On the Use of the Concentration Function in Bayesian Robustness

Sandra Fortini and Fabrizio Ruggeri

ABSTRACT We present applications of the concentration function in both global and local sensitivity analyses, along with its connection with Choquet capacities.

Key words: global sensitivity, local sensitivity, classes of priors.

6.1 Introduction

In this paper, we expose the main properties of the concentration function, defined by Cifarelli and Regazzini (1987), and its application to Bayesian robustness, suggested by Regazzini (1992) and developed, mainly, by Fortini and Ruggeri (1993, 1994, 1995a, 1995b, 1997).

The concentration function allows for the comparison between two probability measures $\Pi$ and $\Pi_0$, either directly by looking at the range spanned by the probability, under $\Pi$, of all the subsets with a given probability under $\Pi_0$ or by considering summarising indices. Such a feature of the concentration function makes its use in Bayesian robustness very suitable.

Properties of the concentration function are presented in Section 2. Some applications of the concentration function are illustrated in the paper; in Section 3 it is used to define classes of prior measures, whereas Sections 4 and 5 deal with global and local sensitivity, respectively. An example in Section 6 describes how to use the results presented in previous sections. Finally, Section 7 illustrates connections between the concentration function and 2-alternating Choquet capacities, described in Wasserman and Kadane (1990, 1992).

6.2 Concentration function

Cifarelli and Regazzini (1987) defined the concentration function (c.f.) as a generalisation of the well-known Lorenz curve, whose description can be
found, for example in Marshall and Olkin (1979, p. 5): “Consider a population of \( n \) individuals, and let \( x_i \) be the wealth of individual \( i, i = 1, \ldots, n \). Order the individuals from poorest to richest to obtain \( x^{(1)}_1, \ldots, x^{(n)}_n \). Now plot the points \((k/n, S_k/S_n)\), \( k = 0, \ldots, n \), where \( S_k = 0 \) and \( S_k = \sum_{i=1}^{k} x^{(i)} \) is the total wealth of the poorest \( k \) individuals in the population. Join these points by line segments to obtain a curve connecting the origin with the point \((1, 1), \ldots\). Notice that if total wealth is uniformly distributed in the population, then the Lorenz curve is a straight line. Otherwise, the curve is convex and lies under the straight line.”

The classical definition of concentration refers to the discrepancy between a probability measure \( \Pi \) (the “wealth”) and a uniform one (the “individuals”), say \( \Pi_0 \), and allows for their comparison, looking for subsets where the former is much more concentrated than the latter. The definition can be extended to non-uniform discrete distributions; we use data analysed by DiBona et al. (1993), who addressed the issue of racial segregation in the public schools of North Carolina, USA. The authors proposed a method to check if students tend to be uniformly distributed across the schools in a district or, otherwise, if they tend to be segregated according to their race. The proposed segregation index allowed the authors to state that segregation was an actual problem for all grades (K–12) and it had increased from 1982 to 1992. Lorenz curves are helpful in analysing segregation for each grade in a school.

<table>
<thead>
<tr>
<th></th>
<th>Native Americans</th>
<th>Asians</th>
<th>Hispanics</th>
<th>Blacks</th>
<th>Whites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durham (82)</td>
<td>0.002</td>
<td>0.012</td>
<td>0.002</td>
<td>0.332</td>
<td>0.652</td>
</tr>
<tr>
<td>School 332</td>
<td>0.011</td>
<td>0.043</td>
<td>0.022</td>
<td>0.403</td>
<td>0.521</td>
</tr>
<tr>
<td>Ratio (S/D)</td>
<td>5.500</td>
<td>3.500</td>
<td>11.000</td>
<td>1.210</td>
<td>0.799</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Native Americans</th>
<th>Asians</th>
<th>Hispanics</th>
<th>Blacks</th>
<th>Whites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durham (92)</td>
<td>0.002</td>
<td>0.026</td>
<td>0.009</td>
<td>0.345</td>
<td>0.618</td>
</tr>
<tr>
<td>School 332</td>
<td>0.007</td>
<td>0.106</td>
<td>0.013</td>
<td>0.344</td>
<td>0.530</td>
</tr>
<tr>
<td>Ratio (S/D)</td>
<td>3.500</td>
<td>4.070</td>
<td>1.440</td>
<td>0.997</td>
<td>0.858</td>
</tr>
</tbody>
</table>

**TABLE 1. Public Kindergartens in Durham, NC**

In Table 1 we present the distribution of students, according to their race, in a school (labelled as 332) in the city of Durham and compare it with the race distribution in the city public school system. We consider the ratios between percentages in the school \( (S_i) \) and the city \( (D_i) \), ordering the races according to their (ascending) ratios. Similarly to Marshall and Olkin, we plot (Fig. 1) the straight line connecting \((0,0)\) and the points \((\sum_{j=1}^{5} D^{(j)}, \sum_{j=1}^{5} S^{(j)})\), \( i = 1, \ldots, 5 \), where \( D^{(j)} \) and \( S^{(j)} \) correspond to the race with the \( j \)th ratio (in ascending order). The distance between the straight line and the other two denote an equal distribution of students in the school with respect to \((w.r.t.)\) the city, and its largest value is one of the proposed segregation indexes. Moreover, the 1992 line lies above the 1982 one up to 0.9 (approx.), denoting an increase in adherence to the city population among the kids from the largest groups (White and Black) and, conversely, a decrease among other groups. Therefore, segregation at school 332 is decreasing over the 10 year period, in contrast with the general tendency in North Carolina, at each grade (see Di Bona et al., 1993, for more details).

