6

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 Π and Π_0, either directly by looking at the range spanned by the probability, under Π, of *all* the subsets with a given probability under Π_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 $x_{(1)}, \ldots, x_{(n)}$. Now plot the points $(k/n, S_k/S_n), k = 0, \ldots, n$, where $S_0 = 0$ and $S_k = \Sigma_{i=1}^k x_{(i)}$ 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 $(1,1)$.... 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 Π (the "wealth") and a uniform one (the "individuals"), say Π_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.

	Native Americans	Asians	Hispanics	Blacks	Whites
Durham (82)	0.002	0.012	0.002	0.332	0.652
School 332	0.011	0.043	0.022	0.403	0.521
Ratio (S/D)	5.500	3.580	11.000	1.210	0.799
Durham (92)	0.002	0.026	0.009	0.345	0.618
School 332	0.007	0.106	0.013	0.344	0.530
Ratio (S/D)	3.500	4.070	1.440	0.997	0.858

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 $\left(\Sigma_{j=1}^i D^{(j)}, \Sigma_{j=1}^i S^{(j)}\right), i = 1, \ldots, 5$, where $D^{(j)}$ and $S^{(j)}$ correspond to the race with the jth ratio (in ascending order). The distance between the

straight line and the other two denotes an unequal 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).

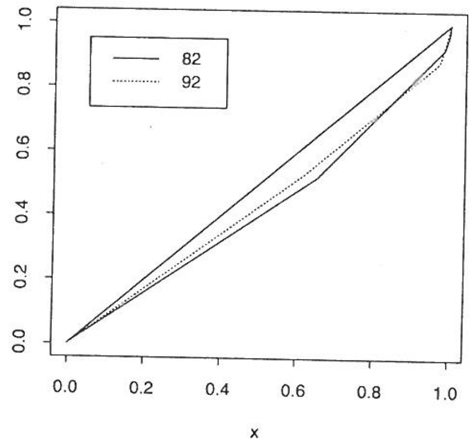

FIGURE 1. Lorenz curve for races at school 332 vs. Durham schools

As an extension of the Lorenz curve, Cifarelli and Regazzini (1987) defined and studied the c.f. of Π w.r.t. Π_0, where Π and Π_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 Θ and a nonnegative function h on N^C such that

$$\Pi(E) = \Pi_a(E \cap N^C) + \Pi_s(E \cap N), \ \forall E \in \mathcal{F},$$

where Π_a and Π_s are, respectively, the absolutely continuous and the singular part of Π w.r.t. Π_0, that is, such that

$$\Pi_a(E \cap N^C) = \int_{E \cap N^C} h(\theta)\Pi_0(d\theta), \Pi_0(N) = 0, \Pi_s(N) = \Pi_s(\Theta).$$

Set $h(\theta) = \infty$ all over N and define $H(y) = \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\})$, $c_x = \inf\{y \in \Re : H(y) \geq x\}$ and $c_x^- = \lim_{t \to x^-} c_t$. Finally, let $L_x = \{\theta \in \Theta : h(\theta) \leq c_x\}$ and $L_x^- = \{\theta \in \Theta : h(\theta) < c_x\}$.

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

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

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

It is worth mentioning that $\varphi_\Pi(1) = 1$ implies that Π is absolutely continuous w.r.t. Π_0 while $\varphi_\Pi(x) = 0, 0 \le x \le \alpha$, means that Π 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 $\mathcal{C}(0,1)$. The Radon–Nikodym derivative $h(\theta)$ is plotted in Fig. 2, a horizontal line is drawn and the subset of Θ 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.

