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 \( \Gamma \). 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 \( \Gamma \), restricting then our attention to some \( \varepsilon \)-contamination classes of priors.

**Keywords:** Concentration function, Bayesian robustness, \( \varepsilon \)-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 \( \Gamma \) 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 \( \Gamma \). Small ranges suggest that the inference (or decision) is not actually affected by a particular choice in \( \Gamma \).

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 \( \Pi^* \) themselves, by means of their concentration function with respect to a base posterior \( \Pi_0 \). 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 \( \Pi^* \), of all the sets with equal probability \( x \) under \( \Pi_0 \). Robust analyses can be performed considering the largest range, for any \( x \) in \([0, 1]\), as the prior varies in \( \Gamma \). The width of such intervals, expressed by means of the concentration function, gives a distance between the posteriors \( \Pi^* \) and \( \Pi_0 \). 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 \( \varepsilon \)-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 \( \Pi \) and a uniform one, say \( \Pi_0 \), and allows for the comparison of both probability measures, looking for subsets where \( \Pi \) is much more concentrated than \( \Pi_0 \) (and vice versa). Cifarelli and Regazzini (1987) defined and studied the concentration function of \( \Pi \) with respect to \( \Pi_0 \), where \( \Pi \) and \( \Pi_0 \) 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 \( \Theta \) and a nonnegative function \( h \) on \( N^c \) such that,

\[
\forall \Theta \in \mathcal{F}, \quad \Pi(E) = \Pi_0(E \cap N^c) + \Pi_0(E \cap N) \quad \text{(with } \Pi_0(E \cap N^c) = \int_{E \cap N^c} h(\theta) d\Pi_0(d\theta)) \quad \Pi_0(N) = 0, \quad \Pi_0(N) = \Pi_0(\Theta), \quad \text{where } \Pi_0 \text{ and } \Pi_0 \text{ denote the absolutely continuous and the singular part of } \Pi \text{ with respect to } \Pi_0, \text{ respectively. Set } h(\theta) = \infty \text{ all over } N \text{ and define } H(\theta) = \Pi_0(\Theta : h(\theta) < \theta), \quad c^*_Y = \inf\{y \in \mathbb{R} : H(y) \geq y\}, \quad c^*_Y = \inf\{y \in \mathbb{R} : H(y) \geq y\} \text{ and } L^*_X = \{\theta \in \Theta : h(\theta) < c^*_Y\}. \]

Definition 1. The function \( \varphi : [0, 1] \rightarrow [0, 1] \) is said to be the concentration function of \( \Pi \) with respect to \( \Pi_0 \) if \( \varphi(x) = \Pi(L^*_X + c^*_Y(x - H(c^*_Y))) \) for \( x \in (0, 1) \), \( \varphi(0) = 0 \) and \( \varphi(1) = \Pi_0(\Theta) \).

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

\[
\varphi(x) = \int_0^x [x - H(t)] dt = \int_0^x c^*_Y dt.
\]

To facilitate the understanding of the behaviour of the concentration function, we describe two cases: \( \varphi(1) = 1 \) means that \( \Pi \) is absolutely continuous with respect to \( \Pi_0 \) while \( \varphi(0) = 0 \), \( 0 \leq x \leq \alpha \), means that \( \Pi \) 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 \( \Pi_0 \) and \( \Pi \) are absolutely continuous with respect to Lebesgue measure on \( \mathbb{R}^+ \), 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 \( h(\theta) = 4e^2 \exp(-\theta) \). The concentration function \( \varphi(x) \) is obtained by evaluating \( x = \Pi_0(L^*_X) \) and \( \varphi(x) = \Pi(L^*_X) \), where \( L^*_X = \{\theta \in \Theta : h(\theta) \leq \theta\} \) and \( \theta \) 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 \( \mathcal{F} \) of all probability measure, hence allowing for their comparison.

Definition 2. Let \( \varphi_1, \varphi_2 \) be the concentration functions of \( \Pi_1 \) and \( \Pi_2 \) with respect to \( \Pi_0 \). We say that \( \Pi_2 \) is not less concentrated than \( \Pi_1 \) with respect to \( \Pi_0 \), 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: \( \varphi_{1h}(\Pi) = 2 \int_0^1 (x - \varphi(x)) dx \) and \( \varphi_{1h}(\Pi) = \sup \{x - \varphi(x)\} \) 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

\[
\varphi_{1h}(\Pi) = 0 \Leftrightarrow G_{1h}(\Pi) = 0 \Leftrightarrow \Pi = \Pi_0, \quad G_{1h}(\Pi) = 1 \Leftrightarrow \Pi = \Pi_0.
\]