6. On the Use of the Concentration Function in Bayesian Robustness

**FIGURE 1. Lorenz curve for races at school 332 vs. Durham schools**

As an extension of the Lorenz curve, Cifarelli and Regazzini (1987) defined and studied the c.f. of \( \Pi \) w.r.t. \( \Pi_0 \), where \( \Pi \) and \( \Pi_0 \) are two probability measures on the same measurable space \((\Theta, \mathcal{F})\). According to the 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

\[
\Pi(E) = \Pi_N(E \cap N^C) + \Pi_E(E \cap N), \quad \forall E \in \mathcal{F},
\]

where \( \Pi_N \) and \( \Pi_E \) are, respectively, the absolutely continuous and the singular part of \( \Pi \) w.r.t. \( \Pi_0 \), that is, such that

\[
\Pi_E(E \cap N^C) = \int_{\theta \in N^C} h(\theta) \Pi_0(d\theta), \Pi_0(N) = 0, \Pi_N(N) = \Pi_\theta(\Theta).
\]

Set \( h(\theta) = \infty \) all over \( N \) and define \( H(y) = \Pi_0\left(\{\theta \in \Theta : h(\theta) \leq y\}\right) \), \( c_\theta = \inf\{y \in \mathbb{R} : H(y) \geq \alpha\} \) and \( c_\theta^+ = \lim_{\alpha \to \infty} c_\alpha^- \). Finally, let \( L_\theta^+ = \{\theta \in \Theta : h(\theta) \leq c_\theta^+\} \) and \( L_\theta^- = \{\theta \in \Theta : h(\theta) < c_\theta^-\} \).
Definition 1 The function $\varphi_\Pi : [0,1] \to [0,1]$ is the concentration function of $\Pi$ with respect to $\Pi_0$ if $\varphi_\Pi(x) = \Pi(L_x^+) + c_\varepsilon[x - H(c_\varepsilon)]$ for $x \in (0,1)$, $\varphi_\Pi(0) = 0$ and $\varphi_\Pi(1) = \Pi_0(\emptyset)$.

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

$$\varphi_\Pi(x) = \int_0^x [x - H(t)]dt = \int_0^x c_\varepsilon dt. \quad (1)$$

It is worth mentioning that $\varphi_\Pi(1) = 1$ implies that $\Pi$ is absolutely continuous w.r.t. $\Pi_0$ while $\varphi_\Pi(x) = 0, 0 \leq x \leq \alpha$, means that $\Pi$ gives no mass to a subset $A \in \mathcal{F}$ such that $\Pi_0(A) = \alpha$.

We present two examples to illustrate how to compute c.f.s. Consider the c.f. of a normal distribution $\mathcal{N}(0,1)$ w.r.t. a Cauchy $C(0,1)$. The Radon-Nikodym derivative $h(\theta)$ is plotted in Fig. 2, a horizontal line is drawn and the subset of $\Theta$ with Radon-Nikodym derivative below the line becomes $L_x$ (for an adequate $x$). This procedure is equivalent to computing and ordering ratios as in the example about school 332. The c.f. is obtained by plotting the points $(x, \varphi_\Pi(x))$, where $x$ and $\varphi_\Pi(x)$ are the probabilities of $L_x$ under the Cauchy and the normal distributions, respectively.

![Figure 2](image.png)

As another example, consider a gamma distribution $\Pi \sim \mathcal{G}(2,1)$ and an exponential one $\Pi_0 \sim \mathcal{E}(1)$. Their densities on $\mathbb{R}^+$ are, respectively, $\pi(\theta) = \theta e^{-\theta}$ and $\pi_0(\theta) = e^{-\theta}$, so that $h(\theta) = \theta, \theta \geq 0$. For any $x \in [0,1]$, we compute the c.f. by finding the value $y$ such that $x = \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\})$.

6. On the Use of the Concentration Function in Bayesian Robustness

It follows that $y = -\log(1-x)$ since $x = \int_0^x e^{-\theta}d\theta = 1 - e^{-y}$. Finally, we get

$$\varphi_\Pi(x) = \Pi(\{\theta \in \Theta : h(\theta) \leq -\log(1-x)\}) = \int_0^{-\log(1-x)} e^{-\theta}d\theta = 1 - (1-x)(1 - \log(1-x)).$$

The comparison of probability measures in a class is made possible by the partial order induced by the c.f. over the space $\mathcal{P}$ of all probability measures, when considering c.f.s lying above others. Total orderings, consistent with the partial one, are discussed in Regazzini (1992); they are achieved when considering synthetic measures of concentration as Gini's (1914) concentration index $C_{\Pi_0}(\Pi) = 2 \int_0^1 (x - \varphi_\Pi(x))dx$ and Pietra's (1915) index $C_{\Pi_0}(\Pi) = \sup_{\Pi \in \Pi_0}(x - \varphi_\Pi(x))$. The latter coincides with the total variation distance between $\Pi$ and $\Pi_0$, as proved by Cifarelli and Regazzini (1987).

