BAYESIAN ANALYSIS UNDER DISTRIBUTION BANDS: THE ROLE OF THE LOSS FUNCTION *

by

Sanjib Basu, University of California, Santa Barbara

Anirban DasGupta, Purdue University

Technical Report #90-48

Department of Statistics
Purdue University

August 1990
Revised March 1992

* Research supported by NSF Grant DMS:89-230-71. Some part of this work was part of the Ph.D. dissertation of the first author at Purdue University. This report also appears as Technical Report No. 208 in the Department of Statistics and Applied Probability at University of California, Santa Barbara.

o

BAYESIAN ANALYSIS WITH DISTRIBUTION BANDS : THE ROLE OF THE LOSS FUNCTION [1]

by

Sanjib Basu

University of California, Santa Barbara

and

Anirban DasGupta

Purdue University

March 1992

Abstract

Given a random variable with distribution indexed by a one dimensional parameter θ, robustness of Bayesian analysis with respect to variations in both the prior and the loss function is considered when the prior cdf of θ lies in the band $\Gamma = \{F : F_L \leq F \leq F_U\}$. Such a prior 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 ranges of ratio-linear posterior quantities over Γ with applications to several examples. In special cases, determination of the ranges involves almost no numerical work. This method is used to compare the sensitivity of different Bayes estimates arising out of different loss functions, again with illustration in an example.

Key words : cdf, distribution band, robust Bayes, Kolmogorov metric, Lèvy metric, posterior mean, posterior fractile, squared error, absolute error, linear loss

AMS 1991 subject classifications. 62F15, 62G35, 62C10

---

[1]Research supported by NSF Grant DMS:89-230-71. Some part of this work was part of the Ph.D. dissertation of the first author at Purdue University.

1 Introduction

1.1 The problem and motivation.

Sensitivity of Bayesian analysis to the choice of prior has recently received considerable attention. It is now generally agreed that a convenient initial prior may be formed incorporating the features that are easy to elicit, followed by a sensitivity analysis when the prior is allowed to vary in a "neighborhood" of the initial prior. Neighborhoods have been specified in the literature in a variety of ways and indeed need not be metric neighborhoods. Past work in this area has concentrated on few different types of prior families such as parametric families (like conjugate priors), contamination classes, density bands and densities with a few specified percentiles.

The class Γ of cumulative prior distributions to be considered in this article is

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

where F_L and F_U are two fixed cdfs (or more generally, F_L and F_U are nondecreasing right continuous functions satisfying $0 \leq F_L, F_U \leq 1$) and, $F_L(\theta) \leq F_U(\theta)$ for all θ. We will refer to the family Γ in (1) as a *distribution band.*

In Bayesian analysis, the likelihood $\ell(\theta)$ and the prior $\pi(\theta)$ are combined through Bayes theorem to obtain the posterior distribution for the parameter(s) of interest. Thus, it requires specification of two parametric models, the likelihood and the prior. Bayesian robustness studies, so far however, are directed almost exclusively towards sensitivity to the choice of the prior. In some sense, sensitivity to the prior and the likelihood can be treated in a unified manner, by considering the prior as a probability measure on the space of possible sampling distributions; recent work on this includes Lavine (1991).

In a decision-theoretic framework, a third component, namely the loss function $L(\theta, a)$, enters the picture, specification of which is again of concern. Surprisingly enough, the literature on Bayesian robustness so far has ignored the question of sensitivity of Bayesian analysis to perturbations in the specified loss structure. Use of squared error loss $(L(\theta, a) = (\theta - a)^2)$ and conjugate priors are quite common, mainly because of their mathematical ease. Whereas much work has been done on the sensitivity of the Bayes estimator and the Bayes risk (under squared error loss) to deviations from the conjugate prior formulation, such sensitivity analysis towards deviations from the specific $L(\theta, a) = (\theta - a)^2$ loss structure is lacking.

A formal approach towards robustness analysis to variations in both the prior $\pi(\theta)$ (or prior cdf $F(\theta)$) and the loss $L(\theta, a)$ would be to let the prior vary in a class Γ and, to take a family (probably nonparametric) of loss functions $\mathcal{L}$ and, find the ranges of interesting posterior quantities or Bayes rules. We take a less formal approach and consider a few plausible loss structures (such as squared error, absolute error and others), let the prior cdf $F(\theta)$ lie in a nonparametric family of priors, namely the distribution band defined in (1), and study the ranges of the Bayes rules w.r.t. the different specified losses. If robustness is achieved for the Bayes rule $\delta_{L_1}(X)$ over the prior class Γ for a specified loss 'L_1', by comparing the behavior of $\delta_{L_1}(X)$ with Bayes rule $\delta_{L_2}(X)$ w.r.t. the other loss 'L_2', we can see if the exact functional form of the loss 'L_1' is important/crucial in achieving robustness and, whether some types of losses behave better than others in this respect.

We will restrict our attention to the following three standard losses :

$$\begin{aligned}
L_1(\theta, a) &= (\theta - a)^2, \quad \text{the squared error loss,} \\
L_2(\theta, a) &= |\theta - a|, \quad \text{the absolute error loss, and} \\
L_3(\theta, a) &= \begin{cases} K_0(\theta - a) & \text{if} \quad \theta - a \geq 0 \\ K_1(a - \theta) & \text{if} \quad \theta - a < 0 \end{cases} \qquad (2) \\
&\quad \text{the linear loss (with } K_0 \text{ and } K_1 \text{ appropriately chosen).}
\end{aligned}$$

Note that L_2 is a special case of L_3 with $K_0 = K_1 = 1$. Obviously, these are not the only possible loss functions that may occur. We choose them mainly because of their mathematical tractability and also, real-life loss functions arising out of a utility analysis can often be suitably approximated by them. The Bayes rules w.r.t. these losses for a specified prior F and a fixed likelihood $\ell_X(\theta)$ are respectively : $\delta_{L_1}(X) = E^{F(\theta|X)}(\theta)$, the posterior mean; $\delta_{L_2}(X) =$ any median of the posterior distribution $F(\theta|X)$; and $\delta_{L_3}(X) =$ any $\frac{K_0}{K_0+K_1}(=\gamma)$ fractile of $F(\theta|X)$ (cf. Berger (1985)).

The problem of finding the ranges of Bayes rules for the three losses L_1, L_2 and

L_3 over a prior class Γ thus reduces to finding ranges of the posterior mean $E^{F(\theta|X)}(\theta)$ and posterior fractiles over the same class Γ. In section 2, we describe a general method for finding the ranges of $E^{F(\theta|X)}(h(\theta))$, for any arbitrary functional $h(\theta)$ which is continuously differentiable. This methodology is then applied to three specific examples for finding the ranges of posterior mean. For finding the ranges of posterior fractiles, however, some additional definitions and results are needed and this development is deferred till section 3.

1.2 Discussion on the class Γ:

An important aspect of prior quantification is that prior probabilities of sets and shape features of the prior are much easier to elicit, while features such as moments and functional forms are much harder to determine. Previous approaches to Bayesian robustness attempt to incorporate the uncertainty in terms of classes of densities. The important distinction of the distribution band is that here, we model the uncertainty of elicitation directly in terms of prior probabilities of sets of the form $(-\infty, \theta]$ (which we commented are easier to elicit), and let it lie between $F_L(\theta)$ and $F_U(\theta)$. We feel this arises more naturally from the elicitation mechanism.

The distribution band Γ is also very flexible in the sense that it can easily be adapted to meet specific subjective inputs from the user in a straightforward manner. As an example, suppose elicitation process specifies the prior median and quartiles of an one dimensional parameter θ at 0 and ± 1 respectively. Several priors, for example, $F_N = N(0, 2.19)$, $F_C = \text{Cauchy}(0,1)$ and many others (see Table 1), meet these quantile specifications. A formal or nonparametric approach would be to define a class of priors subject to these quantile specifications and do a sensitivity analysis, but this approach often suffers from the inclusion of unreasonable priors in the prior class. Many Bayesians argue for an informal approach, of considering a few plausible priors $F_1, \ldots, F_n$ (i.e., in this example, N(0,2.19), Cauchy(0,1) and few others) and checking for sensitivity only among these priors. One way to combine these two approaches is to define $F_L = \min(F_1, \ldots, F_n)$ and $F_U = \max(F_1, \ldots, F_n)$ and do sensitivity studies with the resulting distribution band Γ, which has the flavor of the informal approach while preserving the essence of the nonparametric approach.

Another important aspect of the distribution band is that neighborhoods under various metrics on the space of distribution functions are often of the form (1). Suppose we combine the outputs of a finite elicitation process through a reasonable (and preferably tractable) distribution function F_0. A natural way to incorporate the uncertainty in the elicitation process may then be to allow an error of ϵ in the specification, leading thus to the class (1), with $F_L^K(\theta) = \max[F_0(\theta) - \epsilon, 0]$ and $F_U^K(\theta) = \min[F_0(\theta) - \epsilon, 1]$. Indeed, this is 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 (1) is the Lèvy metric, defined as $d_L(F_1, F_2) = \inf\{\epsilon : F_1(\theta - \epsilon) - \epsilon \leq F_2(\theta) \leq F_1(\theta + \epsilon) + \epsilon\}$. In this case, using the notation of (1), $F_L^L(\theta) = \max[F_0(\theta - \epsilon) - \epsilon, 0]$ and $F_U^L(\theta) = \min[F_0(\theta + \epsilon) + \epsilon, 1]$. Notice that F_L^L and F_U^L thus defined, are not cdfs, since $\lim_{\theta \to \infty} F_L^L(\theta) \neq 1$ and $\lim_{\theta \to -\infty} F_U^L(\theta) \neq 0$ (same is true for F_L^K and F_U^K). But our methodology for finding the ranges of Bayes rules and posterior quantities can be suitably adapted to treat the class Γ (in (1)) generated by F_L^L and F_U^L (or by F_L^K and F_U^K).