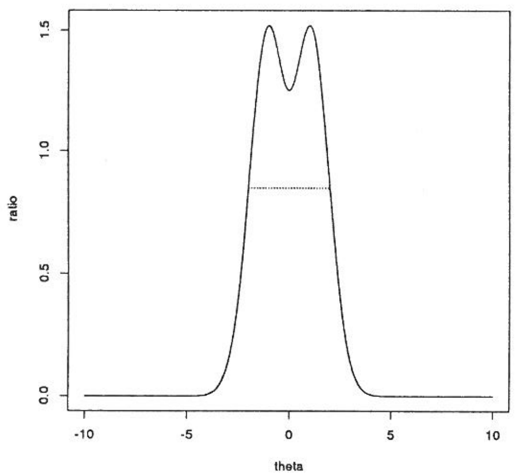

FIGURE 2. Radon–Nikodym derivative of $\mathcal{N}(0,1)$ vs. $\mathcal{C}(0,1)$

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 $\Re^+$ are, respectively, $\pi(\theta) = \theta e^{-\theta}$ and $\pi_0(\theta) = e^{-\theta}$, so that $h(\theta) = \theta$, $\theta \ge 0$. For any $x \in [0,1]$, we compute the c.f. by finding the value y such that $x = \Pi_0(\{\theta \in \Theta : h(\theta) \le y\})$.

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

$$\begin{aligned} \varphi_\Pi(x) &= \Pi(\{\theta \in \Theta : h(\theta) \le -\log(1-x)\}) \\ &= \int_0^{-\log(1-x)} \theta e^{-\theta} d\theta \\ &= 1 - (1-x)(1 - \log(1-x)). \end{aligned}$$

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 $G_{\Pi_0}(\Pi) = \sup_{x \in [0,1]}(x - \varphi_\Pi(x))$. The latter coincides with the total variation distance between Π and Π_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 Π on the measurable subsets of Θ with Π_0-measure not smaller than x.

Theorem 1 *If* $A \in \mathcal{F}$ *and* $\Pi_0(A) = x$, *then* $\varphi_\Pi(x) \le \Pi_a(A)$. *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*

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

If Π_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 Π, of all the subsets A with measure x under Π_0, satisfies

$$\varphi_\Pi(x) \le \Pi(A) \le 1 - \varphi_\Pi(1-x). \tag{3}$$

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 Π, of the sets A with $\Pi_0(A) = 0.4$ (see Fig3).

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 $\Re$: their means are very close, their variances might be the same, but their c.f. is 0 in $[0,1)$.

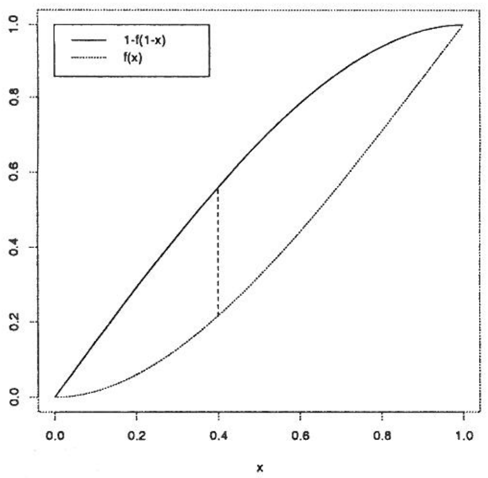

FIGURE 3. Range of $\Pi(A)$ spanned by A s.t. $\Pi_0(A) = .4$ ($\mathcal{G}(2,2)$ vs. $\mathcal{E}(1)$)

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 Π with functional forms close to a nonatomic baseline measure Π_0. In particular, they defined neighbourhoods of Π_0 by imposing constraints on the probability of all measurable subsets, such as requiring $|\Pi_0(A) - \Pi(A)| \le \Pi_0(A)(1 - \Pi_0(A))$, for any $A \in \mathcal{F}$. By observing that the above relation can be written $\Pi(A) \ge g(\Pi_0(A))$, with $g(x) = x^2$, Fortini and Ruggeri gave the following definitions.

Definition 2 *A function $g : [0,1] \to [0,1]$ 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*

$$K_g \doteq \{\Pi : \Pi(A) \ge g(\Pi_0(A)), \quad \forall A \in \mathcal{F}\}$$

will be a g-neighbourhood of Π_0.

Observe that, if $\Pi \in K_g$, then $g(\Pi_0(A)) \le \Pi(A) \le 1 - g(1 - \Pi_0(A))$, for any $A \in \mathcal{F}$. The requirement $g(0) = 0$ is needed to avoid $\Pi(\Theta) \le 1 - g(0) < 1$, 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 Π_0, when g belongs to an adequate class G of compatible functions.