The following theorem, proved in Cifarelli and Regazzini (1987), states that $\varphi_\Pi(x)$ 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 $A \in \mathcal{F}$ and $\Pi_0(A) = x$, then $\varphi_\Pi(x) \leq \Pi_0(A)$. Moreover, if $x \in [0,1]$ is adherent to the range of $H$, then there exists a $B_x$ such that $\Pi_0(B_x) = x$ and

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

If $\Pi_0$ is nonatomic, then (2) holds for any $x \in [0,1]$.

This theorem is relevant when applying the c.f. to robust Bayesian analysis: for any $x \in [0,1]$, the probability, under $\Pi_0$ of all the subsets $A$ with measure $x$ under $\Pi_0$, satisfies

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

As an example, we can consider the c.f. of $\Pi \sim \mathcal{G}(2,2)$ w.r.t. $\Pi_0 \sim \mathcal{E}(1)$, showing that $[0.216, 0.559]$ is the range spanned by the probability, under $\Pi_0$, of the sets $A$ with $\Pi_0(A) = 0.4$ (see Fig 3).

Finally, we mention that the c.f., far from substituting other usual distribution summaries, e.g. the mean, furnishes different information about probability measures. As an example, consider two measures concentrated on disjoint, very close sets in $\mathbb{R}$: their means are very close, their variances might be the same, but their c.f. is 0 in $[0,1]$. 

---

Note: The content is a natural representation of the text in the image. The conversion into LaTeX or another structured format is not necessary for this task.
FIGURE 3. Range of $\Pi(A)$ spanned by $A$ s.t. $\Pi_0(A) = .4$ ($g(2,2)$ vs. $E(1)$)

6.3 Classes of priors

Fortini and Ruggeri (1995a) presented a method, based upon the c.f., to define neighbourhoods of probability measures and applied it in robust Bayesian analyses in Fortini and Ruggeri (1994). Their approach allows the construction of probability measures $\Pi$ with functional forms close to a nonatomic baseline measure $\Pi_0$. In particular, they defined neighbourhoods of $\Pi_0$ by imposing constraints on the probability of all measurable subsets, such as requiring $|\Pi_0(A) - \Pi(A)| \leq \Pi_0(A) (1 - \Pi_0(A))$, for any $A \in \mathcal{F}$. By observing that the above relation can be written $\Pi(A) \geq g(\Pi_0(A))$, with $g(x) = x^2$, Fortini and Ruggeri gave the following definitions.

**Definition 2** A function $g : [0,1] \rightarrow [0,1]$ is said to be compatible if $g$ is a monotone nondecreasing, continuous, convex function, with $g(0) = 0$.

**Definition 3** If $g$ is compatible, then the set

$$K_g = \{\Pi : \Pi(A) \geq g(\Pi_0(A)), \forall A \in \mathcal{F}\}$$

will be a $g$-neighbourhood of $\Pi_0$.

Observe that, if $\Pi \in K_g$, then $g(\Pi_0(A)) \leq \Pi(A) \leq 1 - g(1 - \Pi_0(A))$, for any $A \in \mathcal{F}$. The requirement $g(0) = 0$ is needed to avoid $\Pi(\emptyset) \leq 1 - g(0) < 1$, while monotonicity, continuity and convexity are thoroughly discussed in Fortini and Ruggeri (1995).

As proved in Fortini and Ruggeri (1995), $K_g$ generates a topology since it becomes a fundamental system of neighbourhoods of $\Pi_0$, when $g$ belongs to an adequate class $G$ of compatible functions.

6. On the Use of the Concentration Function in Bayesian Robustness

The definition of a $g$-neighbourhood of $\Pi_0$ can be reformulated by means of the c.f. w.r.t. $\Pi_0$, as stated in the following:

**Theorem 2** The set $K_g = \{\Pi : \varphi_{\Pi}(x) \geq g(x), \forall x \in [0,1]\}$ is a $g$-neighbourhood of $\Pi_0$.

Fortini and Ruggeri (1995) proved that any compatible $g$ is a c.f.

**Theorem 3** Given a function $g : [0,1] \rightarrow [0,1]$, there exists at least one measure $\Pi$ such that $g$ is the c.f. of $\Pi$ w.r.t. $\Pi_0$ if and only if $g$ is compatible.

6.3.1 Main results

Consider the space $\mathcal{P}$ of all probability measures on $\Theta$ endowed with the weak topology. $\mathcal{P}$ can be metrized as a complete separable metric space. Consider the set $F_g$ of extremal points of $K_g$, that is, the probability measures $\Pi \in K_g$ such that

$$\Pi = a\Pi_1 + (1 - a)\Pi_2, \Pi_1 \in K_g, \Pi_2 \in K_g, 0 < a < 1 \Rightarrow \Pi = \Pi_1 = \Pi_2.$$

The following results were proved by Fortini and Ruggeri (1995).

