Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

LOCAL ROBUSTNESS AND INFLUENCE FOR CONTAMINATION CLASSES OF PRIOR DISTRIBUTIONS

BY ELÍAS MORENO, CARMEN MARTÍNEZ AND J. ANTONIO CANO
University of Granada and University of Murcia

In the well-known ε-contamination class of prior distributions $\Gamma = \{\pi(\theta) : \pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta),\ q \in Q\}$, ε represents the degree of uncertainty on the base prior $\pi_0(\theta)$ and Q the allowed class of contaminations. We argue here that the uncertainty we have on $\pi_0(\theta)$ typically is stronger on its tails than on its body. This idea is formalized through a more general $\varepsilon(\theta)$-contamination class that might be seen as a local robustification of $\pi_0(\theta)$. When Q is defined by quantile constraints, the admissible classes of functions $\varepsilon(\theta)$ capable of maintaining the prior information for the resulting priors $\pi(\theta)$ are characterized and robust posterior analysis is carried out. Influence analysis is also considered. In this setting Fréchet derivatives are useful tools: they are easily interpreted and easily computed. Interactive robustness based on influence analysis is discussed.

1. Introduction. Let $\mathbf{x}$ be a set of data which will be assumed to arise from a density $f(\mathbf{x} \mid \theta)$, where θ denotes unknown parameters in the space Θ. Robust Bayesian Analysis assumes uncertainty on the prior distribution $\pi(\theta)$ and models such an uncertainty by considering classes of priors for which robustness analyses are carried out. One of the most interesting classes is the contamination class

$$\Gamma = \{\pi(\theta) : \pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta),\ q \in Q\}, \tag{1}$$

which is proposed as follows. Some prior beliefs are established and a base prior $\pi_0(\theta)$ matching these requirements is elicited. A constant ε, $0 < \varepsilon < 1$, reflecting our degree of uncertainty on the functional form of $\pi_0(\theta)$, is specified. Finally, the class Q of all possible priors compatible with the prior beliefs is considered.

Prior beliefs are expressed by the probabilities of some sets C_i, $i \geq 1$, which form a partition of the parameter space Θ. Therefore, the prior should be any probability measure $\pi(\theta)$ such that $P^{\pi}(C_i) = \alpha_i$, $i \geq 1$, where α_i are known. The base prior $\pi_0(\theta)$ is then chosen such that $P^{\pi_0}(C_i) = \alpha_i$, $i \geq 1$, and the class Q is

$$Q = \{q(\theta) : P^{q}(C_i) = P^{\pi_0}(C_i),\ i \geq 1\}. \tag{2}$$

AMS Subject classification: Primary 62F15; secondary 62A15.
Key words: ε-contamination class, Fréchet derivative, influence analysis, interactive robustness, local uncertainty.

A more general quantile information than that given in (2) might be natural. Suppose that $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ and we are confident on the marginal $\pi_{01}(\theta_1)$ and on the conditional probabilities

$$P^{\pi_0}(C_i \mid \theta_1) = \int_{C_i} \pi_0(d\theta_2 \mid \theta_1),\ i = 1, \ldots, n,$$

where $\{C_i,\ i = 1, \ldots, n\}$ is a partition of $\mathbb{R}$. In this case the prior probabilities of the sets in the σ-field $\mathcal{B} = \sigma(C \times C_i;\ C$ is any Borel set, $\ i = 1, \ldots, n)$ which is non-generated by a countable partition of $\mathbb{R}^2$, are completely specified. Class Q would then be

$$Q_{\mathcal{B}} = \{q(\theta) : \int_B q(d\theta) = \int_B \pi_0(d\theta),\ B \in \mathcal{B}\}. \tag{3}$$

Symmetry or sphericity, are also usual constraints that we also could bring to this analysis.

References on contamination classes where several types of prior beliefs are considered include Berger (1994), Berger and Berliner (1986), Berger and Moreno (1994), Berger and O'Hagan (1988), Bose (1994), Delampady and Dey (1994), Lavine, Wasserman and Wolpert (1991), Liseo, Moreno and Salinetti (1996), Moreno and Cano (1991, 1995), Moreno and Pericchi (1993), Sivaganesan (1988, 1989), Sivaganesan and Berger (1989), and O'Hagan and Berger (1988), among others.

In the class Γ given in (1) for Q as in (2) or (3) the confidence on $\pi_0(\theta)\mathbf{1}_{C_i}(\theta)$ or $\pi_0(\theta_1, \theta_2)\mathbf{1}_{C \times C_i}(\theta_1, \theta_2)$ is the same for any $i \geq 1$. It seems, however, more reasonable to assume that the confidence degree on $\pi_0(\theta)\mathbf{1}_{C_i}(\theta)$ or $\pi_0(\theta_1, \theta_2)\mathbf{1}_{C \times C_i}(\theta_1, \theta_2)$ might depend on the position of the set C_i. That is, on the tails of $\pi_0(\theta)$ we will generally be more uncertain than on its body. This implies that ε should instead be a function of θ. To replace ε with $\varepsilon(\theta)$ in (1) will result not only in a more realistic model of our prior uncertainty but it will also have an important impact on the size of the class and consequently on posterior robustness. Notice that if Q is a convex class, then $\varepsilon_1 \leq \varepsilon_2$ implies that the ε_1-contamination class is contained in the ε_2-contamination class.

It is clear, however, that arbitrary functions $\varepsilon(\theta)$, $0 \leq \varepsilon(\theta) \leq 1$, cannot be used if the prior beliefs, given by the conditions (2) or (3), have to be satisfied. In Section 2 we characterize the admissible class of functions $\varepsilon(\theta)$ such that the resulting priors in Γ satisfy the constraints (2) or (3). For a given quantity of interest $\varphi(\theta)$ the posterior range as the prior ranges over this $\varepsilon(\theta)$-contamination class is also given.

A procedure to analyze which of the sets C_i, $i \geq 1$, is the most influential on the posterior range is given in Section 3. In this aspect restricted Fréchet derivatives are shown to be simple and useful tools. Based on this influence analysis and ideas in Berger (1994) interactive robustness is discussed in Section 4. Finally, Section 5 gives some concluding remarks.

2. Contamination classes with quantile constraints. An extension to the class Γ given in (1), where Q is the class of priors defined by (3) is the $\varepsilon(\theta)$-contamination class

$$\Gamma = \{\pi(\theta) : \pi(\theta) = (1 - \varepsilon(\theta))\pi_0(\theta) + \varepsilon(\theta)q(\theta),\ q \in Q\}, \tag{4}$$

where the function $(1 - \varepsilon(\theta))$, $0 \leq \varepsilon(\theta) \leq 1$, expresses our confidence on $\pi_0(\theta)$ for each of the points θ. For an arbitrary $\varepsilon(\theta)$ the resulting prior $\pi(\theta)$, however, will not satisfy the prior beliefs stated in (3). The question is to characterize the class of functions $\varepsilon(\theta)$ that satisfies those beliefs. The next theorem gives the solution in the case of prior information as in (3). The case (2) is obtained as Corollary 1.