Kolmogorov and Lèvy neighborhoods have been widely used in classical statistics, see for example, Huber (1981) and Wiens (1986). As par our knowledge, this is the first such attempt in Bayesian analysis. It is well known that $d_L(F_1, F_2) \leq d_K(F_1, F_2)$ (for any F_1, F_2) so that Lèvy neighborhoods are bigger than Kolmogorov neighborhoods. An advantage of the Kolmogorov and Lèvy neighborhoods is that these families often allow for flexibility in the prior tail for small values of ϵ. For example, the Kolmogorov distance between F_N and F_C is 0.081. Thus, a Kolmogorov neighborhood of $F_N = N(0, 2.19)$ with $\epsilon = 0.081$ includes the flat tailed F_C = Cauchy(0,1) prior. Table 1 gives the Kolmogorov and Lèvy distances between $N(0, 2.19)$ and several other natural priors, each having median = 0 and quartiles = ± 1. It is interesting to see that ϵ as small as 0.03 suffices for covering flat priors like a t prior with 3 degrees of freedom.

Yet another interesting feature of the family (1) is that <u>any</u> prior class Γ' can be embedded into a family Γ of the form (1) by defining $F_L(\theta) = \inf\{F(\theta) : F \in \Gamma'\}$ and $F_U(\theta) = \sup\{F(\theta) : F \in \Gamma'\}$. If F_L and F_U defined in this manner are distribution functions, the corresponding distribution band Γ will include Γ' and thus the ranges of $\rho(x, F)$ obtained from Γ can be used as bounds on $\rho(x, F)$ for Γ'. This may be particularly useful when the original family Γ' is mathematically intractable. Nevertheless if robustness obtains with the bigger family Γ, we know robustness is present for Γ' too.

The principal goals of this paper are then the following :

(i) to use neighborhoods of cdf's as models for prior uncertainty; such neighborhoods have been used in the classical robustness literature before, but their use in Bayes theory is new;

(ii) to investigate the formal effect of a loss function in this setup within a class of commonly used loss functions; we believe this is rather important;

(iii) to bridge together the ideas of informal and formal robustness checks by studying CDF bands of the type $\min(F_1, F_2) \leq F \leq \max(F_1, F_2)$, where F_1, F_2 may be adhoc candidates for a robustness check.

Moreover, at a technical level, our results demonstrate further potential of moment theory methods in this area.

1.3 History

The "robust Bayesian" view of quantifying subjective information through a class of priors has been exposed by many, which is too broad to be reviewed here. Excellent reviews are done in Berger (1984, 1985 and 1990) and Walley (1990). Commonly used classes can broadly be divided into five groups : conjugate classes, classes with approximately specified moments, contamination classes, density bands and sub-sigma field classes. The parametric analysis with conjugate prior classes includes Leamer (1978, 1982), Polasek (1985), DasGupta and Studden (1988) etc. Classes with moment constraints have been considered in Stone (1963), Hartigan (1969) and Goldstein (1980). Contamination classes have been studied extensively in the literature, among others, see Huber(1981), Berger (1984, 1985), Berger and Berliner (1986), Sivaganesan and Berger (1989), DasGupta and Delampady (1990a,b). The density band was first proposed by DeRobertis (1978) and DeRobertis and Hartigan (1981); recent work on this includes Lavine (1991), DasGupta and Studden (1988, 1990), Bose (1990) etc. The sub-sigma field approach has been considered in Fine (1973), DeRobertis (1978), Berger and O'Hagan (1989), Moreno and Cano (1989), Berliner and Goel (1990), etc. Srinivasan and Truszczynska (1990) and Wasserman and Kadane (1990) take a more general approach, based on functional derivatives and theory of capacities. The approach of stable estimation is considered in Meczarski and Zieliński (1991). Another robust Bayesian viewpoint is to perform an "objective" analysis with a noninformative prior. Work on this includes Jeffreys (1981) and the "reference prior approach" of Bernardo (1979) and Berger and Bernardo (1989).

2 Ranges of Bayes rules for $h(\theta)$ under squared error loss.

2.1 Notations, assumptions and preliminaries.

It is well known that, under squared error loss $L_1(\theta, a) = (\theta - a)^2$, with prior cdf $F(\theta)$, the Bayes estimate of θ is given by $\delta_{L_1}(X) = E^{F(\theta|X)}(\theta)$, the posterior expectation of θ. In a more general setting, suppose interest lies in $h(\theta)$, some known function of θ, defined on the parameter space Θ. The Bayes estimate for $h(\theta)$ (under loss L_1 and prior F), is then given by ($\ell(\theta)$ here denotes the likelihood function)

$$\rho(h, F) = E^{F(\theta|x)}(h(\theta)) = \frac{\int_\Theta h(\theta)\ell(\theta)dF(\theta)}{\int_\Theta \ell(\theta))dF(\theta)} \tag{3}$$

A natural goal of robustness investigation to the choice of the prior F would be to find the ranges of the Bayes estimate $\rho(h, F)$, as F varies over Γ. Thus, we seek

$$\underline{\rho}(h) = \inf_{F \epsilon \Gamma} \rho(h, F) \quad \text{and} \quad \overline{\rho}(h) = \sup_{F \epsilon \Gamma} \rho(h, F).$$

Notice the ratio linear posterior quantity $\rho(h, F)$ is, in fact, well defined for all monotone functions F, bounded between 0 and 1. 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 are made to elaborate on it. Some common choices of $h(\theta)$ are $h(\theta) = \theta$, $h(\theta) = I_C(\theta)$, $h(\theta) = L(\theta, a)$ 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 probability of the set C, 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).

(2) $\ell(a) = \ell(b) = 0$ $(\lim_{\theta \to \pm\infty} \ell(\theta) = 0$ in case $\Theta = \Re)$ and $\ell(\theta)$ is continuously differentiable on (a, b). Cases where this assumption is not satisfied are rare; our arguments can be adapted to the case when $\ell(a)$ or $\ell(b)$ may be nonzero, but at the expense of significant technical complexity.

(3) $h(\theta)$ is continuously differentiable on (a, b). In case $\Theta = \Re$, we also assume $\lim_{\theta \to \pm\infty} h(\theta)\ell(\theta) = 0$.

Notice that $h(\theta) = I_C(\theta)$ does not satisfy assumption (3) if, for example, C is an interval $[u, v]$. This case needs special treatment and is deferred till Section 3.2.

We first give an example to illustrate the basic technical idea used in more generality later in this article.

Example 1 : Suppose we sample X from $N(\theta, 1)$, yielding the likelihood $\ell(\theta) = \exp\left(-\frac{1}{2}(\theta - X)^2\right)$, $\theta \in \Re$. The prior cdf for θ, $F(\theta)$, is known to lie in the band Γ (cf. (1)), generated by two fixed cdfs F_L and F_U $(F_L \leq F_U)$, otherwise, it is unknown.

Suppose we are interested in finding the supremum of the posterior mean. Using the notation of (3), then,

$$\rho(h,F)=\frac{\int_{\Re}\theta\ell(\theta)dF(\theta)}{\int_{\Re}\ell(\theta)dF(\theta)} \;=\; \frac{\int_{-\infty}^{\infty}\frac{d}{d\theta}(\theta\ell(\theta))F(\theta)d\theta}{\int_{-\infty}^{\infty}\frac{d}{d\theta}(\ell(\theta))F(\theta)d\theta} \tag{4}$$

(on integration by parts).

Define $f_1(\theta)=\frac{d}{d\theta}(\theta\ell(\theta))$ and $f_2(\theta)=\frac{d}{d\theta}\ell(\theta)$. Assume for the moment that $\overline{\rho}(h)=\sup_{F\in\Gamma^*}\rho(h,F)$ is finite, equals λ and is attained at $\overline{F}\in\Gamma$. By a standard linearization argument, we then have,

$$\int_{-\infty}^{\infty}\{f_1(\theta)-\lambda f_2(\theta)\}F(\theta)d\theta\geq 0 \quad \text{for every } F\in\Gamma$$

with equality for $F=\overline{F}$. Equivalently, the infimum of $\int_{-\infty}^{\infty}\{f_1(\theta)-\lambda f_2(\theta)\}F(\theta)d\theta$ over $F\in\Gamma$ is attained at $F=\overline{F}$. (Notice this standard linearization technique reduces the original problem of maximizing a nonlinear functional of F to minimizing a functional linear in F. Such linearization techniques are also discussed in Srinivasan and Truszczynska (1990) and Lavine, Wasserman and Wolpert (1988). A formal proof of the validity of this linearization argument is given in Lemma 1 .)

Observe now that for any fixed λ,

$$f_1(\theta)-\lambda f_2(\theta)=e^{-\frac{1}{2}(\theta-x)^2}\{-\theta^2+(x+\lambda)\theta+(1-x\lambda)\}, \tag{5}$$

which changes sign at at most 2 points, say α_1 and α_2 ($\alpha_1<\alpha_2$), with $f_1-\lambda f_2\geq 0$ iff $\theta\in(\alpha_1,\alpha_2)$.

Hence, for the purpose of minimizing $\int\{f_1(\theta)-\lambda f_2(\theta)\}F(\theta)d\theta$ over $F\in\Gamma$, we will like to make F as small as possible in the interval (α_1,α_2) and as large as possible outside, in a way consistent with the conditions required for belonging to Γ.

Towards this end, fix $F(\alpha_1)=h_1$, where h_1 is any number satisfying $F_L(\alpha_1)\leq h_1\leq F_U(\alpha_1)$. Subject to this constraint, the function $\overline{F}_{h_1}\in\Gamma$ minimizing $\int\{f_1(\theta)-\lambda f_2(\theta)\}F(\theta)d\theta$ is then clearly given by

$$\overline{F}_{h_1}(\theta)=\begin{cases}\min[h_1,F_U(\theta)] & \text{if } -\infty<\theta<\alpha_1\\ \max[h_1,F_L(\theta)] & \text{if } \alpha_1\leq\theta<\alpha_2\\ F_U(\theta) & \text{if } \alpha_2\leq\theta<\infty.\end{cases} \tag{6}$$

Note that α_1 and α_2 are solutions of $f_1(\theta) - \lambda f_2(\theta) = 0$ and are therefore determined by λ. Thus, the problem of finding the overall optimal $\overline{F}$ over the class Γ is reduced to a 2–dimensional maximization, namely maximizing $\rho(h, \overline{F}_{h_1})$ over λ and h_1.

