begin{cases} \frac{1}{z}\int_{\theta_0}^{\theta_0+z} 1_C(\theta) f_x(\theta)d\theta & \text{if } z \neq 0 \\ f_x(\theta_0) & \text{if } z = 0 \text{ and } \theta_0 \in C \\ 0 & \text{if } z = 0 \text{ and } \theta_0 \notin C. \end{cases} \tag{3.4}$$

3.3 Range of the Posterior Mean

Let Γ be as in (1.1), and $\Gamma_1 \subseteq \Gamma$ be given by

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

<u>Theorem</u> 3.3.1: For Γ, Γ_1 as above,

$$\sup_{\pi\epsilon\Gamma} \delta^\pi(x) = \sup_{\pi\epsilon\Gamma_1} \delta^\pi(x) = \sup_z \frac{a_0 + H_1(z)}{a + H(z)},$$

and

$$\inf_{\pi\epsilon\Gamma} \delta^\pi(x) = \inf_{\pi\epsilon\Gamma_1} \delta^\pi(x) = \inf_z \frac{a_0 + H_1(z)}{a + H(z)},$$

where a,a_0 are as given in Theorem 2.3.1 of section 2.3.

<u>Example</u> 3.3.1: 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

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

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

Letting $t = \varphi(\frac{\mu-x}{\sigma}) - \varphi(\frac{z+\mu-x}{\sigma})$, $u = \Phi(\frac{z+\mu-x}{\sigma}) - \Phi(\frac{\mu-x}{\sigma})$ and $v = \varphi(\frac{z+\mu-x}{\sigma})$, the values of z which maximize and minimize

$$\frac{a_0 + H_1(z)}{a + H(z)}$$

are given by the solutions of the equation (obtained by differentiating the above)

$$z = \frac{(\sigma t + ux)(\sigma a + v) - q(\sigma a_0 + r\mu)}{v(za + a\mu + u - a_0)}. \tag{3.5}$$

As a specific example, let $\mu = 0$, $\sigma^2 = 1$, $\tau^2 = 1$ and $\varepsilon = 0.1$. In Figure 3.1 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 2.1, as would be expected.

3.4 Range of the Posterior Variance For Fixed Posterior Mean

3.4.1 Determining the Supremum of the Posterior Variance

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 $U(\theta_0 - y, \theta_0)$.

<u>THEOREM</u> 3.4.1: Suppose $\frac{1}{f_x(\theta)}$ is convex, and $\frac{d}{d\theta}\left(\frac{1}{f_x}\right)$ is concave in $(-\infty, t)$ and convex in (t, ∞) for some t. Then

$$\sup_{\pi\epsilon\Gamma} V^\pi = \sup_{\pi\in\Gamma_{02}} V^\pi.$$

<u>PROOF</u> :The proof is similar to that of Theorem 2.4.1, using Lemma A3 of the Appendix to show that the support of μ_0 has at most two points. ||

3.4.2 Determining the Infimum of the Posterior Variance

Let Γ_{01} be defined by

$$\Gamma_{01} = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q \epsilon \Gamma_0 : q \text{ is } \alpha U_z \text{ for } 0 \le \alpha \le 1 \text{ and } z \ge 0\};$$

here U_z again denotes a $U(\theta_0, \theta_0 + z)$ or $U(\theta_0 - z, \theta_0)$ density, and α is the mass of q.

THEOREM 3.4.2: If $1/f_x(\theta)$ is convex in θ, then

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

(As in Theorem 2.4.2, this infinum can be expressed as the minimum of a function of one variable over a specified range.)

PROOF : The proof follows the lines of the proof of Theorem 2.4.2 and uses Lemma A4 of the Appendix. ||

EXAMPLE 3.4.1: Normal Distribution

Let $X|\theta \sim N(\theta, \sigma^2)$, $\pi_0(\theta)$ be $N(\mu, \tau^2)$ and ε=0.1. It is easy to check that $1/f_x(\theta)$ is convex in θ and that $\frac{d}{d\theta}\left(\frac{1}{f_x(\theta)}\right)$ is concave in $(-\infty, x)$, and convex in (x, ∞), and hence Theorems 3.4.1 and 3.4.2 are applicable. Ranges of the posterior variance, when the posterior mean is held fixed at different values, are displayed in Figures 3.2 and 3.3.

