VARIATIONS OF POSTERIOR EXPECTATIONS FOR SYMMETRIC UNIMODAL PRIORS IN A DISTRIBUTION BAND*

by

Sanjib Basu
Purdue University

Technical Report #91-45

Department of Statistics
Purdue University

August 1991

* This research was supported by NSF Grant DMS-8923071.

◇

Abstract

Given a random variable with distribution indexed by a one dimensional parameter θ, we consider the problem of robustness of given Bayesian posterior criteria when the prior cdf lies in the class $\Gamma_{SU} = \{F : F_L \leq F \leq F_U$ and F is symmetric and unimodal$\}$. Such a class includes as special cases well known metric neighborhoods of a fixed cdf such as Kolmogorov and Lèvy neighborhoods. A general method is described for finding the extremum of posterior expectation of a function $h(\theta)$ as the prior varies in Γ_{SU}. Finally, the method is illustrated with two examples. The use of this family in subjective prior elicitation is also discussed.

Key words : cdf, distribution band, symmetry, unimodal, prior, posterior, likelihood, robustness, range, Kolmogorov metric, Lèvy metric

1 Introduction

Study of sensitivity of Bayesian quantities to possible perturbations or misspecifications of the prior is crucial to any complete Bayesian analysis. It is now generally recognized that any elicitation process, leading to a prior π_0, is to some extent arbitrary and, therefore, any prior in a "neighborhood" of the elicited prior would also be a reasonable representation of the prior beliefs. Hence a Bayesian analysis will be reliable if the ranges of posterior quantities over a neighborhood Γ of the initial prior are not significantly wide. Past work in this area has mainly concentrated on four different types of prior neighborhoods; parametric families (like conjugate priors), contamination classes, density bands and priors with specified quantiles; significant among them are, Huber (1973), Leamer (1978,1982), DeRobertis and Hartigan (1981), Polasek (1984), Berger and Berliner (1986), Berger and O'Hagan (1989), Sivaganesan and Berger (1989), DasGupta and Studden (1988,1989), Lavine (1987,1991), Berliner and Goel (1990), Basu and DasGupta (1990), Wasserman and Kadane (1990), DasGupta and Delampady (1990), Srinivasan and Truszczynska (1990) and Basu (1991). The review paper of Berger (1990) contains a detailed discussion as well as other references.

The class Γ of cumulative prior distributions to be considered here is what we call the "distribution band", namely

$$\Gamma = \left\{ \begin{array}{ll} F: & F \text{ is a cumulative distribution function} \\ & \text{and } F_L(\theta) \leq F(\theta) \leq F_u(\theta) \quad \forall \theta \end{array} \right\} \tag{1}$$

where F_L and F_u are two fixed cdf's, satisfying $F_L(\theta) \leq F_u(\theta)$ for all θ. Bayesian robustness investigation over this class was carried out in Basu and DasGupta (1990). It was commented there that a "distribution band" is very intuitive, rich, easily can be adjusted to meet the subjective specification of the user and also allows a wide flexibility in the prior tails. It is also mathematically tractable and the ranges of interesting posterior quantities can be determined with limited numerical work.

Unfortunately, as pointed out in Berger and Berliner (1986), the ranges of posterior quantities will often be exceedingly large if reasonable shape restrictions are not imposed on the prior family. Indeed, the distribution band contains many unreasonable priors and the extremal prior cdfs are found to assign point masses to several points and also have regions with zero masses. It is argued by Berger and Berliner (1986) that a more reasonable choice is to restrict only to symmetric and unimodal priors in the specified family, which leads to the new class,

$$\Gamma_{SU} = \{F \in \Gamma : \text{F is symmetric and unimodal } \} \tag{2}$$

Similar to the distribution band Γ, the restricted class Γ_{SU} also allows wide variations in the functional form and tails of $F \in \Gamma_{SU}$, while deleting many unreasonable priors, and retaining an overall uniform shape feature. The overall shape of the prior is often rather confidently known, so it is not desirable to allow priors in the class with widely different shapes.

Another important aspect of the above class is that neighborhoods of various metrics on the space of distribution functions are often of the form (2). Suppose we believe that the prior will be unimodal about some point θ_0, and through the elicitation process, decide on F_0, unimodal about θ_0, as the prior cdf. A natural way to incorporate the uncertainty in the elicitation process would be to allow an error of ϵ in the specification, leading thus to the class (2), with $F_L(\theta) = \max[F_0(\theta)-\epsilon, 0]$ and $F_U(\theta) = \min[F_0(\theta)-\epsilon, 1]$. This is, in fact, the class of cdf's, unimodal about θ_0, in the closed ϵ-neighborhood of F_0 under the Kolmogorov metric on the space of distribution functions, defined by $d_k(F_1, F_2) = \sup_{\theta\in\Re} |F_1(\theta) - F_2(\theta)|$. Another commonly used metric that also leads to a family of the form (2) is the Lèvy metric, defined as $d_L(F_1, F_2) = \inf\{\epsilon : F_1(\theta-\epsilon)-\epsilon \leq F_2(\theta) \leq F_1(\theta+\epsilon)+\epsilon\}$. More discussions on Kolmogorv and Lèvy neighborhoods, and their applications in Bayesian robustness studies are given in Basu and DasGupta (1990).

2 Notations, assumptions and preliminaries

Suppose interest lies in $h(\theta)$, some known function of θ, defined on the parameter space Θ. For robustness investigation, we seek

$$\begin{gathered}\underline{\rho}(h) = \inf_{F\in\Gamma_{SU}} \rho(h, F) \text{ and } \overline{\rho}(h) = \sup_{F\in\Gamma_{SU}} \rho(h, F), \\ \text{where } \rho(h, F) = E^{F(\theta|x)}(h(\theta)) = \frac{\int_\Theta h(\theta)\ell(\theta)dF(\theta)}{\int_\Theta \ell(\theta))dF(\theta)},\end{gathered} \tag{3}$$

where $\ell(\theta)$ $(= f(x|\theta))$ denotes the likelihood function.