THEOREM 1. *Let $(\Theta, \mathcal{A})$ be the measurable parameter space of the statistical problem, and let $\mathcal{B}$ be a sub sigma field of $\mathcal{A}$. Consider the class*

$$\Gamma_{\mathcal{B}} = \{\pi(\theta) : \pi(\theta) = (1 - \varepsilon(\theta))\pi_0(\theta) + \varepsilon(\theta)q(\theta),\ q \in Q_{\mathcal{B}}\}$$

where $\varepsilon(\theta)$ is an $\mathcal{A}$-measurable function $0 \leq \varepsilon(\theta) \leq 1$, and

$$Q_{\mathcal{B}} = \{q(\theta) : \int_B q(d\theta) = \int_B \pi_0(d\theta),\ B \in \mathcal{B}\}.$$

Then, (i) if $\varepsilon(\theta)$ is $\mathcal{B}$-measurable it follows that for any $\pi \in \Gamma_{\mathcal{B}}$ and $B \in \mathcal{B}$, the equality $\int_B \pi(d\theta) = \int_B \pi_0(d\theta)$, holds.

Conversely (ii) suppose that for any $\pi \in \Gamma_{\mathcal{B}}$ and $B \in \mathcal{B}$ the equality $\int_B \pi(d\theta) = \int_B \pi_0(d\theta)$ holds. Then, if $\mathcal{B}$ is such that for any set $A \in \mathcal{A} - \mathcal{B}$ there exist $q_1(d\theta)$ and $q_2(d\theta)$ belonging to $Q_{\mathcal{B}}$ such that

$$\int_A q_1(d\theta) = (\pi_0^{\mathcal{B}})_*(A), \int_A q_2(d\theta) = (\pi_0^{\mathcal{B}})^*(A),$$

where $(\pi_0^{\mathcal{B}})_*$, $(\pi_0^{\mathcal{B}})^*$ are the inner and outer measures of the restriction of π_0 to $\mathcal{B}$ respectively, it follows that $\varepsilon(\theta)$ is a $\mathcal{B}$-measurable function a.s. $[\pi_0^{\mathcal{B}}]$.

PROOF. (i) If $\varepsilon(\theta)$ is $\mathcal{B}$-measurable, the conditional expectation to $\mathcal{B}$ with respect to $q \in Q_{\mathcal{B}}$ satisfies $E^q[\varepsilon(\theta) \mid \mathcal{B}] = \varepsilon(\theta)$, a.s. $[\pi_0^{\mathcal{B}}]$. From the definition of conditional expectation we have that for any $B \in \mathcal{B}$

$$\int_B \varepsilon(\theta)q(d\theta) = \int_B E^q[\varepsilon(\theta) \mid \mathcal{B}]q^{\mathcal{B}}(d\theta) = \int_B E^q[\varepsilon(\theta) \mid \mathcal{B}]\pi_0^{\mathcal{B}}(d\theta) = \int_B \varepsilon(\theta)\pi_0(d\theta),$$

and this proves assertion (i).

(ii) Suppose that $\varepsilon(\theta)$ is not an a.s. $\mathcal{B}$-measurable function. Then there exists an interval $[0, \alpha]$, $\alpha \geq 0$, such that $\varepsilon^{-1}([0,\alpha]) = A \in \mathcal{A} - \mathcal{B}$ and $(\pi_0^{\mathcal{B}})_*(A) < (\pi_0^{\mathcal{B}})^*(A)$. Let B_1, B_2 be sets in $\mathcal{B}$ such that $B_1 \subset A \subset B_2$, $\pi_0(B_1) = (\pi_0^{\mathcal{B}})_*(A)$ and $\pi_0(B_2) = (\pi_0^{\mathcal{B}})^*(A)$. Consider the set $B_2 - B_1 \in \mathcal{B}$. Then,

$$\int_{B_2-B_1} \varepsilon(\theta)q_1(d\theta) = \int_{B_2-A} \varepsilon(\theta)q_1(d\theta) + \int_{A-B_1} \varepsilon(\theta)q_1(d\theta) = \int_{B_2-A} \varepsilon(\theta)q_1(d\theta)$$

$$> \alpha(\pi_0(B_2) - \pi_0(B_1)),$$

$$\int_{B_2-B_1} \varepsilon(\theta)q_2(d\theta) = \int_{B_2-A} \varepsilon(\theta)q_2(d\theta) + \int_{A-B_1} \varepsilon(\theta)q_2(d\theta) = \int_{A-B_1} \varepsilon(\theta)q_2(d\theta)$$

$$\leq \alpha(\pi_0(B_2) - \pi_0(B_1)),$$

but since that $q_1, q_2 \in Q_{\mathcal{B}}$ the inequalities above are not possible. So that, $\varepsilon(\theta)$ has to be a.s. $\mathcal{B}$-measurable. This proves assertion (ii). □

COROLLARY 1. *Let $\mathcal{B} = \sigma(C_i,\ i \geq 1)$ be the sub sigma field of $\mathcal{A}$ generated by the partition $\{C_i,\ i \geq 1\}$. Then, any $\pi \in \Gamma_{\mathcal{B}}$ and $B \in \mathcal{B}$ satisfy $\int_B \pi(\theta)d\theta = \int_B \pi_0(\theta)d\theta$, if and only if $\varepsilon(\theta)$ is a $\mathcal{B}$-measurable function.*

PROOF. For any set A, we have that

$$(\pi_0^{\mathcal{B}})_*(A) = \begin{cases} \pi_0(C), & \text{if } A \supset C = \cup_{i \in I} C_i, \\ 0, & \text{otherwise,} \end{cases}$$

where I is some subset of indices of $\{1, 2, ...\}$.

Therefore, there exists $q_1 \in Q_{\mathcal{B}}$ such that $\int_A q_1(d\theta) = (\pi_0^{\mathcal{B}})_*(A)$. The existence of $q_2 \in Q_{\mathcal{B}}$ can be similarly proved and Theorem 1 applies. This completes the proof. □

Theorem 1 means that the expert can choose $\varepsilon(\theta)$ in the class of $\mathcal{B}$-measurable functions to express his degree of uncertainty on different parts of $\pi_0(\theta)$, maintaining at the same time fixed the probabilities of every set in $\mathcal{B}$.

COROLLARY 2. *If Q is the class of all prior distributions the only possible contamination model Γ is that given in (1), ε being a constant in the interval* $[0, 1]$.

