Sankhya: The Indian Journal of Statistics
1994, Volume 56, Series A, Pt. 3, pp. 444-457

ON DEFINING NEIGHBOURHOODS OF MEASURES THROUGH THE CONCENTRATION FUNCTION

By SANDRA FORTINI and FABRIZIO RUGGERI
CNR-IAMI

SUMMARY. Statistical procedures are often interested in comparing probability measures by means of distances defined over the space $\mathcal{P}$ of all probability measures, endowed with some classical topology, like the variational or the Prohorov ones. Other topologies can be obtained by means of the concentration function, which extends the notion of Lorenz curve. Hence, neighbourhood classes Γ of probability measures, including well-known ones, are defined and a representation theorem is proved. Finally, ranges of functionals over Γ are found, restricting the search among the extremal measures in Γ.

1. Comparison of probability measures

Many statistical problems require the specification of topologies or distances on the space $\mathcal{P}$ of all probability measures, e.g. to study similarities among populations or to consider neighbourhoods of a given probability measure. Suppose that we are interested in comparing the functional forms of two probability measures, say P and P_0, on the same measurable space $(\Theta, \mathcal{F})$, Θ being a Polish space and $\mathcal{F}$ its Borel σ-field. We could use the variational distance $d_V(P, P_0) = \sup_{A \in \mathcal{F}} |P(A) - P_0(A)|$ or the Prohorov distance $d_P(P, P_0) = \inf\{\varepsilon > 0 : P(A) \leq P_0(A^\varepsilon) + \varepsilon \ \forall \ A \in \mathcal{F}\}$, where $A^\varepsilon = \{\theta \in \Theta : d(\theta, A) \leq \varepsilon\}$ and d is a metric on Θ.

However, such rules do not seem sufficiently sensitive on the sets with small probability under P_0. For example, if the variational metric is considered, then a ε-neighbourhood of P_0 contains all the probability measures P such that, for any $A \in \mathcal{F}, |P(A) - P_0(A)| \leq \varepsilon$. Consider now a set E such that $P_0(E) = \varepsilon/10$. Given P in the ε-neighbourhood of P_0, it follows that $P(E) \leq 11\varepsilon/10$ is the only restriction about P on E; i.e. P is considered close to P_0 even if its value on E is eleven times greater than $P_0(E)$. A similar reasoning holds for the ε-contamination class of probability measures, described in Huber (1981), which contains all the probability measures P such that, for any $A \in \mathcal{F}, (1-\varepsilon)P_0(A) \leq P(A) \leq (1-\varepsilon)P_0(A)+\varepsilon$. When such a consequence is deemed inconvenient, then different bounds on $P(A)$ could be

Paper received. September 1992.

AMS (1980) *subject classification.* Primary 62E10; secondary 60E05.

Key words and phrases. Neighbourhoods of probability measures; concentration function; mixtures of probability measures; extremal probability measures.

considered and the concentration function (c.f.) is a flexible tool to get them. As an example, require, for any $A \in \mathcal{F}$, either $|P_0(A) - P(A)| \leq \varepsilon \min\{P_0(A), 1 - P_0(A)\}$ or $|P_0(A) - P(A)| \leq P_0(A)(1 - P_0(A))$, so that more stringent bounds are found on $P(E)$; in the former case, we have $\frac{\varepsilon}{10}(1 - \varepsilon) \leq P(E) \leq \frac{\varepsilon}{10}(1 + \varepsilon)$, while the latter implies $\frac{\varepsilon^2}{100} \leq P(E) \leq \frac{\varepsilon}{10}(2 - \frac{\varepsilon}{10})$, i.e. $P(E)$ does not exceed twice $P_0(E)$.

In this paper, we develop a method which enables us to define neighbourhoods of a probability measure, specifying bounds on the probability of any measurable subset $A \in \mathcal{F}$. Such a method is essentially based on the concentration function defined by Cifarelli and Regazzini (1987) as an extension of the classical notion of the Lorenz-Gini curve. A g-neighbourhood of a probability measure P_0, defined in Section 2, is made of all the probability measures whose concentration function with respect to P_0 lies above a specified continuous, convex, monotone nondecreasing function g. In Section 3, it will be shown that g-neighbourhoods determine a topology over $\mathcal{P}$, while a representation theorem will be proved in Section 4. Computations of upper and lower bounds on functionals over g-neighbourhoods are simplified by the results in Section 5, while some final remarks are presented in Section 6.