For the sake of brevity, only the problem of evaluating $\overline{\rho}(h)$ will be described. The infimum problem is technically exactly similar and consequently, no attempts will be made to elaborate on it. Some useful examples of $h(\theta)$ are $h(\theta) = \theta$, $h(\theta) = L(\theta, a)$ where L is a loss function and a is an action, and $h(\theta) = f(x_0|\theta)$. These make $E^{F(\theta|x)}(h(\theta))$ respectively equal to the posterior mean, the posterior expected loss of an action 'a', and the predictive density at x_0.

Assumptions:

The following technical assumptions will be made:

(1) The parameter space Θ is either a compact interval $[a, b]$ on the real line $\Re$ or $\Re$ itself (in which case we take $a = -\infty$, $b = +\infty$ and interpret the interval to be open). Let M be the mid-point of the parameter space Θ.

(2) $F_L(\theta)$ and $F_U(\theta)$ are continuous cumulative distribution functions on Θ, and they are unimodal about M, i.e., $F_L(\theta)$ and $F_U(\theta)$ are convex for $\theta \in [a, M)$ and concave for $\theta \in (M, b]$.

(3) $F_U(\theta) = 1 - F_L(2M - \theta) \quad \forall \theta \in \Theta$. This assumption is not necessary, but the effective band generated by symmetric F's between F_L and F_U will automatically satisfy this requirement.

(4) $F_L(M) \leq 0.5$ and, $F_U(M) \geq 0.5$.

(5) We further assume that $F_L(\theta)$ is smooth for $\theta \in (M, b)$. Since F_L is concave in $(M, b]$, the right and left derivative of $F_L(\theta)$ exists for $\theta \in (M, b)$. We are further assuming that they are equal.

(6) Likelihood $\ell(\theta)$ is continuously differentiable on (a, b) and $\ell(a) = \ell(b) = 0$.

(7) $h(\theta)$ is continuously differentiable on (a, b).

The class Γ_{SU} is now defined by

$$\Gamma_{SU} = \left\{ \begin{array}{ll} F: & F_L(\theta) \leq F(\theta) \leq F_U(\theta) \quad \forall \theta \in \Theta, \\ & \text{F is a cumulative distribution function and} \\ & \text{F is symmetric and unimodal about } M \end{array} \right\}. \tag{4}$$

Let $b_0 = \inf_{\theta \in (M,b]} \{\theta : F_L(\theta) = F_U(\theta)\}$. Since $F_L(b) = F_U(b) = 1$, $b_0 \leq b$. Also, as F_L and F_U are both concave on $(M, b]$, and $F_L(\theta) \leq F_U(\theta)$ $\forall \theta \in [a, b]$; hence $F_L(\theta) = F_U(\theta)$ $\forall \theta \geq b_0$.

<u>Effective band if $F_L(M) < 0.5$</u>

Since any $F \in \Gamma_{SU}$, being symmetric about M, satisfies $F(M) \geq 0.5$, it is clear that a sharper lower band than F_L exists on $\theta \in [M, b]$ in this case.

Let

$$t_0(\theta) = 0.5 + \frac{F_L(b_0) - 0.5}{b_0 - M}(\theta - M)$$

be the straight line joining the points $(M, 0.5)$ and $(b_0, F_L(b_0))$.

(A) If the straight line $t_0(\theta)$ always lies above or on $F_L(\theta)$ for $\theta \in [M, b_0]$, then we define

$$G_L(\theta) = \begin{cases} t_0(\theta) & M \leq \theta < b_0, \\ F_L(\theta) & b_0 \leq \theta \leq b. \end{cases}$$

(B) Else, if $t_0(\theta)$ crosses F_L in $[M, b_0]$, draw a tangent $t_1(\theta)$ from the point $(M, 0.5)$ on F_L between $\theta \in (M, b_0)$. Such a tangent will exist if $t_0(\theta)$ crosses F_L. Let the tangent $t_1(\theta)$ meet F_L at the point η. Now define

$$G_L(\theta) = \begin{cases} t_1(\theta) & M \leq \theta < \eta, \\ F_L(\theta) & \eta \leq \theta \leq b. \end{cases}$$

It is easy to see that any $F \in \Gamma_{SU}$ will lie between G_L and F_U for $\theta \in [M, b]$. The effective band, between $[M, b]$ is thus G_L and F_U. For notational convenience, we will keep on denoting them by F_L and F_U.

3 Linearization

For any $F \in \Gamma_{SU}$, we now have

$$
\begin{aligned}
\rho(h,F) &= \frac{\int_{[a,b]} h(\theta)\ell(\theta)dF(\theta)}{\int_{[a,b]} \ell(\theta)dF(\theta)} \\
&= \frac{\int_{(M,b]} \{h(\theta)\ell(\theta) + h(2M-\theta)\ell(2M-\theta)\}\, dF(\theta) \ + h(M)\ell(M)\{F(M) - F(M-)\}}{\int_{(M,b]} \{\ell(\theta) + \ell(2M-\theta)\}\, dF(\theta) \ + \ell(M)\{F(M) - F(M-)\}} \\
&\quad \text{(by symmetry about } M\text{)} \\
&= \frac{\int_M^b \frac{d}{d\theta} \{h(\theta)\ell(\theta) + h(2M-\theta)\ell(2M-\theta)\}\, F(\theta)d\theta \ + h(M)\ell(M)}{\int_M^b \frac{d}{d\theta} \{\ell(\theta) + \ell(2M-\theta)\}\, F(\theta)d\theta \ + \ell(M)} \qquad (5) \\
&\quad \text{(integration by parts)}
\end{aligned}
$$

By using a standard linearization argument, the problem of finding $\sup_{F \in \Gamma_{SU}} \rho(h,F)$ can be reduced to finding

$$\inf_{F \in \Gamma_{SU}} \int_M^b f_\lambda(\theta)F(\theta)d\theta \qquad (6)$$

where

$$f_\lambda(\theta) = \frac{d}{d\theta}\{h(\theta)\ell(\theta) + h(2m-\theta)\ell(2M-\theta)\} - \lambda\frac{d}{d\theta}\{\ell(\theta) + \ell(2M-\theta)\} \qquad (7)$$

and λ is any real number (see Lemma 2.2.1 of Basu and DasGupta (1990) for a formal proof).