PROOF. The proof follows from Corollary 1 by observing that $\mathcal{B} = \sigma(\emptyset, \Theta)$, so that the $\mathcal{B}$-measurable functions $\varepsilon(\theta)$ are now constant functions. □

The sup and inf of $E^{\pi}[\varphi(\theta) \mid \mathbf{x}]$, as π ranges over $\Gamma_{\mathcal{B}}$ are derived with minor changes from results in Moreno and Cano (1991). For simplicity we give the sup and inf of $E^{\pi}[\varphi(\theta) \mid \mathbf{x}]$ as π varies over $\Gamma_{\mathcal{B}}$ where $\mathcal{B} = \sigma(C_i,\ 1 \leq i \leq n)$. Notice that in this case the admissible class of $\varepsilon(\theta)$'s are step functions, i.e., $\varepsilon(\theta) = \sum \varepsilon_i \mathbf{1}_{C_i}(\theta)$, where $0 \leq \varepsilon_i \leq 1$, $i = 1, \ldots, n$.

THEOREM 2. *For any integrable function $\varphi(\theta)$ with respect to $\pi \in \Gamma_{\mathcal{B}}$, the supremum of the posterior expectation of $\varphi(\theta)$ as π ranges over $\Gamma_{\mathcal{B}}$, is given by*

$\sup_{\pi \in \Gamma_{\mathcal{B}}} E^{\pi}[\varphi(\theta) \mid \mathbf{x}] =$

$$\sup_{\theta_i \in C_i, i=1,\ldots,n} \frac{\sum_{i=1}^{n}(1-\varepsilon_i)\int_{C_i} \varphi(\theta) f(\mathbf{x}|\theta)\pi_0(d\theta) + \sum_{i=1}^{n} \varepsilon_i \varphi(\theta_i) f(\mathbf{x}|\theta_i)\alpha_i}{\sum_{i=1}^{n}(1-\varepsilon_i)\int_{C_i} f(\mathbf{x}|\theta)\pi_0(d\theta) + \sum_{i=1}^{n} \varepsilon_i f(\mathbf{x}|\theta_i)\alpha_i}$$

where $\alpha_i = \int_{C_i} \pi_0(d\theta)$, $i \geq 1$.
The infimum is obtained replacing sup with inf.

The case where $\varphi(\theta) = \mathbf{1}_A(\theta)$ is more simple as Theorem 3 shows.

THEOREM 3. *Let $A \in \mathcal{A}$ be an arbitrary set. Then, the extreme values of the posterior probability of A as π ranges over $\Gamma_{\mathcal{B}}$ are given by*
$\sup_{\pi \in \Gamma_{\mathcal{B}}} P^{\pi}(A|\mathbf{x}) =$

$$\left[1 + \frac{\sum_{i=1}^{n}(1-\varepsilon_i)\int_{A^c \cap C_i} f(\mathbf{x}|\theta)\pi_0(d\theta) + \sum_{i \in K} \varepsilon_i \alpha_i \inf_{\theta \in C_i} f(\mathbf{x}|\theta)}{\sum_{i=1}^{n}(1-\varepsilon_i)\int_{A \cap C_i} f(\mathbf{x}|\theta)\pi_0(d\theta) + \sum_{i \in J} \varepsilon_i \alpha_i \sup_{\theta \in A \cap C_i} f(\mathbf{x}|\theta)}\right]^{-1},$$

$\inf_{\pi \in \Gamma_{\mathcal{B}}} P^{\pi}(A|\mathbf{x}) =$

$$\left[1 + \frac{\sum_{i=1}^{n}(1-\varepsilon_i)\int_{A^c \cap C_i} f(\mathbf{x}|\theta)\pi_0(d\theta) + \sum_{i=1}^{n} \varepsilon_i \alpha_i \sup_{\theta \in A^c \cap C_i} f(\mathbf{x}|\theta)}{\sum_{i=1}^{n}(1-\varepsilon_i)\int_{A \cap C_i} f(\mathbf{x}|\theta)\pi_0(d\theta) + \sum_{i \in I} \varepsilon_i \alpha_i \inf_{\theta \in C_i} f(\mathbf{x}|\theta)}\right]^{-1},$$

where the subsets of indices I, J, K of the set $\{1, 2, \ldots, n\}$ are defined by $i \in I$ if and only if $C_i \subseteq A$, $i \in J$ if and only if $C_i \cap A \neq \emptyset$, $i \in K$ if and only if $C_i \cap A = \emptyset$.

EXAMPLE 1. Let X be a random variable $\mathcal{N}(\theta, 1)$ distributed. Suppose we are interested in testing $H_0 : \theta \leq 0$. It is elicited that the distribution of θ is approximately symmetric around zero and that the probabilities of the sets $C_1 = (-\infty, -0.954]$, $C_2 = (-0.954, 0.954]$, and $C_3 = (0.954, \infty)$ are

$$\int_{-\infty}^{-0.954} \pi(\theta)d\theta = 0.25, \int_{-0.954}^{0.954} \pi(\theta)d\theta = 0.5, \int_{0.954}^{\infty} \pi(\theta)d\theta = 0.25.$$

The base prior $\pi_0(\theta) = \mathcal{N}(0, 2)$ matches these quantiles and is typically used to form the class Γ_0

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

where $\varepsilon = 0.2$ and

$$Q = \{q : \int_{-\infty}^{-0.954} q(\theta)d\theta = 0.25, \int_{-0.954}^{0.954} q(\theta)d\theta = 0.5, \int_{0.954}^{\infty} q(\theta)d\theta = 0.25\}.$$

The posterior imprecision of H_0 with respect to this class for various values of $\mathbf{x}$ are displayed in the second column of Table 1. This imprecision is defined as

$$\Delta_{\Gamma_0} P^\pi(H_0 \mid \mathbf{x}) = \sup_{\pi \in \Gamma_0} P^\pi(H_0 \mid \mathbf{x}) - \inf_{\pi \in \Gamma_0} P^\pi(H_0 \mid \mathbf{x}).$$

If in the class Γ_0 ε is replaced with $\varepsilon(\theta) = 0\mathbf{1}_{C_2}(\theta) + 0.5\mathbf{1}_{C_2^c}(\theta)$, where C_2^c denotes the complement of C_2, only an uncertainty of 0.5 on the tails of $\pi_0(\theta)$ is allowed. Let us denote by $\Delta_{\Gamma_{0.5}} P^\pi(H_0 \mid \mathbf{x})$ the posterior imprecision of H_0 with respect to the class associated with this $\varepsilon(\theta)$. Values of this imprecision for various observations $\mathbf{x}$ are given in the third column of Table 1.