3.5 Posterior Probability of a Set

3.5.1 Credible Set

THEOREM 3.5.1: For any measurable subset C of Θ,

$$\sup_{\pi\epsilon\Gamma} P^{\pi(\theta|x)}(\theta\epsilon C) = \sup_z \frac{k_1 + H_C(z)}{k + H(z)}$$

and

$$\inf_{\pi\epsilon\Gamma} P^{\pi\theta|x)}(\theta\epsilon C) = \inf_z \frac{k_1 + H_C(z)}{k + H(z)},$$

where $H(z)$ and $H_C(z)$ are defined in section 3.2, $k = (\frac{1-\varepsilon}{\varepsilon})m(x|\pi_0)$, and $k_1 = kP^{\pi_0(\theta|x)}(\theta \in C)$.

Bounds for an Interval C

Let $C = (a, b)$, and $\theta_0 < \hat{z}$, for $\hat{z}$ given by the solution of

$$f_x(\theta_0 + z) = \frac{1}{z}\int_{\theta_0}^{\theta_0+z} f_x(\theta)d\theta. \tag{3.6}$$

Then the range of $P^{\pi(\theta|x)}(\theta \epsilon C)$ can be found using the following equations. We need to consider various cases of a and b.

(1) When $a < \theta_0 < b$:

$$\inf_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \frac{k_1}{k + \max\{f_x(z_1), f_x(z_2)\}},$$

and

$$\sup_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \frac{k_1 + H(z_s)}{k + H(z_s)},$$

where $z_s = \min\{\hat{z}, b\}$ and z_1, z_2 respectively are the solutions of the following equations (defined as ∞ when the solution to an equation does not exist):

$$z = \frac{\int_{\theta_0}^{\theta_0+z} f_x(\theta)d\theta}{f_x(z)} - \frac{k + f_x(z)}{k_1 f_x(z)} \int_{\theta_0}^{z} f_x(\theta)d\theta \qquad z > b \tag{3.7}$$

and

$$z = -\frac{\int_z^{\theta_0} f_x(\theta)d\theta}{f_x(z)} - \frac{k + f_x(z)}{k_1 f_x(z)} \int_a^{\theta_0} f_x(\theta)d\theta \qquad z < a. \tag{3.8}$$

(2) When $\theta_0 < a < b < \hat{z}$:

$$\inf_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \frac{k_1}{k + \max\{H(a), f_x(z_i)\}},$$

where z_i is the solution of the following equation:

$$z = \frac{1}{f_x(z)} \int_{\theta_0}^{z} f_x(\theta)d\theta - \frac{k + f_x(z)}{k_1 f_x(z)} \int_a^{b} f_x(\theta)d\theta \qquad z > \hat{z}; \tag{3.9}$$

and

$$\sup_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \frac{bk_1 + \int_a^b f_x(\theta)d\theta}{bk + \int_{\theta_0}^b f_x(\theta)d\theta}.$$

(3) When $\theta_0 < a < \hat{z} < b$:

$$\inf_{\pi \epsilon \Gamma} P^{\pi(\theta|x)}(\theta \epsilon C) = \frac{k_1}{k + \max\{H(a), f_x(z_i)\}},$$

where z_i is the (largest) solution to the equation (3.9). Also,

$$\sup_{\pi\epsilon\Gamma} P^{\pi(\theta|x)}(\theta\epsilon C) = \max\left\{\frac{k_1 + f_x(z_s)}{k + f_x(z_s)}, \frac{k_1}{k + f_x(z_s)}\right\},$$

where z_i is the (smallest) solution to the equation (3.9) and z_s is the solution (=∞ if solution does not exist) of the equation:

$$z = \frac{1}{f_x(z)}\int_a^z f_x(\theta)d\theta - \frac{k_1 + f_x(z)}{(k-k_1)f_x(z)}\int_{\theta_0}^a f_x(\theta)d\theta \qquad a < z < b.$$

<u>Example</u> 3.5.1: (Normal Distribution) 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

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

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

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

These are attained when the respective contaminations are $U(0,6.3)$ and $U(0,.97)$.