Assumption (8) : For each fixed λ, $f_\lambda(\theta)$ changes sign a finite number of times in (M, b_0).

Suppose, for $\theta \in (M, b_0)$, $f_\lambda(\theta)$ changes sign n times, at $M < \alpha_1 < \ldots < \alpha_n < b_0$. Let $\alpha_0 = M$ and $\alpha_{n+1} = b_0$.

$$\text{Define } I_i = \begin{cases} [M, \alpha_1) & \text{for } i = 1, \\ (\alpha_n, b_0] & \text{for } i = n+1, \\ (\alpha_{i-1}, \alpha_i) & \text{for } i = 2, \ldots, n. \end{cases} \qquad (8)$$

Furthermore, label I_i as '+' if $f_\lambda(\theta)$ is nonnegative on I_i, and as '−' if it is nonpositive on I_i.

Next, for each $i = 1, \ldots, n$, fix h_i satisfying

1. $F_L(\alpha_i) \le h_i \le F_U(\alpha_i)$, and

2. $h_i \le h_j$ if $i < j$.

We will find $\inf_F \int_M^b f_\lambda(\theta)F(\theta)d\theta$ over the restricted class

$$\Gamma^*_{SU} = \{F \in \Gamma_{SU} : F(\alpha_i) = h_i, i = 1, \ldots, n\}. \tag{9}$$

Further restrictions on the h_i's :

We need at least one unimodal F to pass through the points $\{(\alpha_i, h_i), i = 1, \ldots, n\}$.

For any concave function g, defined on $[u, v]$, and for any two points $s < t \in [u, v]$; let $r(\theta)$ be the straight line joining the points $(s, g(s))$ and $(t, g(t))$. Then g satisfies

$$\begin{aligned} g(\theta) &\leq r(\theta) \quad \text{for } \theta \in [u, s), \\ &\geq r(\theta) \quad \text{for } \theta \in [s, t), \\ &\leq r(\theta) \quad \text{for } \theta \in [t, v]. \end{aligned}$$

Thus, for any pair $j < k \in \{1, \ldots, n\}$, let $r_{j,k}(\theta)$ denote the straight line joining the points (α_j, h_j) and (α_k, h_k). Then the set of h_i's ($i = 1, \ldots, n$) must satisfy

(i) $r_{j,k}(\theta) \geq F_L(\theta)$ for $\theta \in [M, \alpha_j] \cup [\alpha_k, b_0]$.

(ii) $r_{j,k}(\alpha_i) \geq h_i$ if $i < j$ or $i > k$, and
$r_{j,k}(\alpha_i) \leq h_i$ if $j \leq i \leq k$.

The method for finding the extremal prior is best illustrated through an example.

Example 1

Suppose we sample X from $N(\theta, 1)$ yielding the likelihood,

$$\ell(\theta) = \exp\left(-\frac{1}{2}(\theta - X)^2\right), \quad \theta \in \Re.$$

Let F_0 be the conjugate $N(0, 1)$ prior distribution, and we consider the class of all symmetric unimodal distributions in the ϵ-Kolmogorov neighborhood of F_0. Thus

$$\Gamma_{SU} = \left\{ \begin{array}{l} F: \ \max(0, F_0(\theta) - \epsilon) \leq F(\theta) \leq \min(1, F_0(\theta) - \epsilon), \ \ \theta \in \Re, \\ \qquad \text{F is a cumulative distribution function and} \\ \qquad \text{F is symmetric and unimodal about zero} \end{array} \right\}. \tag{10}$$

Suppose we are interested in finding the supremum of the posterior mean over Γ_{SU}. Assume $X > 0$. For $h(\theta) = \theta$, it can be shown that $f_\lambda(\theta)$ (cf. (7)) changes sign at most 2 times, say at α_1 and α_2 ($0 < \alpha_1 < \alpha_2 < \infty$), with $f_\lambda(\theta) \geq 0$ for $\theta \in (\alpha_1, \alpha_2)$ if $\lambda > 0$.

Hence for the purpose of minimizing $\int f_\lambda(\theta)F(\theta)d\theta$ over $F \in \Gamma_{SU}$, we would like to make F as small as possible in the interval (α_1, α_2) and as large as possible outside, in a way such that the resulting F is nondecreasing and concave on $(0, \infty)$.

Fix h_1 and h_2 as mentioned before, satisfying the required conditions. Let $t_1(\theta)$ be the straight line joining the points (α_1, h_1) and (α_2, h_2).

<u>Case A</u> : $t_1(\theta)$ lies above or on F_L (cf. Figure 1).

$$\text{Define } \overline{F}(\theta) = \begin{cases} \min[F_U(\theta), t_1(\theta)] & \text{if } \theta \in [0, \alpha_1), \\ t_1(\theta) & \text{if } \theta \in [\alpha_1, \alpha_2), \\ \min[F_U(\theta), t_1(\theta)] & \text{if } \theta \in [\alpha_2, \infty). \end{cases} \qquad (11)$$

We claim that $\overline{F}$ is the extremal prior in this case, which minimizes $\int f_\lambda(\theta)F(\theta)d\theta$ over Γ_{SU}.

<u>Proof</u> : We aim to make F as small as possible in (α_1, α_2) and as large as possible outside, subject to the restrictions that $F(\alpha_1) = h_1$ and $F(\alpha_2) = h_2$ and that F is nondecreasing and concave in $(0, \infty)$.

Because of concavity, any $F \in \Gamma_{SU}$ must lie on or above $t_1(\theta)$ for $\theta \in (\alpha_1, \alpha_2)$, and on or below $t_1(\theta)$ outside. Also, the class Γ_{SU} restricts F between F_L and F_U.

This clearly indicates $\overline{F}$ to be the extremal prior, subject to it being concave on $(0, \infty)$, which, indeed, it is.

<u>Case B</u> : $t_1(\theta)$ crosses F_L (twice) (cf. Figure 2).