TABLE 1

Posterior imprecisions of H_0 for Γ_0 and $\Gamma_{0.5}$

$\mathbf{x}$	$\Delta_{\Gamma_0} P^\pi(H_0 \mid \mathbf{x})$	$\Delta_{\Gamma_{0.5}} P^\pi(H_0 \mid \mathbf{x})$
0	0.21	0.13
0.5	0.18	0.12
1.0	0.14	0.09
1.5	0.09	0.04

Table 1 shows that a significant reduction of posterior imprecision is obtained if only uncertainty on the tails of $\pi_0(\theta)$ is considered, even when this prior uncertainty is as big as 0.5. Probably the situation considered in the last column is a better reflection of our posterior uncertainty on $\pi_0(\theta)$ than that of the second column and therefore those numbers would be a more realistic measure of robustness of $\pi_0(\theta)$.

3. Influence and sensitivity. In the $\varepsilon(\theta)$-contamination class $\Gamma_{\varepsilon(\theta)}$, where $\varepsilon(\theta) = \sum \varepsilon_i \mathbf{1}_{C_i}(\theta)$, the uncertainty on $\pi_0(\theta)$ has been decomposed into local uncertainty on each of the elements of the partition $\{C_i,\ i \geq 1\}$. A natural question is which set C_i of the partition has the largest effect on the posterior range of our quantity of interest $\varphi(\theta)$ as the prior varies over $\Gamma_{\varepsilon(\theta)}$.

A way to answer this question is as follows. Consider the class

$$\Gamma^i = \{\pi(\theta) : \pi(\theta) = \pi_0(\theta) + \varepsilon_i \mathbf{1}_{C_i}(\theta)[q(\theta) - \pi_0(\theta)],\ q \in Q\}, \tag{5}$$

where only uncertainty on C_i is allowed. This class is derived from $\Gamma_{\varepsilon(\theta)}$ for $\varepsilon_j = 0, j \neq i$. Let $R_i(\mathbf{x})$ be the posterior imprecision of $\varphi(\theta)$ as π varies over Γ^i, i.e., $R_i(\mathbf{x}) = \sup_{\pi \in \Gamma^i} E^\pi(\varphi(\theta) \mid \mathbf{x}) - \inf_{\pi \in \Gamma^i} E^\pi(\varphi(\theta) \mid \mathbf{x})$ and $R(\mathbf{x})$ the corresponding one as the prior varies over $\Gamma_{\varepsilon(\theta)}$.

LEMMA 1. *For any sample observation* $\mathbf{x}$, *the inequality*

$$R(\mathbf{x}) \geq \max_{i \geq 1} R_i(\mathbf{x})$$

holds.

PROOF. Any $\pi(\theta)$ in Γ^i can be written as

$$\pi(\theta) = \pi_0(\theta) + \sum \varepsilon_i \mathbf{1}_{C_i}(\theta)[q(\theta) - \pi_0(\theta)],$$

where

$$q(\theta) = \begin{cases} q(\theta) & \text{if } \theta \in C_i, \\ \pi_0(\theta) & \text{otherwise.} \end{cases}$$

Since $\pi_0(\theta) \in Q$ it follows that $\Gamma^i \subset \Gamma_{\varepsilon(\theta)}$ for any i. This proves the assertion. □

Lemma 1 means that posterior robustness related to the class $\Gamma_{\varepsilon(\theta)}$ cannot be achieved if the classes Γ^i, $i \geq 1$, are not robust. Therefore, if $R_k(\mathbf{x}) = \max_{i \geq 1} R_i(\mathbf{x})$ is not sufficiently small, the prior elicitation effort should be concentrated on the set C_k.

EXAMPLE 1 (continued). In the situation of Example 1 consider the class Γ_0. For various values of $\mathbf{x}$ the posterior imprecisions $R(\mathbf{x})$, $R_i(\mathbf{x})$, $i = 1, 2, 3$, of $\varphi(\theta) = \mathbf{1}_{H_0}(\theta)$, are given in Table 2.

TABLE 2

Values of $R_i(\mathbf{x})$ and $R(\mathbf{x})$

$\mathbf{x}$	$R_1(\mathbf{x})$	$R_2(\mathbf{x})$	$R_3(\mathbf{x})$	$R(\mathbf{x})$
0	0.03	0.17	0.03	0.21
0.5	0.02	0.16	0.03	0.18
1.0	0.01	0.14	0.02	0.14
1.5	0.00	0.09	0.01	0.09

Table 2 shows that the biggest posterior imprecision corresponds to Γ^2 which is associated with $C_2 = (-0.954, 0.954)$, for all the considered sample points. Notice that the values of $R_2(\mathbf{x})$ are very close to those of $R(\mathbf{x})$. Thus, we should try to reduce uncertainty on C_2. This explains the dramatic reduction we obtained in Example 1 on the posterior ranges of H_0 with respect to the ε-contamination class Γ_0 ($\varepsilon = 0.2$) when no prior uncertainty on C_2 were considered, even when the uncertainty on $C_1 \cup C_3$ was assumed to be very big, i.e., $\varepsilon(\theta) = 0\mathbf{1}_{C_2}(\theta) + 0.5\mathbf{1}_{C_1 \cup C_3}(\theta)$. Note also that as $\mathbf{x}$ increases the set C_3 is becoming more influential than C_1 which is intuitively obvious.

The way shown above to study the posterior influence of the sets $\{C_i,\ i \geq 1\}$ is very understandable and only requires the computation of the posterior ranges for the sub classes Γ^i, $i \geq 1$.

Other tools even easier to compute are the Fréchet derivatives (see Cuevas and Sanz (1988), Ruggeri and Wasserman (1993) and Sivaganesan (1993)).

It is felt, however, that the main drawback of the functional derivative is that it is not clear how to interpret its value (see for instance Berger's reply to Gustafson and Wasserman, in Berger (1994) p. 121). In this section we consider restricted derivatives of the Bayes operator for which somehow the difficulty of interpretation is avoided as long as we used them comparatively.

For a given bounded quantity of interest $\varphi(\theta)$, a prior π in $\Gamma_{\varepsilon(\theta)}$ with Q the class of all discrete prior distributions satisfying (2), and $\varepsilon(\theta) = \sum \varepsilon_i \mathbf{1}_{C_i}(\theta)$, consider the following functional

$$T_\varphi(\pi) = \frac{\int \varphi(\theta) f(\mathbf{x} \mid \theta)\pi(d\theta)}{\int f(\mathbf{x} \mid \theta)\pi(d\theta)}. \tag{6}$$

Let us denote $N(\pi) = \int_\Theta \varphi(\theta) f(\mathbf{x} \mid \theta)\pi(d\theta)$, and $D(\pi) = \int_\Theta f(\mathbf{x} \mid \theta)\pi(d\theta)$. Q is taken as the discrete probability measures because the posterior ranges of $T_\varphi(\pi)$ are attained at discrete priors.