The basic argument given in the above example applies to more general situations. We first prove that the linearization technique described in Example 1 works in general.

2.2 Linearization.

For a general function $h(\theta)$ and any $F \in \Gamma$, integration by parts yields

$$\rho(h, F) = \frac{\int_{[a,b]} h(\theta)\ell(\theta)dF(\theta)}{\int_{[a,b]} \ell(\theta)dF(\theta)} = \frac{\int_a^b \frac{d}{d\theta}(h(\theta)\ell(\theta))F(\theta)d\theta}{\int_a^b \frac{d}{d\theta}(\ell(\theta))F(\theta)d\theta}. \tag{7}$$

Let $f_1(\theta) = \frac{d}{d\theta}(h(\theta)\ell(\theta))$ and $f_2(\theta) = \frac{d}{d\theta}(\ell(\theta))$.

Lemma 1 *Let Γ^* be any class of functions $\mathcal{G}$, which are nondecreasing, right continuous and bounded between 0 and 1. Assume that for all θ , $\ell(\theta) \le M < \infty$ and for all $\mathcal{G} \in \Gamma^*$, $\int_{[a,b]} \ell(\theta)\, d\mathcal{G}(\theta) \ge \delta > 0$. Denote $\rho(h, \mathcal{G})$ by $s(\mathcal{G})$.*

Suppose, for every real λ, $\inf_{\mathcal{G}\in\Gamma^*} \int_a^b \{f_1(\theta) - \lambda f_2(\theta)\}\mathcal{G}(\theta)d\theta > -\infty$ and is attained at $\overline{\mathcal{G}}_\lambda \in \Gamma^*$. Then $\lambda_0 \overset{def}{=} \sup_{\mathcal{G}\in\Gamma^*} s(\mathcal{G}) < \infty$ and is attained at $\overline{\mathcal{G}}_{\lambda_0}$.

Proof: This lemma is proved in 2 steps.

Step 1. First assume that $\sup_{\mathcal{G}\in\Gamma^*} s(\mathcal{G})$ is finite and equals λ_0.

Then for any particular sequence $\epsilon_n \downarrow 0$, there exists $\mathcal{G}_n \in \Gamma^*$ such that

$$s(\mathcal{G}_n) = \frac{\int_a^b f_1(\theta)\mathcal{G}_n(\theta)d\theta}{\int_a^b f_2(\theta)\mathcal{G}_n(\theta)d\theta} \ge \lambda_0 - \epsilon_n \tag{8}$$

$$\Longleftrightarrow \int_a^b \{f_1(\theta) - \lambda_0 f_2(\theta)\}\mathcal{G}_n(\theta) \le \epsilon_n \int_{[a,b]} \ell(\theta)d\mathcal{G}_n(\theta).$$

Since $\ell(\theta)$ is bounded above, it now follows from (9) that

$$\int_a^b \{f_1(\theta) - \lambda_0 f_2(\theta)\}\overline{\mathcal{G}}_{\lambda_0}(\theta) \le 0$$

$$\Rightarrow \quad s(\overline{\mathcal{G}}_{\lambda_0}) \ge \lambda_0, \text{ i.e., } \sup_{\mathcal{G}\in\Gamma^*} s(\mathcal{G}) \text{ is attained at } \overline{\mathcal{G}}_{\lambda_0}.$$

Step 2. We now prove that $\sup_{\mathcal{G}\in\Gamma^*} s(\mathcal{G})$ must be finite. For, if $\sup_{\mathcal{G}\in\Gamma^*} s(\mathcal{G}) = \infty$, then for any sequence $M_n \uparrow \infty$, we can find $\mathcal{G}_n \in \Gamma^*$ such that $s(\mathcal{G}_n) \ge M_n$, implying

$$\int_a^b f_1(\theta)\mathcal{G}_n(\theta)d\theta \le M_n \int_a^b f_2(\theta)\mathcal{G}_n(\theta)d\theta. \tag{9}$$

Then for any real λ, using (9),

$$\begin{aligned}\int_a^b f_1(\theta)\mathcal{G}_n(\theta)d\theta - \lambda\int_a^b f_2(\theta)\mathcal{G}_n(\theta)d\theta &\leq (M_n - \lambda)\int_a^b f_2(\theta)\mathcal{G}_n(\theta)d\theta \\ &= -(M_n-\lambda)\int_{[a,b]}\ell(\theta)d\mathcal{G}_n(\theta) \\ &\leq -(M_n-\lambda)\delta. \end{aligned} \tag{10}$$

Since the right side of (10) goes to $-\infty$, we have a contradiction ∎

¿From the above lemma therefore, for the purpose of maximizing $\rho(h, F)$ over Γ, it is enough to find, for real λ, $\inf_{F\in\Gamma}\int_a^b\{f_1(\theta)-\lambda f_2(\theta)\}F(\theta)d\theta$.

Using earlier notation, we assume that for any real λ, $f_1(\theta)-\lambda f_2(\theta)$ changes sign a finite number of times. This assumption is satisfied in all cases of practical importance. In fact, for $h(\theta) = \theta^m$, the number of sign changes is at most $(m+1)$ for common likelihoods such as Normal, Gamma, Binomial, Poisson and Negative Binomial.

Suppose now we are in the special case where $f_1 - \lambda f_2$ changes sign only twice, at α_1 and α_2, with $f_1 - \lambda f_2 \leq 0$ on $[a, \alpha_1]$ and $[\alpha_2, b]$ and ≥ 0 on (α_1, α_2). If we now fix the value of F at α_1, say, $F(\alpha_1) = h_1$ (where $F_L(\alpha_1) \leq h_1 \leq F_U(\alpha_1)$), then as in Example 1, we get the following general form for the extremal $\overline{F}$:

$$\overline{F}(\theta) = \begin{cases} \min[h_1, F_U(\theta)] & a \leq \theta < \alpha_1 \\ \max[h_1, F_L(\theta)] & \alpha_1 \leq \theta < \alpha_2 \\ F_U(\theta) & \alpha_2 \leq \theta \leq b. \end{cases} \tag{11}$$

$\rho(h, \overline{F})$ is then to be maximized over $\overline{F}$ of the form (11).

2.3 The Extremal Prior.

Finally, we present here, the form of the extremal prior $\overline{F}$ when the function $f_1 - \lambda f_2$ changes sign n times, say, at $\alpha_1 < \alpha_2 < \ldots < \alpha_n$.

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

Furthermore, label I_i as '+' if $f_1 - \lambda f_2$ is nonnegative on I_i, and as '−' if it is nonpositive on I_i.

Theorem 1 $\sup_{F\in\Gamma} \rho(h,F) = \sup_{F\in\Gamma_s} \rho(h,F)$, *where*

$$\Gamma_s = \left\{ \begin{array}{rcll} F : F(\theta) & = & h_i & \text{if } \theta = \alpha_i, i = 1,\ldots,n \\ & = & \max[h_{i-1}, F_L(\theta)] & \text{if } \theta \in I_i \text{ and } I_i \text{ is '+'} \\ & = & \min[h_i, F_U(\theta)] & \text{if } \theta \in I_i \text{ and } I_i \text{ is '}-\text{'}. \end{array} \right.$$

If, for a particular i, I_i is as labeled '+' and I_{i+1} as '$-$', the corresponding h_i is fixed and equals $\min[h_{i+1}, F_U(\alpha_i)]$. *Else, the h_i's are not fixed and can take any values subject to the restrictions $F_L(\alpha_i) \leq h_i \leq F_U(\alpha_i)$ and $h_i \leq h_j$ if $i < j$* $\Big\}$.

Remark 1. For finding the optimizing $\overline{F} \in \Gamma_s$, the optimal 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 numerical optimization would be over λ and the h_i's.

Remark 2. In the above theorem, we assume that both F_L and F_U are distribution functions on the real line. ε neighborhoods of Kolmogorov and Lèvy metrices of a fixed cdf F_0 are of the form of the distribution band Γ, but as we mentioned earlier, F_L and F_U are no longer cdfs ($\lim_{\theta\to\infty} F_L(\theta) < 1$ and $\lim_{\theta\to-\infty} F_U(\theta) > 0$), but nondecreasing, right continuous functions bounded between 0 and 1. Interestingly, with minor modifications, Theorem 1 can still be applied to these neighborhoods. For illustration, let us consider the $\varepsilon-$Kolmogorov neighborhood of a fixed cdf F_0,

$$\Gamma = \{F : F \text{ is a cdf and } \max(0, F_0(\theta) - \epsilon) \leq F(\theta) \leq \min(1, F_0(\theta) + \epsilon), \quad \theta \in \Re\}$$

It turns out that, for evaluating $\overline{\rho}(h)$, it is easier to deal with the following enlarged family.

$$\Gamma^* = \left\{ \begin{array}{l} G: \ \max(0, F_0(\theta) - \epsilon) \leq G(\theta) \leq \min(1, F_0(\theta) + \epsilon), \ \theta \in \Re, \\ \quad 0 \leq G \leq 1, \ G \text{ nondecreasing and right continuous} \end{array} \right\}$$