3.5.2 Testing Hypotheses

The discussion in section 2.5.2 applies here, though now we will constrain g to be in

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

Then Theorem 2.5.2 is valid here, with $\hat{H}$ replaced by

$$\hat{H} = f_x(\theta_1 + \hat{z}),$$

where $\hat{z}$ is the solution of (3.6).

EXAMPLE 3.5.2: 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 3.5.1.

Table 3.5.1 Posterior Probabilities of H_0

x	$\theta_1 = 0$	$\theta_1 = 1.0$	$\theta_1 = 2.0$
$x = 0.5$	(.46, .69)	(.46, .69)	(.57, .79)
$x = 1.0$	(.41, .65)	(.35, .59)	(.41, .65)
$x = 1.5$	(.35, .57)	(.24, .45)	(.24, .45)
$x = 4.0$	(.004, .04)	(.002, .007)	(.0007, .002)

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 zero 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 prior are not very appropriate (see Berger(1985)).

As mentioned in subsection 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 $Q = \{$ all distributions $\}$ is generally "too big". Further evidence of this comes from noting that, if $X_1, ..., X_n$ are i.i.d. $N(\theta, 1)$ and $Q = \{$ all distributions $\}$ is used, then $|\sup_{\pi \epsilon \Gamma} \delta^{\pi}(x) - \bar{x}|$ converges to zero as $n \to \infty$ at the rate of $\sqrt{ln(n)/n}$. For the classes we consider in this paper, the rate of convergence can be shown to be the correct rate $\sqrt{1/n}$ (see Sivaganesan (1986a)), 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.

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 paper 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, e.g. the class of normal or maybe Cauchy distributions. Indeed this will work for the range of the posterior mean, the only real change being the replacement of (2.2) through (2.4) or (3.2) through (3.4) by the corresponding mixtures over the new base parametric family. Substantial additional work might be necessary, however, to develop the analogue of the posterior variance material for these other mixture classes.

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 small 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 paper 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 π_0 or Q may lead to robustness; knowledge of the priors in Γ, at which the extremes occur, can be invaluable in suggesting where to concentrate such efforts at refinement.

ACKNOWLEDGMENT

We are grateful to Prem Goel, Herman Rubin, Stephen Samuels, and Bill Studden for helpful comments, and to the Associate Editor and three referees for very valuable suggestions.

Bibliography

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

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

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

[4] Berger, J. (1987). Robust Bayesian analysis: sensitivity to the prior. Technical report #87-10, Department of Statistics, Purdue Univ., W. Lafayette.

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

[6] Berger, J. and Delampady, M. (1987). Testing precise hypotheses. To appear in *Statistical Science.*

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

[8] Berger, J. and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of significance levels and evidence. *J. Amer. Statist. Assoc.* 82, 112–139.

[9] Berliner, L. M. and Goel, P. (1986). Incorporating partial prior information: ranges of posterior probabilities. Technical Report #357, Department of Statistics, Ohio State University.

[10] Bickel, P. (1984). Parametric robustness: small biases can be worthwhile. *Ann. Statist.* 12, 864–879.

[11] Blum, J. R. and Rosenblatt, J. (1967). On partial a priori information in statistical inference. *Ann. Math. Statist.* 38, 1671–1678.

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

[13] Delampady, M. (1986). Testing a Precise Hypothesis: Interpreting P-Values from a Robust Bayesian Viewpoint. *Ph.D. Thesis, Purdue University, West Lafayette.*

[14] Dempster, A. P. (1975). A subjectivist look at robustness. *Bull. Int. Statist. Inst.* 46, 349–374.