**Theorem 4** $F_g \subseteq E_g$, where $E_g = \{\Pi : \varphi_{\Pi}(x) = g(x), \forall x \in [0,1]\}$. If $g(1) = 1$, then $F_g$ coincides with $E_g$.

Furthermore, every probability measure whose c.f. is greater than $g$ can be represented as a mixture of probability measures having $g$ as c.f., applying Choquet's Theorem (Phelps, 1966).

**Theorem 5** Let the function $g : [0,1] \rightarrow [0,1]$ be compatible. Then, for any probability measure $\Pi \in K_g$, there exists a probability measure $\mu_{\Pi}$ on $\mathcal{P}$ such that $\mu_{\Pi}(F_g) = 1$ and $\Pi = \int_{\mathcal{P}} \mu_{\Pi}(d\Pi)$.

The supremum (and infimum) of ratio-linear functionals of $\Pi$ is found in $E_g$, as shown in

**Theorem 6** Let $f$ and $m$ be real-valued functions on $\Theta$ such that $\int_{\Theta} f(\theta) m(d\theta) < \infty$ and $0 \leq \int_{\Theta} m(\theta) \Pi(d\theta) < \infty$ for any $\Pi \in K_g$. Then

$$\sup_{\Pi \in K_g} \int_{\Theta} f(\theta) m(d\theta) \Pi(d\theta) = \sup_{\Pi \in E_g} \int_{\Theta} f(\theta) m(d\theta) \Pi(d\theta).$$

Computations of bounds on prior expectations are simplified by taking in account

**Theorem 7** Let $H_f(\gamma) = \Pi_0 (\{\theta \in \Theta : f(\theta) \leq \gamma\}$, $c_f(x) = \inf \{y : H_f(y) \geq x\}$. Then $\sup_{\Pi \in K_g} \int_{\Theta} f(\theta) m(d\theta) \int_{c_f(x)}^{\infty} f(z) g(z) dz$. 

The result can be applied to find bounds on posterior expectations, too, using the linearization technique presented by Lavine (1988) and Lavine et al. (2000). Finally, the result was used in Ruggeri (1994) to compute bounds on the posterior probability of sets.

Corollary 1

\[
\sup_{\Pi \in K} \Pi(A|x) = \left\{ \frac{-\int_{A} \Pi_{0}(A') e_{-I_{A}}(x) c(x) dx}{\int_{\Pi_{0}(A')} e_{I_{A}}(x) c(x) dx} \right\}^{-1},
\]

where \(I_{A}\) is the indicator function of the subset \(A\), \(I_{A}(\theta)\) is the likelihood function and, for any \(\theta \in \Theta\), \(I_{A}(\theta) = I_{A}(\theta)I_{AC}(\theta)\) and \(I_{x}(\theta) = I_{x}(\theta)I_{A}(\theta)\).

6.3.2 Classes of priors as concentration function neighbourhoods

Fortini and Ruggeri (1994) considered classes of prior measures \(K_{g}\) such that their c.f.s w.r.t. a nonatomic base one, say \(\Pi_{0}\), are pointwise not smaller than a specified compatible function \(g\). The function \(g\) gives the maximum concentration of a measure w.r.t. a base one which is deemed compatible with our knowledge. Note that, assuming \(\Pi_{0}\) nonatomic, the discrete measures can be ruled out or not by choosing \(g(1) = 1\) or \(< 1\), respectively. The posterior expectation of any function \(f(\theta)\), say \(E^{*}(f)\), can be maximised all over \(K_{g}\) applying Theorems 6 and 7. The results are consistent with those found in the literature.

Here we review the collection of classes, including some that are well known, defined by Fortini and Ruggeri (1994). Note that \(F_{g}\) can be a proper subset of \(E_{g}\), as in the case of \(\epsilon\)-contamination and total variation neighbourhoods.

\(\epsilon\)-contamination. The \(\epsilon\)-contamination class \(\Gamma_{\epsilon} = \{\Pi_{Q} = (1-\epsilon)\Pi_{0} + \epsilon\Pi, \Pi \in \mathcal{P}\}\) is defined by \(g(x) = (1-\epsilon)x, \forall x \in [0,1]\). The sets \(E_{g}\) and \(F_{g}\) are obtained, respectively, considering singular w.r.t. \(\Pi_{0}\) and Dirac contaminating measures. As shown in Berger (1990), \(E^{*}(f)\) is maximised by contaminating Dirac measures, i.e. over \(F_{g}\).

DENSITY BOUNDED CLASS. Given a probability measure \(\Pi_{0}\) and \(k > 0\), consider the class

\[
\Gamma_{k}^{D} = \{\Pi : (1/k)\Pi_{0}(A) \leq \Pi(A) \leq k\Pi_{0}(A), \forall A \in \mathcal{F}\},
\]

studied by Ruggeri and Wasserman (1991). This class, a special case of the density bounded classes defined by Lavine (1991), is a c.f. neighbourhood \(K_{g}\), with \(g(x) = \max\{x/k, k(x-1) + 1\}\).