For this enlarged family Γ^*, the linearization technique of Lemma 1 still applies and Theorem 1 reduces the infinite dimensional problem of finding $\sup_{G\in\Gamma^*} \rho(h,G)$ to a finite dimensional numerical optimization. The only problem is, the extremal $\overline{G} \in \Gamma^*$ thus obtained, might not be a distribution function. To exemplify, let us go back to the problem of finding the supremum of posterior mean in Example 1 (cf. section 2.1), but now consider $F_L(\theta) = \max(0, F_0(\theta) - \epsilon)$ and $F_U(\theta) = \min(1, F_0(\theta) + \epsilon)$, where F_0 is a fixed cdf on $\Re$, say for example, the conjugate $N(0,1)$ prior. The form of the extremal $\overline{G}$ is given by (6), and it is clear from (6) that $\lim_{\theta \to -\infty} \overline{G}(\theta) = \lim_{\theta \to -\infty} \min(h_1, F_0(\theta) + \epsilon) > 0$, hence $\overline{G}$ is not a cdf. However, this problem can be circumvented by truncating $\overline{G}$ at the left tail, i.e., by defining $\overline{G}_n(\theta) = 0$ if $-\infty < \theta < -n$, and $\overline{G}_n(\theta) = \overline{G}(\theta)$ otherwise. For large n, $\overline{G}_n$ is a cdf and $\in \Gamma$. Let μ and μ_n be the measures corresponding to $\overline{G}$ and $\overline{G}_n$ on the space $\overline{\Re} = [-\infty, \infty]$ (both $-\infty$ and $+\infty$ inclusive). For any bounded continuous function f on $\Re$, satisfying $\lim_{\theta \to \pm\infty} f(\theta) = 0$,

$$\begin{aligned}
\int_{\overline{\Re}} f(\theta) d\overline{G}(\theta) &= \lim_{n \to \infty} \left\{ f(-n) d\mu([-\infty, n)) + \int_{[-n,\infty]} f(\theta) d\mu(\theta) \right\} \\
&= \lim_{n \to \infty} \int_{[-n,\infty]} f(\theta) d\mu(\theta) \quad \text{since } \lim_{n \to \infty} f(-n) = 0 \\
&= \lim_{n \to \infty} \int_{[-n,\infty]} f(\theta) d\mu_n(\theta) = \lim_{n \to \infty} \int_{-\infty}^{\infty} f(\theta) dG_n(\theta)
\end{aligned}$$

It is now an easy exercise to show that $\lim_{n \to \infty} E^{\overline{G}_n(\theta|X)}(\theta) = E^{\overline{G}(\theta|X)}(\theta)$, which establishes $E^{\overline{G}(\theta|X)}(\theta)$ to be the supremum of $E^{F(\theta|X)}(\theta)$ over the smaller family Γ as well. This limiting argument can be modified for the case of general $h(\theta)$ easily.

2.4 Application: Ranges of Posterior Mean.

In this section, we apply Theorem 1 to several examples on finding the ranges of posterior mean over specific prior classes. Note that the posterior mean is the Bayes rule for θ under the squared error loss $L_1(\theta, a) = (\theta - a)^2$.

Example 2 (continuation of Example 1).

Suppose we observe $X \sim N(\theta, 1)$, and a finite elicitation process specify the prior median and quartiles at 0 and ± 1 respectively. As we mentioned before, $F_N = N(0, 2.19)$, $F_C =$Cauchy$(0,1)$, and several other priors (see Table 1) satisfy these specifications. Combining the informal and the formal approach to Bayesian robustness (see section 1.2), let us define $F_L^1 = \min(F_N, F_C)$, $F_U^1 = \max(F_N, F_C)$ and let Γ_1 be the resulting distribution band. Also define, $F_L^2 = \min(F_N, F_C, F_E)$ and $F_U^2 = \max(F_N, F_C, F_E)$, where $F_E =$ Double Exponential$(0, 1.44)$ (which also satisfies the quantile restrictions), and let Γ_2 be the distribution band generated by F_L^2 and F_U^2. In fact, Γ_2 contains all the priors that are mentioned in Table 1. In Table 2, we present the suprema ($\overline{E}$), infima ($\underline{E}$) and the ranges ($\overline{E} - \underline{E}$) of the posterior mean $E^{F(\theta|X)}(\theta)$ over these classes. Γ_1 and Γ_2 yield almost identical ranges, indeed, these two classes are same except for $\theta \in (-0.396, 0.396)$ where Γ_2 is slightly wider. From Table 2, the ranges of $E^{F(\theta|X)}(\theta)$ for small values of $|X|$ are small, showing insensitivity to the choice of the prior, whereas for large $|X|$ values, the ranges are rather large. Such lack of robustness for large X is often typical in Bayesian robustness studies. More discussions along this line follow in Example 3.

Example 3 (continuation of Example 1).

Again, assume $X \sim N(\theta, 1)$, and suppose we decide, combining elicitation information and mathematical convenience, on the conjugate prior $F_0 = N(0, 1)$. Instead of quantile specification, we now directly model our uncertainty in F_0 in terms of Γ_K and Γ_L, the $\varepsilon-$Kolmogorov and $\varepsilon-$Lèvy neighborhoods of F_0. Figure 2 shows the suprema ($\overline{E}_K$) and the infima ($\underline{E}_K$) of $E^F(\theta|X)$ as F varies in Γ_K, whereas Figure 3 compares the ranges ($\overline{E}_K - \underline{E}_K$ and $\overline{E}_L - \underline{E}_L$) over Γ_K and Γ_L. Again, significant lack of robustness is observed for large $|X|$ values and/or large ϵ. Two possible reasons could be suggested for this lack of robustness. One is that we are allowing prior distributions with point masses to be included in Γ_K and Γ_L. Secondly, if F_0 is a $N(0, 1)$ prior, then an ϵ–Kolmogorov neighborhood of F_0 contains priors with extremely thin as well as extremely thick tails, in comparison to the tail of F_0. If F_0 itself was thicker tailed, better robustness would have been achieved. Another surprising feature is that though $\Gamma_K \subset \Gamma_L$, they yield almost similar ranges. The base prior F_0 is nearly flat for $|\theta| > 2$. Hence, even though the band Γ_L is quite a bit wider than Γ_K in the center, towards the tail Γ_K and Γ_L are almost the same. This makes the suprema and infima for the two prior classes rather similar.

As noted before, the Kolmogorov distance between $F_N = N(0, 2.19)$ and $F_C =$ Cauchy$(0, 1)$ is 0.081. Thus, an $\varepsilon-$Kolmogorov neighborhood of F_N with $\epsilon = 0.081$, which we denote by Γ_3, contains F_C, all the other priors in Table 1, and $\Gamma_3 \supset \Gamma_2 \supset \Gamma_1$ (cf. Example 2). ¿From Table 2, the ranges of the posterior mean over the class Γ_3 are much larger compared to ranges over the classes Γ_1 and Γ_2. This is expected, Γ_3 is much wider than the other two, has no quantile restrictions and, while in Γ_1 and Γ_2, the prior tail is bounded between the exponential tail of Normal and polynomial tail of Cauchy, the priors in Γ_3 have wider flexibility in their tails.

Example 4.

Let X be discrete $\sim$ Binomial (n, θ), yielding the likelihood $\ell(\theta) = \theta^X(1-\theta)^{n-X}$, $\theta \in [0,1]$, $X \epsilon \{0, 1, \ldots, n\}$. We take F_0 to be a conjugate Beta (α, β) cdf and again, consider Γ_K and Γ_L, an ϵ–Kolmogorov and an ϵ–Lèvy neighborhood of F_0 respectively.

For $n = 5$ and $\alpha = \beta = 2$, Figure 4 compares the suprema $(\overline{E})$ and infima $(\underline{E})$ of $E^F(\theta|X)$ as F as F varies in $\varepsilon-$Kolmogorov and $\varepsilon-$Lèvy neighborhoods of the Beta (2,2) prior. Figure 5 shows the ranges of $(\overline{E}_K - \underline{E}_K)$ for Kolmogorov neighborhoods with different combinations of (α, β). These figures, especially Figure 5, clearly show that for X compatible with the base prior F_0, the range of the mean is small, whereas the ranges for extreme X values are much larger. For a neighborhood of the Uniform (0,1) prior, the range is approximately the same (≈ 0.2) for all X's, and in fact, $\overline{E}_K$ and $\underline{E}_K$ are each approximately 0.1 unit away from the posterior mean under the Uniform (0,1) prior. But for skewed base priors (such as Beta (0.5,2) and Beta (2,1)), or even for some (non–uniform) symmetric priors (such as Beta (0.5, 0.5) and Beta (2,2)), extreme X–values give a much larger range. Again, this lack of robustness can be attributed to the inclusion of a large variety of priors.

¿From Figure 4, it is seen that unlike in Example 3, the Lèvy class Γ_L yields much larger ranges than Γ_K. This is expected because, unlike in the normal case, Γ_L is, in fact, much wider than Γ_K. For example, when F_0 is Uniform (0,1), the ϵ–Lèvy class is exactly equal to the 2ϵ–Kolmogorov class.

3 Ranges of Bayes rules under linear loss.

3.1 The algorithm

In section 1.1, we introduced three different loss structures, namely L_1, the squared error loss, L_2, the absolute error loss, and L_3, the linear loss (cf. (2)). We also mentioned that L_2 is a special case of L_3, and the Bayes rules under L_1 and L_3 are respectively : $\delta_{L_1}(X)$ = the posterior mean; and $\delta_{L_3}(X)$ = any $\frac{K_0}{K_0+K_1}(=\gamma)$ fractile of the posterior distribution $F(\theta|X)$. A general method is described in the previous section for finding the ranges of $\delta_{L_1}(X)$. In this section, we take up the problem of finding the ranges of $\delta_{L_3}(X)$, i.e., posterior fractiles, when the prior cdf F varies in a distribution band Γ (cf. (1)). As before, for brevity, only the problem of finding the infimum of the posterior fractiles will be described (The supremum problem being exactly similar). For this problem, we only consider a unique fractile (in case there are several), namely the smallest one, which we define below.

Definition 1 *For fixed $0 < \gamma < 1$, we define the γ-th fractile of a cdf F on $\Re$ to be*

$$Q_F^\gamma = \inf_{a\in\Re} \{a : P_F\{(-\infty, a]\} \geq \gamma\}. \tag{12}$$

Next, we prove that the infimum of Q_F^γ over a family of distributions $\mathcal{F}$, can in fact be described, in terms of suprema of probability of intervals (over the same family $\mathcal{F}$).