2. Definition of g-neighbourhoods

In this section we consider classes K_g of probability measures which can be defined through the c.f.'s, as neighbourhoods around a base measure P_0.

Definition 1. If $g : [0,1] \to [0,1]$ is a continuous, convex, monotone nondecreasing function such that $g(0) = 0$, then the set

$$K_g = \{P \in \mathcal{P} : P(A) \geq g(P_0(A)) \; \forall \; A \in \mathcal{F}\} \qquad \dots (1)$$

will be said a g-neighbourhood of P_0.

Observe that, if $P \in K_g$, then $g(P_0(A)) \leq P(A) \leq 1 - g(1 - P_0(A))$.

We give now the reasons for g to be continuous, monotone nondecreasing and convex. The requirement $g(0) = 0$ is obvious to get $P(\emptyset) = 0$.

Monotonicity. Let $g(x)$ belong to the range of a measure $P \in K_g$ and let $A \in \mathcal{F}$ be such that $P_0(A) = x$ and $P(A) = g(x)$. If $B \subset A$, then $P(B) \leq P(A)$. Hence we choose a monotone nondecreasing function g.

Continuity. Let $g(x)$ belong to the range of a measure $P \in K_g$ and let $A \in \mathcal{F}$ be such that $P_0(A) = x$ and $P(A) = g(x)$. Since P is a regular measure, it follows that $P(A) = \inf_{\{G \text{ open}: A \subset G\}} P(G)$. Therefore, for any $\varepsilon > 0$ there exists an open set G such that $P(G) < P(A) + \varepsilon \leq g(x) + \varepsilon$. Since $P(G) \geq g(P_0(G))$, g must be right-continuous in order to be meaningful. Analogously, left-continuity is required.

Convexity. Let $g(x_1)$ and $g(x_2)$ belong to the range of a measure $P \in K_g$ and suppose there exist $A_1, A_2 \in \mathcal{F}$ such that $A_1 \subset A_2, P_0(A_i) = x_i$ and $P(A_i) =$

$g(x_i), i = 1, 2$. If B_1 and B_2 partition $A_2 \backslash A_1$, then there exists B_i ($i = 1$ or 2) such that $P_0(B_i) = \lambda P_0(A_2 \backslash A_1)$ and $P(B_i) \leq \lambda P(A_2 \backslash A_1) \leq \lambda(g(x_2) - g(x_1))$. Since

$$\inf_{P_0(C) \geq x_1 + \lambda(x_2 - x_1)} P(C) \leq P(A_1) + P(B_i) \leq g(x_1) + \lambda(g(x_2) - g(x_1)),$$

it follows that $g((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)g(x_1) + \lambda g(x_2)$. Hence the convexity is another reasonable requirement.

The definition of g-neighbourhood can be reformulated by means of the concentration function, defined by Cifarelli and Regazzini (1987), as a generalisation of the Lorenz curve. Marshall and Olkin (1979, p.5) give the following definition of Lorenz concentration curve (also known as the Lorenz-Gini curve): "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 $S_k = \sum_{i=1}^{k} x_{(i)}$ 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)$\cdots$ 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 P, which gives mass $x_{(i)}/S_n$ to $\theta_i, i = 1, \ldots, n$, and the uniform distribution P_0 on $\Theta = \{\theta_1, \ldots, \theta_n\}$. Cifarelli and Regazzini (1987) defined the c.f. of P with respect to (w.r.t.) P_0, where P and P_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 $P(E) = \int_{E \cap N^C} h(\theta) P_0(d\theta) + P_s(E \cap N)$, $\forall\, E \in \mathcal{F}, P_0(N) = 0$, $P_s(N) = P_s(\Theta)$, where $P_a(\cdot) = \int_{\cdot \cap N^C} h(\theta) P_0(d\theta)$ and P_s denote the absolutely continuous and the singular part of P w.r.t. P_0, respectively. Set $h(\theta) = \infty$ all over N and define $H(y) = P_0(\{\theta \in \Theta : h(\theta) \leq y\}), c(x) = \inf\{y \in \Re : H(y) \geq x\}$. Finally, let $L(x) = \{\theta \in \Theta : h(\theta) \leq c(x)\}$ and $L^-(x) = \{\theta \in \Theta : h(\theta) < c(x)\}$.

