JOURNAL OF THE ITALIAN STATISTICAL SOCIETY

VOLUME 4, NO. 3, 1995

SIS SOCIETÀ ITALIANA DI STATISTICA

GIARDINI EDITORI · PISA

J. Ital. Statist. Soc. (1995)
3, pp. 283-297

CONCENTRATION FUNCTION AND SENSITIVITY TO THE PRIOR

Sandra Fortini

Istituto di Metodi Quantitativi, Università «Luigi Bocconi»

Fabrizio Ruggeri

Istituto per le Applicazioni della Matematica e dell'Informatica, Consiglio Nazionale delle Ricerche

Summary

In robust Bayesian analysis, ranges of quantities of interest (e.g. posterior means) are usually considered when the prior probability measure varies in a class Γ. Such quantities describe the variation of just one aspect of the posterior measure. The concentration function describes changes in the posterior probability measure more globally, detecting differences in probability concentration and providing, simultaneously, bounds on the posterior probability of all measurable subsets. In this paper, we present a novel use of the concentration function, and two concentration indices, to study such posterior changes for a general class Γ, restricting then our attention to some ε-contamination classes of priors.

Keywords: Concentration function, Bayesian robustness, ε-contaminations, Gini's area of concentration, Pietra's index.

1. Introduction

As described in Berger (1985, 1990, 1994), the robust Bayesian approach usually deals with the uncertainty in specifying the prior probability measure. Sometimes, a class Γ of probability measures seems to be the most plausible result of an elicitation process (e.g. it might consists of all the priors elicited by some experts).

Posterior ranges of functions of the parameters (e.g. set probabilities, means) are usually considered, as the prior measure varies in Γ. Small ranges suggest that the inference (or decision) is not actually affected by a particular choice in Γ.

The paper complements the work in Fortini and Ruggeri (1994) where a new approach to Bayesian robustness, based on the concentration function, was proposed and in which the concentration function was mainly used to define classes of priors and compute bounds on posterior quantities over such classes. This paper studies a different robustness problem, that of comparing the posterior probability measures Π^* themselves, by means of their concentration function with respect to a base posterior Π_0^*. The concentration function, as defined by Cifarelli and Regazzini (1987), extends the notion of Lorenz curve and allows finding the range of the probabilities, under Π^*, of all the sets with equal probability x under Π_0^*. Robust analyses can be performed considering the largest range, for any x in [0, 1], as the prior varies in Γ. The width of such intervals, expressed by means of the concentration function, gives a distance between the posteriors Π^* and Π_0^*. Moreover, the concentration function satisfies the need, as demanded by Wasserman (1992), of providing graphical summaries of robust Bayesian analyses. This paper presents a general framework which holds for many classes of priors, applying it to some ε-contamination classes, very relevant in Bayesian robustness literature. Finally, we apply the results to some well-known examples and to an ongoing study about accidents for a Spanish insurance company.

2. Concentration function

Cifarelli and Regazzini (1987) defined the concentration function, as a generalization of the well-known Lorenz curve (see, e.g., Marshall and Olkin, 1979, p. 5). The classical definition of concentration refers to the discrepancy between a probability measure Π and a uniform one, say Π_0, and allows for the comparison of both probability measures, looking for subsets where Π is much more concentrated than Π_0 (and vice versa). Cifarelli and Regazzini (1987) defined and studied the concentration function of Π with respect to Π_0, where Π and Π_0 are two probability measures on the same measurable space $(\Theta, \mathcal{F})$. According to Radon-Nikodym theorem, there is a unique partition $\{N, N^C\} \subset \mathcal{F}$ of Θ and a nonnegative function h on N^C such that, $\forall E \in \mathcal{F}$, $\Pi(E) = \Pi_a(E \cap N^C) + \Pi_s(E \cap N)$ (with $\Pi_a(E \cap N^C) = \int_{E \cap N^C} h(\vartheta)\,\Pi_0(d\vartheta))$, $\Pi_0(N) = 0$, $\Pi_s(N) = \Pi_s(\Theta)$, where Π_a and Π_s denote the absolutely continuous and the singular part of Π with respect to Π_0, respectively. Set $h(\vartheta) = \infty$ all over N and define $H(y) = \Pi_0(\{\vartheta \in \Theta : h(\vartheta) \leq y\})$, $c_x = inf\{y \in \mathcal{R} : H(y) \geq x\}$ and $c_x^- = \lim_{t \to x^-} c_t$. Finally, let $L_x = \{\vartheta \in \Theta : h(\vartheta) \leq c_x\}$ and $L_x^- = \{\vartheta \in \Theta : h(\vartheta) < c_x\}$.

Definition 1. The function $\varphi : [0, 1] \to [0, 1]$ is said to be the concentration function of Π with respect to Π_0 if $\varphi(x) = \Pi(L_x^-) + c_x\{x - H(c_x^-)\}$ for $x \in (0, 1)$, $\varphi(0) = 0$ and $\varphi(1) = \Pi_a(\Theta)$.