Lemma 2 *Let $\mathcal{F}$ be a family of distributions and let $\underline{Q}^\gamma = \inf_{F\in\mathcal{F}} Q_F^\gamma$. Then $\underline{Q}^\gamma$ satisfies*

$$\underline{Q}^\gamma = \inf_{a\in\Re}\left\{a : \sup_{F\in\mathcal{F}} P_F\{(-\infty, a]\} \geq \gamma\right\}.$$

Proof : The lemma is proved in 2 steps. For step 1, the statement

$$\forall F \in \mathcal{F}, \quad \inf_{a\in\Re} \{a : P_F\{(-\infty, a]\} \geq \gamma\} \ \geq \underline{Q}^\gamma$$

implies that $\forall F \in \mathcal{F}$ and any $a \in \Re$, $P_F\{(-\infty, a]\} \geq \gamma \ \Rightarrow \ a \geq \underline{Q}^\gamma$. This, in turn, implies $\sup_{F\in\mathcal{F}} P_F\{(-\infty, a]\} \geq \gamma \ \Rightarrow \ a \geq \underline{Q}^\gamma$. It now follows that

$$\inf_{a\in\Re}\left\{a : \sup_{F\in\mathcal{F}} P_F\{(-\infty, a]\} \geq \gamma\right\} \ \geq \ \underline{Q}^\gamma.$$

For step 2, fix $\delta > 0$, and find $F_\delta \in \mathcal{F}$ such that

$$\underline{Q}^\gamma + \delta \; > \; Q^\gamma_{F_\delta} = \inf_{a \in \Re} \{a : P_{F_\delta}\{(-\infty, a]\} \geq \gamma\} \; \geq \; \underline{Q}^\gamma.$$

It follows that, $\exists\, a^* < \underline{Q}^\gamma + \delta$ such that $P_{F_\delta}\{(-\infty, a^*]\} \geq \gamma$. Hence $\sup_{F \in \mathcal{F}} P_F\{(-\infty, a^*]\} \geq \gamma$. Thus, $\inf_{a \in \Re} \left\{ a : \sup_{F \in \mathcal{F}} P_F\{(-\infty, a]\} \geq \gamma \right\} \leq a^* \; < \; \underline{Q}^\gamma + \delta$.

$\delta > 0$ being arbitrary, the last statement, alongwith step 1, proves the lemma ■

Lemma 2 gives us a clearcut algorithm for finding the infimum of the γ-th posterior fractile over any prior class Γ, along the following steps.

(i) For $a \in \Re$, find $\sup_{F \in \Gamma} P^{F(\theta|X)}\{(-\infty, a]\}$.

(ii) Find the collection $\mathcal{A} = \left\{ a : \sup_{F \in \Gamma} P^{F(\theta|X)}\{(-\infty, a]\} \; \geq \; \gamma \right\}$.

(iii) Find infimum over the set $\mathcal{A}$, i.e., $\inf\{a : a \in \mathcal{A}\}$.

The problem of finding the ranges of posterior fractiles over the distribution band Γ is thus settled, as long as we can carry out step (i), i.e., if we are able to find $\sup_{F \in \Gamma} P^{F(\theta|X)}\{(-\infty, a]\}$ for any $a \in \Re$. This is done (with some additional assumptions) in Basu and DasGupta (1990), and is briefly described in the next subsection. Most surprisingly, as it turns out, the extremal priors $\overline{F}_a \in \Gamma$ which gives $\sup_{F \in \Gamma} P^{F(\theta|X)}\{(-\infty, a]\}$, can be found with <u>almost</u> no numerical work. The numerical work is concentrated on step (iii), which involves finding the infimum of an interval, a relatively easy job.

3.2 Ranges of posterior probability of an interval

In this section, we very briefly overview the methodology for finding the ranges of posterior probability of an interval $[u, v]$, contained in the parameter space $\Theta = [a, b]$ (or $\Re$), over the distribution band Γ defined in (1). As noted before, $P^{F(\theta|X)}([u, v]) = E^{F(\theta|X)}(h(\theta))$ with $h(\theta) = I_{[u,v]}(\theta)$, but $I_{[u,v]}(\theta)$ does not satisfy assumption 3 of section 2.1, namely it is not continuously differentiable. Hence the methodology of section 2 and Theorem 1 is not directly applicable here. The problem of finding $\sup_{F \in \Gamma} P^{F(\theta|X)}([u, v])$ (and $\inf_{F \in \Gamma} P^{F(\theta|X)}$ involves some unique features, which need careful mathematical treatment. For length considerations, we refrain from the details, but simply sketch the additional assumptions that are required and the form of the extremal prior for the supremum case. The details are given in Basu and DasGupta (1990).

In addition to assumption (1) and (2) of section 2.1, we need the following extra assumptions here.

(3) F_L and F_U are absolutely continuous on (a,b) with densities (w.r.t. Lebesgue measure) π_L and π_U, and are differentiable at all but at most finitely many points.

(4) The likelihood function $\ell(\theta)$ is unimodal about some $w = w(X)$.

The linearization technique of section 2.2 still applies, and it reduces the problem of finding $\sup_{F\in\Gamma} P^{F(\theta|X)}([u,v])$ to evaluating $\inf_{F\in\Gamma} \int_a^b \{f_1(\theta) - \lambda f_2(\theta)\} F(\theta) d\theta$ where $f_1(\theta) - \lambda f_2(\theta) = (I_{[u,v]}(\theta) - \lambda)\ell'(\theta)$, and λ is any (nonnegative) real number.

The function $f_1 - \lambda f_2$ defined above, changes sign at u, v and w; but the exact sequence of sign changes depends on the relative position of u, v and w with respect to one another. We thus have the following three cases: (A) $w\ (\alpha_1) < u\ (\alpha_2) < v\ (\alpha_3)$, (B) $u\ (\alpha_1) < v\ (\alpha_2) < w\ (\alpha_3)$ and (C) $u\ (\alpha_1) < w\ (\alpha_2) < v\ (\alpha_3)$.

Skipping the details of the derivation, the extremal prior $\overline{F}$ (at which the $\sup_{F\in\Gamma} P^{F(\theta|X)}([u,v])$ attains), in general, is given by

$$\overline{F}(\theta) = \begin{cases} h_i & \text{if } \theta = \alpha_i \\ \min[h_1, F_u(\theta)] & \text{if } a \le \theta < \alpha_1 \\ \max[h_1, F_L(\theta)] & \text{if } \alpha_1 < \theta < \alpha_2 \\ \min[h_3, F_u(\theta)] & \text{if } \alpha_2 < \theta < \alpha_3 \\ \max[h_3, F_L(\theta)] & \text{if } \alpha_3 < \theta \le b. \end{cases} \tag{13}$$

The optimal values of h_u and h_v are completely determined, ($h_u = h_2 = F_U(u)$ in (A), and $h_u = h_1 = F_L(u)$ in (B) and (C); $h_v = h_3 = F_U(v)$ in (A) and (C), and $h_v = h_2 = \min[h_w, F_U(v)]$ in (B)). In case (C), the optimal value of $h_w = h_2$ is also determined to be $F_U(w)$, whereas, in cases (A) and (B), the determination of optimal h_w involves numerical optimization over only finitely many points (for details, see Basu and DasGupta (1990)). Hence, the extremal prior $\overline{F}$ can be found here in a more or less closed analytic form, as opposed to the numerical determination of α_i's and h_i's in section 2.

3.3 Application.

Example 5 (continuation of Example 2).

As before, let $X \sim N(\theta, 1)$, and consider the prior class Γ_1 to be the distribution band generated by $F_L^1 = \min(F_N, F_C)$ and $F_U^1 = \max(F_N, F_C)$ where $F_N = N(0, 2.19)$ and $F_C = \text{Cauchy}(0,1)$. We consider four different losses, $L^1 = L_1 =$ squared error, $L^2 = L_2 =$ absolute error, $L^3 = L_3$ with $K_0 = 1$ and $K_1 = 3$, and $L^4 = L_3$ with $K_0 = 1$ and $K_1 = 9$. In Table 3, we present the ranges of Bayes rules with respect to these 4 losses, namely, $\delta_1 = E^{F(\theta|X)}(\theta)$, $\delta_2 =$ posterior median, $\delta_3 = \frac{1}{4}$-th fractile or the first quartile of $F(\theta|X)$, and $\delta_4 = \frac{1}{10}$-th fractile of $F(\theta|X)$. Even without different loss considerations, these 4 measures give a quantitative summary of the posterior distribution $F(\theta|X)$, so their ranges are of importance (The ranges of the $\frac{3}{4}$-th fractile and the $\frac{9}{10}$-th fractile can be obtained from the values in Table 3 by using symmetry).

In classical robustness studies, the sample median often turns out to be more robust than the sample mean. In contrast, in Bayesian robustness context (at least in this example), the posterior median δ_2 yields wider ranges than the posterior mean δ_1, except for $X = 0$, as can be seen from Table 3. Another surprising observation is that, though δ_3 is further in the left tail of the posterior $F(\theta|X)$ than δ_2, δ_4 being the furthest, all the three fractiles δ_2, δ_3 and δ_4 yield almost similar ranges. In conclusion, we find that more robustness is achieved in terms of the Bayes rule by using the L^1 loss function instead of the other 3 losses L^2, L^3, L^4.

4 Summary and concluding remarks

Much of previous work on Bayesian robustness has dealt with densities of the prior 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 be directly assessed from prior probability considerations. No assumptions are made about absolute continuity or existence of densities. Surprisingly, such a class is mathematically tractable and indeed in some cases even closed form expressions for the extremal priors are possible. This family can also be used to reconcile formal and informal approaches to Bayesian robustness, and flexibility in the prior tail is achieved very easily, as illustrated in the examples.

We have also proposed an investigation of the sensitivity of the Bayes estimate to the choice of both the loss function and the prior. In this context, we have considered a mixture of the informal and the formal approach, letting the loss function vary among a few commonly used loss functions, whereas the prior cdf lies in a distribution band Γ. Comparing different estimators arising out of different loss structures, such as the mean and the median, are quite common in classical literature; such comparison in Bayesian analysis is rare. Contrary to the results of classical statistics, we find that the Bayes estimate under squared error loss, namely the posterior mean, is more robust than the posterior fractiles, the Bayes estimates under linear losses.