1. Draw a tangent $t_1^{(1)}(\theta)$ from the point (α_1, h_1) on $F_L(\theta)$ between $\theta \in [\alpha_1, \alpha_2]$. Let $t_1^{(1)}(\theta)$ and $F_L(\theta)$ meet at α_{12}.

2. Similarly, draw another tangent $t_1^{(2)}(\theta)$ from the point (α_2, h_2) on $F_L(\theta)$ between $\theta \in [\alpha_1, \alpha_2]$. Let $t_1^{(2)}(\theta)$ and $F_L(\theta)$ meet at α_{21}.

It can be shown that there indeed exist $t_1^{(1)}(\theta)$ and $t_1^{(2)}(\theta)$ satisfying the above, and $\alpha_1 \leq \alpha_{12} < \alpha_{21} \leq \alpha_2$.

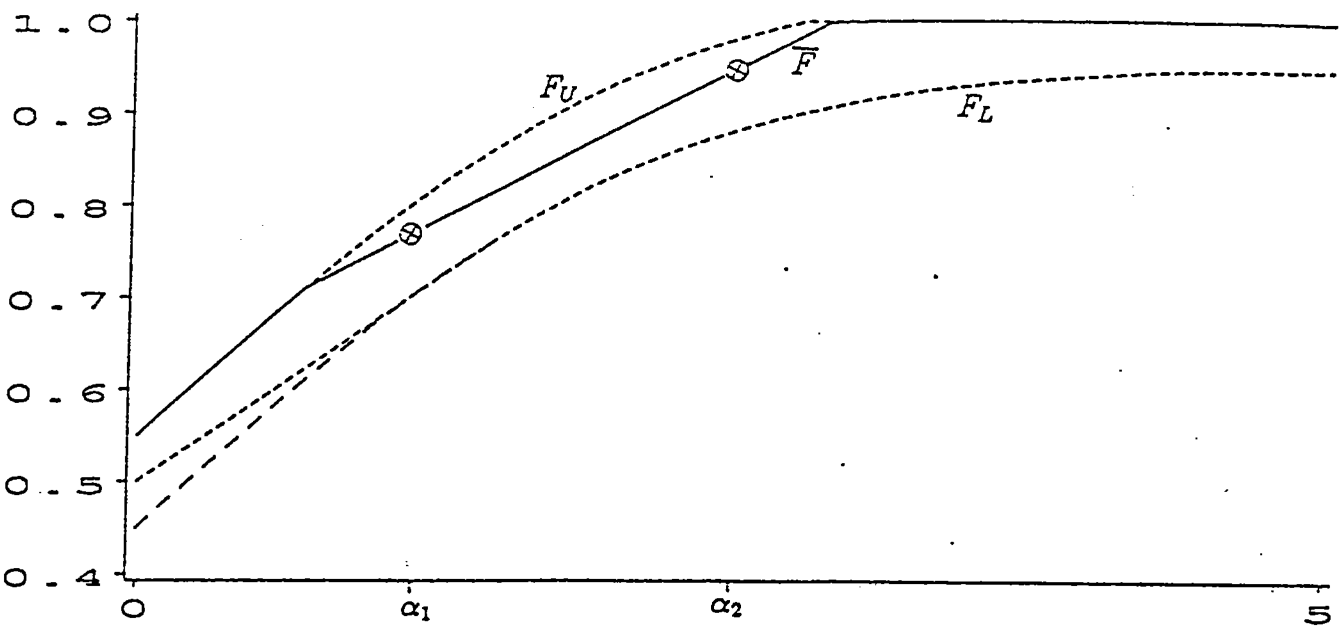

Figure 1: Extremal $\overline{F}$ in Case A. The dashed lines are F_L and F_U. The solid line is the extremal $\overline{F}$.

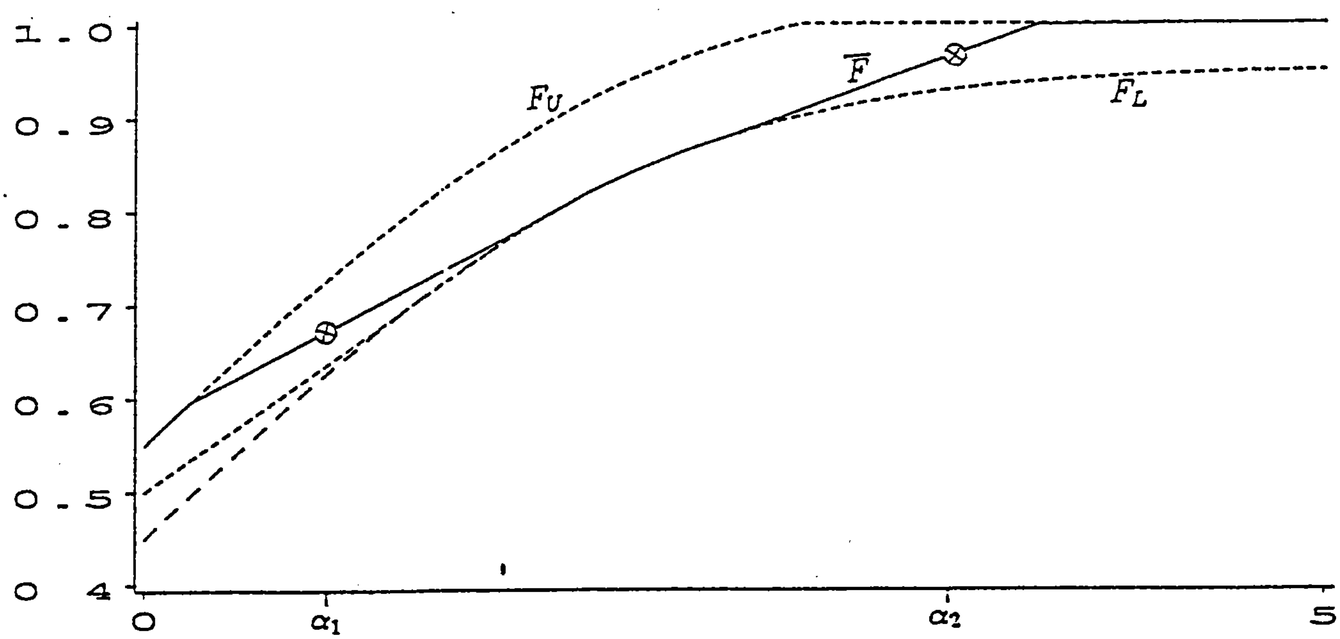

Figure 2: Extremal $\overline{F}$ in Case B. The dashed lines are F_L and F_U. The solid line is the extremal $\overline{F}$.