The following Theorem, proved in Cifarelli and Regazzini (1987), states that \( \varphi(x) \) substantially coincides with the minimum value of \( \Pi \) on the measurable subsets of \( \Theta \) with \( \Pi_0 \)-measure not smaller than \( x \).
**Theorem 1.** If \( A \in \mathcal{T} \), \( \Pi_0(A) = x \), then \( \hat{q}(x) \leq \Pi_0(A) \). Moreover if \( x \in [0, 1] \) is adherent to the range of \( H \), then there exists a \( B_x \) such that \( \Pi(B_x) = x \) and

\[
\hat{q}(x) = \Pi_0(B_x) = \min \{ \Pi(A) : A \in \mathcal{T} \text{ and } \Pi_0(A) \geq x \}.
\]

(2)

If \( \Pi_0 \) 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 \( \Pi \), of all the subsets \( A \) with \( \Pi_0 \)-measure \( x \) satisfies

\[
\hat{q}(x) \leq \Pi(A) \leq 1 - \hat{q}(1 - x).
\]

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 \( \mathcal{F} \), 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.

### 4. General approach

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

\[
\hat{q}(x) \leq \Pi^*(A) \leq 1 - \hat{q}(1 - x),
\]

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

The interpretation of \( \hat{q} \), in terms of Bayesian robustness, is straightforward: the closest \( \hat{q}(x) \) and \( 1 - \hat{q}(1 - x) \) 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 \( \hat{q}(x) \) 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

\[
\sup_{A \in \mathcal{F}} \left[ \Pi_0(A) - \Pi^*(A) \right] \leq \Pi_0(A) (1 - \Pi_0(A)).
\]

If \( \Pi_0(A) = x \), then

\[
A \in \mathcal{F}, \Pi_0(A) = x \Rightarrow \Pi^*(A) \geq \Pi_0(A) (1 - \Pi_0(A)).
\]

Theorem 1 implies that \( \Pi^*(A) \geq \Psi_0(x) \), so that such criterion ensures robustness if \( \hat{q}(x) \geq g(x) \) for all \( x \in [0, 1] \).
5. Concentration function of ε-contaminations

We now apply the general framework, described in the previous Section, to the class $T_0$ of ε-contaminated priors, one of the most relevant classes in robust Bayesian analysis (see Berger, 1994).

**Definition 3.** Let $P_0$ be a fixed prior probability measure and let $\varepsilon \in [0, 1]$. The class $T_0 = \{P_0 + \epsilon Q, Q \in \mathcal{G}\}$, where $\mathcal{G} \subseteq \mathcal{P}$, is said to be an ε-contamination class of priors.

Some classes $\mathcal{G}$ have been proposed and their properties are discussed in Berger (1990). In this paper, we consider the class $\mathcal{G}_a$ of all the probability measures, the class $\mathcal{G}_a$ of all probability measures defined by means of generalised moments conditions (see Betrò et al., 1994) and, if $P_0$ is unimodal, the class $\mathcal{G}_a$ of all unimodal probability measures, with the same mode as $P_0$.

We consider some derivatives of $q(x)$, 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 $P_0(A)$ is sufficiently close to 0.

**Proposition 1.** Let $P_0$ and $Q$ be probability measures on $(\Theta, \mathcal{F})$; for any $a \in [0, 1]$, define $P_0 = (1 - a)P_0 + aQ$. If $q_0$ and $q_0$ are the concentration functions of $P_0$ and $Q$ with respect to $P_0$ respectively, then $q(x) = (1 - a)x + aq_0(x)$.

Let $q_0'(x)$ and $q_0''(x)$ be, respectively, the right-hand and the left-hand derivatives of $q(x)$, then it follows that

$$q_0'(0) = (1 - a) + a \inf_{\theta \in \Theta} h_0(\theta)$$
$$q_0''(1) = (1 - a) + a \sup_{\theta \in \Theta} h_0(\theta),$$

where $h_0$ is the Radon-Nikodym derivative of $Q$ with respect to $P_0$.

**Proof.** The relation between $q$ and $q_0$ 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 $(q_0)'(1) = c$. In the same way, $(q_0)'(0)$ is computed, observing that $c$ is right continuous at the origin.