Observe that $\varphi(x)$ is a nondecreasing, continuous and convex function, such that $\varphi(x) \equiv 0 \Leftrightarrow \Pi \perp \Pi_0$, $\varphi(x) = x$, $\forall x \in [0, 1] \Leftrightarrow \Pi = \Pi_0$, and

$$\varphi(x) = \int_0^{c_x} [x - H(t)]dt = \int_0^x c_t dt. \quad (1)$$

To facilitate the understanding of the behaviour of the concentration function, we describe two cases: $\varphi(1) = 1$ means that Π is absolutely continuous with respect to Π_0 while $\varphi(x) = 0$, $0 \leq x \leq \alpha$, means that Π gives no mass to a subset $A \in \mathcal{F}$ such that $\Pi_0(A) = \alpha$. To show how to practically draw the concentration function when Π_0 and Π are absolutely continuous with respect to Lebesgue measure on $\mathcal{R}^+$, consider a gamma distribution $\Pi \sim \mathcal{G}(2, 2)$ and an exponential one $\Pi_0 \sim \mathcal{E}(1)$. Then it follows that the Radon-Nikodym derivative is $h(\vartheta) = 4\vartheta^2 exp(-\vartheta)$. The concentration function $\varphi(x)$ is obtained by evaluating $x = \Pi_0(L_q)$ and $\varphi(x) = \Pi(L_q)$, where $L_q = \{\vartheta \in \Theta : h(\vartheta) \leq q\}$ and q takes as many as possible values to draw the concentration function within the required accuracy.

The concentration function induces a partial order in the space $\mathcal{P}$ of all probability measure, hence allowing for their comparison.

Definition 2. Let φ_1, φ_2 be the concentration functions of Π_1 and Π_2 with respect to Π_0. We say that Π_2 is not less concentrated than Π_1 with respect to Π_0, and denote it by $\Pi_2 \succeq \Pi_1$, if $\varphi_2(x) \leq \varphi_1(x)$, $\forall x \in [0, 1]$.

Total orderings, consistent with the previous partial one, are achieved when considering coefficients of divergence (see Regazzini, 1992). Here we consider two particular case: $C_{\Pi_0}(\Pi) = 2\int_0^1 \{x - \varphi(x)\}dx$ and $G_{\Pi_0}(\Pi) = \sup_{x \in [0,1]} (x - \varphi(x))$ which are, respectively, the Gini's area of concentration (Gini, 1914) and an index proposed by Pietra (1915), which equals the total variation norm, as proved by Cifarelli and Regazzini (1987). Observe that

$$C_{\Pi_0}(\Pi) = 0 \Leftrightarrow G_{\Pi_0}(\Pi) = 0 \Leftrightarrow \Pi = \Pi_0;$$

$$C_{\Pi_0}(\Pi) = 1 \Leftrightarrow G_{\Pi_0}(\Pi) = 1 \Leftrightarrow \Pi \perp \Pi_0.$$

The following Theorem, proved in Cifarelli and Regazzini (1987), states that $\varphi(x)$ substantially coincides with the minimum value of Π on the measurable subsets of Θ with Π_0-measure not smaller than x.

Theorem 1. If $A \in \mathcal{F}$, $\Pi_0(A) = x$, then $\varphi(x) \leq \Pi_a(A)$. Moreover if $x \in [0, 1]$ is adherent to the range of H, then there exists a B_x such that $\Pi_0(B_x) = x$ and

$$\varphi(x) = \Pi_a(B_x) = min\{\Pi(A) : A \in \mathcal{F} \text{ and } \Pi_0(A) \geq x\}. \quad (2)$$

If Π_0 is nonatomic, then (2) holds for any $x \in [0, 1]$.

Such a Theorem is relevant when applying the concentration function to robust Bayesian analysis; for *any* $x \in [0, 1]$, the probability, under Π, of all the subsets A with Π_0-measure x satisfies

$$\varphi(x) \leq \Pi(A) \leq 1 - \varphi(1 - x). \quad (3)$$

Two examples are presented to show how the concentration function, far from substituting the other usual functions (e.g. the mean), furnishes different information about the probability measures. As a first example, we mention the fact that comparisons among probability measures are sometimes made through their moments; it is well-known that different measures could be found such that they share the first n moments, while their difference could be detected by the concentration function. As another example, consider two measures concentrated on disjoint but very close sets in $\mathcal{R}^n$, say E_1 and E_2. In this case, the concentration function shows the difference between them, which is large on the subsets which contain just one E_i, $i = 1, 2$, although their means are very close.

3. Statement of the problem

Following a suggestion in Regazzini (1992), we shall compare a class of posterior probability measures, formed from a class Γ of priors, with a base posterior measure with the aid of the concentration function. Hence, we provide a novel approach to Bayesian robustness studies, based on an innovative use of the concentration function. We will apply our general approach to one of the most relevant classes in robust Bayesian analysis: the ε-contaminations.

Two methods will be presented here. The first one considers the class of concentration functions of the posterior probability measures with respect to the base one and looks either for the lowest concentration function, if it exists, or, pointwise, for the infimum of the concentration functions. The obtained function is then considered, by looking how far apart are it and the line connecting (0, 0) and (1, 1) (corresponding to equal measures) or if it lies above an adequate continuous, convex, nondecreasing function g, which imposes bounds on the variation of the set probabilities, as described in Fortini and Ruggeri (1994, 1995a). Results are provided for well-known contaminating classes (arbitrary, unimodal and generalised moments priors). In the second method, we measure robustness considering the distance between probability measures by means of some indices, related to the concentration function, as described in Regazzini (1992). In particular, we consider Gini's and Pietra's indices.