[15] De Robertis, L. (1978). The use of partial prior knowledge in Bayesian inference. *Ph.D. Thesis, Yale University, New Haven.*

[16] De Robertis, L. and Hartigan, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* 1, 235–244.

[17] Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review* 70, 193–242.

[18] Goldstein, M. (1980). The linear Bayes regression estimator under weak prior assumptions. *Biometrica* 67, 621–628.

[19] Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications.* University of Minnesota Press, Minneapolis.

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

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

[22] Huber, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Int. Statist. Inst.* 45, 181–191.

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

[24] Kemperman, J. H. B. (1968). The general moment problem; a geometric approach. *Ann. Math. Statist.* 39, 93–122.

[25] Kudō, H. (1967). On partial prior information and the property of parametric sufficiency. *Proc. Fifth Berkeley Symp. Statist. and Prob.* 1, University of California Press, Berkeley.

[26] Lambert, D. and Duncan, G. (1986). Single-parameter inference based on partial prior information. The *Canadian J. of Statist.* 14, 297–305.

[27] Leamer, E. E. (1978). *Specification Searches.* Wiley, New York.

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

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

[30] Marazzi, A. (1985). On constrained minimization of the Bayes risk for the linear model. *Statistics and Decisions* 3, 277–296.

[31] O'Hagan, A. and Berger, J. (1987). Ranges of posterior probabilities for quasi-unimodal priors with specified quantiles. To appear in *J. Amer. Statist. Assoc.*, 1988.

[32] Polasek, W. (1985). Sensitivity analysis for general and hierarchical linear regression models. *Bayesian Inference and Decision Techniques with Applications*, (P. K. Goel and A. Zellner, Eds.). North-Holland, Amsterdam.

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

[34] Sivaganesan, S. (1986a). Robust Bayesian Analysis with ε-Contaminated Classes. *Ph.D. Thesis, Purdue University, West Lafayette.*

[35] Sivaganesan, S. (1986b). Range of posterior measures for priors with arbitrary contaminations. *Manuscript in Preparation.*

[36] Sivaganesan, S. (1987). Sensitivity of posterior mean to unimodality preserving contamination. Technical Report, Dept. of Statistics, Southern Methodist University, Dallas.

[37] Walley, P. (1986). *Rationality and Vagueness.* Manuscript in preparation.

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

[39] Wolfenson, M. and Fine, T. L. (1982). Bayes-like decision making with upper and lower probabilities. *J. Amer. Statist. Assoc.* 77, 80–88.

APPENDIX

LEMMA A1: Suppose $F \in \mathcal{F}$, a class of probability distributions which contains all one point distributions, and that $B + g(z) > 0$ for all z. Then

$$\begin{aligned} \sup_{F\in\mathcal{F}} \frac{A+\int f(z)dF(z)}{B+\int g(z)dF(z)} &= \sup_z \frac{A+f(z)}{B+g(z)}, \\ \inf_{F\in\mathcal{F}} \frac{A+\int f(z)dF(z)}{B+\int g(z)dF(z)} &= \inf_z \frac{A+f(z)}{B+g(z)}. \end{aligned} \tag{A.1}$$

PROOF: Clearly

$$\begin{aligned} A+\int f(z)dF(z) &= \int \frac{[A+f(z)]}{[B+g(z)]}\cdot[B+g(z)]dF(z) \\ &\leq [\sup_z \frac{A+f(z)}{B+g(z)}]\int[B+g(z)]dF(z). \end{aligned}$$

Thus

$$\sup_{F\in\mathcal{F}} \frac{A+\int f(z)dF(z)}{B+\int g(z)dF(z)} \leq \sup_z \frac{A+f(z)}{B+g(z)}.$$

But, since $\mathcal{F}$ contains all one point distributions, equality must obtain. The proof for the "inf" is similar. ||