Now, define

$$\overline{F}(\theta) = \begin{cases} \min[F_U(\theta), t_1^{(1)}(\theta)] & \text{if } \theta \in [0, \alpha_1), \\ t_1^{(1)}(\theta) & \text{if } \theta \in [\alpha_1, \alpha_{12}), \\ F_L(\theta) & \text{if } \theta \in [\alpha_{12}, \alpha_{21}), \\ t_1^{(2)}(\theta) & \text{if } \theta \in [\alpha_{21}, \alpha_2), \\ \min[F_U(\theta), t_1^{(2)}(\theta)] & \text{if } \theta \in [\alpha_2, \infty). \end{cases} \tag{12}$$

We now claim that $\overline{F}$ is the extremal prior in this case.

Proof : For minimizing $\int f_\lambda(\theta)F(\theta)d\theta$, we want to make $F(\theta)$ as small as possible for $\theta \in (\alpha_1, \alpha_2)$. Hence, to beat the performance of $\overline{F}$ in the interval (α_1, α_{12}), an F passing through (α_1, h_1) must lie below $t_1^{(1)}(\theta)$ for at least a non-empty interval $\subseteq [\alpha_1, \alpha_{12}]$. Such an F will surely meet $t_1^{(1)}(\theta)$ on or before α_{12}, say at β_1. Thus, the straight line $t_1^{(1)}(\theta)$ meets F at two points, α_1 and β_1, and lies above F in between. This clearly violates the concavity of F.

Any concave F thus, must lie on or above $\overline{F}$ in the interval $[\alpha_1, \alpha_{12}]$. Now, to beat the performance of $\overline{F}$ in the interval $[0, \alpha_1)$, an F must lie above $\overline{F}$ for a non-empty interval $\subseteq [0, \alpha_1)$, pass through the point (α_1, h_1), and lie on or above $\overline{F}$ in the interval (α_1, α_{12}). It is easy to see that such an F cannot be concave.

The arguments for showing that $\overline{F}$ is the best in the other intervals too are very similar.

We have, thus, solved the problem of minimizing $\int f_\lambda(\theta)F(\theta)d\theta$ in the restricted class $\Gamma_{SU}^* = \{F \in \Gamma_{SU} : F(\alpha_1) = h_1, F(\alpha_2) = h_2\}$. For finding the overall optimal $\overline{\overline{F}}$ which maximizes the posterior mean $\rho(h, F) = E^{F(\theta|X)}(\theta)$ over the class Γ_{SU}, what remains is a 3-dimensional numerical maximization, namely maximizing $\rho\left(h, \overline{F}(\lambda, h_1, h_2)\right)$ over λ, h_1 and h_2.

Remark : The ϵ-Kolmogorov neighborhood of $F_0 = N(0,1)$ does not satisfy assumption (2), namely $\lim_{\theta\to\infty} F_L(\theta) = \lim_{\theta\to\infty} \max(F_0(\theta) - \epsilon, 0) \neq 1$, and $\lim_{\theta\to-\infty} F_U(\theta) \neq 0$. However, this problem can easily be circumvented by standard limiting arguments.

4 The General Case

The basic argument given in the above example applies to the general situation. We present here the form of the extremal prior in the general set-up, without proof.

Recall that the original general problem was reduced to finding $\inf_{F \in \Gamma^*_{SU}} \int_M^b f_\lambda(\theta)F(\theta)d\theta$ (cf. (9)), and that we assumed that $f_\lambda(\theta)$ changes sign n times, at $M < \alpha_1 < \ldots < \alpha_n < b_0$.

The determination of the extremal prior is done in a few steps.

Step 1 : $I_i = (\alpha_{i-1}, \alpha_i)$ (cf. (8)) has label '+', and I_i is not at end, i.e. $i \neq 1$ or $n+1$.

Let $t_i^{(0)}(\theta)$ be the straight line joining the points (α_{i-1}, h_{i-1}) and (α_i, h_i).

Case A : $t_i^{(0)}(\theta) \geq F_L(\theta)$ for $\theta \in [\alpha_{i-1}, \alpha_i]$.

Define $G_i(\theta) = t_i^{(0)}(\theta) \quad \theta \in [M, b_0]$.

Case B : $t_i^{(0)}(\theta) < F_L(\theta)$ for some $\theta \in [\alpha_{i-1}, \alpha_i]$.

1. Draw a tangent $t_i^{(1)}(\theta)$ from the point (α_{i-1}, h_{i-1}) on $F_L(\theta)$ between $\theta \in [\alpha_{i-1}, \alpha_i]$. Let $t_i^{(1)}(\theta)$ and $F_L(\theta)$ meet at $\alpha_i^{(1)} \in [\alpha_{i-1}, \alpha_i]$.

2. Similarly, draw another tangent $t_i^{(2)}(\theta)$ from the point (α_i, h_i) on $F_L(\theta)$ between $\theta \in [\alpha_{i-1}, \alpha_i]$. Let $t_i^{(2)}(\theta)$ and $F_L(\theta)$ meet at $\alpha_i^{(2)} \in [\alpha_{i-1}, \alpha_i]$.

It can again be shown that $t_i^{(1)}(\theta)$ and $t_i^{(2)}(\theta)$ can indeed be drawn and $\alpha_{i-1} \leq \alpha_i^{(1)} < \alpha_i^{(2)} \leq \alpha_i$. Now, define

$$G_i(\theta) = \begin{cases} t_i^{(1)}(\theta) & \text{for } \theta \in [M, \alpha_i^{(1)}), \\ F_L(\theta) & \text{for } \theta \in [\alpha_i^{(1)}, \alpha_i^{(2)}), \\ t_i^{(2)}(\theta) & \text{for } \theta \in [\alpha_i^{(2)}, b_0]. \end{cases} \tag{13}$$

This ends Step 1.