The Fréchet derivative $\dot{T}_\varphi(\pi)$ at point π_0 can be expressed as (see Diaconis and Freedman (1986), Ruggeri and Wasserman (1993)),

$$\dot{T}_\varphi(\pi_0) = D(\pi_0)^{-1}\{N(\delta) - T_\varphi(\pi_0)D(\delta)\}. \tag{7}$$

For the signed measures $\delta(\theta) = \sum_i \varepsilon_i \mathbf{1}_{C_i}(\theta)(q(\theta) - \pi_0(\theta))$, (7) turns out to be

$$\dot{T}_\varphi(\pi_0) = D(\pi_0)^{-1} \sum_i \varepsilon_i \int_{C_i} \{[\varphi(\theta) - T_\varphi(\pi_0)] f(\mathbf{x} \mid \theta) + \frac{k_{i0}}{\alpha_i}\} q(d\theta), \tag{8}$$

where $\alpha_i = P^{\pi_0}(C_i)$, and $k_{i0} = \int_{C_i} f(\mathbf{x} \mid \theta)\{T_\varphi(\pi_0) - \varphi(\theta)\}\pi_0(d\theta)$.

If π_0 is non-atomic, then the norm of δ related to the total variation distance is $||\delta|| = \sum_i \varepsilon_i \alpha_i$, and from (8) we have

$$||\dot{T}_\varphi(\pi_0)|| = \sup_{q \in Q} \frac{|\dot{T}_\varphi(\pi_0)|}{||\delta||} = D(\pi_0)^{-1} \frac{\max\{\sum_i \varepsilon_i \alpha_i \overline{L}_i(\theta), -\sum_i \varepsilon_i \alpha_i \underline{L}_i(\theta)\}}{\sum_i \varepsilon_i \alpha_i},$$

where $\overline{L}_i(\theta) = \sup_{\theta \in C_i} L_i(\theta)$, $\underline{L}_i(\theta) = \inf_{\theta \in C_i} L_i(\theta)$ and

$$L_i(\theta) = f(\mathbf{x} \mid \theta)\{\varphi(\theta) - T_\varphi(\pi_0)\} + \frac{k_{i0}}{\alpha_i}.$$

Consider the operator

$$T_\varphi^i(\pi) = T_\varphi(\pi)_{|\varepsilon_j = 0, j \neq i}.$$

This is the Bayes operator related to the class

(9) $\Gamma^i = \{\pi(\theta) : \pi(\theta) = \pi_0(\theta) + \varepsilon_i \mathbf{1}_{C_i}(\theta)[q(\theta) - \pi_0(\theta)],\ q \in Q(\text{discrete})\}$,

For $\delta_i(\theta) = \varepsilon_i \mathbf{1}_{C_i}(\theta)[q(\theta) - \pi_0(\theta)]$ the norm of the Fréchet derivative

$$\|\dot{T}^i_\varphi(\pi_0)\| = \sup_{q \in Q} \frac{|\dot{T}^i_\varphi(\pi_0)|}{\|\delta_i\|},$$

is going to be taken as the measure of sensitivity to small changes in π_0 in the class Γ^i. This would give an indication on the posterior influence of the set C_i.

These derivatives turn out to be the restrictions of $\|\dot{T}_\varphi(\pi_0)\|$ to $\varepsilon_j = 0, j \neq i$, i.e., $\|\dot{T}^i_\varphi(\pi_0)\| = \|\dot{T}_\varphi(\pi_0)\||_{\varepsilon_j=0,j\neq i}$, $i = 1, \ldots, n$, and therefore, their expressions are

$$\|\dot{T}^i_\varphi(\pi_0)\| = \frac{1}{D(\pi_0)} \max\{\overline{L}_i(\theta),\ -\underline{L}_i(\theta)\}. \tag{10}$$

Note that $\|\dot{T}^i_\varphi(\pi_0)\|$ does not depend on $\varepsilon_i \neq 0$, the degree of contamination allowed, although it depends on the prior mass that the class puts on C_i.

EXAMPLE 1 (continued). For the classes Γ^i given in (9) associated with the three sets C_1, C_2 and C_3 in Example 1 and several values of $\mathbf{x}$, the norms of the derivatives for $\varphi(\theta) = \mathbf{1}_{H_0}(\theta)$, say $\|\dot{T}^i_{H_0}(\pi_0)\|$, calculated from (10) are displayed in Table 3.

TABLE 3

Norms of derivatives $\|\dot{T}^i_\varphi(\pi_0)\|$; $\varphi(\theta) = \mathbf{1}_{H_0}(\theta)$

$\mathbf{x}$	$\|\dot{T}^1_{H_0}(\pi_0)\|$	$\|\dot{T}^2_{H_0}(\pi_0)\|$	$\|\dot{T}^3_{H_0}(\pi_0)\|$
0	0.31	0.87	0.31
0.5	0.26	0.97	0.31
1.0	0.17	0.87	0.30
1.5	0.08	0.63	0.23

Table 3 shows that the biggest values of the norms of the derivatives correspond to the class Γ^2 for any of the considered sample values. It is also observed that as $\mathbf{x}$ increases $\|\dot{T}^3_{H_0}(\pi_0)\|$ becomes bigger than $\|\dot{T}^1_{H_0}(\pi_0)\|$.

The posterior ranges of H_0 as π ranges over Γ^i, $i = 1, 2, 3$, given in Table 2, are in agreement with the corresponding values of the norms given in Table 3. These derivatives are also easier to compute than the posterior ranges.

4. Interactive robustness. The idea of using Bayesian robustness to guide the elicitation process has been pointed out in the thoughtful paper by Berger (1994). He argues that *Since we are eliciting in terms of quantiles, this means that a new quantile θ^* must be chosen with the associated α^* (for* *the new interval created) being elicited.* This idea will be implemented in the ε-contamination context.

It is relevant to appropriately choose in which set we put the new quantile. In fact, it might be that we do not gain in robustness even when we add an infinite number of quantiles. This is a consequence of Lemma 1.

For instance, if in Example 1 we consider the observation $\mathbf{x} = 1$ we can create as many new intervals inside either C_1 or C_3 as we want but the posterior ranges of H_0 remain 0.14 (see third row in Table 2).

The influence analysis in Section 3 gives clear indications on which set $C_k = (\theta_{k-1}, \ \theta_k]$ the new quantile must be chosen. Furthermore, if our confidence on $\pi_0(\theta)\mathbf{1}_{C_k}(\theta)$ is not small it seems reasonable to take α^* such that

$$\alpha^* = \int_{\theta_{k-1}}^{\theta^*} \pi_0(d\theta).$$

If $R^*(x, \theta^*)$ denotes the posterior range of our quantity of interest as the prior varies over the contamination class

$$\Gamma^* = \{\pi(\theta) : \pi(\theta) \in \Gamma_{\varepsilon(\theta)}, \int_{\theta_{k-1}}^{\theta^*} \pi(d\theta) = \int_{\theta_{k-1}}^{\theta^*} \pi_0(d\theta)\},$$

the point θ^* is determined as the most favorable point in C_k, that is

$$R^*(\mathbf{x}, \theta^*) = \inf_{b \in C_k} R^*(\mathbf{x}, b).$$

When our prior confidence on $\pi_0(\theta)\mathbf{1}_{C_k}(\theta)$ is small, $\pi_0(\theta)\mathbf{1}_{C_k}(\theta)$ should be replaced by some other base $\pi_1(\theta)\mathbf{1}_{C_k}(\theta)$. This means that we are acting as if we were starting again the problem of modelling prior uncertainty but instead of considering the whole parameter Θ we consider the set C_k where the quantile θ^* is subjectively elicited and the new $\pi_1(\theta)\mathbf{1}_{C_k}(\theta)$ is chosen in agreement with it.