The definition of a g-neighbourhood of Π_0 can be reformulated by means of the c.f. w.r.t. Π_0, as stated in the following,

Theorem 2 *The set $K_g = \{\Pi : \varphi_\Pi(x) \ge g(x), \forall x \in [0,1]\}$ is a g-neighbourhood of Π_0.*

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

Theorem 3 *Given a function $g : [0,1] \to [0,1]$, there exists at least one measure Π such that g is the c.f. of Π w.r.t. Π_0 if and only if g is compatible.*

6.3.1 Main results

Consider the space $\mathcal{P}$ of all probability measures on Θ 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

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

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

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

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

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

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

Theorem 7 *Let $H_f(y) = \Pi_0(\{\theta \in \Theta : f(\theta) \le y\})$, $c_f(x) = \inf\{y : H_f(y) \ge x\}$. Then $\sup_{\Pi \in K_g} \int_\Theta f(\theta)\Pi(d\theta) = \int_0^1 c_f(x) g'(x) dx$.*

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

$$\sup_{\Pi \in K_g} \Pi(A|x) = \left\{ 1 + \frac{-\displaystyle\int_0^{\Pi_0(A^C)} c_{-l_1}(x)c(x)dx}{\displaystyle\int_{\Pi_0(A^C)}^1 c_{l_2}(x)c(x)dx} \right\}^{-1},$$

where I_A is the indicator function of the subset A, $l_x(\theta)$ is the likelihood function and, for any $\theta \in \Theta$, $l_1(\theta) = l_x(\theta)I_{A^C}(\theta)$ and $l_2(\theta) = l_x(\theta)I_A(\theta)$.

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 Π_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 Π_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 cases of $\varepsilon-$contamination and total variation neighbourhoods.

ε-CONTAMINATIONS. The ε-contamination class $\Gamma_\varepsilon = \{\Pi_Q = (1-\varepsilon)\Pi_0 + \varepsilon\Pi, \Pi \in \mathcal{P}\}$ is defined by $g(x) = (1-\varepsilon)x$, $\forall x \in [0,1]$. The sets E_g and F_g are obtained, respectively, considering singular (w.r.t. Π_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 Π_0 and $k > 0$, consider the class

$$\Gamma_k^B = \{\Pi : (1/k)\Pi_0(A) \le \Pi(A) \le k\Pi_0(A), \forall A \in \mathcal{F}\},$$

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 $g(x) = \max\{x/k, k(x-1)+1\}$.

DENSITY RATIO CLASS. Density ratio classes were defined by DeRobertis and Hartigan (1981); Ruggeri and Wasserman (1995) considered Γ_k^{DR}, the density ratio neighbourhood around Π_0 (with density $\pi_0(\theta)$), given by all the probability measures whose densities $\pi(\theta)$ are such that there exists $c > 0$ so that $\pi_0(\theta) \le c\pi(\theta) \le k\pi_0(\theta)$ for almost all θ. It can be shown that Γ_k^{DR} is the class of the probability measures such that their c.f.s w.r.t. Π_0 are inside any triangle with vertices $(0,0)$, $(1,1)$ and a point on the curve $g(x) = x/(k-(k-1)x)$, $0 < x < 1$.

TOTAL VARIATION NEIGHBOURHOOD. A class Γ^T is said to be a total variation neighbourhood of a probability measure Π_0 if it contains all the probability measures Π that satisfy $\sup_{A\in\mathcal{F}} |\Pi(A) - \Pi_0(A)| \le \varepsilon$, given a fixed $\varepsilon \in [0,1]$. Assuming Π_0 nonatomic, then the class is a c.f. neighbourhood K_g with $g(x) = \max\{0, x-\varepsilon\}$.

OTHER NEIGHBOURHOODS. As discussed earlier, we can consider many neighbourhoods, like the class of all the probability measures Π satisfying

$$|\Pi_0(A) - \Pi(A)| \le \Pi_0(A)(1-\Pi_0(A)), \forall A \in \mathcal{F}.$$

In this case, the neighbourhood K_g is given by $g(x) = x^2$.

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 $\varepsilon-$contaminations and compared the c.f.s of the posterior probability measures w.r.t. a base posterior measure Π_0^*. In computing, pointwise, the infimum $\hat{\varphi_\Pi}(x)$ of the c.f., their interest was twofold: providing a measure of the distance between the distributions in the class and Π_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 Γ of probability measures Π and a base prior Π_0, as in the $\varepsilon-$contamination class given by

$$\Gamma_\varepsilon = \{\Pi_Q = (1-\varepsilon)\Pi_0 + \varepsilon Q, Q \in \mathcal{Q}\},$$

where $\mathcal{Q} \subseteq \mathcal{P}$ and $0 \le \varepsilon \le 1$.