6.4 Global sensitivity

As discussed by Moreno (2000), global sensitivity addresses the issue of computing ranges for quantities of interest as the prior measure varies in a class. Usually, quantities like posterior means and set probabilities have been considered, whereas less attention has been paid to changes in the functional form of the posterior measures (see Boratynska, 1996, for the study of the “radius” in the class of posterior measures, endowed with the total variation metric). C.f.s have been used in such a context, and the main reference is the paper by Fortini and Ruggeri (1995b), who considered \(\epsilon\)-contaminations and compared the c.f.s of the posterior probability measures w.r.t. a base posterior measure \(\Pi_{0}\). In computing, pointwise, the infimum \(\varphi_{f}(x)\) of the c.f., their interest was twofold: providing a measure of the distance between the distributions in the class and \(\Pi_{0}\) and checking if the probability of all measurable sets would satisfy bounds like those used in the previous section to define classes of measures.

Consider a class \(\Gamma\) of probability measures \(\Pi\) and a base prior \(\Pi_{0}\), as in the \(\epsilon\)-contamination class given by

\[
\Gamma_{\epsilon} = \{\Pi_{Q} = (1-\epsilon)\Pi_{0} + \epsilon Q, Q \in \mathcal{Q}\},
\]

where \(\mathcal{Q} \subseteq \mathcal{P}\) and \(0 \leq \epsilon \leq 1\).
Let $\Pi^*$ denote the posterior measure corresponding to the prior $\Pi$. Consider the class

$$\Psi = \{\phi_\Pi : \phi_\Pi \text{ is the c.f. of } \Pi^* \text{ w.r.t. } \Pi_0, \Pi \in \Gamma\}.$$ 

From Theorem 1 and (3), it follows, for any $\Pi \in \Gamma$ and $A \in \mathcal{F}$ with $\Pi_0^*(A) = x$, that

$$\widehat{\phi}(x) \leq \Pi^*(A) \leq 1 - \widehat{\phi}(1 - x),$$

where $\widehat{\phi}(x) = \inf_{\Pi \in \Gamma} \phi_\Pi(x)$, for any $x \in [0, 1]$.

The interpretation of $\widehat{\phi}$, in terms of Bayesian robustness, is straightforward: the closest $\widehat{\phi}(x)$ and $1 - \widehat{\phi}(1 - x)$ are for all $x \in [0, 1]$, the closest the posterior measures are. It is then possible to make judgments on robustness by measuring the distance between $\widehat{\phi}(x)$ and the line $y = x$, for example, by Gini and Pietra's indices as in Carota and Ruggeri (1994) and Fortini and Ruggeri (1995b).

Fortini and Ruggeri (1995b) proved, for $\epsilon$-contaminations, the following:

**Theorem 8** If $\phi$ and $\phi_0$ denote the c.f.'s of $\Pi_0$ and $Q^*$ w.r.t. $\Pi_0^*$, respectively, then it follows that

$$\phi(x) = \lambda_Q x + (1 - \lambda_Q)\phi_0(x),$$

where

$$\lambda_Q = (1 - \epsilon)D_0/[(1 - \epsilon)D_0 + \epsilon D_Q],$$

with $D_0 = \int_\Theta l_\theta \Pi_0(\theta) d\theta$ and $D_Q = \int_\Theta l_\theta \Pi_Q(\theta) d\theta$.

They were able to find $\epsilon$-contaminations of a nonatomic prior $\Pi_0$ leading to the lowest c.f., when considering arbitrary contaminations and those given by generalised moment conditions; they found the lowest c.f. in the unimodal case when $\sup_{\mathcal{Q}} D_Q \leq l(\theta_0)$ holds.

A similar approach was followed by Carota and Ruggeri (1993), who considered the class of mixtures of probability measures defined on disjoint sets with weights known to vary in an interval. The class is suitable to describe, with some approximation, the case of two (or more) populations, depending on the same parameter $\theta$, which are strongly concentrated in disjoint subsets.

Finally, it is worth mentioning that Fortini and Ruggeri (1995b) used c.f.s and compatible functions $g$ in checking posterior robustness as well. They considered $g$ as a threshold function, denoting how much the posterior set probabilities were allowed to vary (for example, $\Pi^*(A) \geq g(\Pi_0^*(A))$, for any $A \in \mathcal{F}$). Therefore, robustness is achieved when $\widehat{\phi}(x) \geq g(x)$ for all $x \in [0, 1]$.

### 6.5 Local sensitivity

Fortini and Ruggeri (1997) studied functional derivatives of the c.f. and mentioned they could be used in Bayesian robustness to perform local sensitivity analysis (see Gustafson, 2000, on the latter). An example is presented in the next section. Here we present some results based on Gâteaux differentials.

**Definition 4** Let $X$ and $Y$ be linear topological spaces. The Gâteaux differential in the direction of $h \in X$ and at a point $x_0$ of a mapping $f : X \to Y$ is given by

$$\lim_{\lambda \to 0^+} \frac{f(x_0 + \lambda h) - f(x_0)}{\lambda}$$

if the limit exists.

Fortini and Ruggeri (1993) extended the definition of c.f. given by Cifarelli and Regazzini (1987), considering the c.f. between a signed measure and a probability. The extended version of the c.f. allows for the computation of the limit

$$L(x, \Delta) = \lim_{\lambda \to 0} \frac{\phi((\Pi_0 + \lambda \Delta)^*, x) - \phi(\Pi_0, x)}{\lambda},$$

where $\phi(\Pi, \cdot)$ denotes the c.f. of $\Pi$ w.r.t. a baseline measure and $\Delta$ is a signed measure such that $\Delta(\Theta) = 0$ and $||\Delta|| \leq 1$ for a suitable norm $|| \cdot ||$.