4. General approach

Let $(\mathcal{X}, \mathcal{F}_\mathcal{X}, \{P_\vartheta, \vartheta \in (\Theta, \mathcal{F})\})$ be a dominated statistical space, where $(\Theta, \mathcal{F})$ is any measurable space. Given a sample s from $\mathcal{X}$, the experimental evidence about ϑ will be expressed by the likelihood function $l(\vartheta)$, which we assume $\mathcal{F}_\mathcal{X} \otimes \mathcal{F}$-measurable. Let $\mathcal{P}$ denote the space of the probability measures on the parameter space $(\Theta, \mathcal{F})$. Given a prior $\Pi \in \mathcal{P}$, the posterior measure is defined by $\Pi^*(A) = \dfrac{\int_A l(\vartheta)\Pi(d\vartheta)}{\int_\Theta l(\vartheta)\Pi(d\vartheta)}$, for any $A \in \mathcal{F}$. Following the robust Bayesian viewpoint, we consider a class Γ of probability measures Π, rather than just one. Suppose that there exists a base prior Π_0, as in the ε-contamination class, and consider the class Ψ of concentration functions φ_Π of Π^*, $\Pi \in \Gamma$, with respect to Π_0^*. Because of Theorem 1 and (3), it follows that, for any $\Pi \in \Gamma$ and $A \in \mathcal{F}$ with $\Pi_0^*(A) = x$,

$$\hat{\varphi}(x) \leq \Pi^*(A) \leq 1 - \hat{\varphi}(1 - x),$$

where $\hat{\varphi}(x) = \inf_{\Pi \in \Gamma} \varphi_\Pi(x)$, for any $x \in [0, 1]$.

The interpretation of $\hat{\varphi}$, in terms of Bayesian robustness, is straightforward: the closest $\hat{\varphi}(x)$ and $1 - \hat{\varphi}(1 - x)$ are for all $x \in [0, 1]$, the closest the posterior measures are. It is then possible to make judgements on robustness by looking how far apart the plots of $\hat{\varphi}(x)$ and $y = x$ (equal measures) are. Fortini and Ruggeri (1995a) suggested checking, for any $A \in \mathcal{F}$, if $\Pi^*(A) \geq g(\Pi_0^*(A))$, where g is a given continuous, convex, nondecreasing function such that $g(0) = 0$. As an example, the choice $g(x) = x^2$ is equivalent to $\sup_{A \in \mathcal{F}: \Pi_0^*(A) = x} |\Pi_0^*(A) - \Pi^*(A)| \leq \Pi_0^*(A)\,(1 - \Pi_0^*(A))$. If $\Pi_0^*(A) = x$, then Theorem 1 implies that $\Pi^*(A) \geq \varphi_\Pi(x)$, so that such criterion ensures robustness if $\hat{\varphi}(x) \geq g(x)$ for all $x \in [0, 1]$.

The computation of $\hat{\varphi}$ is quite simple when there exists $\hat{\Pi} \in \Gamma$ such that $\hat{\varphi} \equiv \varphi_{\hat{\Pi}}$. The paper gives conditions ensuring its existence for some classes of priors; otherwise, $\hat{\varphi}(x)$ can be obtained numerically. Besides, the paper compares probability measures by means of some concentration indices, considering $\bar{C}_{\Pi_0} = \sup_{\Pi \in \Gamma} C_{\Pi_0}(\Pi)$ and $\bar{G}_{\Pi_0} = \sup_{\Pi \in \Gamma} G_{\Pi_0}(\Pi)$.

The concentration function is well suited to satisfy the need, pointed out by Wasserman (1992), of summarising graphically the results of a robust Bayesian analysis. In fact, the plot of $\hat{\varphi}(x)$ and $1 - \hat{\varphi}(1 - x)$ gives a very intuitive description of the discrepancy between the base prior and the other measures. As an example $\varphi(x) = 1 - (1 - x)(1 - \log(1 - x))$ is the concentration function of $\Pi \sim \mathcal{G}(2, 1)$ with respect to $\Pi_0 \sim \mathcal{E}(1)$, showing that [0.094, 0.767] is the range spanned by the probability, under Π, of the sets A with $\Pi_0(A) = 0.4$.

5. Concentration function of ε-contaminations

We now apply the general framework, described in the previous Section, to the class Γ_ε of ε-contaminated priors, one of the most relevant classes in robust Bayesian analysis (see Berger, 1994).

Definition 3. Let Π_0 be a fixed prior probability measure and let $\varepsilon \in [0, 1]$. The class $\Gamma_\varepsilon = \{\Pi_Q = (1 - \varepsilon)\Pi_0 + \varepsilon Q,\ Q \in \mathcal{Q}\}$, where $\mathcal{Q} \subseteq \mathcal{P}$, is said to be an ε-contamination class of priors.

Some classes $\mathcal{Q}$ have been proposed and their properties are discussed in Berger (1990). In this paper, we consider the class $\mathcal{Q}_{\mathcal{A}}$ of all the probability measures, the class $\mathcal{Q}_{\mathcal{M}}$ of all probability measures defined by means of generalised moments conditions (see Betrò *et al.*, 1994) and, if Π_0 is unimodal, the class $\mathcal{Q}_{\mathcal{U}}$ of all unimodal probability measures, with the same mode as Π_0.

We consider some derivatives of $\varphi(x)$, since they are needed in proving some of the next Proposition. Moreover, it is worth observing that such derivatives allow the approximation of lower and upper bounds on the probability of the subsets A such that $\Pi_0^*(A)$ is sufficiently close to 0.

Proposition 1. Let Π_0 and Q be probability measures on $(\Theta, \mathcal{F})$; for any $a \in [0, 1]$, define $\Pi_Q = (1 - a)\Pi_0 + aQ$. If φ and φ_0 are the concentration functions of Π_Q and Q with respect to Π_0 respectively, then $\varphi(x) = (1 - a)x + a\varphi_0(x)$.