For any measure $Q$, let $\lambda_0 = (1 - \varepsilon)D_0 + (1 + \varepsilon)D_0 + \varepsilon D_Q$, with

$$D_0 = \int_\Theta l(\theta)P_0(d\theta)$$
$$D_Q = \int_\Theta l(\theta)Q(d\theta).$$

**Proposition 2.** Consider $P_0 = (1 - \varepsilon)P_0 + \varepsilon Q$ so that $P_0 = \lambda_0P_0 + (1 - \lambda_0)Q$. If $q_0$ and $q_0$ denote the concentration functions of $P_0$ and $Q$ with respect to $P_0$, respectively, then it follows that $q_0(x) = \lambda_0q_0 + (1 - \lambda_0)q_0(x)$. Moreover, it follows that

$$q_0'(0) = \lambda_0 + (1 - \lambda_0) \inf_{\theta \in \Theta} h_0(\theta)$$
$$= \lambda_0 \{1 + \varepsilon \inf_{\theta \in \Theta} h_0(\theta)/(1 - \varepsilon)\},$$
$$q_0''(1) = \lambda_0 + (1 - \lambda_0) \sup_{\theta \in \Theta} h_0(\theta)$$
$$= \lambda_0 \{1 + \varepsilon \sup_{\theta \in \Theta} h_0(\theta)/(1 - \varepsilon)\},$$

where $h_0$ and $h_0$ are, respectively, the Radon-Nikodym derivatives of $Q$ with respect to $P_0$ and of $Q$ with respect to $P_0$. 

**Proof.** The expression about $P_0$ is well known (see Berger, 1985, p. 206). The other results follow from Proposition 1 and the fact that $h_0(\theta) = (D_0/D_0) \times h_0(\theta)$. 

\[ \square \]
The previous Propositions are now applied to robustness analysis, when considering $\epsilon$-contaminations of a nonatomic prior $\Pi_0$. It is worth mentioning that the results hold, with slight changes, for $\epsilon = 1$, 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 $\mathcal{A}_\epsilon$ of all contaminations. Let $\hat{\theta}$ be the maximum likelihood estimate of $\theta$ and $\hat{Q}$ be the Dirac measure concentrated at $\hat{\theta}$. It follows that $\Pi_0 \succeq \Pi_0$ for any $Q \in \mathcal{A}_\epsilon$. Moreover $\mathcal{C}_n = \mathcal{G}_n = \{1 + [(1 - \epsilon)/\epsilon]D_0/D_0(\hat{\theta})\}^{-1}$.

Proof. Proposition 2 and (1) imply that $\varphi_Q(x) = \lambda_Q x$. From Kemperman (1968), it follows that $D_Q$ is maximised by a discrete measure, concentrated in at most $n + 1$ points, so that $\varphi_Q(x) / \lambda_Q x$ is maximised by $\varphi_{Q_0}(x)$, for any $Q \in \mathcal{A}_\epsilon$. The result about Gini’s and Pietra’s indices follows immediately.

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

8. Unimodal contaminations

Suppose that $\Pi_0$ is unimodal with mode $\theta_0$. Consider the class $\mathcal{A}_\epsilon$ of all unimodal probability measures with mode $\theta_0$. Such a class contains just one discrete measure, the Dirac measure $\hat{Q}$ concentrated at $\theta_0$, the only one which could lead to the lowest concentration function, as shown in the next Proposition.

Proposition 5. If

$$\sup_{Q \in \mathcal{A}_\epsilon} D_Q \leq l(\theta_0),$$

then it follows that $\Pi_0 \succeq \Pi_0$ for any $Q \in \mathcal{A}_\epsilon$.

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

$$\inf_{\theta \in \Theta} h_0(\theta) > 0 \text{ and } \inf_{\theta \in \Theta} q(\theta) \Pi_0(\theta) > (D_Q - l(\theta_0))[D_0 + \epsilon l(\theta_0)/(1 - \epsilon)]$$

for every $Q \in \mathcal{A}_\epsilon$ such that $D_Q > l(\theta_0)$.

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

If $\Pi_0 \succeq \Pi_0$ for any $Q \in \mathcal{A}_\epsilon$, it follows that $\mathcal{C}_n = \mathcal{G}_n = \{1 + [(1 - \epsilon)/\epsilon]D_0/D_0(\theta_0)\}^{-1}$.
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 $Q \in \mathcal{Q}_n$ such that $D_{Q} > l(\theta_0)$. We now prove that there is no $Q$ such that $q_{Q}(x) < q_{Q'}(x)$ for all $x \in [0, 1]$ and $Q \in \mathcal{Q}_n$, $\theta \in \Theta$ for any $Q \in \mathcal{Q}_n$. Since $Q$ is a discrete measure, it follows that $q_{Q}(1) = \sum_{\theta \in \Theta} q_{Q}(\theta) < \sum_{\theta \in \Theta} q_{Q}(\theta) = q_{Q}(1) = 1$ for all $Q \neq Q$. Let $\hat{Q} \in \mathcal{Q}_n$ such that $D_{\hat{Q}} > l(\theta_0)$; then Proposition 2 implies that $(q_{\hat{Q}})'_{\theta}(0)$. Therefore, since the concentration function is continuous, there exists $x^*$ such that $q_{\hat{Q}}(x) < q_{\hat{Q}}(x)$, for any $x \in (0, x^*)$.