Step 2 : I_i has label '+' and I_i is at end, i.e. $i = 1$ or $n+1$.

Suppose $i = 1$, i.e. $I_i = [M, \alpha_1)$.

Let $t_1^{(0)}(\theta)$ be the straight line joining the points $(M, 0.5)$ and (α_1, h_1).

If $t_1^{(0)}(\theta) \geq F_L(\theta) \ \ \forall \theta \in [M, \alpha_1]$, then define $G_1(\theta) = t_1^{(0)}(\theta), \theta \in [M, b_0]$.

Else, if $t_1^{(0)}(\theta)$ crosses $F_L(\theta)$ between $[M, \alpha_1]$, draw a tangent $t_1^{(2)}(\theta)$ from the point (α_1, h_1) on F_L between $\theta \in [M, \alpha_1]$. Let $t_1^{(2)}(\theta)$ and F_L meet at $\alpha_1^{(2)} \in [M, \alpha_1]$. Now define,

$$G_1(\theta) = \begin{cases} F_L(\theta) & \text{for } \theta \in [M, \alpha_1^{(2)}), \\ t_1^{(2)}(\theta) & \text{for } \theta \in [\alpha_1^{(2)}, b_0]. \end{cases}$$

The case of $i = n+1$ is handled similarly.

Step 3 : I_i has label '$-$'.

If I_i is not at end ($i \neq 1, n+1$), then I_{i-1} and I_{i+1} have labels '+', and for those, G_{i-1} and G_{i+1}, are already defined by Step 1 and 2.

Define $G_0(\theta) = G_{n+2}(\theta) = F_U(\theta), \;\; \theta \in [M, b_0]$, and now define

$$G_i(\theta) = \min\left[G_{i-1}(\theta), F_U(\theta), G_{i+1}(\theta)\right], \quad \theta \in [M, b_0].$$

Finally, the extremal prior $\overline{F}$ which minimizes $\int_M^b f_\lambda(\theta) F(\theta) d\theta$ over Γ_{SU}^* is given by,

$$\overline{F}(\theta) = \begin{cases} h_i & \text{if } \theta = \alpha_i, \;\; i = 1, \ldots, n, \\ G_i(\theta) & \text{if } \theta \in I_i, \;\; i = 1, \ldots, n+1, \\ F_U(\theta) & \text{if } \theta \in (b_0, b]. \end{cases} \tag{14}$$

Remark : For finding the optimizing $\overline{\overline{F}}$ which maximizes $\rho(h, F)$ over $F \in \Gamma_{SU}$, the optimum values of α_i's and h_i's have to be determined through numerical optimization. The α_i's are determined by λ (see definition of α_i's), so the finite-dimensional numerical maximization would be over λ and the h_i's.

5 Application: Ranges of Posterior Mean

In this section, we apply the method described earlier to two examples to find the ranges of the posterior mean over two specific prior classes of symmetric and unimodal distributions.

5.1 Example 1 continued

Let $X \sim N(\theta, 1)$. Let F_0, as before, denote the conjugate $N(0,1)$ prior cdf.

In Table 1, we list down the values of the supremum $(\overline{E}, \overline{E}_S, \overline{E}_{SU})$ of the posterior mean $E^F(\theta|X)$ as F varies in the following 3 prior classes :

$$\begin{aligned} \Gamma &= \epsilon\text{-Kolmogorov neighborhood of } F_0. \\ \Gamma_S &= \{F \in \Gamma : \text{F is symmetric about zero}\}. \\ \Gamma_{SU} &= \{F \in \Gamma : \text{F is symmetric and unimodal about zero}\}. \end{aligned}$$

Only the values of the supremum are reported. Due to the underlying symmetry of the classes, it follows that

$$\inf_{F \in \mathcal{F}} E^F(\theta|X) = -\sup_{F \in \mathcal{F}} E^F(\theta|-X) \quad \text{for } \mathcal{F} = \Gamma, \Gamma_S, \Gamma_{SU}. \tag{15}$$

The first observation from the table is, the imposed constraints of symmetry and symmetry and unimodality, do not help much in reducing the ranges. As commented in Basu and Dasgupta (1990), when the absolute value of the observed X is large, the difference between the supremum $(\overline{E})$ and infimum $(\underline{E})$ of the posterior mean $E^F(\theta|X)$ over the ϵ-Kolmogorov neighborhood Γ is rather large. For example, if $X = 2$ and $\epsilon = 0.1$, then $E^{F_0}(\theta|X) = 1, \underline{E} = 0.258$ and $\overline{E} = 1.623$, giving a range of 1.365. For $X = 4$ and the same ϵ, the values are respectively 0.483 and 5.815, giving a range of 5.332. We expected that the additional constraints of symmetry and symmetry and unimodality will greatly reduce the ranges, by deleting the unreasonable priors from the considered prior class. But as can be seen from table 1, such is not the case. For $X = 2$ and $\epsilon = 0.1$, the values in the class Γ_S are $\underline{E}_S = 0.2764$ and $\overline{E}_S = 1.521$, with a range of 1.2446. For the symmetric unimodal class Γ_{SU}, the values are $\underline{E}_{SU} = 0.2764$ and $\overline{E}_{SU} = 1.4768$, with a range of 1.2004. Thus, imposing the added constraints only lead to a minor reduction in the range. For $X = 4$, $\underline{E}_{SU} = 0.4957$ and $\overline{E}_{SU} = 3.9462$ with a range of 3.4505 in the class Γ_{SU}. Hence, the imposed shape constraints reduces the range quite a bit for large $|X|$ values, but still they fail to achieve robustness. This is primarily due to our choice of F_0 as $N(0,1)$, the resulting class Γ_{SU} contains priors with extremely thin as well as extremely thick tails, even after the imposed shape restrictions. Robustness is generally lacking when there is such a wide variety of tails and the likelihood $\ell(\theta)$ is concentrated in the tail.