Let $\varphi'_+(x)$ and $\varphi'_-(x)$ be, respectively, the right-hand and the left-hand derivatives of $\varphi(x)$, then it follows that

$$\varphi'_+(0) = (1 - a) + a \inf_{\vartheta \in \Theta} \hat{h}_0(\vartheta) \quad \text{and} \quad \varphi'_-(1) = (1 - a) + a \sup_{\vartheta \in \Theta} \hat{h}_0(\vartheta),$$

where $\hat{h}_0$ is the Radon-Nikodym derivative of Q with respect to Π_0.

Proof. The relation between φ and φ_0 follows easily from the definition of concentration function. As pointed out by Cifarelli and Regazzini (1987), the right-hand derivative can be computed, because (1) implies that $(\varphi_0)'_-(1) = c_1$. In the same way, $(\varphi_0)'_+(0)$ is computed, observing that c_t is right continuous at the origin. □

For any measure Q, let $\lambda_Q = (1 - \varepsilon)D_0/[(1 - \varepsilon)D_0 + \varepsilon D_Q]$, with

$$D_0 = \int_\Theta l(\vartheta)\Pi_0\,(d\vartheta) \text{ and } D_Q = \int_\Theta l(\vartheta)Q\,(d\vartheta).$$

Proposition 2. Consider $\Pi_Q = (1 - \varepsilon)\Pi_0 + \varepsilon Q$ so that $\Pi_Q^* = \lambda_Q \Pi_0^* + (1 - \lambda_Q)Q^*$. If φ and φ_0 denote the concentration functions of Π_Q^* and Q^* with respect to Π_0^*, respectively, then it follows that $\varphi(x) = \lambda_Q x + (1 - \lambda_Q)\varphi_0(x)$. Moreover, it follows that

$$\begin{aligned}\varphi'_+(0) &= \lambda_Q + (1 - \lambda_Q) \inf_{\vartheta \in \Theta} h_0(\vartheta) \\ &= \lambda_Q\,[1 + \varepsilon \inf_{\vartheta \in \Theta} \hat{h}_0(\vartheta)/(1 - \varepsilon)],\\ \varphi'_+(1) &= \lambda_Q + (1 - \lambda_Q) \sup_{\vartheta \in \Theta} h_0(\vartheta) \\ &= \lambda_Q\,[1 + \varepsilon \sup_{\vartheta \in \Theta} \hat{h}_0(\vartheta)/(1 - \varepsilon)],\end{aligned}$$

where $\hat{h}_0$ and h_0 are, respectively, the Radon-Nikodym derivatives of Q with respect to Π_0 and of Q^* with respect to Π_0^*.

Proof. The expression about Π^* is well known (see Berger, 1985, p. 206). The other results follow from Proposition 1 and the fact that $\hat{h}_0(\vartheta) = \left(\frac{D_Q}{D_0}\right) \times h_0(\vartheta)$. □

The previous Propositions are now applied to robustness analysis, when considering ε-contaminations of a nonatomic prior Π_0. It is worth mentioning that the results hold, with slight changes, for $\varepsilon = 1$, so that they essentially apply also to arbitrary, unimodal and generalised moments priors.

6. Arbitrary contaminations

In the case of arbitrary contaminations, there exists a contaminated measure which is not less concentrated than the others, i.e. with lowest concentration function.

Proposition 3. Consider the class $\mathcal{Q}_{\mathcal{A}}$ of all contaminations. Let $\hat{\vartheta}$ be the maximum likelihood estimate of ϑ and $\hat{Q}$ be the Dirac measure concentrated at $\hat{\vartheta}$. It follows that $\Pi^*_{\hat{Q}} \succeq \Pi^*_Q$ for any $Q \in \mathcal{Q}_{\mathcal{A}}$. Moreover $\bar{C}_{\Pi_0} = \bar{G}_{\Pi_0} = \{1 + [(1 - \varepsilon)/\varepsilon] D_0/l(\hat{\vartheta})\}^{-1}$.

Proof. Proposition 2 and (1) imply that $\varphi_{\hat{Q}}(x) = \lambda_{\hat{Q}} x$. For any $Q \in \mathcal{Q}_{\mathcal{A}}$, it follows that $\varphi_{\hat{Q}}(x) \leq \lambda_Q x \leq \varphi_Q(x)$, since $l(\hat{\vartheta}) \geq D_Q$. The result about Gini's and Pietra's indices follows immediately. □

The previous result can be extended to all classes in which there exists a Dirac contamination $Q_{\hat{\vartheta}}$ maximising D_Q.

7. Generalised moments contaminations

We now consider the class $\mathcal{Q}_{\mathcal{M}}$ defined by generalised moments conditions (see Betrò *et al.*, 1994), which contains, as particular cases, the classes defined by quantiles of either the prior probability measure on ϑ (see Berger, 1990) or the (prior) marginal on a sample from $\mathcal{X}$. Let us define

$$\mathcal{Q}_{\mathcal{M}} = \left\{Q : \int_\Theta H_i(\vartheta)\, Q(d\vartheta) \leq \alpha_i,\ i = 1, \dots, n\right\}$$

where H_i are given Π_Q-integrable functions and α_i, $i = 1, \dots, n$, are fixed real numbers. Suppose that $\mathcal{Q}_{\mathcal{M}} \neq \emptyset$. As in the case of arbitrary contaminations, there exists a contaminated measure which is not less concentrated than the others, i.e. with the lowest concentration function. Such measure maximises D_Q, a linear functional in Q, rather than the more complicate ratio-linear posterior quantities considered in Betrò *et al.* (1994).