Let us suppose that there exists at least one $Q \in \mathcal{Q}_n$ such that $\int_{\Theta} h_{Q}(\theta) > 0$.

From Proposition 2, it follows that $(q_{Q})_{x}(0) > (q_{Q})_{x}(0)$ if and only if (5) holds. Therefore, $\Pi_{Q} \geq \Pi_{Q}$ for any $Q \in \mathcal{Q}_n$ 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. \hfill \Box

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

9. Examples

We now concentrate functions in analysing the robustness in some examples, showing the applicability of our proposal, even when considering classes different from the $\epsilon$-contaminations.

Example 1. (Berger, 1985, p. 212). Assume that $P_0 \sim N(\theta_0, \sigma^2)$, $\sigma^2$ known, and $\Pi_k = N(\theta_0, \sigma^2)$, $\theta_0$ and $\sigma^2$ known, with density function $\pi_k(\theta)$. Let $\Pi$ be the class of the probability measures which are either $\Pi_k \sim \pi_k(\theta_0 - k, \theta_0 + k)$, $k > 0$, or $Q_m$, which assigns probability one to the point $\theta_0$ (note that Berger, 1985, does not consider $Q_m$, but we can add it with no changes in his results). Define $\Pi_k = (1 - \epsilon)\Pi_0 + \epsilon\Pi_k$. Given a sample $s$ from $\Pi$, the likelihood function is $l(\theta) = [1/(2\pi \sigma^2)] \exp(-((\theta - s)^2/(2\sigma^2)))$. The density of $Q_m$, $k > 0$, is given by $q_{Q_m}(\theta) = (1/2k)I_{[\theta_0-k, \theta_0+k]}(\theta)$ where $I_A$ is the indicator function of the set $A$. Note that $Q_m$ converges in distribution to $Q_m$ as $k \to 0$.

It can be seen that $D_{Q_m} = (1/(2\pi \sigma^2 + \sigma^2_0)) \int (\theta_0 - s)^2(2(\sigma^2 + \sigma^2_0)) d\theta.$

929

Result 1. Given a sample $s$, then $D_{Q_m} \leq l(\theta_0)$ for any $k > 0$ if and only if $|s - \theta_0|/\sigma \leq 1$.

Proof. It holds that $\lim_{k \to \infty} D_{Q_m} = l(\theta_0)$ and $\lim_{k \to \infty} D_{Q_m} = 0^+$. Set $h = k\sigma$ and $v = (s - \theta_0)/\sigma$. Then, there exists the derivative of $D_{Q_m}$ with respect to $h$:

$D'_{Q_m}(h) = [1/(2\pi \sigma^2)] f(h),$

with

$f(h) = -(1/h) \int_{-h}^{h} \exp(\sqrt{-2\pi \sigma^2} v^2) dv.$

and, consequently, $\lim_{h \to 0} f(h) = 0$ and $f'(h) = h([v - h] \exp(-v^2/2) - [v + h] \exp(-v^2/2)).$ For any $h > 0$, $f'(h) < 0$ if and only if $g(h) = (v - h) \exp(-v^2/2) - (v + h) \exp(-v^2/2) < 0$.

Such a condition is obviously satisfied if $h \geq v$, so that $D_{Q_m} \leq l(\theta_0)$ for any $k \geq |s - \theta_0|$. Otherwise, suppose that $v \leq 1$, then

$g(h) = 2 \sum_{j=0}^{\infty} \frac{\sigma^2 \exp(-v^2/2)}{(2j+1)!} \left(\frac{\sigma^2 \exp(-v^2/2)}{(2j)!}\right) \leq 0.$

The convexity of $D_{Q_m}$ implies that $D_{Q_m} \leq l(\theta_0)$ for any $k > 0$ if $v < 1$. Suppose now that $v > 1$, then $\lim_{h \to 0^+} D_{Q_m}(h) = \lim_{h \to 0^+} f'(h)/4h = \exp(-v^2)2 \cdot \lim_{h \to 0^+} h(2\pi - 1) = 0^+$ so that $D_{Q_m} > l(\theta_0)$, for $k$ close enough to 0. \hfill \Box