5.2 Example 2

Let $X \sim$ Binomial(n, θ) yielding the likelihood $\ell(\theta) \approx \theta^X(1-\theta)^{n-X}$, $\theta \in [0,1]$, $X \in \{0,1,\ldots,n\}$. Take F_0 to be the conjugate symmetric Beta(β,β) cdf. Let Γ denote the ϵ-Kolmogorov neighborhood of F_0 and let

$$\Gamma_{SU} = \{F \in \Gamma : \text{F is symmetric and unimodal at } 0.5\}.$$

Remark : The resulting $F_L(\theta) = \min(1, F_0(\theta) - \epsilon)$ does not satisfy assumption (2), namely F_L is not concave on $(0.5, 1]$, since it has a point mass at 1. Same is true for F_U, due to its point mass at 0. But this problem can easily be circumvented by considering the effective band generated by symmetric and unimodal distributions in between F_L and F_U (similar to what has been done for $F_L(M) < 0.5$ in the general setup).

For $n = 5$ and $\beta = 2$, Table 2 shows the supremum of the posterior mean $E^F(\theta|X)$ over the two classes, Γ ($\overline{E}$) and Γ_{SU} ($\overline{E}_{SU}$), for two different values of ϵ. Table 3 shows the supremum ($\overline{E}_{SU}$) for the class Γ_{SU} with different choices of the β parameter for the base prior $F_0 =$ Beta(β, β). As before, only the supremum values are reported, since

$$\inf_{F\in\mathcal{F}} E^F(\theta|X) = 1 - \sup_{F\in\mathcal{F}} E^F(\theta|n-X) \quad \text{for } \mathcal{F} = \Gamma, \Gamma_{SU}. \tag{16}$$

In our earlier discussion on this example in Basu and DasGupta (1990), we observed that the ranges of posterior mean $E^F(\theta|X)$ over the Kolmogorov neighborhood Γ is, in fact, large for extreme X-values, whereas for X-values in the middle, which are compatible with the base prior F_0, the ranges are small. From table 2, we observe that, contrary to the results of Example 1, the added constraints of symmetry and unimodality, indeed, reduce the ranges significantly in this example. Whereas in the Kolmogorov neighborhood Γ, the infimum ($\underline{E}$) and supremum ($\overline{E}$) of $E^F(\theta|X)$ for $X = 2$ and $\epsilon = 0.10$ are 0.376 and 0.527 respectively, giving a range of 0.151, with the additional shape constraints, the infimum ($\underline{E}_{SU}$) and supremum ($\overline{E}_{SU}$) over the class Γ_{SU} are respectively 0.4286 and 0.4702, thus reducing the range from 0.151 to 0.0416. Such a reduction can be observed through out the table. But still we fail to achieve robustness for extreme X-values even in the constrained class Γ_{SU}. For example, for $X = 1$, the range of $E^F(\theta|X)$ over Γ_{SU} is 0.0662 even for $\epsilon = 0.05$, though it is reduced from the range of 0.0980 of the unconstrained class Γ.

From table 3, it can be seen that as the tail of the base prior F_0 gets sharper (as β increases), the ranges of $E^F(\theta|X)$ for X values in the middle gets smaller. But for $X = 1$ or 4, the reduction of the ranges is in a much smaller scale. Had we had more extreme X-values, it is expected that the ranges will increase with sharper tails of F_0.

6 Discussion and summary

Much of previous work on Bayesian robustness has dealt with prior densities such as density bands and their modifications. In this article we have proposed specifying the prior through its cumulative distribution function, which we feel is more intuitive and can directly be assessed from prior probability considerations. An advantage of the family described here is that flexibility in the prior tail is achieved very easily. Indeed, it seems that for Kolmogorov and Lèvy neighborhoods, even a small ϵ gives a wide variation in prior tails.

The choice of a prior class Γ depends on the two competing goals of including all reasonable priors as well as not including the unreasonable priors. As mentioned in section 1, a distribution band, while containing the reasonable priors, can be "too big", in the sense of containing unreasonable priors which artificially inflate the ranges of the posterior criteria. The further restriction of symmetry and unimodality on the distribution band seems 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 such a class, of course, is dependent on believing that symmetry and unimodality are reasonable.

As is clear from example 1 and 2, the degree of robustness present, in any given situation, can depend heavily on the observed value of X. Indeed, in example 1, the further shape restrictions of symmetry and unimodality, fail to remove the lack of robustness for extreme X values, and even for moderate X values, the reduction in the ranges of the posterior mean is minor. In example 2 though, the restriction of the distribution band to symmetric and unimodal distributions, indeed, leads to major reductions in the ranges of the posterior mean.

When robustness fails to achieve for a given class Γ, one must reconsider the subjective inputs. In particular, further refinement of Γ 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.

Acknowledgement I am deeply grateful to Anirban DasGupta for his valuable advice and suggestions throughout this work.

References

[1] Basu, S. and DasGupta, A. (1990), "Bayesian analysis under distribution bands," Technical Report, **90-48**, Department of Statistics, Purdue University.

[2] Basu, S. (1991), "Robustness of Bayesian and Classical Inference under distribution bands and shape restricted families," Ph.D. Thesis, Purdue University.

[3] Berger, J. and Berliner, L.M. (1986), "Robust Bayes and empirical Bayes Analysis with ϵ-contaminated priors," *Ann. Statist.*, **14**, pp. 461-486.

[4] Berger, J. and O'Hagan, A. (1989), "Ranges of posterior probabilities for the class of unimodal priors with specified quantiles," in *Bayesian Statistics 3*,J.M. Bernardo, M.H. Degroot, D.V. Lindley and A.F.M. Smith (ed.), New York: Oxford University Press.

[5] Berger, J. (1990), "Robust Bayesian Analysis : sensitivity to the prior," *Jour. Stat. Planning and Inf.*, **25**, pp. 303-328.