Proposition 4. Consider the class $\mathcal{Q}_{\mathcal{M}}$. Let $\hat{Q}$ be the discrete measure (concentrated in at most $n + 1$ points) which maximises D_Q. It follows that $\Pi^*_{\hat{Q}} \succeq \Pi^*_Q$ for any $Q \in \mathcal{Q}_{\mathcal{M}}$. Moreover $\bar{C}_{\Pi_0} = \bar{G}_{\Pi_0} = \{1 + [(1 - \varepsilon/\varepsilon] D_0/D_{\hat{Q}}\}^{-1}$.

Proof. Proposition 2 and (1) imply that $\varphi_{\hat{Q}}(x) = \lambda_{\hat{Q}} x$. From Kemperman (1968), it follows that D_Q is maximised by a discrete measure, concentrated in at most $n + 1$ points, so that, $\varphi_{\hat{Q}}(x) \leq \lambda_{\hat{Q}} x \leq \varphi_Q(x)$, for any $Q \in \mathcal{Q}_{\mathcal{M}}$. The result about Gini's and Pietra's indices follows immediately. □

We can compute $D_{\hat{Q}}$ very easily in the quantile class, defined by the probabilities $Q(A_i) = p_i$, $i = 1, \dots, n$, of a partition $\{A_i\}$ of Θ. In such case, it follows that $D_{\hat{Q}} = \sum_{i=1}^{n} p_i \sup_{\vartheta_i \in A_i} l(\vartheta_i)$, which can be easily computed if, e.g., $l(\vartheta)$ is unimodal and the subsets A_i are intervals.

8. Unimodal contaminations

Suppose that Π_0 is unimodal with mode ϑ_0. Consider the class $\mathcal{Q}_{\mathcal{U}}$ of all unimodal probability measures with mode ϑ_0. Such a class contains just one discrete measure, the Dirac measure $\hat{Q}$ concentrated at ϑ_0, the only one which could lead to the lowest concentration function, as shown in the next Proposition.

Proposition 5. If

$$\sup_{Q \in \mathcal{Q}_{\mathcal{U}}} D_Q \leq l(\vartheta_0), \tag{4}$$

then it follows that $\Pi^*_{\hat{Q}} \succeq \Pi^*_Q$ for any $Q \in \mathcal{Q}_{\mathcal{U}}$.

No contaminated measure, different from $\hat{Q}$, leads to the lowest concentration function when

$$\inf_{\vartheta \in \Theta} h_Q(\vartheta) > 0 \text{ and } \inf_{\vartheta \in \Theta} q(\vartheta)/\pi_0(\vartheta) > (D_Q - l(\vartheta_0))/[D_0 + \varepsilon l(\vartheta_0)/(1 - \varepsilon)] \tag{5}$$

for every $Q \in \mathcal{Q}_{\mathcal{U}}$ such that $D_Q > l(\vartheta_0)$.

If conditions (4) and (5) are not satisfied, then there exists no contaminated measure whose concentration function is below all the others.

If $\Pi^*_{\hat{Q}} \succeq \Pi^*_Q$ for any $Q \in \mathcal{Q}_{\mathcal{U}}$, it follows that $\bar{C}_{\Pi_0} = \bar{G}_{\Pi_0} = \{1 + [(1 - \varepsilon)/\varepsilon] D_0/l(\vartheta_0)\}^{-1}$.

Proof. The result about condition (4) can be proved as in Proposition 3.
From now on, let us suppose that there exists at least one $Q \in \mathfrak{Q}_{\mathcal{U}}$ such that $D_Q > l(\vartheta_0)$. We now prove that there is no $\bar{Q}$ such that $\varphi_{\bar{Q}}(x) \leq \varphi_Q(x)$ for all $x \in [0, 1]$ and $Q \in \mathfrak{Q}_{\mathcal{U}}$, when $\inf_{\vartheta \in \Theta} h_Q(\vartheta) = 0$ for any $Q \in \mathfrak{Q}_{\mathcal{U}}$. Since $\hat{Q}$ is a discrete measure, it follows that $\varphi_{\hat{Q}}(1) = (\Pi^*_{\hat{Q}})_a(\Theta) < (\Pi^*_Q)_a(\Theta) = \varphi_Q(1) = 1$ for all $Q \neq \hat{Q}$. Let $\tilde{Q} \in \mathfrak{Q}_{\mathcal{U}}$ such that $D_{\tilde{Q}} > l(\vartheta_0)$; then Proposition 2 implies that $(\varphi_{\tilde{Q}})'_+(0)$. Therefore, since the concentration function is continuous, there exists x^* such that $\varphi_{\tilde{Q}}(x) < \varphi_{\hat{Q}}(x)$, for any $x \in (0, x^*)$.
Let us suppose that there exists at least one $Q \in \mathfrak{Q}_{\mathcal{U}}$ such that $\inf_{\vartheta \in \Theta} h_Q(\vartheta) > 0$. From Proposition 2, it follows that $(\varphi_Q)'_+(0) > (\varphi_{Q_{\hat{Q}}})'_+(0)$ if and only if (5) holds. Therefore, $\Pi^*_{\hat{Q}} \succeq \Pi^*_Q$ for any $Q \in \mathfrak{Q}_{\mathcal{U}}$ when (5) holds. If (5) does not hold, there exists no contaminated measure whose concentration function is below all the others. The result about Gini's and Pietra's indices follows immediately. □

The previous result shows that just one contaminations, $\hat{Q}$, might lead to the lowest concentration function. Therefore, numerical computation is needed to compute $\hat{\varphi}$, unless (4) holds.