THEOREM A0: Suppose that $g_0 \equiv 1$, $g_i (i = 1, \ldots, m)$ and h are real-valued measurable functions defined on a fixed measurable space $(T, \mathcal{A})$; and that $g_i (i = 0, \ldots, m)$ are linearly independent in the sense that

$$\sum_{i=0}^{m} a_i g_i(x) = 0 \ \forall x \in T \quad \Longleftrightarrow \quad a_i = 0 \text{ for } i = 0, \ldots, m.$$

Suppose further that the set of measures Γ on T, given by

$$\Gamma = \left\{\mu : \int g_i d\mu = \lambda_i \ for\ i = 0, \ldots, m\right\} \tag{A.2}$$

where the λ_i's are given constants with $\lambda_0 = 1$, is non-empty. Define V to be the convex hull generated by $\{(g_1(t), \ldots, g_m(t)) : t \in T\}$, and

$$\Lambda = \left\{\underline{a} = (a_0, \ldots, a_m) \in R^{m+1} : h(t) \le \sum_{i=0}^{m} a_i g_i(t), \forall t \in T\right\}.$$

Then, if $(\lambda_1, \ldots, \lambda_m)$ is in the interior of V and Λ is not empty, it is true that

$$\sup_{\mu\in\Gamma} \int h d\mu = \sum_{i=0}^{m} \hat{a}_i \lambda_i = \inf\left\{\sum_{i=0}^{m} a_i \lambda_i : \underline{a} \in \Lambda\right\}. \tag{A.3}$$

Furthermore, if T is compact Hausdorff with respect to some topology, and $g_i(i = 0, \ldots, m)$ and h are continuous, then there exists $\mu_0 \in \Gamma$ such that

$$\int h d\mu_0 = \sup_{\mu\in\Gamma} \int h d\mu.$$

<u>PROOF</u>:Follows from the results in pp. 98-99 of Kemperman (1968). ||

<u>THEOREM</u> A1: Consider the situation in Theorem A0, with h, $g_i(i = 1, \ldots, m) \in C_0(R)$. Then Λ is not empty and, for $(\lambda_1, \ldots, \lambda_m)$ in the interior of V, (A.3) holds and there exists a measure ν_0 on R, $\nu_0(R) \le 1$, such that

$$\int g_i d\nu_0 = \lambda_i \quad i = 1, \ldots, m \tag{A.4}$$

$$\int h d\nu_0 = \sup_{\mu\in\Gamma} \int h d\mu. \tag{A.5}$$

Furthermore $\nu_0(R) = 1$ when $\hat{a}_0 > 0$.

<u>PROOF</u>:Only the existence of ν_0 is not immediate from Theorem A0. To prove (A.4) and (A.5), let $T = R \cup \{\infty\}$ be the one-point compactification of R with the usual metric topology. Then T is a compact Hausdorff space, and by Theorem A0 there exists a measure μ_0 on T such that $\mu_0 \in \Gamma$ and

$$\sup_{\mu\in\Gamma} \int h d\mu = \int h d\mu_0.$$

Now, taking ν_0 as the restriction of μ_0 to R, we have (A.4) and (A.5) since $h,\ g_i(i=1,\ldots,m) \in C_0(R)$. Now, let $\hat{a}_0 > 0$ and suppose that $\mu_0(R) = \lambda_0' < 1$. Consider the problem of maximizing $\int h d\mu$ over Γ as in (A.2) but with the λ_i now replaced by the λ_i', where $\lambda_i' = \lambda_i$ for $i = 1,\ldots,m$. Then, if $\sum_{i=0}^m \hat{a}_i' \lambda_i' = \inf\{\sum_{i=0}^m a_i \lambda_i' : \underline{a} \in \Lambda\}$, we have, as before, that

$$\sum_{i=0}^m \hat{a}_i' \lambda_i' = \sup_{\Gamma} \int h d\mu \geq \int h d\mu_0 = \sum_{i=0}^m \hat{a}_i \lambda_i. \tag{A.6}$$