Result 2. Given a sample $s$, then the concentration function $q_{Q_m}$ of $\Pi_m$ with respect to $\Pi_k$ is such that $q_{Q_m}(x) \leq q_{Q_m}(x)$, for any $x \in [0, 1]$ and for any concentration function $q_{Q_m}$ of $\Pi_m$ with respect to $\Pi_k$, if and only if $|s - \theta_0|/\sigma \leq 1$. Given such a sample $s$, it follows that

$\tilde{C}_{\Pi_k} = \tilde{C}_{\Pi_k} = \left[1 + \frac{(1-\epsilon)}{\epsilon} \sqrt{\frac{\sigma^2}{(\sigma^2 + \sigma^2_0)}} \exp\left(-\frac{\sigma^2_0 (s - \theta_0)^2}{2(\sigma^2 + \sigma^2_0)^2}\right)\right].$

Otherwise, there is no $Q_{Q_m}(x)$ such that $q_{Q_m}(x) \leq q_{Q_m}(x)$, for any $x \in [0, 1]$ and $q_{Q_m}$.

Proof. From Result 1 and Proposition 5, it follows that $q_{Q_m}$ lies under any other concentration function $q_{Q_m}$ if $|s - \theta_0|/\sigma \leq 1$. Otherwise, there are no $\Pi_m$,.
on \( \lambda \) was chosen and an expert from the insurance company provided some quantities of the prior measure, leading to the choice \( \alpha_0 = 1.59 \) and \( \beta_0 = 2.22 \).

We know that the posterior measure on \( \lambda \) 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 \( \Gamma_\alpha = \{ \Pi_\alpha : \Pi_\alpha \sim \mathcal{G}(\alpha, \beta_0), \alpha \leq 1 \} \), which is not an \( \varepsilon \)-contaminations. We compare the probability measures with respect to \( \Pi_\alpha \) (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 \( \Pi_\alpha \) (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 \( \theta \) is normal distributed \( \mathcal{N}(\theta, \sigma^2) \); at the i-th level, \( \theta_i \) given \( \theta_{i-1} \) is normal distributed \( \mathcal{N}(\theta_{i-1}, \sigma_{i-1}^2) \), \( i = 1, \ldots, k-1 \) and the variances \( \sigma_i \), \( i = 1, \ldots, k \), are known. We know that, for any \( i = 1, \ldots, k-1 \), the posterior distribution of \( \theta_i \) is \( \mathcal{N}(\alpha_i x + \beta_i \theta_{i-1}, \eta_i) \), where \( \alpha_i = \eta_i / \sum_{j=1}^{i} \sigma_j^2 \) and \( \beta_i = 1 - \alpha_i \) and \( \eta_i = \sum_{j=i+1}^{k} \sigma_j^2 / \sigma_i^2 \).

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 \( \mathcal{N}(\theta, \sigma^2) \) IQ test (note that there is a misprint in Berger’s values). We suppose that \( \theta_1 \) comes from a normal distribution whose mean, \( \theta_2 \), is the “true” IQ, on which a normal prior \( \theta_0 \) is finally elicited. As in Berger, let \( \sigma_0^2 = 100, \sigma_3^2 = 225 \) and \( \sigma_2^2 = 100 \). We suppose that \( \sigma_3^2 = 1 \) (strong belief that \( \theta_3 \) is close to the “true” IQ \( \theta_2 \)) and that \( \sigma_3^2 \) is specified with uncertainty, so that we consider an \( \varepsilon \)-contamination class of priors on \( \theta_3 \), given by \( \Pi_k = (1 - \varepsilon) \Pi_0 + \varepsilon \mathcal{Q}_k \), where \( \mathcal{Q}_k \) is \( \mathcal{N}(0, \sigma^2) \), with \( 49 \leq k \leq 400 \).

Consider the corresponding concentration functions \( \phi_k \); it can be shown numerically that \( \Pi_{k,0} \leq \Pi_Q \) for any \( \mathcal{Q} \in \mathcal{Q}_1 \) and that the corresponding Pietra’s index is equal to 0.385. Numerically, it is possible to show that \( \Pi_{k,0} \) is the measure leading to the smallest posterior mean. In comparing posterior mean \( \mu \) and 95% credible set \( C \) for \( \Pi_{k,0} \) and \( \Pi_0 \), we find that they are \( \mu = \ldots \)
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 interpreted 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