9. Examples

We use now concentration functions in analysing the robustness in some examples, showing the applicability of our proposal, even when considering classes different from the ε-contaminations.

Example 1. (Berger, 1985, p. 212). Assume that $P_\vartheta \sim \mathcal{N}(\vartheta, \sigma^2)$, σ^2 known, and $\Pi_0 \sim \mathcal{N}(\vartheta_0, \sigma_0^2)$, ϑ_0 and σ_0^2 known, with density function $\pi_0(\vartheta)$. Let $\mathfrak{Q}$ be the class of the probability measures which are either $Q_k \sim \mathcal{U}(\vartheta_0 - k, \vartheta_0 + k)$, $k > 0$, or Q_∞, which assigns probability one to the point ϑ_0 (note that Berger, 1985, does not consider Q_∞, but we can add it with no changes in his results). Define $\Pi_k = (1 - \varepsilon)\Pi_0 + \varepsilon Q_k$. Given a sample s from $\mathcal{X}$, the likelihood function is $l(\vartheta) = [1/(\sqrt{2\pi}\,\sigma)]\, exp(-(\vartheta - s)^2/(2\sigma^2))$. The density of Q_k, $k > 0$, is given by $q_k(\vartheta) = (1/2k) I_{[\vartheta_0 - k, \vartheta_0 + k]}(\vartheta)$ where I_A is the indicator function of the set A. Note that Q_k converges in distribution to Q_∞ as $k \to 0$.
It can be seen that $D_0 = (1/\sqrt{2\pi(\sigma^2 + \sigma_0^2)})\, exp(-(\vartheta_0 - s)^2/(2(\sigma^2 + \sigma_0^2)))$ and $D_{Q_k} = (1/2k) \int_{\vartheta_0 - k}^{\vartheta_0 + k} l(\vartheta) d\vartheta$.

Result 1. Given a sample s, then $D_{Q_k} \leq l(\vartheta_0)$ for any $k > 0$ if and only if $|s - \vartheta_0|/\sigma \leq 1$.

Proof. It holds that $\lim_{k \to 0^+} D_{Q_k} = l(\vartheta_0)$ and $\lim_{k \to +\infty} D_{Q_k} = 0^+$. Set $h = k/\sigma$ and $v = |s - \vartheta_0|/\sigma$. Then, there exists the derivative of D_{Q_k} with respect to h: $D'_{Q_k}(h) = [1/(2\sqrt{2\pi}\, \sigma h^2)]\, f(h)$, with

$$f(h) = -(1/h) \int_{v-h}^{v+h} exp(-t^2/2)dt + exp((v - h)^2/2) + exp(-(v + h)^2/2)$$

and, consequently, $\lim_{h \to 0^+} f(h) = 0$ and $f'(h) = h[(v - h)\, exp(-(v - h)^2/2) - (v + h)\, exp(-(v + h)^2/2)]$. For any $h > 0$, $f'(h) < 0$ if and only if $g(h) = (v - h)\, exp(vh) - (v + h)\, exp(-vh) < 0$.
Such a condition is obviously satisfied if $h \geq v$, so that $D_{Q_k} \leq l(\vartheta_0)$ for any $k \geq |s - \vartheta_0|$. Otherwise, suppose that $v \leq 1$, then

$$g(h) = 2\left\{\sum_0^\infty \frac{v^{2j+2}h^{2j+1}}{(2j+1)!} - \sum_0^\infty \frac{v^{2j}h^{2j+1}}{(2j)!}\right\} \leq$$

$$\leq 2\left\{\sum_0^\infty \frac{v^{2j}h^{2j+1}}{(2j+1)!} - \sum_0^\infty \frac{v^{2j}h^{2j+1}}{(2j)!}\right\} \leq 0.$$

The convexity of D_{Q_k} implies that $D_{Q_k} \leq l(\vartheta_0)$ for any $k > 0$ if $v \leq 1$. Suppose now that $v > 1$, then $\lim_{h \to 0^+} D'_{Q_k}(h) = \lim_{h \to 0^+} f'(h)/4h = exp(-v^2)/2 \cdot \lim_{h \to 0^+} h(v^2 - 1) = 0^+$ so that $D_{Q_k} > l(\vartheta_0)$, for k close enough to 0. □

Result 2. Given a sample s, then the concentration function φ_∞ of Π^*_∞ with respect to Π^*_0 is such that $\varphi_\infty(x) \leq \varphi_k(x)$, for any $x \in [0, 1]$ and for any concentration function φ_k of Π^*_k with respect to Π^*_0, if and only if $|s - \vartheta_0|/\sigma \leq 1$. Given such a sample s, it follows that

$$\bar{C}_{\Pi_0} = \bar{G}_{\Pi_0} = \left\{1 + \left[\frac{(1-\varepsilon)}{\varepsilon}\right] \sqrt{\frac{\sigma^2}{(\sigma^2 + \sigma_0^2)}}\; exp\left(\frac{\sigma_0^2 (s - \vartheta_0)^2}{[2\sigma^2(\sigma^2 + \sigma_0^2)]}\right)\right\}^{-1}.$$

Otherwise, there is no $Q_{\mathcal{E}}(x)$ such that $\varphi_{\mathcal{E}}(x) \leq \varphi_k(x)$, for any $x \in [0, 1]$ and φ_k.