As mentioned in Berger (1990), the selection of a prior class Γ depends mainly on the following two competing goals : (*i*) Γ should contain as many "reasonable" priors as possible, and (*ii*) Γ should not contain unreasonable priors. While the "Distribution Band" satisfies (i), it often contains unreasonable priors, such as priors with point masses, leading to lack of robustness for moderate X. Indeed, the extremal $\overline{\mathcal{G}}$'s were found to assign point masses to several points and also have flat pieces, i.e. regions with zero masses. One possible remedy is to put further reasonable restrictions on the prior, such as requiring the prior to be absolutely continuous w.r.t. Lebesgue measure. But, this restriction alone will usually not change the values of $\overline{\rho}(h)$ and $\underline{\rho}(h)$, as absolutely continuous cdfs in a band will usually be weakly dense in the band itself. Other standard shape restrictions include symmetry and unimodality. Surprisingly, we have found that even Γ restricted to symmetric and unimodal priors does not lead to significantly better robustness. The construction of the extremal priors under these added restrictions is reported in Basu (1991).

Summarizing, the "Distribution band" Γ is very intuitive, rich and can easily be adapted to meet the specific subjective information of the user. It is also mathematically tractable, the ranges of interesting posterior quantities can be determined with limited numerical work, and its use to study the sensitivity of Bayes rules w.r.t. joint variations in the prior and the loss function is appealing. The main use of the distribution band in Bayesian sensitivity analysis, however, seems to be in the case when the data are compatible with the prior(s) because then one can feel comfortable knowing that robustness is present even when a rich family of priors is used in the sensitivity study.

Acknowledgement We are deeply indebted to James Berger and Rajeeva Karandikar for valuable advice and helpful discussions.

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] Basu, S. (1991), "Variations of posterior expectations for symmetric unimodal priors in a distribution band", Technical Report, **91-45**, Department of Statistics, Purdue University.

[4] Berger, J. (1984), "The robust Bayesian viewpoint (with discussion)", in *Robustness of Bayesian Analysis*,Kadane, J. (ed.), Amsterdam: North Holland.

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

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

[7] Berger, J.O. and Bernardo, J. (1989), "Estimating a product of means: Bayesian analysis with reference priors", *J. of the Amer. Statist. Assoc.*, **84**, pp. 200-207.

[8] 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.

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

[10] 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).

[11] Bernardo, J. (1979), "Reference posterior distribution for Bayesian inference", *J. Roy. Statist. Soc., Series B*, **41**, pp. 113-147.

[12] Bose, S. (1990), "Bayesian Robustness with shape-constrained priors and mixture priors", Ph.D. Thesis., Purdue University.

[13] 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.

[14] DasGupta, A. and Studden, W.J. (1990), "Robust Bayes designs in normal linear models", *Ann. Statist.*, **19**, pp. 1244-1256.

[15] DasGupta, A. and Delampady, M. (1990a), "Γ-minimax estimation of vector parameters in restricted parameter spaces", Technical Report, **90-42**, Department of Statistics, Purdue University.

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

[17] DeRobertis, L. (1978) "The use of partial prior knowledge in Bayesian inference", Ph.D. Thesis, Yale University.

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

[19] Fine, T. (1973), *Theories of Probability*, New York: Academic Press.

[20] Goldstein, M. (1980), "The linear Bayes regression estimator under weak prior assumptions", *Biometrika*, **67**, pp. 621-628.

[21] Hartigan, J.A. (1969), "Linear Bayesian Methods", *J.Roy. Statist. Soc. Ser. B*, **31**, pp. 446-454.

[22] Huber, P.J. (1981), *Robust Statistics*, London: John Wiley.

[23] Jeffreys, H. (1981), *Theory of Probability* , London: Oxford University Press.

[24] Lavine, M., Wasserman, L., and Wolpert, R. (1991), "Bayesian inference with specified prior marginals", *J. of the Amer. Statist. Assoc.*, **86**, pp. 964-971.

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

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

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

[28] Meczarski, M. and Zieliński, R. (1991), "Stability of Bayesian estimator of the Poisson mean under the inexactly specified gamma prior", *Stat. and Prob. Letters*, **12**, pp. 329-333.

[29] Moreno, E. and Cano, J.A. (1989), "Testing a point null hypothesis : Asymptotic Robust Bayesian analysis with respect to the priors given in a subsigma field", *Internat. Statist. review*, **57**, pp. 221-232.

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

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

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

[33] Stone, M. (1963), "Robustness of nonideal decision procedures", *J. Amer. Statist. Assoc.*, **58**, pp. 480-486.

[34] Walley, P. (1990), *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.

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

[36] Wiens, D. (1986), "Minimax variance M-estimators of location in Kolmogorov neighborhoods", *Ann. Statist.*, **14**, pp. 724-732.

Table 1: Distances from $N(0, 2.19)$

Distribution	Parameters	Kolmogorov distance (d_K)	Lèvy distance (d_L)
Logistic $(0, \tau^2)$	$\tau^2 = (\log 3)^{-1}$	0.015	0.014
Double Exponential $(0, \tau^2)$	$\tau^2 = (\log 2)^{-1}$	0.043	0.041
Cauchy $(0, 1) = t(1, \tau^2)$	$\tau^2 = 1$	0.081	0.079
$t(2, \tau^2)$	$\tau^2 = 1.499$	0.046	0.044
$t(3, \tau^2)$	$\tau^2 = 1.709$	0.031	0.030
$t(4, \tau^2)$	$\tau^2 = 1.823$	0.024	0.023
$t(5, \tau^2)$	$\tau^2 = 1.894$	0.019	0.018
$t(10, \tau^2)$	$\tau^2 = 2.042$	0.010	0.009
$t(15, \tau^2)$	$\tau^2 = 2.093$	0.007	0.006
$t(20, \tau^2)$	$\tau^2 = 2.119$	0.005	0.005

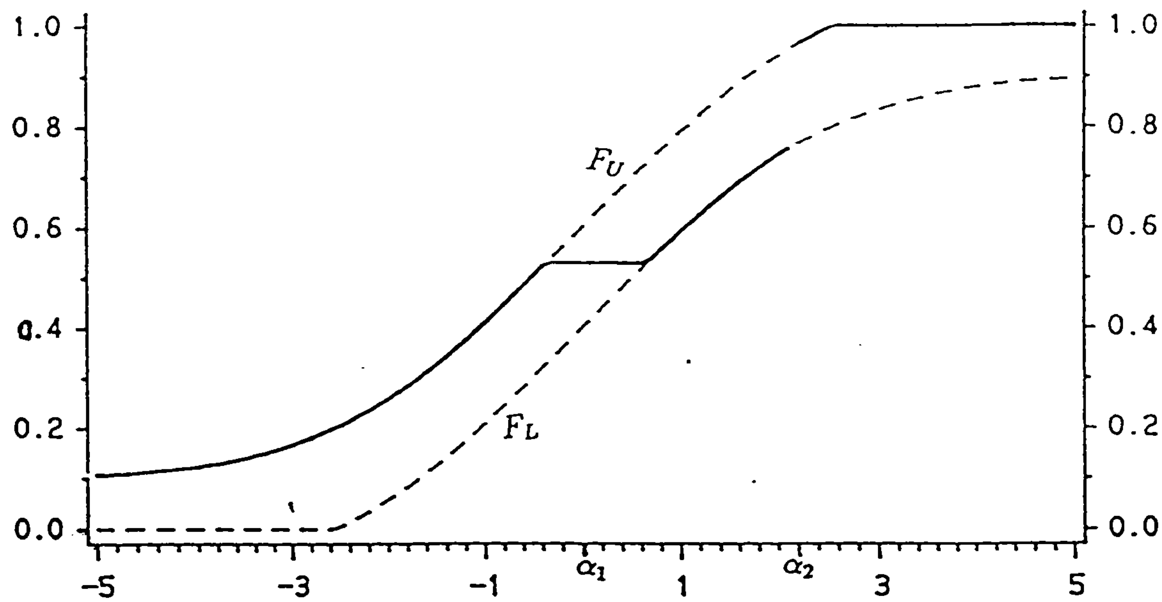

Figure 1: ϵ-Kolmogorov neighborhood of $N(0,4)$ and the maximizing G for posterior mean. The dashed lines are F_L and F_U. The solid line is the extremal prior $\overline{G}$.

Table 2: Ranges of Posterior Mean $E^{F(\theta|X)}(\theta)$ (likelihood = $N(\theta,1)$)

				Γ_3		
X	$E^{F_N(\theta\|X)}(\theta)$	$E^{F_C(\theta\|X)}(\theta)$	$E^{F_E(\theta\|X)}(\theta)$	Sup ($\overline{E}$)	Inf ($\underline{E}$)	Range ($\overline{E}-\underline{E}$)
−5	−3.4366	−4.5632	−4.3069	−1.1135	−7.4248	6.3113
−4	−2.7493	−3.4351	−3.3073	−1.0281	−5.1226	4.0945
−3	−2.0619	−2.2851	−2.3167	−0.8952	−3.0454	2.1502
−2	−1.3746	−1.2822	−1.3885	−0.6649	−1.8459	1.1810
−1	−0.6873	−0.5542	−0.6197	−0.2514	−1.0263	0.7749
0	0.0000	0.0000	0.0000	0.3407	−0.3407	0.6814
1	0.6873	0.5542	0.6197	1.0263	0.2514	0.7749
2	1.3746	1.2822	1.3885	1.8459	0.6649	1.1810
3	2.0619	2.2851	2.3167	3.0454	0.8952	2.1502
4	2.7493	3.4351	3.3073	5.1226	1.0281	4.0945
5	3.4366	4.5632	4.3069	7.4248	1.1135	6.3113