Definition 2. The function $\varphi : [0,1] \to [0,1]$ is said to be the concentration function of P w.r.t. P_0 if $\varphi(x) = P(L^-(x)) + c(x)\{x - H(c(x)^-)\}$ for $x \in (0,1)$, $\varphi(0) = 0$ and $\varphi(1) = P_a(\Theta)$.

When the dependence on P is to be emphasized, we will use the notations $h_P(\theta)$, $H_P(x)$, $c_P(x)$, $L_P(x)$ and $\varphi_P(x)$.

Observe that

$$\varphi(x) = \begin{cases} P(L(x)) & x = H(c(x)) \quad = \quad P_0(L(x)) \\ P(L^-(x)) & x = H(c(x)^-) \quad = \quad P_0(L^-(x)) \end{cases}$$

while $\varphi(x)$ is defined by linear interpolation on $\{x : H(c(x)\ ) < x < H(c(x))\}$, if it is not empty. Furthermore, as proved in Cifarelli and Regazzini (1987), $\varphi(x)$ is a nondecreasing, continuous and convex function such that $\varphi(x) \equiv 0 \Leftrightarrow P \perp P_0$, $\varphi(x) = x \;\forall\, x \in [0,1] \Leftrightarrow P = P_0$ and $\varphi(x) = \int_0^{c(x)} \{x - H(t)\}dt = \int_0^x c(t)dt$.

As pointed out by Cifarelli and Regazzini (1987, Remark 2.1), the definition of the c.f. can be extended to bounded positive measures which need not be probability measures. If P is such a measure, its concentration function w.r.t. P_0 coincides with that of P_a w.r.t. P_0.