Proof. From Result 1 and Proposition 5, it follows that φ_∞ lies under any other concentration function φ_k if $|s - \vartheta_0|/\sigma \leq 1$. Otherwise, there are no Π_k,

$k > 0$, less concentrated that the other measures, since $\inf_\vartheta h_{0k}(\vartheta) = 0$, for any k, where $h_{0k}(\vartheta) = [q_k(\vartheta|s)/\pi_0(\vartheta|s)] = [q_k(\vartheta)/\pi_0(\vartheta)] \cdot (D_0/D_{Q_k})$. The value of the indices follows from Proposition 5. □

As in Berger (1985), take $\vartheta_0 = 0$, $\sigma^2 = 1$, $\sigma_0^2 = 2$, $\varepsilon = 0.1$ and $s = 1$. For Pietra's index, it follows that $\underline{G}_{\Pi_0} = 0.121$. Therefore, there exists a subset A such that $\sup_{Q\in\mathcal{Q}} \Pi_0^*(A) - \Pi_Q^*(A) = 0.121$. Such a difference is sensibly larger than the one found by Berger (1985) about the 95% credible interval (for Π_0^*) $C = (-0.93, 2.27)$ which attains 0.945 as minimum value (at $k = 3.4$) as Q_k varies in $\mathcal{Q}$. Therefore, the concentration function gives a different information, with respect to the range of $\Pi_k(C)$, about the changes in the posterior measures induced by the class $\mathcal{Q}$.

Here $\hat{\varphi}(x) = 0.879x$ and the inspection of the plots of $\hat{\varphi}(x)$ and $y = x$ gives us an idea of the changes in the prior. If we choose to compare $\hat{\varphi}(x)$ with, say, the function $g(x) = x^2$, described in Section 4, it is evident that $\varphi(x) < g(x)$ for some x and, therefore, robustness is not achieved. A different choice of $g(x)$, which allows for discrete contaminations (i.e. such that $g(1) < 1$), might have lead to a different situation.

Example 2. (Berger and Berliner, 1986, Moreno and Cano, 1991, Betrò *et al.*, 1994). Assume that $P_\vartheta \sim \mathcal{N}(\vartheta, 1)$, and $\Pi_0 \sim \mathcal{N}(0, 2)$. Consider Π_0 contaminated, with $\varepsilon = 0.1$, by the class $\mathcal{Q}_\mathcal{M}$ of probability measures which have the same median as Π_0. Observe the sample $s = 1$. From Proposition 4, the lowest concentration function is obtained by considering the contamination $(\delta_0 + \delta_1)/2$, where δ_x denotes the Dirac measure at x. Considering Pietra's index, it follows that $\underline{G}_{\Pi_0} = 0.154$. Comparing such a result with Table 1 in Betrò *et al.* (1994), it follows that the credible interval C (the same as in Example 1) attains 0.889 as minimum value, while there exists a subset A such that $\sup_{Q\in\mathcal{Q}_\mathcal{M}} \Pi_0^*(A) - \Pi_Q^*(A) = 0.154$. In this case, we can see that the behaviour of C is not sensibly different from the one of the «worst» subset A.

Example 3. (Rios Insua *et al.*, 1995). The following consulting problem has been considered by some of the authors in Rios Insua *et al.* (1995) for a Spanish insurance company, being then studied by means of dynamical models in such paper (where the problem is throughly presented, along with data).

Given the accident history D_k of a company (number of workers n_k, number of accidents X_k) at the period k, we want to make inference on λ, the individual accident proneness rate. We consider a Poisson model for the number of accidents, i.e. $X_k|\lambda, n_k \sim \mathcal{P}(n_k\lambda)$. A gamma prior $\Pi_0 \sim \mathcal{G}(\alpha_0, \beta_0)$ on λ was chosen and an expert from the insurance company provided some quantiles of the prior measure, leading to the choice $\alpha_0 = 1.59$ and $\beta_0 = 2.22$. We know that the posterior measure on λ is still gamma distributed. Data are recorded monthly, from January 1988 to November 1990, and the number of accidents in the years 1988, 1989 and 1990 are, respectively, 54, 68, 60, while the number of workers oscillates between 286 and 401.

Here we consider the class of gamma priors $\Gamma_\alpha = \{\Pi_\alpha : \Pi_\alpha \sim \mathcal{G}(\alpha, \beta_0), 1 \leq \alpha \leq 3\}$, which is not an ε-contaminations. We compare the probability measures with respect to Π_0 (and its updates) at four stages: a priori, after one year, after 2 years, at the end of the period, i.e. after November 1990. We can see that Π_3 (and its updates) provides the lowest concentration in all cases and that the Pietra's indices are 0.395, 0.075, 0.051 and 0.041, respectively. Therefore, we start with probability measures which are quite far apart from each other and then we get very similar measures, becoming the weight of the data overwhelming with respect to the (different) priors.