	Γ_1			Γ_2		
X	Sup ($\overline{E}$)	Inf ($\underline{E}$)	Range ($\overline{E}-\underline{E}$)	Sup ($\overline{E}$)	Inf ($\underline{E}$)	Range ($\overline{E}-\underline{E}$)
−5	−2.1994	−6.2418	4.0424	−2.1994	−6.2418	4.0424
−4	−1.9261	−4.3476	2.4215	−1.9261	−4.3477	2.4216
−3	−1.5534	−2.7697	1.2163	−1.5532	−2.7701	1.2169
−2	−1.0826	−1.5826	0.5000	−1.0822	−1.5831	0.5009
−1	−0.5221	−0.7218	0.1997	−0.5214	−0.7214	0.2000
0	0.0703	−0.0703	0.1406	0.0712	−0.0712	0.1424
1	0.7218	0.5221	0.1997	0.7214	0.5214	0.2000
2	1.5826	1.0826	0.5000	1.5831	1.0822	0.5009
3	2.7697	1.5534	1.2163	2.7701	1.5532	1.2169
4	4.3476	1.9261	2.4215	4.3477	1.9261	2.4216
5	6.2418	2.1994	4.0424	6.2418	2.1994	4.0424

Table 3: Ranges of Bayes rules w.r.t. different Loss functions over prior class Γ_1

X	$E^{F_N(\theta\|X)}(\theta)$	Posterior Mean (δ_1) Sup	Inf	Range	Posterior Median (δ_2) Sup	Inf	Range
−5	−3.4366	−2.199	−6.242	4.042	−2.062	−7.548	5.486
−4	−2.7493	−1.926	−4.348	2.422	−1.849	−5.642	3.793
−3	−2.0619	−1.553	−2.770	1.217	−1.547	−3.674	2.217
−2	−1.3746	−1.083	−1.583	0.500	−1.102	−1.678	0.576
−1	−0.6873	−0.522	−0.722	0.200	−0.482	−0.703	0.221
0	0.0000	0.070	−0.070	0.140	0.049	−0.049	0.098
1	0.6873	0.722	0.522	0.200	0.703	0.482	0.221
2	1.3746	1.583	1.083	0.500	1.678	1.102	0.576
3	2.0619	2.770	1.553	1.217	3.674	1.547	2.217
4	2.7493	4.348	1.926	2.422	5.642	1.849	3.793
5	3.4366	6.242	2.199	4.042	7.548	2.062	5.486

X	$\delta_3^{F_N(\theta\|X)}$	Posterior $\frac{1}{4}$-th fractile (δ_3) Sup	Inf	Range	$\delta_4^{F_N(\theta\|X)}$	Posterior $\frac{1}{10}$-th fractile (δ_4) Sup	Inf	Range
−5	−3.996	−2.287	−7.921	5.634	−4.498	−2.445	−8.261	5.816
−4	−3.309	−2.106	−6.149	4.043	−3.810	−2.276	−6.580	4.304
−3	−2.622	−1.856	−4.447	2.591	−3.123	−2.050	−5.006	2.956
−2	−1.935	−1.491	−2.833	1.342	−2.436	−1.740	−3.576	1.836
−1	−1.247	−1.015	−1.390	0.375	−1.748	−1.354	−2.267	0.913
0	−0.560	−0.420	−0.593	0.173	−1.061	−0.879	−1.113	0.234
1	0.127	0.143	0.014	0.129	−0.374	−0.297	−0.465	0.168
2	0.815	0.864	0.562	0.302	0.314	0.340	0.126	0.214
3	1.502	2.279	1.163	1.116	1.000	1.241	0.689	0.552
4	2.189	4.992	1.544	3.448	1.688	3.759	1.222	2.537
5	2.877	7.133	1.806	5.327	2.376	6.641	1.545	5.096

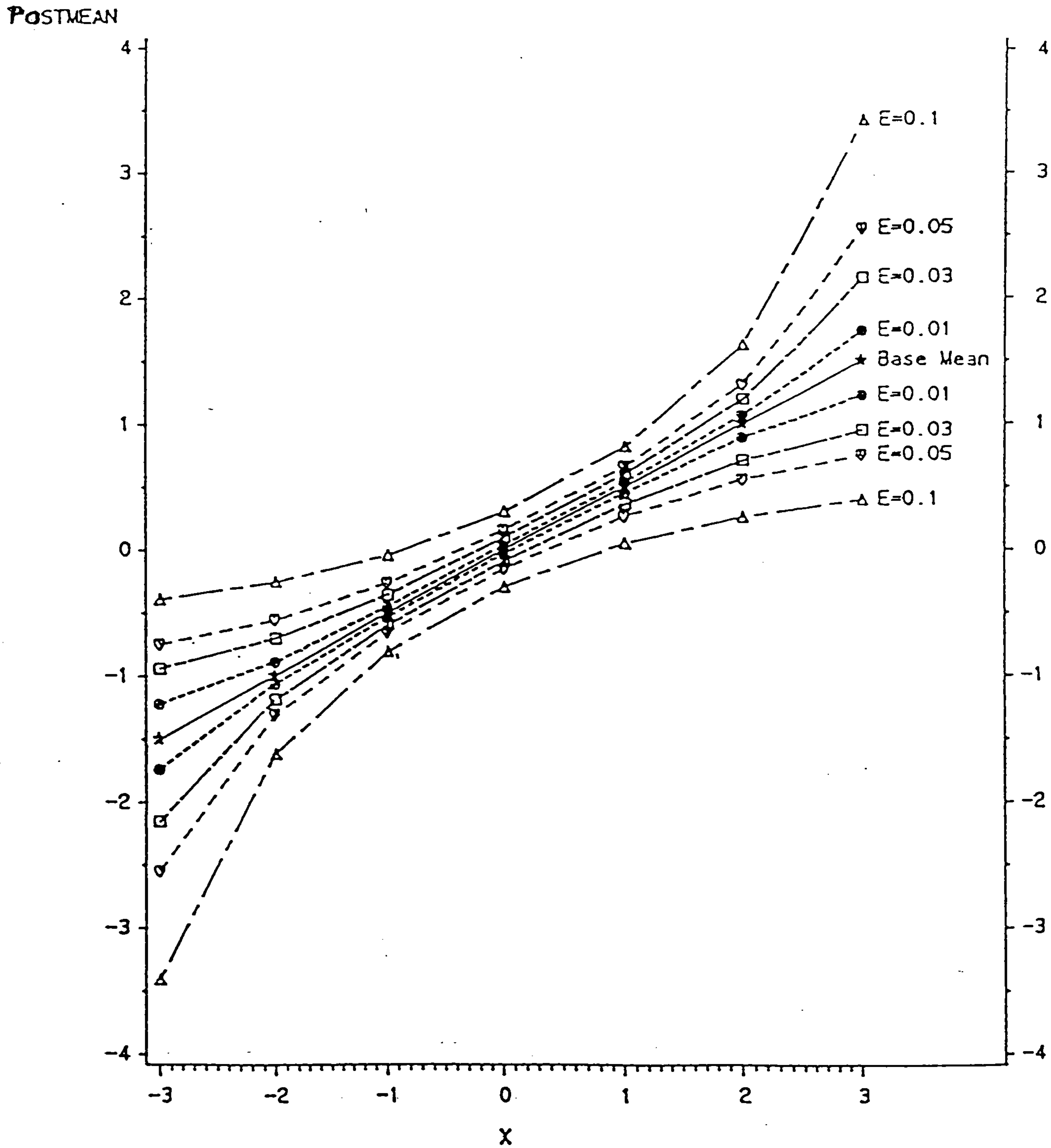

Figure 2: Supremum and infimum of posterior mean in ε-Kolmogorv neighborhood of $N(0,1)$. The likelihood is $N(X,1)$ and 'E' denotes ε. Base mean is the posterior mean w.r.t. $N(0,1)$ prior