It is clear that the set in which the new quantile has to be chosen is depending on the observation $\mathbf{x}$ we have. Thus, it can be argued that the new class Γ^* is designed depending upon the observation which somehow might be seen as a data-dependent elicitation. However, interactive robustness means that we jump from samples to priors and therefore some dependence is inherent in the idea.

EXAMPLE 1 (continued). Consider Example 1 and the class Γ_0. We saw in Table 2 and Table 3 that for all the samples considered C_2 was the most influential set. The values for $\theta^* \in C_2$ and the corresponding α^* turn out to be $\theta^* = 0$ and $\alpha^* = 0.25$ for all the considered values of $\mathbf{x}$. The posterior imprecisions $R^*(\mathbf{x})$, $R_i^*(\mathbf{x})$, $i = 1, 2, 3, 4$ are given in Table 4.

TABLE 4

Values of $R_i^*(\mathbf{x})$ and $R^*(\mathbf{x})$

$\mathbf{x}$	$R_1^*(\mathbf{x})$	$R_2^*(\mathbf{x})$	$R_3^*(\mathbf{x})$	$R_4^*(\mathbf{x})$	$R^*(\mathbf{x})$
0	0.03	0.016	0.016	0.03	0.087
0.5	0.02	0.032	0.004	0.03	0.082
1.0	0.01	0.037	0.008	0.02	0.081
1.5	0.00	0.031	0.007	0.01	0.062

From Table 4 it follows that the most influential set is now either C_1^*, or C_4^* for $\mathbf{x} = 0$, and $C_2^* = (-0.954, 0)$ for the other sample values (notice that $P^{\pi_0}(C_2^* \mid \mathbf{x})$ is not the biggest $P^{\pi_0}(C_i^* \mid \mathbf{x})$, $i \geq 1$, $\mathbf{x} = 0.5, 1, 1.5$).

The norms of the restricted derivatives for the class Γ^* are displayed in Table 5.

TABLE 5

Norms of derivatives $||\dot{T}_\varphi^i(\pi_0)||$; $\varphi(\theta) = \mathbf{1}_{H_0}(\theta)$

$\mathbf{x}$	$\lvert\lvert \dot{T}_{H_0}^1(\pi_0)\rvert\rvert$	$\lvert\lvert \dot{T}_{H_0}^2(\pi_0)\rvert\rvert$	$\lvert\lvert \dot{T}_{H_0}^3(\pi_0)\rvert\rvert$	$\lvert\lvert \dot{T}_{H_0}^4(\pi_0)\rvert\rvert$
0	0.31	0.21	0.21	0.31
0.5	0.26	0.33	0.05	0.31
1.0	0.17	0.40	0.10	0.30
1.5	0.08	0.36	0.08	0.23

From Table 5 follows that the biggest norm of the derivatives corresponds to Γ_1^*, Γ_4^*, for $\mathbf{x} = 0$ and Γ_2^* for the other sample values. They give the same indication on the influential set than that given by Table 4. For complex problems, however, these norms of the derivatives are easier to compute than posterior ranges.

5. Conclusions. In this paper an extension of the ε-contamination class of priors suitable to model local robustification of the base prior has been introduced. The motivation is that elicitation difficulties are typically on the tails of the prior distributions. If C_i, $i \geq 1$ are the sets in the partition assumed to be ordered from the body to the tail, our recommendation is to increase the corresponding values ε_i's. A reasonable choice would be to start with $\varepsilon = 0.1$ for sets in the body to $\varepsilon = 0.4$ for sets in the tails.

It should be remarked that in the formulation of this concept no new difficulties arise either in computation or interpretation with respect to the Global Robustness Analysis. The analyses in the first three Sections of the paper are valid regardless the dimension of the parameter space.

A procedure for analyzing the influence on the posterior ranges of the sets in the partition and interactive robustness have been given. In this setting, Fréchet restricted derivatives are interpreted comparatively and are proved to be useful tools for a better understanding of some facts of the priors that provoke imprecision on the posterior answers. The suggestions derived from the comparisons of the values of the norm of the derivatives are very reasonable in the examples we have considered. The hope is that for local and interactive robustness these derivatives are in general good indicators, even when we are aware of the asymptotic poor behaviour of the Fréchet derivatives (see Gustafson, 1994). However, this general relationship between infinitesimal sensitivity and posterior robustness still deserves more research.

REFERENCES

BERGER, J.O. (1994). An overview of Robust Bayesian Analysis (with discussion). *Test* **3** 1 5-124.

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

BERGER, J.O. and MORENO, E. (1994). Bayesian robustness in bidimensional models: prior independence (with discussion). *J. Statist. Planning and Inference* **40** 161-176.

BERGER, J.O. and O'HAGAN, A. (1988). Ranges of posterior probabilities for unimodal priors with specified quantiles. *Bayesian Statistics 3* (J.M. Bernardo, M.H.Degroot, D.V. Lindley and A.F.M. Smith, eds.) Oxford: University Press 45-66.

BOSE, S. (1994). Bayesian robustness with more than one class of contaminations (with discussion). *J. Statist. Planning and Inference* **40** 177-187.

CUEVAS, A. and SANZ, P. (1988). On differentiability properties of Bayes Operators. *Bayesian Statistics 3* (J.M. Bernardo, M.H.Degroot, D.V. Lindley and A.F.M. Smith, eds.) Oxford: University Press 569-577.

DELAMPADY, M. and DEY, D. (1991). Bayesian robustness for multiparameter problems. *J. Statist. Planning and Inference* **40** 375-382.

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1-67.

GELFAND, A. and DEY, D. (1991). On Bayesian robustness of contaminated classes of priors. *Statistics and Decisions* **9** 63-80.

GUSTAFSON, P. (1994). Local Sensitivity of Posterior Expectations. *Ph.D. Dissertation.* Department of Statistics, Carnegie Mellon University.

LISEO, B., MORENO, E. and SALINETTI, G. (1996). Bayesian Robustness of bidimensional priors with given marginals. (to appear in this volume).

LAVINE, M., WASSERMAN, L. and WOLPERT, R.L. (1991). Bayesian inference with specified prior marginals. *Journal of the American Statistical Association* **86** 964-971.