This limit coincides with the differential $\psi'_\Delta(\Pi_0, x)$ of the functional $\psi(\Pi, \cdot)$ in $\Pi_0$ in the direction of $\Delta$. The following theorem gives an explicit expression for $\psi'_\Delta(\Pi_0, x)$.

**Theorem 9**

$$\psi'_\Delta(\Pi_0, x) = \begin{cases} \frac{D_\Delta}{D_{\Pi_0}} (\phi(\Delta^*, x) - x) & \text{if } D_\Delta > 0 \\ \frac{D_\Delta}{D_{\Pi_0}} (\phi(\Delta^*, 1 - x) - \phi(\Delta^*, 1 - x) & \text{if } D_\Delta < 0 \\ \frac{D_\Delta}{D_{\Pi_0}} (\phi(\Delta^*, 1) - \phi(\Delta^*, 1) & \text{if } D_\Delta = 0 \end{cases}$$

where $\Delta^*$ is defined by $\Delta^*(B) = \int_B l_\theta \Pi(\theta) \Delta(\theta) d\theta$, for any $B \in \mathcal{F}$ and $D_\Delta = \int_\Theta l_\theta \Pi(\theta) \Delta(\theta) d\theta \neq 0$.

**Proof.** Along the lines of the proof of Theorem 2 in Fortini and Ruggeri (1993), it can be shown that, for any real $\lambda$ such that $\lambda D_\Delta \geq 0$, $\phi((\Pi_0 + \lambda \Delta)^*, x) = \frac{D_{\Pi_0}}{D_{\Pi_0} + \lambda D_\Delta} x + \frac{\lambda D_\Delta}{D_{\Pi_0} + \lambda D_\Delta} \phi(\Delta^*, x)$. Otherwise, $\lambda$ is taken so that $-D_{\Pi_0} < \lambda D_\Delta < 0$, and it follows, from Lemma 1 in Fortini and Ruggeri (1993), that

$$\phi((\Pi_0 + \lambda \Delta)^*, x) = \frac{D_{\Pi_0}}{D_{\Pi_0} + \lambda D_\Delta} x - \frac{\lambda D_\Delta}{D_{\Pi_0} + \lambda D_\Delta} \phi(\Delta^*, 1 - x) - \phi(\Delta^*, 1).$$
Applying the definition of Gâteaux differential, then \( \psi'_{\Delta}(\Pi_0, x) \) is easily computed.

Given \( \varepsilon \in [0, 1] \) and a probability measure \( Q \), the choice \( \Delta = \varepsilon(Q - \Pi_0) \) implies that \( \Pi_0 + \lambda \Delta \) is a contaminated measure for any \( \lambda \in [0, 1] \). In this case, the Gâteaux differential is given by

\[
\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0, x) = \varepsilon \frac{D_Q}{D_{\Pi_0}} (\varphi(Q^*, x) - x).
\]

The previous Gâteaux differential mainly depends on three terms: \( \varepsilon \), \( D_Q / D_{\Pi_0} \), and \( \varphi(Q^*, x) \). Because of their interpretation, they justify the use of the Gâteaux differential to measure the sensitivity of the concentration function to infinitesimal changes in the baseline prior. In fact, the first term measures how contaminated the prior is with respect to the baseline one, while the third says how far the contaminating posterior is from the baseline one. Besides, the second term can be interpreted as the Bayes factor of the contaminating prior with respect to the baseline one. Hence, the Gâteaux differential stresses any possible aspect which might lead to nonrobust situations.

We consider \( \|\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0)\| = \sup_{0 \leq \varepsilon \leq 1} \|\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0, x)\| \) as a concise index of the sensitivity of \( \Pi_0 \) to contaminations with \( Q \).

**Theorem 10** Given \( Q \) and \( \Pi_0 \) as above, it follows that

\[
\|\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0)\| = \frac{D_Q}{D_{\Pi_0}} G_{\Pi_0}(Q^*),
\]

where \( G_{\Pi_0}(Q^*) = \sup_{0 \leq \varepsilon \leq 1} \{ x - \varphi(Q^*, x) \} \) is the Pietra concentration index.

When contaminating \( \Pi_0 \) with the probability measures in a class \( Q \), we can assume

\[
\|\psi(\Pi_0, Q)\| = \sup_{Q \in Q} \|\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0)\|
\]

as a measure of local robustness.