But $\sum_{i=0}^m \hat{a}_i' \lambda_i' \leq \sum_{i=0}^m \hat{a}_i \lambda_i' < \sum_{i=0}^m \hat{a}_i \lambda_i$ since $\underline{\hat{a}} \in \Lambda$ and $\lambda_0' < \lambda_0 = 1$, contradicting (A.6). Hence $\mu_0(R) = 1$ when $\hat{a}_0 > 0$. ||

<u>Lemma</u>A2: Suppose that for any constant a, $h(z) = (z-a)f_x(z)$ is such that $h(\theta_0+z)+h(\theta_0-z)$ has at most two positive local extrema. Then for $H(z), H_1(z)$ as in section 2, the extrema of $\int H_1(z)\mu(dz)$ (respectively of $\int H(z)\mu(dz)$), taken over all probability measures $\mu(.)$ having a fixed value for $\int H(z)\mu(dz)$ (respectively, for $\int H_1(z)\mu(dz)$), are attained by measures with support having one point. For $H(z), H_1(z)$ as in section 3, the condition on h can be replaced by the condition that $1/f_x(\theta)$ be bowl-shaped.

<u>Proof :</u> Uses Theorem A1 in a way similar to the proof of Theorem 2.4.1. ||

<u>Lemma</u> A3: Let $f_x(\theta)$ be such that $1/f_x(\theta)$ is convex, and $[1/f_x(\theta)]'$ is concave in $(-\infty, t)$ and convex in (t, ∞) for some t. Suppose there exist a, b, c such that

$$H_2(z) \leq aH_1(z) + bH(z) + c \qquad \forall z, \tag{A.7}$$

where H_2, H_1 and H are as in section 3.2. Then, equality can hold in (A.7) for at most two values of z.

<u>Proof :</u> We will only prove the lemma for the case $\theta_0 = 0$, since the proof for the other general case is similar. One can write (A.7) as

$$\frac{1}{z}\int_0^z (\theta^2 - a\theta - b) f_x(\theta) d\theta \ \leq\ c \ \ \forall z. \tag{A.8}$$

Then, z for which equality holds satisfy

$$\frac{1}{z}\int_0^z (\theta^2 - a\theta - b) f_x(\theta) d\theta = c \tag{A.9}$$

and (since z is a point of maximum for the function on the left hand side of (A.8))

$$\frac{1}{z}(z^2 - az - b) f_x(z) - \frac{1}{z^2}\int_0^z (\theta^2 - a\theta - b) f_x(\theta) d\theta = 0. \tag{A.10}$$

Now, letting $g(z)$ denote the left hand side of the equation (A.9), we have $g'(0) = -\frac{1}{2}(af_x(0) + bf_x'(0))$. So, if equality holds at 0, then $g'(0) = 0$ and hence $((\theta^2 - a\theta - b)f_x(\theta))'$ at 0 equals $-(af_x(0) + bf_x'(0))$ which is 0; thus equality cannot hold at any other point. It is, therefore, sufficient to consider only $z \neq 0$. Now, the $z \neq 0$ for which equality holds satisfy (by (A.10))

$$(z^2 - az - b) f_x(z) = c. \tag{A.11}$$

If z_1 is the smallest positive solution to both (A.9) and (A.11), then there exists $z_1', 0 < z_1' < z_1$, satisfying (A.11); for otherwise, $(z^2 - az - b)f_x(z) < c, 0 < z < z_1$, and hence $(1/z)\int_0^z (a\theta^2 - a\theta - b) f_x(\theta) d\theta < c$, which is a contradiction. It can be similarly verified that between any two consecutive solutions of (A.9) there exists at least one solution to (A.11). Hence

$$\begin{aligned} &\text{Number of solutions satisfying both (A.9) and (A.11)} \\ &\quad \leq \tfrac{1}{2}(\text{ the number of solutions of (A.11)}). \end{aligned}$$