MORENO, E. and CANO, J.A. (1991). Robust Bayesian analysis with ε-contaminations partially known. *J. Roy. Statist. Soc. B* **53** 143-155.

MORENO, E. and CANO, J.A. (1995). Classes of bidimensional priors specified on a collection of sets: Bayesian Robustness. *J. Statist. Planning and Inference* **46** 325-334.

MORENO, E. and PERICCHI, L.R. (1993). On ε- contaminated priors with quantile and piece-wise unimodality constraints. *Comm. in Statist. Theory and Methods* **22** 7 1963-1978.

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

RUGGERI, F. and WASSERMAN, L. (1993). Infinitesimal sensitivity of posterior distributions. *Canadian J. Statist.* **21** 195-203.

SIVAGANESAN, S. (1988). Range of posterior measures for priors with arbitrary contaminations. *Comm. Statist.* **17** 1591-1612.

SIVAGANESAN, S. (1989). Sensitivity of posterior mean to unimodality preserving contaminations. *Statistics and Decisions* **7** 77-93.

SIVAGANESAN, S. (1993). Range of the posterior probability of interval for priors with unimodality preserving contaminations. *Ann. Inst. of Statist. Math.* **45** 187-199.

SIVAGANESAN, S. and BERGER, J. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17** 868-889.

E. MORENO and C. MARTÍNEZ
DPTO. DE ESTADÍSTICA E I.O.
UNIVERSIDAD DE GRANADA
GRANADA 18071
SPAIN

J. A. CANO
DPTO. DE MAT. APLICADA Y EST.
UNIVERSIDAD DE MURCIA
MURCIA 30100
SPAIN

Local Robustness And Influence For Contamination Classes Of Prior Distributions

discussion by
LARRY WASSERMAN
Carnegie Mellon University

I would like to begin by congratulating the authors on an interesting paper. Their main point is that uncertainty about a prior need not be constant across the parameter space. We might, for example, be less confident about the tails than the center of the distribution. The authors recommend that we generalize the ϵ-contamination class to explicitly account for this.

To begin, let us briefly review the class of priors under consideration. Let $\mathcal{C} = \{C_1, \ldots, C_n\}$ be a fixed partition of the parameter space, let $\epsilon_1, \ldots, \epsilon_n$ be such that $0 \leq \epsilon_i \leq 1$, $i = 1, \ldots, n$ and let $\epsilon(\theta) = \sum_i \epsilon_i I_{C_i}(\theta)$. The authors define

$$(1) \qquad \Gamma = \{\pi(\theta); \pi(\theta) = (1 - \epsilon(\theta))\pi_0(\theta) + \epsilon(\theta)q(\theta), q \in \mathcal{Q}\}$$

where $\mathcal{Q} = \{q; \int_C q(\theta)d\theta = \int_C \pi_0(\theta)d\theta, C \in \mathcal{C}\}$.

The idea is to set ϵ_i large if we have great uncertainty about the form of the prior over C_i. Table 1 of their paper verifies that if our uncertainty is only in the tails, then the posterior bounds will be narrower than a standard ϵ-contamination class. Table 2 of Section 3 contains the surprising result that the center of the prior, not the tails, produces the greatest effect on the posterior. This may be due to the fact that the table stops at $X = 1.5$. When n is large, $\overline{X}_n$ could easily be far in the tail of the prior. In this case, Table 2 might look quite different. This leads me to my first question: Why is R_2 (the sensitivity due to perturbations to the center) greater than R_1 and R_3 (the sensitivity to the tails)? Specifically, what are the conditions that imply greater sensitivity in the tails?

At this point I would like to mention an alternative method for quantifying local uncertainty which was developed in Wasserman (1990, section 6). There I define a "local perturbation" to a prior π using the notion of a random set. Associate with each θ, a set N_θ such that $\theta \in N_\theta$. Then Γ is defined to be the set of priors formed by moving mass from θ to any point in N_θ. Loosely, we replace the random point θ with the random set N_θ. More formally, let $\mathcal{F}$ be the set of measurable mappings $f : \Omega \to \Omega$ such that $f(\theta) \in N_\theta$ and let $\Gamma = \{\pi f^{-1}; f \in \mathcal{F}\}$. Then $\sup_{P \in \Gamma} \int h dP = \int \overline{h}(\theta)\pi(d\theta)$ where $\overline{h}(\theta) = \sup_{u \in N_\theta} h(u)$. Bounds on posterior expectations can be found using linearization. The bound on the posterior probability of a set A is given by

$$\overline{P}(A|x) = \frac{\int \sup_{u \in N_\theta \cap A} L(u)\pi(d\theta)}{\int \sup_{u \in N_\theta \cap A} L(u)\pi(d\theta) + \int \inf_{u \in N_\theta \cap A^c} L(u)\pi(d\theta).}$$

One might set $N_\theta = [\theta - \epsilon(\theta), \theta + \epsilon(\theta)]$, say. This approach seems to offer slightly more flexibility than the authors' method though it does not preserve fixed quantiles. It would be interesting to compare this approach with the authors'.

In section 3, the authors also consider the Frechet derivative of the posterior with respect to the prior, as various pieces of the prior are perturbed. The authors observe qualitatively similar behavior here as with the global analysis.

I like the idea of using these derivatives. Unfortunately, these derivatives have some problems. As noted in Gustafson and Wasserman (1995), the norm of the derivative, $||\dot{T}||$ often has poor asymptotic behavior. The norm of the Frechet derivative of the posterior with respect to the prior is

$$||\dot{T}||_n = \frac{L(\hat{\theta})}{\int L(\theta)\pi(d\theta)}$$

where L is the likelihood and $\hat{\theta}$ is the maximum likelihood estimator. Typically, we find that $||\dot{T}||_n = O(n^{1/2})$ assuming that the parameter is scalar. This diagnostic is clearly inappropriate since it implies that the posterior has increasing sensitivity in n. If we compute the derivative of the posterior expectation of a given function, rather than the whole posterior, we find that $||\dot{T}||_n$ is bounded but still does not go to zero. I expect that the authors' will find that the norm does not tend to zero at least for the partition element containing the true value. Note that restrictions like quantile restrictions are generally not strong enough to correct the asymptotic behavior. Unless the diagnostic is $o(1)$ its value is questionable; see Gustafson (1994). This suggests two things. First, the derivative is an inappropriate measure of sensitivity (unless some modification is made). Second, the agreement between Table 2 and Table 3 may break down for large n. Have the authors investigated examples where n is large?

Finally, the authors suggest that their diagnostics could be useful for interactive elicitation. This sounds like a fine idea and I look forward to further developments in this area. This leads me to my final question. To ask this question, I must first step back and ask: what is the purpose of robust Bayesian analysis? In my opinion, the main purpose of robust Bayesian inference is to simplify the process of specifying assumptions. To clarify this point, consider two different types of Bayesian analyses:

Type (I) analysis:

(Step 1): Carefully and fully elicit a prior π_0.