The index (4) can be found analytically in some cases. Let \( \Pi_0 \) be absolutely continuous w.r.t. the Lebesgue measure on \( R \). If the contaminating class is the class \( Q_\alpha \) of all the probability measures over \( \Theta \), then \( \|\psi(\Pi_0, Q_\alpha)\| = \varepsilon \frac{D_{\alpha}}{D_{\Pi_0}} (\varphi(\hat{\theta}|, x) / D_{\Pi_0}) \), where \( \hat{\theta} \in \Theta \) is the maximum likelihood estimator of \( \theta \). Considering the class \( Q_\alpha \) of all probability measures sharing \( m - 1 \) given quantiles, then \( \|\psi(\Pi_0, Q_\alpha)\| = \varepsilon \sum_{i=1}^{m-1} q_i(\hat{\theta}) / D_{\Pi_0} \), where \( \hat{\theta} \in I_i \) is the maximum, for \( \hat{\theta} \), over any interval \( I_i \) of \( \Theta \) determined by the quantiles, while \( q_i \) is the probability of \( I_i \), \( i = 1, \ldots, m \).

### 6.6 Example

Consider the model \( P_\theta \sim \mathcal{N}(\theta, 1) \), the prior \( \Pi_0 \sim \mathcal{N}(0, 2) \) and an \( \epsilon \)-contamination class of probability measures around \( \Pi_0 \). Let \( \epsilon = 0.1 \). This example has been widely used in Bayesian robustness by Berger and Berliner (1986) and many other authors since then.

#### 6.6.1 Global sensitivity

Consider \( \Pi_0 \) contaminated either by the class \( Q_{1/2} \) of probability measures which have the same median as \( \Pi_0 \) or by \( Q_\alpha \), the class of the probability measures which are either \( \delta_k \sim \mathcal{N}(\theta_0 - k, \theta_0 + k), \) \( k > 0 \), or \( \delta_{\infty} \) the Dirac measure at \( \theta_0 \). Observe the sample \( s = 1 \). In the former case, the lowest c.f. is \( \varphi_0 \equiv 0 \), given by the contamination \( (\delta_0 + \delta_k)/2 \), whereas, in the latter case, the lowest c.f. is \( \varphi(x) = 0.879x \). Should we decide to compare \( \varphi(x) \) with, say, the function \( g(x) = x^2 \), 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 led to a different situation.

#### 6.6.2 Local sensitivity

Consider \( \Pi_0 \) to be contaminated either by the class \( Q_\alpha \) or by \( Q_{1/2} \). It can be easily shown that \( \|\psi\| \) is achieved for a Dirac prior concentrated at the sample \( s \) in the former case and for a two-point mass prior, which gives equal probability to 0 and \( s \), in the latter. The values of \( \|\psi(\Pi_0, Q_\alpha)\| \) and \( \|\psi(\Pi_0, Q_{1/2})\| \) are shown in Table 2, for different samples \( s \). As expected, the class \( Q_{1/2} \) induces smaller changes than \( Q_\alpha \). It is worth noting that the difference is negligible for small values of \( s \) while it increases when observing larger values of \( s \) (in absolute value). The finding is coherent with Table 1 in Betrò et al. (1994). While their table was obtained by numerical solution of a constrained nonlinear optimisation problem, here the use of the Gâteaux differential requires just a simple analytical computation.

This example shows that sometimes local sensitivity analysis can give information on the global problem as well, but favoured by simpler computations.

Notice that

\[
\|\psi(\Pi_0, Q_\alpha)\| = \varepsilon \frac{D_{\alpha}}{D_{\Pi_0}} \sup_{Q \in Q_\alpha} \varepsilon \frac{D_Q}{D_{\Pi_0}},
\]

so that the Pietra index does not seem to have an important part in (4). The same happens when \( Q_{1/2} \) is considered. As shown in the following example, there are contaminating classes for which (5) does not hold. Consider, for example, the class \( Q_N = \{ \mathcal{N}(0, \tau^2) \mid 1 \leq \tau \leq 2 \} \). The values of
sup_{Q \in Q_n} \varepsilon D_Q / D_{D_n}$, are shown in Table 2 for different samples $s$'s. They are quite large, especially if compared with those of $\|\psi'(\Pi_0, Q_n)\|$.

<table>
<thead>
<tr>
<th>$s$</th>
<th>$|\psi'(\Pi_0, Q_n)|$</th>
<th>$|\psi'(\Pi_0, Q_1/2)|$</th>
<th>$|\psi'(\Pi_0, Q_n)|$</th>
<th>$\sup_{Q \in Q_n} \varepsilon D_Q / D_{D_n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ex</td>
<td>0.5</td>
<td>0.1805</td>
<td>0.1660</td>
<td>0.0127</td>
</tr>
<tr>
<td>1</td>
<td>0.2046</td>
<td>0.1399</td>
<td>0.0170</td>
<td>0.1126</td>
</tr>
<tr>
<td>1.5</td>
<td>0.2820</td>
<td>0.1392</td>
<td>0.0202</td>
<td>0.1019</td>
</tr>
<tr>
<td>2</td>
<td>0.3373</td>
<td>0.1717</td>
<td>0.0218</td>
<td>0.1023</td>
</tr>
<tr>
<td>2.5</td>
<td>0.4098</td>
<td>0.2458</td>
<td>0.0225</td>
<td>0.1174</td>
</tr>
<tr>
<td>3</td>
<td>0.7762</td>
<td>0.3881</td>
<td>0.0317</td>
<td>0.1411</td>
</tr>
<tr>
<td>3.5</td>
<td>1.3342</td>
<td>0.6671</td>
<td>0.0453</td>
<td>0.1752</td>
</tr>
<tr>
<td>4</td>
<td>2.4927</td>
<td>1.2463</td>
<td>0.0656</td>
<td>0.2250</td>
</tr>
</tbody>
</table>

**TABLE 2.** Gâteaux differentials

### 6.7 Connections with Choquet capacities

We conclude the paper by observing that $g$-neighbourhoods can be considered as an example of 2-alternating Choquet capacities. We refer to Wasserman and Kadane (1990, 1992) for a thorough description of the properties of the latter, their application in Bayesian robustness and their links with other notions, like special capacities (see, for example, Bednarski, 1981, and Buja, 1986) and upper and lower probabilities in Walley's (1991) approach. Details on capacities can be found in Choquet (1955) and Huber and Strassen (1973).