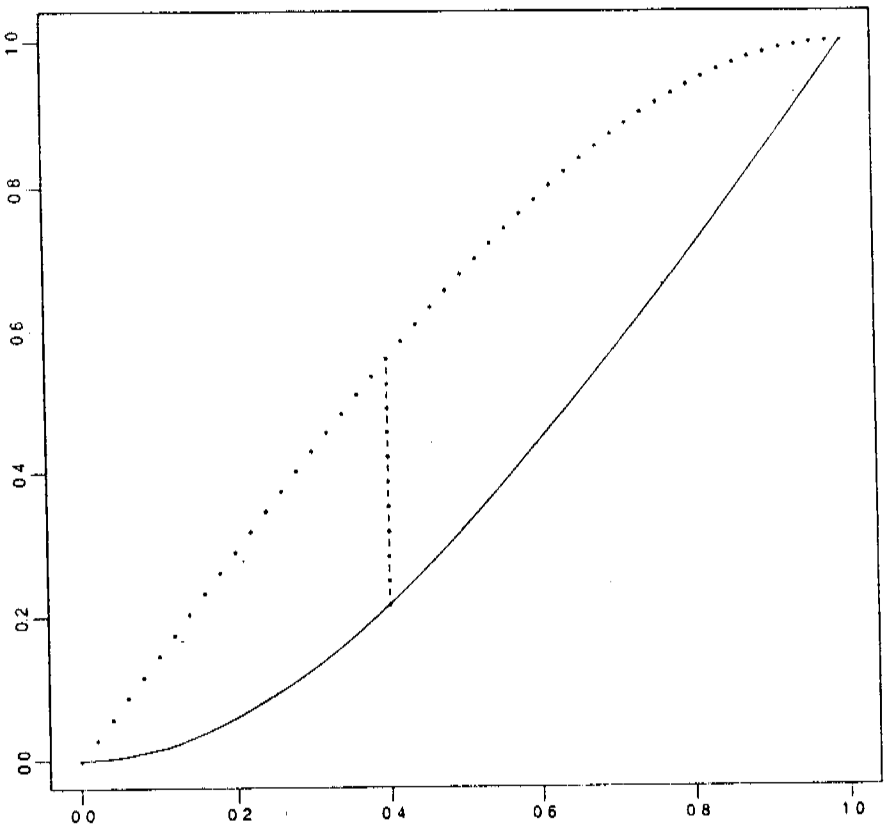

Figure 1. Concentration function $\varphi(x)(—)$ and $1 - \varphi(1-x)(\cdots)$ of $P \sim \mathcal{G}(2,\ 2)$ w.r.t. $P_0 \sim \mathcal{E}(1)$.

As an example, the c.f. $\varphi(x)$ of $P \sim \mathcal{G}$ (2,2) w.r.t.$P_0 \sim \mathcal{E}(1)$ is plotted in Fig.1, and it is shown, e.g., that [.216, .559] is the range spanned by the probability, under P, of the sets A with $P_0(A) = .4$. Such a range is a consequence of the following theorem, due to Cifarelli and Regazzini (1987), which provides an interesting interpretation of the c.f.; in fact, given *any* $x \in [0,1]$, then the probability, under P, of any A with P_0-measure x, is such that $\varphi(x) \le P(A) \le 1 - \varphi(1-x)$.

Theorem 1. *If $A \in \mathcal{F}$, $P_0(A) = x$, then $\varphi(x) \le P_a(A)$. Moreover if $x \in [0,1]$ is adherent to the range of H, then B_x exists such that $P_0(B_x) = x$ and*

$$\varphi(x) = P_a(B_x) = min\{P(A) : A \in \mathcal{F} \text{ and } P_0(A) \ge x\}. \qquad \dots (2)$$

If P_0 is nonatomic, then (2) holds for any $x \in [0,1]$.

Theorem 1 allows to express g-neighbourhoods by means of c.f.'s.

Proposition 1. *The set $K_g = \{P \in \mathcal{P} : \phi_P(x) \ge g(x),\ \forall\ x \in [0,1]\}$ is a g-neighbourhood of P_0 as defined in (1).*

3. Topology over $\mathcal{P}$

If G is a suitable class of monotone nondecreasing, convex continuous functions on $[0,1]$, then the class of neighbourhoods $\{K_g\}_{\{g \in G\}}$ can be used to define a topology on the space $\mathcal{P}$ of all the probability measures on $(\Theta, \mathcal{F})$.

When the dependence on P_0 has to be emphasized, $\varphi_P^{P_0}$ denotes the c.f. of P w.r.t. P_0 and let $K_g(P_0) = \{P \in \mathcal{P} : \varphi_P^{P_0}(x) \ge g(x), \forall\ x \in [0,1]\}$.

Proposition 2. *Let G be a class of monotone nondecreasing, continuous, convex functions $g : [0,1] \to [0,1]$, with $g(0) = 0$ and let G be such that, for any $g \in G$, there exists $\bar{g}, \tilde{g} \in G$ such that $\tilde{g}(\bar{g}(x)) \ge g(x), \forall\ x \in [0,1]$. Then there exists a topology $\mathcal{T}$ on P such that the class $\{K_g(P_0)\}_{g \in G}$ is a fundamental system of neighbourhoods of P_0.*

Before proving Proposition 2, we need the following lemma.

Lemma 1. *Consider $\mathcal{N}(P_0) = \{U \subset \mathcal{P} : K_g(P_0) \subseteq U$ for some $g \in G\}$. The following properties hold :*

(I) If $U_1 \subset U_2$ and $U_1 \in \mathcal{N}(P_0)$, then $U_2 \in \mathcal{N}(P_0)$.

(II) If $U_1, U_2, \dots, U_n \in \mathcal{N}(P_0)$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{N}(P_0)$.

(III) For any $U \in \mathcal{N}(P_0), P_0 \in U$.