Example 4. (Goel and DeGroot, 1981). We present a simple k-level hierarchical model in which X given ϑ_1 is normal distributed $\mathcal{N}(\vartheta_1, \sigma_1^2)$; at the i-th level, ϑ_i given ϑ_{i+1} is normal distributed $\mathcal{N}(\vartheta_{i+1}, \sigma_{i+1}^2)$, $i = 1, \ldots, k-1$. ϑ_k and the variances σ_i, $i = 1, \ldots, k$, are known. We know that, for any $i = 1, \ldots, k-1$, the posterior distribution of ϑ_i, is $\mathcal{N}(\alpha_i x + \beta_i \vartheta_k, \eta_i)$, where $\alpha_i = \eta_i / \sum_1^i \sigma_j^2$, $\beta_i = 1 - \alpha_i$ and $\eta_i = \sum_1^i \sigma_j^2 \sum_{i+1}^k \sigma_j^2 / \sum_1^k \sigma_j^2$.

Consider the data in Example 17 in Berger (1985, p. 181). They can be modelled according to a 2-level hierarchical model. Unlike Berger, we assume the variance at the first level is known (or, to keep the formal equivalence, we consider a Dirac measure as a prior on it). During a 7-years period, a child scores 105, 127, 115, 130, 120, 135, and 115 on a $\mathcal{N}(\vartheta_1, 100)$ IQ test (note that there is a misprint in Berger's values). We suppose that ϑ_1 comes from a normal distribution whose mean, ϑ_2, is the «true» IQ, on which a normal prior Π_0 is finally elicited. As in Berger, let $\sigma_1^2 = 100$, $\sigma_3^2 = 225$ and $\vartheta_3 = 100$. We suppose that $\sigma_2^2 = 1$ (strong belief that ϑ_1 is close to the «true» IQ ϑ_2) and that σ_3^2 is specified with uncertainty, so that we consider an ε-contamination class $\mathcal{Q}_\mathcal{N}$ of priors on ϑ_2, given by $\Pi_k = (1-\varepsilon)\Pi_0 + \varepsilon Q_k$, where Q_k is $\mathcal{N}(100, k)$, with $49 \leq k \leq 400$.

Consider the corresponding concentration functions φ_k: it can be shown numerically that $\Pi_{Q_{400}} \succeq \Pi_Q$ for any $Q \in \mathcal{Q}_\mathcal{N}$ and that the corresponding Pietra's index is equal to 0.385. Numerically, it is possible to show that $\Pi_{Q_{49}}$ is the measure leading to the smallest posterior mean. In comparing posterior mean μ and 95% credible set C for Π_{49} and Π_0, we find that they are $\mu =$

116.223 and $C = (109.679, 122.767)$ for the former, and $\mu = 119.735$ and $C = (112.517, 126.951)$ for the latter. Therefore, a sensitivity analysis focused on posterior means (even supported by credible sets around them) might be intepreted as a robust situation, while it is hard to claim robustness when considering the concentration function. This fact is not contradictory: we are simply looking at two different aspects of the posterior measures.

10. Discussion

In this paper, we have used the concentration function to compare posterior probability measures, corresponding to a class of priors, with respect to a base one. We have considered some classes of contaminated priors; the same approach could be applied to measure robustness with respect to changes in the model, but we expect that computing the lowest concentration function will be a harder task.

In the paper we have studied a problem of global sensitivity, considering the behaviour of a quantity of interest (e.g. Pietra's index) in a class of priors. A different approach, aimed at considering the effects of infinitesimal changes in the prior measure and based on the use of the Gâteaux differential of the concentration function, has been pursued in Fortini and Ruggeri (1995b). Berger (1994) and Wasserman (1992) provide wide discussion about these two approaches.

REFERENCES

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

Berger, J. O. (1990). Robust Bayesian Analysis: sensitivity to the prior. *J. Statist. Plann. Inference*, 25, 303-328.

Berger, J. O. (1994). An overview of Robust Bayesian Analysis. *TEST*, 3, 5-58.

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

Betrò, B., Męczarski, M. and Ruggeri, F. (1994). Robust Bayesian analysis under generalized moments conditions. *J. Statist. Plann. Inference*, 41, 257-266.

Cifarelli, D. M. and Regazzini, E. (1987). On a general definition of concentration function. *Sankhyā B*, **49**, 307-319.

Fortini, S. and Ruggeri, F. (1994). Concentration functions and Bayesian robustness. *J. Statist. Plann. Inference*, 40, 205-220.

Fortini, S. and Ruggeri, F. (1995a). On defining neighbourhoods of measures through the concentration function. *Sankhyā A*, 56, 444-457.

Fortini, S. and Ruggeri, F. (1995b). Infinitesimal properties of the concentration function and Bayesian Robustness. *Quaderno IAMI 96.6*, Milano: CNR-IAMI.

Gini, C. (1914). Sulla misura della concentrazione della variabilità dei caratteri. *Atti del Reale Istituto Veneto di S.L.A., A.A. 1913-1914*, 73, parte II, 1203-1248.

Goel, P. K. and DeGroot, M. H. (1981). Information about hyperparameters in hierarchical models. *J. Amer. Statist. Assoc.*, 76, 140-147.

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

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York.

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

Pietra, G. (1915). Delle relazioni tra gli indici di variabilità. *Atti del Reale Istituto Veneto di S.L.A., A.A. 1914-1915*, 74, parte II, 775-792.

Regazzini, E. (1992). Concentration comparisons between probability measures. *Sankhyā B* **54**, 129-149.

Rios Insua, S., Martin, J., Rios Insua, D. and Ruggeri, F. (1995). Bayesian forecasting for accident proneness evaluation. *Quaderno IAMI 95.7*, Milano: CNR-IAMI.

Wasserman, L. A. (1992). Recent methodological advances in robust Bayesian inference. In *Bayesian Statistics IV* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford Press.