Let Π^* denote the posterior measure corresponding to the prior Π. Consider the class

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

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

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

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

The interpretation of $\widehat{\varphi}$, in terms of Bayesian robustness, is straightforward: the closest $\widehat{\varphi}(x)$ and $1 - \widehat{\varphi}(1-x)$ 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 $\widehat{\varphi}(x)$ 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 ε-contaminations, the following:

Theorem 8 *If φ and φ_0 denote the c.f.'s of Π_Q^* and Q^* w.r.t. Π_0^*, respectively, then it follows that*

$$\varphi(x) = \lambda_Q x + (1 - \lambda_Q)\varphi_0(x),$$

where

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

with $D_0 = \int_\Theta l_x(\theta)\Pi_0(d\theta)$ and $D_Q = \int_\Theta l_x(\theta)Q(d\theta)$.

They were able to find $\varepsilon-$contaminations of a nonatomic prior Π_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 $\sup_Q D_Q \leq l(\theta_0)$ 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 θ, 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 $\widehat{\varphi}(x) \geq g(x)$ 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

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

where $\varphi(\Pi, \cdot)$ denotes the c.f. of Π w.r.t. a baseline measure and Δ 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 $\psi(\Pi) = \varphi(\Pi^*(\Pi), x)$ in Π_0 in the direction of Δ. The following theorem gives an explicit expression for $\psi'_\Delta(\Pi_0, x)$.

Theorem 9

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

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

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

$$\varphi((\Pi_0 + \lambda\Delta)^*, x) = \frac{D_{\Pi_0}}{D_{\Pi_0} + \lambda D_\Delta} x - \frac{\lambda D_\Delta}{D_{\Pi_0} + \lambda D_\Delta}[\varphi(\Delta^*, 1-x) - \varphi(\Delta^*, 1)].$$

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: ε, D_Q/D_{Π_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 $||\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0)|| = \sup_{0\leq x\leq 1} | \psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0, x) |$ as a concise index of the sensitivity of Π_0 to contaminations with Q.

Theorem 10 *Given Q and Π_0 as above, it follows that*

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

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

When contaminating Π_0 with the probability measures in a class $\mathcal{Q}$, we can assume

$$||\psi'(\Pi_0, \mathcal{Q})|| = \sup_{Q\in\mathcal{Q}} ||\psi'_{\varepsilon(Q-\Pi_0)}(\Pi_0)|| \tag{4}$$

as a measure of local robustness.

The index (4) can be found analytically in some cases. Let Π_0 be absolutely continuous w.r.t. the Lebesgue measure on $\Re$. If the contaminating class is the class $\mathcal{Q}_a$ of all the probability measures over Θ, then $||\psi'(\Pi_0, \mathcal{Q}_a)|| = \varepsilon(l_x(\hat\theta)/D_{\Pi_0}$, where $\hat\theta \in \Theta$ is the maximum likelihood estimator of θ. Considering the class $\mathcal{Q}_q$ of all probability measures sharing $m-1$ given quantiles, then $||\psi'(\Pi_0, \mathcal{Q}_q)|| = \varepsilon\sum_{i=1}^m q_i l_x(\hat\theta_i)/D_{\Pi_0}$, where $\hat\theta_i \in I_i$ is the maximum, for $l_x(\theta)$, over any interval I_i of the partition $\{I_i\}$ of Θ 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 ε–contamination class of probability measures around Π_0. Let $\varepsilon = 0.1$. 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 Π_0 contaminated either by the class $\mathcal{Q}_{1/2}$ of probability measures which have the same median as Π_0 or by $\mathcal{Q}$, the class of the probability measures which are either $Q_k \sim \mathcal{U}(\theta_0 - k, \theta_0 + k)$, $k > 0$, or $Q_\infty \equiv \delta_{\theta_0}$, the Dirac measure at θ_0. Observe the sample $s = 1$. In the former case, the lowest c.f. is $\varphi \equiv 0$, given by the contamination $(\delta_0 + \delta_1)/2$, whereas, in the latter case, the lowest c.f. is $\hat\varphi(x) = 0.879x$. Should we decide to compare $\hat\varphi(x)$ with, say, the function $g(x) = x^2$, it is evident that $\hat\varphi(x) < g(x)$ 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 Π_0 to be contaminated either by the class $\mathcal{Q}_a$ or by $\mathcal{Q}_{1/2}$. 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 $||\psi'(\Pi_0, \mathcal{Q}_a)||$ and $||\psi'(\Pi_0, \mathcal{Q}_{1/2})||$ are shown in Table 2, for different samples s's. As expected, the class $\mathcal{Q}_{1/2}$ induces smaller changes than $\mathcal{Q}_a$. 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