Now, (A.11) is equivalent to

$$z^2 - az - b = \frac{c}{f_x(z)}.$$

We note that $c \geq 0$, and that it is enough to consider $c > 0$. Thus, (A.11) is equivalent to

$$g(z) = \frac{1}{c}z^2 + \left(-\frac{a}{c}\right)z + \left(-\frac{b}{c}\right) - \frac{1}{f_x(z)} = 0.$$

Now,

$$\text{Number of solutions to } g(z) = 0$$
$$\leq \text{ (Number of solutions to } g'(z) = 0) + 1.$$

But,

$$g'(z) = \frac{2}{c}z + (-\frac{a}{c}) - \left(\frac{1}{f_x(z)}\right)'.$$

Thus, solutions to $g'(z) = 0$ are the points of intersection of the line $\frac{2}{c}z + (-\frac{a}{c})$ and the curve $[1/f_x(z)]'$. Since $[1/f_x(z)]'$ is concave in $(-\infty, t)$ and convex in (t, ∞), there exist at most two points of intersection in each of the intervals $(-\infty, t)$ and (t, ∞). However, if there exist 2 points of intersection in $(-\infty, t]$, viz $t_1, t_2 (t_1 < t_2 \leq t)$, then the curve $[1/f_x(z)]'$ must be below the line $2az + b$ in the interval $(t_1, t_2 + \varepsilon)$ for some $\varepsilon > 0$. Hence, by the convexity of $[1/f_x(z)]'$ in (t, ∞), there exists at most one point of intersection in (t, ∞). Thus, the number of solutions to $g'(z) = 0$ is less than or equal to 3, and hence the number of solutions to $g(z) = 0$ is less than or equal to 4. Thus, the number of solutions satisfying both (A.9) and (A.11) is less than or equal 1/2 times 4 or 2, proving the lemma. ||

<u>LEMMA</u> A4: Let $f_x(\theta)$ be such that $(1/f_x(\theta))$ is convex. Suppose that there exists a, b, c such that

$$-H_2(z) \leq aH_1(z) + bH(z) + c \quad \forall z, \tag{A.12}$$

where H_1, H_2 are as in section 3.2 . Then, equality can hold in (A.12) for at most one point.

<u>PROOF</u>: We will prove the lemma for the case $\theta_0 = 0$, since the the proof for the other case is similar. (A.12) gives

$$\frac{1}{z}\int_0^z -\theta^2 f_x(\theta) d\theta \leq a\frac{1}{z}\int_0^z \theta f_x(\theta) d\theta + b\frac{1}{z}\int_0 z f_x(\theta) d\theta + c \;\forall z, \tag{A.13}$$

which can be re-written as

$$\frac{1}{z}\int_0^z (\theta^2 f_x(\theta) + a\theta f_x(\theta) + bf_x(\theta) + c)\; d\theta \;\geq\; 0 \;\; \forall\; z.$$

As before, we note that $c \geq 0$, and that it is sufficient to consider $c > 0$ and $z \neq 0$. Further, for equality to hold at $z(\neq 0)$ in (A.13),

$$\int_0^z (\theta^2 f_x(\theta) + a\theta f_x(\theta) + bf_x(\theta) + c) d\theta \;=\; 0, \qquad \text{(A.14)}$$

and, because there must be local minima for the function on the l.h.s. of (A.14) at the points of equality in (A.13), differentiating with respect to z yields that

$$z^2 f_x(z) + azf_x(z) + bf_x(z) + c \;=\; 0. \qquad \text{(A.15)}$$

Now (A.15) is equivalent to

$$\frac{1}{c}z^2 + \frac{a}{c}z + \frac{b}{c} \;=\; -\frac{1}{f_x(z)}.$$

Since az^2+bz+c is strictly convex and $(-\frac{1}{f_x(z)})$ is concave, the number of solutions to (A.15) is less than or equal to 2. By a similar argument to that in Lemma A3, there exists at most one solution to both (A.14) and (A.15). Hence, $support(\mu_0)$ has at most one point. $\qquad \|$