[6] Berliner, L.M. and Goel, P. (1990), "Incorporating partial prior information : ranges of posterior probabilities," in *Bayesian and Likelihood Methods in Statistics and Econometrics*,S. Geisser, J.S. Hodges, S.J. Press and A. Zellner (ed.), , pp. 397-406, Elsevier Science Publishers B.V. (North-Holland).

[7] DasGupta, A. and Studden, W.J. (1988), "Robust Bayesian Analysis and optimal experimental designs in normal linear models with many parameters - I," Technical Report, **88-14**, Department of Statistics, Purdue University.

[8] DasGupta, A. and Studden, W.J. (1989), "Robust Bayes estimates of normal means," *Statist. and Decisions*, **7**, pp. 333-361.

[9] DasGupta, A. and Delampady, M. (1990b), "Bayesian hypothesis testing with symmetric and unimodal priors," Technical Report, **90-43**, Department of Statistics, Purdue University.

[10] DeRobertis, L., Hartigan, J.A., and 1981 "Bayesian inference using intervals of measure," *Ann. Statist.*, **9**, pp. 235-244.

[11] Huber, P.J. (1973), "The use of choquet capacities in statistics," *Bull. Internat. Statist. Inst.*, **45**, pp. 181-191.

[12] Lavine, M. (1987), "Prior influence in Bayesian Statistics," Ph.D. Thesis, University of Minnesota.

[13] Lavine, M. (1991), "Sensitivity in Bayesian Statistics: The Prior and the Likelihood," *J. of the Amer. Statist. Assoc.*, **86**, pp. 396-399.

[14] Lavine, M. (1991), "An Approach to Robust Bayesian Analysis for Multidimensional Parameter Spaces," *J. of the Amer. Statist. Assoc.*, **86**, pp. 400-403.

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

[16] Leamer, E.E. (1982), "Sets of posterior means with bounded variance priors," *Econometrica*, **50**, pp. 725-736.

[17] Polasek, W. (1984), "Multivariate regression systems - estimation and sensitivity analysis of two-dimensional data," in *Robustness of Bayesian Analyses*,J.B. Kadane (ed.)

[18] Sivaganesan, S. and Berger, J. (1989), "Ranges of posterior measures for priors with unimodal contamination," *Ann. Statist.*, **17**, pp. 868-889.

[19] Srinivasan, C. and Truszczynska, H. (1990), "Approximation to the range of a ratio-linear posterior quantity based on Frechet derivative," Technical Report, **289**, Department of Statistics, University of Kentucky.

[20] Srinivasan, C. and Truszczynska, H. (1990), "On the ranges of posterior quantities," Technical Report, **294**, Department of Statistics, University of Kentucky.

[21] Wasserman, L.A. and Kadane, J.B. (1990), "Bayes' theorem for choquet capacities," *Ann. Statist.*, **18**, pp. 1328-1339.

Table 1: Supremum of posterior mean over the prior classes Γ, Γ_S and Γ_{SU}, the three neighborhoods of $N(0,1)$. The likelihood is $N(X,1)$.

		Prior classes								
		$\epsilon = 0.01$			$\epsilon = 0.05$			$\epsilon = 0.10$		
X	E^{F_0}	Γ	Γ_S	Γ_{SU}	Γ	Γ_S	Γ_{SU}	Γ	Γ_S	Γ_{SU}
-4	-2.0	-1.454	-1.456	-1.468	-0.871	-0.882	-0.897	-0.483	-0.496	-0.496
-3	-1.5	-1.222	-1.226	-1.230	-0.748	-0.760	-0.760	-0.393	-0.397	-0.397
-2	-1.0	-0.889	-0.890	-0.890	-0.558	-0.560	-0.560	-0.258	-0.276	-0.276
-1	-0.5	-0.453	-0.456	-0.457	-0.266	-0.294	-0.296	-0.048	-0.140	-0.141
0	0.0	0.033	0.0	0.0	0.154	0.0	0.0	0.298	0.0	0.0
1	0.5	0.537	0.534	0.532	0.658	0.645	0.631	0.814	0.767	0.747
2	1.0	1.066	1.062	1.061	1.312	1.281	1.266	1.623	1.521	1.477
3	1.5	1.738	1.733	1.705	2.548	2.511	2.194	3.409	3.296	2.533
4	2.0	3.214	3.213	2.851	5.010	4.999	3.820	5.815	5.772	3.946
5	2.5	6.031	6.031	4.832	7.424	7.422	4.991	8.085	8.069	4.999

Table 2: Supremum of posterior mean over the prior classes Γ, and Γ_{SU}, the two neighborhoods of Beta(2,2). The likelihood is Binomial$(5,\theta)$.

		Prior classes			
		$\epsilon = 0.05$		$\epsilon = 0.10$	
X	$E^{F_0}(\theta\|X)$	Γ	Γ_{SU}	Γ	Γ_{SU}
1	0.3333	0.3940	0.3695	0.4540	0.4012
2	0.4444	0.4860	0.4605	0.5270	0.4702
3	0.5555	0.5920	0.5690	0.6240	0.5714
4	0.6666	0.7040	0.6967	0.7400	0.7143

Table 3: Supremum of posterior mean ($\overline{E}_{SU}$) over Γ_{SU}, symmetric unimodal Kolmogorov neighborhood, of different Beta(β, β) priors. The likelihood is Binomial($5, \theta$) and $\epsilon = 0.05$.

	$\beta = 1.5$		$\beta = 2.0$		$\beta = 4.0$		$\beta = 6.0$	
X	E^{F_0}	$\overline{E}_{SU}$	E^{F_0}	$\overline{E}_{SU}$	E^{F_0}	$\overline{E}_{SU}$	E^{F_0}	$\overline{E}_{SU}$
1	0.3125	0.3486	0.3333	0.3695	0.3846	0.4180	0.4118	0.4409
2	0.4375	0.4544	0.4444	0.4605	0.4615	0.4736	0.4706	0.4802
3	0.5625	0.5771	0.5555	0.5690	0.5385	0.5487	0.5294	0.5373
4	0.6875	0.7355	0.6666	0.6967	0.6154	0.6411	0.5882	0.6123