(Step 2): Calculate the posterior. Stop.

Type (II) analysis:

(Step 1): Carelessly and quickly construct a prior π_0.

(Step 2): Calculate the posterior.

(Step 3): Carry out a sensitivity analysis. If answers are not robust, go back to Step 1. Otherwise, Stop.

Now a Type II analysis appears to involve more work. But as long as sensitivity analysis is simple, then the work involved in a Type II analysis can be substantially less than the work involved in a Type I analysis even though the former has more steps. But if the work involved in Step 3 is substantial, then little, if anything, is gained by taking the second route. My concern is that the authors' methods require much input, making Step 3 quite difficult. Their method requires that we specify $\{\pi_0, \mathcal{C}, \mathcal{P}, \epsilon\}$ where $\mathcal{C} = \{C_1, \ldots, C_n\}$, $\mathcal{P} = \{P(C_1), \ldots, P(C_n)\}$ and $\epsilon = \{\epsilon_1, \ldots, \epsilon_n\}$. It seems like a lot of work to specify all this information. I am concerned that this might not be feasible in hard problems. I'd be interested in the authors' comments on this point.

In summary, I find the author's methods to be quite interesting and I enjoyed their paper. The idea that uncertainty about the prior should vary across the parameter space is important. The authors deserve much credit for taking an important first step in this direction.

REFERENCES

Gustafson, P. (1994). "Local Sensitivity of Posterior Expectations." Ph.D. Dissertation. Department of Statistics, Carnegie Mellon University.

Gustafson, P. and Wasserman, L. (1995). "Local Sensitivity Diagnostics for Bayesian Inference," to appear: *The Annals of Statistics.*

Wasserman, L. (1990). "Prior Envelopes Based on Belief Functions," *The Annals of Statistics,* 18, 454-464.

REJOINDER

Elias Moreno, Carmen Martinez and Juan A. Cano

We thank Professor Wasserman for his interesting and stimulating comments.

His first comment is on the results in Tables 2 and 3. Values of $R_i(x)$ and $R(x)$ for values of x greater than 1.5 are given in Table 6.

TABLE 6
Values of $R_i(x)$ and $R(x)$

x	$R_1(x)$		$R_2(x)$		$R_3(x)$		$R(x)$
2	2.0	10^{-3}	52.0	10^{-3}	9.7	10^{-3}	0.069
3	1.5	10−4	92.0	10^{-4}	28.0	10^{04}	0.013
4	0.1	10−4	8.5	10^{-4}	4.1	10^{-4}	0.001
6	0.01	10−7	11.0	10^{-7}	5.9	10^{-7}	1.8 10^{-6}
10	1.0	10−20	5.8	10^{-16}	2.0	10^{-16}	8.8 10^{-16}

From table 6 follows that as the data goes far from zero, the influence of the prior becomes smaller and robustness is achieved. However, the relative influence of C_2 still remains bigger than that of C_3 (or C_1). It should be noted that the prior mass on C_2 is twice that of C_3 (or C_1).

The approach with belief functions in Wasserman (1990, section 6) has indeed the same aim than than the approach taken here. However, from the beginning our analysis was derived to maintain the prior beliefs stated in terms of quantile constraints.

We share his perceptive comments on Frechet derivates. In fact the quantile class of prior distributions behaves asymptotically poorly too.

In Moreno and Pericchi (1993) a normal sampling model was assumed and the asymptotic behavior of the posterior probabilities $P^{\pi}(I_\gamma \mid x_1, \ldots, x_n)$, where $I_\gamma = \{\theta : \bar{x} - \frac{z_\gamma}{n} \leq \theta \leq \bar{x} + \frac{z_\gamma}{n}\}$ and $z_\gamma = \Phi^{-1}(\frac{1+\gamma}{2})$, as π ranges over the quantile class

$$\Gamma_C = \{\pi(\theta) : \pi(\theta) = (1-\epsilon)\pi_0(\theta) + \epsilon q(\theta),\ q \in \mathcal{Q}\},$$

was analyzed. There it was found that

$$\lim_{n \longrightarrow \infty} \inf_{\pi \in \Gamma_C} P^{\pi}(I_\gamma \mid x_1, \ldots, x_n) = 0,$$
$$\lim_{n \longrightarrow \infty} \sup_{\pi \in \Gamma_C} P^{\pi}(I_\gamma \mid x_1, \ldots, x_n) = 1,$$

and that the correction term δ_n needed to ensure that

$$\inf_{\pi \in \Gamma_C} P^{\pi}(I'_\gamma \mid x_1, \ldots, x_n) = \gamma,$$

where $I'_\gamma = (\bar{x} - \frac{z_\gamma + \delta_n}{n}, \bar{x} + \frac{z_\gamma + \delta_n}{n})$, is $\delta_n = O(\sqrt{\log n})$. This disturbing behavior from Γ_C can be corrected if, in adition, unimodality is imposed on the contamination class.

The conclusion is that to process so weak information as that stated by quantile constraints has a price to pay, say this non-reasonable asymptotic behavior.

This, however, does not mean that Γ_C is not useful. It contains accesible prior information which might be enough for the inference. Otherwise, by adding shape constraints the asymptotic hehavior is corrected.

On the other hand, when comparison behavior of sets C_i are considered the problem is less important.

We specially like the final question as long as it affects the foundation of Robust Bayesian Analysis. Apparently we have at hand two types of Bayesian Analysis, denoted by Professor Wasserman as Type (I) and Type (II). We would like to be able to deal with Type (I) analysis. However, we have to say that even when it is the most rational way to deal with in statistics, step (1) is so difficult that gives a clear justifaction to Type (II) analysis. The same appplies to the sampling model selection. Thus, in our view, robustness analysis, in its wider possible sense, is absolutely necessary.

Therefore, in the context of contamination we are left in the case of the classical ϵ-contamination class (ϵ being a constant, say 0.2) or as here in the case of inputs $(C_1, \ldots, C_n; \pi_0)$ and $\epsilon_1, \ldots, \epsilon_n$.

Now take Type (II) analysis as follows. Start with a few sets C's. For those in the center of the distribution of π_0 take $\epsilon = 0.2$ as usual. For those in the tails increase this value. If robustness is not present, analyze the influence of the sets C's. Choose the more influential one and split it. Carry out a robustness analysis. Continue until robustness is achieved.

All of us agree that prior elicitation is, in general, a hard task. What we are proposing here is to elicit step by step. In the contamination context considered here this decomposition results in an understandable way with a very little increment in complexity.

REFERENCES

MORENO, E. and PERICCHI, L.R. (1993). On ϵ-contaminated priors with quantile and piece-wise unimodality constraints. *Comm. in Statist. Theory and Methods*, **22**, 7, 1963-1978.