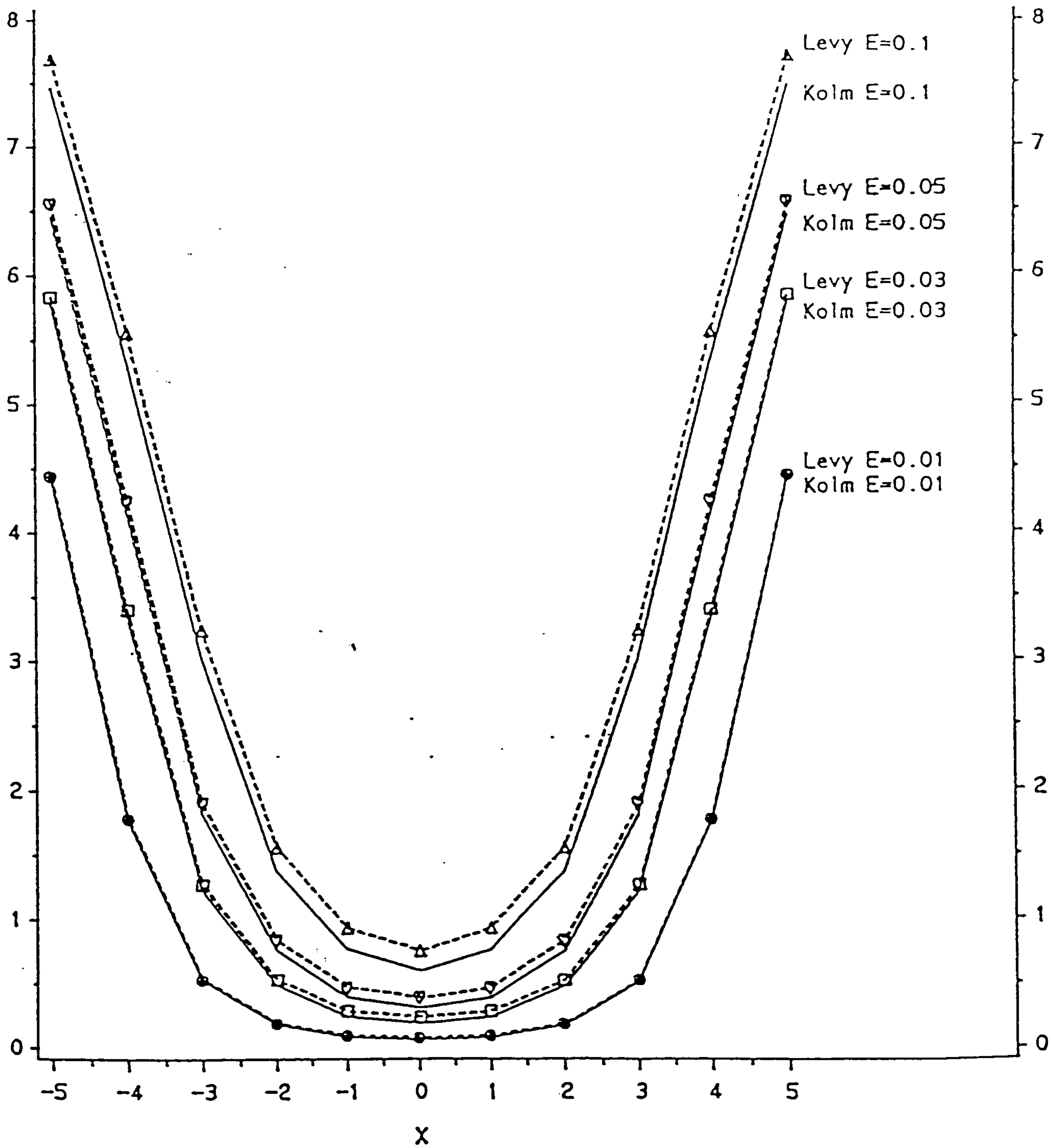

Figure 3: Ranges of posterior mean in ε-Kolmogorov and Levý neighborhoods of $N(0,1)$. The likelihood is $N(X,1)$ and 'E' denotes ε. The solid lines are the ranges in the Kolmogorov classes. The dashed lines are those in the Levý classes.

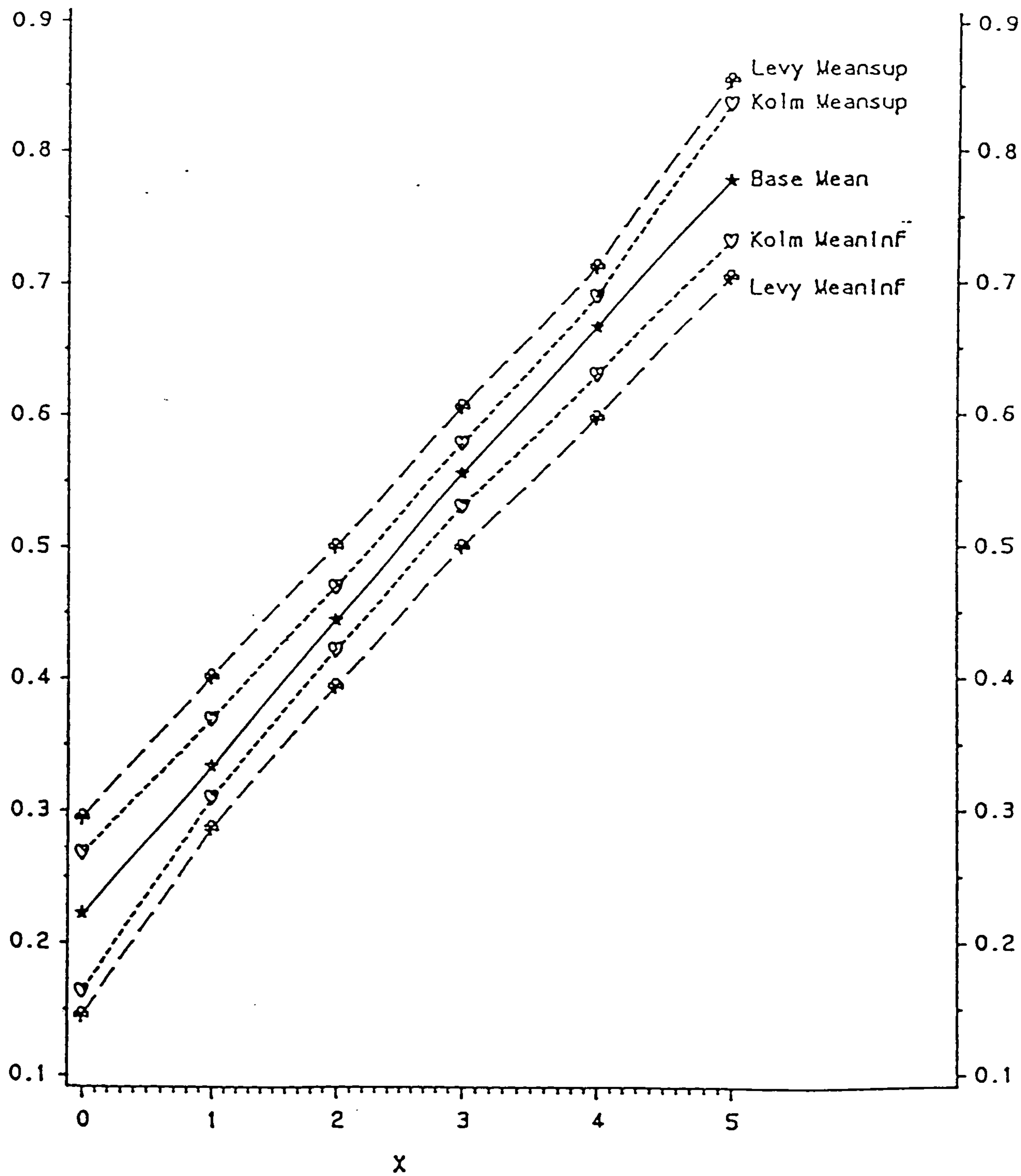

Figure 4: Supremum and infimum of posterior mean in ε-Kolmogorv and ε-Levy neighborhoods of Beta(2,2). The likelihood is Binomial$(5, \theta)$. $\varepsilon = 0.03$. Base mean is the posterior mean w.r.t. Beta(2,2) prior.

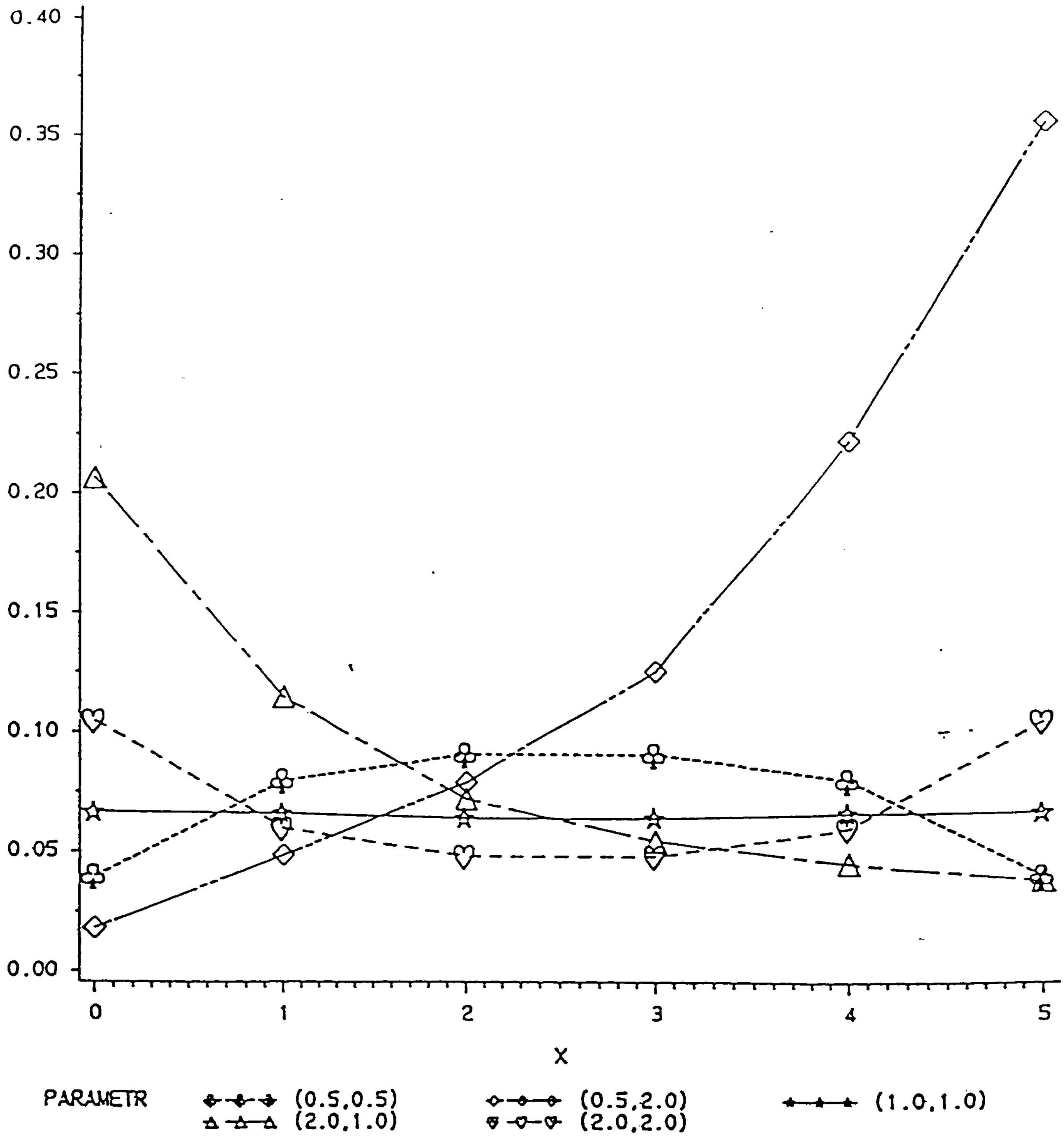

Figure 5: Ranges of posterior mean in ε-Kolmogorov neighborhoods of different Beta base priors. The likelihood is Binomial$(5, \theta)$ and $\varepsilon = 0.03$. The different pairs of Beta parameters are shown at the bottom of the figure.