$$|\psi'(\Pi_0, \mathcal{Q}_a)| = \varepsilon\frac{D_{\delta_s}}{D_{\Pi_0}} = \sup_{Q\in\mathcal{Q}_a}\varepsilon\frac{D_Q}{D_{\Pi_0}}, \tag{5}$$

so that the Pietra index does not seem to have an important part in (4). The same happens when $\mathcal{Q}_{1/2}$ is considered. As shown in the following example, there are contaminating classes for which (5) does not hold. Consider, for example, the class $\mathcal{Q}_N = \{\mathrm{N}(0, \tau^2) : 1 \leq \tau \leq 2\}$. The values of

$\sup_{Q\in\mathcal{Q}_N} \varepsilon D_Q/D_{\Pi_0}$, are shown in Table 2 for different samples s's. They are quite large, especially if compared with those of $||\psi'(\Pi_0, \mathcal{Q}_N)||$.

s	$\|\|\psi'(\Pi_0, \mathcal{Q}_a)\|\|$	$\|\|\psi'(\Pi_0, \mathcal{Q}_{1/2})\|\|$	$\|\|\psi'(\Pi_0, \mathcal{Q}_N)\|\|$	$\sup_{Q\in\mathcal{Q}_N} \varepsilon \frac{D_Q}{D_{\Pi_0}}$
[1ex] 0.5	0.1805	0.1606	0.0127	0.1199
1.0	0.2046	0.1399	0.0170	0.1126
1.5	0.2520	0.1392	0.0202	0.1019
2.0	0.3373	0.1717	0.0218	0.1023
2.5	0.4908	0.2458	0.0225	0.1174
3.0	0.7762	0.3881	0.0317	0.1411
3.5	1.3342	0.6671	0.0453	0.1752
4.0	2.4927	1.2463	0.0656	0.2250

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

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

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

The set $\mathcal{Q}$ is said to be *2-alternating* if

$$\overline{\Pi}(A\cup B) \le \overline{\Pi}(A) + \overline{\Pi}(B) - \overline{\Pi}(A\cap B),$$

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

Consider a $g-$neighbourhood K_g around a nonatomic probability measure Π_0. Let g be such that $g(1) = 1$. For any set $A \in \mathcal{F}$, we show that

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

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

$$g(\Pi_0(A)) \le \underline{\Pi}(A) \le \Pi(A) \le \overline{\Pi}(A) \le 1 - g(1 - \Pi_0(A)),$$

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

Consider

$$\pi_A(\cdot) = \frac{g(\Pi_0(A))}{\Pi_0(A)} \pi_0(\cdot) I_A(\cdot) + \frac{1 - g(\Pi_0(A))}{1 - \Pi_0(A)} \pi_0(\cdot) I_{A^C}(\cdot),$$

where π_A and π_0 are the densities, w.r.t. some dominating measure, of the probability measures Π_A and Π_0, respectively.

Note that Π_A differs from Π_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 Π_A w.r.t. Π_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 $\Pi_A \in K_g$. Besides, $\Pi_A(A) = g(\Pi_0(A)) = \underline{\Pi}(A)$ follows from the construction of Π_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 $\overline{\Pi}(A) = 1 - \underline{\Pi}(A^c)$, the 2-alternating property can be rewritten as

$$\underline{\Pi}(A\cup B) \ge \underline{\Pi}(A) + \underline{\Pi}(B) - \underline{\Pi}(A\cap B),$$

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

In our case, the property becomes

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

which is satisfied because the convexity of g implies

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

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

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

References

BEDNARSKI, T. (1981). On solutions of minimax test problems for special capacities. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **58**, 397–405.

BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd edition). New York: Springer-Verlag.

BERGER, J.O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and 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. *Annals of Statistics*, **14**, 461–486.

BETRÒ, B., MĘCZARSKI, M. and RUGGERI, F. (1994). Robust Bayesian analysis under generalized moments conditions. *Journal of Statistical Planning and Inference*, **41**, 257–266.

BORATYNSKA, A. (1996). On Bayesian robustness with the ε–contamination class of priors. *Statistics & Probability Letters*, **26**, 323–328.

BUJA, A. (1986). On the Huber-Strassen theorem. *Probability Theory and Related Fields*, **73**, 149–152.

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

CHOQUET, G. (1955). Theory of capacities. *Annales de l'Institute Fourier*, **5**, 131–295.

CIFARELLI, D.M. and REGAZZINI, E. (1987). On a general definition of concentration function. *Sankhyā, B*, **49**, 307–319.

DEROBERTIS, L. and HARTIGAN, J. (1981). Bayesian inference using intervals of measures. *Annals of Statistics*, **9**, 235–244.

DI BONA, J., PARMIGIANI, G. and RUGGERI, F. (1993). Are we coming together or coming apart? Racial segregation in North Carolina schools 1982–1992. *Technical Report*, **93.6**, CNR-IAMI.

FORTINI, S. and RUGGERI, F. (1993). Concentration function and coefficients of divergence for signed measures. *Journal of the Italian Statistical Society*, **2**, 17–34.

FORTINI, S. and RUGGERI, F. (1994). Concentration function and Bayesian robustness. *Journal of Statistical Planning and 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). Concentration function and sensitivity to the prior. *Journal of the Italian Statistical Society*, **4**, 283–297.

FORTINI, S. and RUGGERI, F. (1997). Differential properties of the concentration function, *Sankhyā, A*, **59**, 345–364.

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

GUSTAFSON, P. (2000). Local robustness in Bayesian analysis. In *Robust Bayesian Analysis* (D. Ríos Insua and F. Ruggeri, eds.). New York: Springer-Verlag.

HUBER, P.J. and STRASSEN, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Annals of Statistics*, **1**, 251–263.

LAVINE, M. (1988). Prior influence in Bayesian statistics. *Technical Report*, **88–06**, ISDS, Duke University.

LAVINE, M. (1991). Sensitivity in Bayesian statistics: the prior and the likelihood. *Journal of the American Statistical Association*, **86**, 396–399.

LAVINE, M., PERONE PACIFICO, M., SALINETTI, G. and TARDELLA, G. (2000). Linearization techniques in Bayesian robustness. In *Robust Bayesian Analysis* (D. Ríos Insua and F. Ruggeri, eds.). New York: Springer-Verlag.

MARSHALL, A.W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press.

MORENO, E. (2000). Global Bayesian robustness for some classes of prior distributions. In *Robust Bayesian Analysis* (D. Ríos Insua and F. Ruggeri, eds.). New York: Springer-Verlag.

PHELPS, R.R. (1966). *Lectures on Choquet's Theorem.* Princeton: Van Nostrand Company.

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

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

RUGGERI, F. (1994). Local and global sensitivity under some classes of priors. In *Recent Advances in Statistics and Probability* (J.P. Vilaplana and M.L. Puri, eds). Ah Zeist: VSP.

RUGGERI, F. and WASSERMAN, L.A. (1991). Density based classes of priors: infinitesimal properties and approximations. *Technical Report*, **91.4**, CNR-IAMI.

RUGGERI, F. and WASSERMAN, L.A. (1995). Density based classes of priors: infinitesimal properties and approximations. *Journal of Statistical Planning and Inference*, **46**, 311–324.

WALLEY, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* London: Chapman Hall.

WASSERMAN, L.A. (1992). Recent methodological advances in robust Bayesian inference. *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.). Oxford: Oxford University Press.

WASSERMAN, L.A. and KADANE, J. (1990). Bayes' theorem for Choquet capacities. *Annals of Statistics*, **18**, 1328–1339.

WASSERMAN, L.A. and KADANE, J. (1992). Symmetric upper probabilities. *Annals of Statistics*, **20**, 1720–1736.