Let $Q$ be a nonempty set of prior probability measures on $(\Theta, \mathcal{F})$. We define upper and lower prior probability functions by

$$\Pi(A) = \sup_{\Pi \in Q} \Pi(A) \quad \text{and} \quad \Pi(A) = \inf_{\Pi \in Q} \Pi(A),$$

for any $A \in \mathcal{F}$.

The set $Q$ is said to be **2-alternating** if

$$\Pi(A \cup B) \leq \Pi(A) + \Pi(B) - \Pi(A \cap B),$$

for any $A, B$ in $\mathcal{F}$. The set $Q$ is said to generate a **Choquet capacity** if $\Pi(C_n) \downarrow \Pi(C)$ for any sequence of closed sets $C_n \downarrow C$. It can be shown that $Q$ generates a Choquet capacity if and only if the set $C = \{ \Pi : \Pi(A) \leq \Pi(Q), \forall A \in \mathcal{F} \}$ is weakly compact. We say that $Q$ is $m$-closed (or closed w.r.t. majorisation) if $C \subseteq Q$.

Consider a $g$-neighbourhood $K_g$ around a nonatomic probability measure $\Pi_0$. Let $g$ be such that $g(1) = 1$. For any set $A \in \mathcal{F}$, we show that

$$\Pi(A) = 1 - g(1 - \Pi_0(A)) \quad \text{and} \quad \Pi(A) = g(\Pi_0(A)).$$

From the properties of $g$-neighbourhoods and the definition of upper and lower probability functions, it follows that

$$g(\Pi_0(A)) \leq \Pi(A) \leq \Pi_0(A) \leq 1 - g(1 - \Pi_0(A)),$$

for any $\Pi \in Q$. We show that both upper and lower bounds are actually achieved for probability measures in $Q$, so that they coincide with upper and lower probability functions, respectively.

Consider

$$x_A(\cdot) = \frac{g(\Pi_0(A))}{\Pi_0(A)} x_0(\cdot) I_A(\cdot) + \frac{1 - g(\Pi_0(A))}{1 - \Pi_0(A)} x_0(\cdot) I_A^c(\cdot),$$

where $x_A$ and $x_0$ are the densities, w.r.t. some dominating measure, of the probability measures $\Pi_A$ and $\Pi_0$, respectively.

Note that $\Pi_0$ differs from $\Pi_0$ because of two different multiplicative factors (the one on $A$ is smaller than 1, whereas the one on $A^c$ is bigger than 1). The c.f. of $\Pi_0$ w.r.t. $\Pi_0$ is made of two segments joining on the curve $g$ at $(\Pi_0(A), g(\Pi_0(A)))$. Because of the convexity of $g$, the c.f. is above $g$, so that $\Pi_0 \subseteq K_g$. Besides, $\Pi(A) = g(\Pi_0(A)) = \Pi(A)$ follows from the construction of $\Pi_A$. Since the upper probability function is obtained in a similar way, we have proved that upper and lower probability functions in $K_g$ can be expressed by means of $g$ (i.e., respectively, as $1 - g(1 - x)$ and $g(x)$, $x \in [0, 1]$).

Fortini and Ruggeri (1995a) proved that the set $K_g$ is compact in the weak topology; therefore, its definition (see (3)) implies that it generates a Choquet capacity, besides being $m$-closed.

Using the equation $\Pi(A) = 1 - \Pi(A^c)$, the 2-alternating property can be rewritten as

$$\Pi(A \cup B) \geq \Pi(A) + \Pi(B) - \Pi(A \cap B),$$

for any $A, B$ in $\mathcal{F}$.

In our case, the property becomes

$$g(\Pi_0(A) + \Pi_0(B) - \Pi_0(A \cap B)) \geq g(\Pi_0(A)) + g(\Pi_0(B)) - g(\Pi_0(A \cap B)),$$

which is satisfied because the convexity of $g$ implies

$$g(x_1 + x_2 - x_3) + g(x_3) \geq \frac{g(x_1) + g(x_2)}{2},$$

for $x_3 \leq x_1 \leq x_2$; note that $x_2 \leq x_1 + x_2 - x_3$.

Therefore, $K_g$ is an $m$-closed, 2-alternating Choquet capacity, and results in Wasserman and Kadane (1990) apply to it.
References


CAROTA, C. and RUGGERI, F. (1994). Robust Bayesian analysis given priors on partition sets. TEST, 3, 73–86.