(IV) If $U_1 \in \mathcal{N}(P_0)$, then there exists $U_2 \in \mathcal{N}(P_0)$ such that $U_1 \in \mathcal{N}(P_1)$ for any $P_1 \in U_2$.

Proof. The proofs of I, II, III are trivial. Given $U_1 \in \mathcal{N}(P_0)$, there exists $g \in G$ such that

$$\{P \in \mathcal{P} : \varphi_P^{P_0}(x) \ge g(x),\ \forall\ x \in [0,1]\} \subseteq U_1.$$

To prove IV, it is sufficient to show that there exists $\tilde{g} \in G$ such that

$$\{P \in \mathcal{P} : \varphi_P^{P_1}(x) \geq \tilde{g}(x),\ \forall\, x \in [0,1]\} \subseteq \{P \in \mathcal{P} : \varphi_P^{P_0}(x) \geq g(x),\ \forall\, x \in [0,1]\} \subseteq U_1, \qquad \ldots (3)$$

for P_1 belonging to $K_{\bar{g}}(P_0)$, and for a suitable g. Take $\bar{g}$ and $\tilde{g}$ such that $\tilde{g}(\bar{g}(x)) \geq g(x), \forall\, x \in [0,1]$, and $U_2 = K_{\bar{g}}(P_0)$.

If $P_1 \in U_2$, then it follows, from Theorem 1, that

$$\{A \in \mathcal{F} : P_0(A) \geq x\} \subseteq \{A \in \mathcal{F} : P_1(A) \geq \bar{g}(x)\}.$$

Hence if $\varphi_P^{P_1}(x) \geq \tilde{g}(x),\ \forall\, x \in [0,1]$, then

$$\inf_{\{P_0(A) \geq x\}} P(A) \geq \inf_{\{P_1(A) \geq \bar{g}(x)\}} P(A) \geq \tilde{g}(g(x)) \geq g(x)$$

which proves (3). ▱

Let us come to the proof of Proposition 2.

Proof. From I, II, III, IV it follows that there exists a unique topological structure $\mathcal{T}$ on P such that, for each $P_0 \in \mathcal{P}$, $\mathcal{N}(P_0)$ is the set of the neighbourhoods of P_0 in the topology $\mathcal{T}$. Moreover the class of neighbourhoods $\{K_g(P_0)\}_{g \in G}$ is a fundamental system of neighbourhoods of P_0 in $\mathcal{T}$ (see Bourbaki, 1989). ▱

Example 1. The trivial topology $\mathcal{T}_A$, in which any probability measure is an open set, is obtained taking any G such that $g_1 \in G$, where $g_1(x) = x, \forall\, x \in [0,1]$.

Example 2. Considering $G = \{g_\varepsilon(x) = \max\{0, x - \varepsilon\}, \forall\, x \in [0,1], 0 < \varepsilon \leq 1\}$, the topology $\mathcal{T}_V$ of the variational metric in $\mathcal{P}$ is obtained. In such a case, all the requirements about the functions g_ε are satisfied, along with the property $\tilde{g}(\bar{g}(x)) \geq g_\varepsilon(x), \forall\, x \in [0,1]$, for any $\varepsilon, 0 < \varepsilon \leq 1$, e.g. taking $\tilde{g} = g_{\varepsilon_1}$, $\bar{g} = \bar{g}_{\varepsilon_2}$ with $\varepsilon \geq \varepsilon_1 + \varepsilon_2$.

Example 3. A topology $\mathcal{T}_\alpha$ is obtained when taking $G = \{g_\alpha : g_\alpha(x) = x^\alpha,\ \forall\, x \in [0,1], 1 < \alpha < \infty\}$. In such a case, all the requirements about the functions g_α are satisfied, along with the property $\tilde{g}(g(x)) \geq g_\alpha(x),\ \forall\, x \in [0,1]$, for any $\alpha, 1 < \alpha < \infty$, e.g. taking $\tilde{g} = \bar{g} = g_{\sqrt{\alpha}}$. The topology $\mathcal{T}_\alpha$ is finer than the topology $\mathcal{T}_V$. In fact, let U_V be a neighbourhood of $P_0 \in \mathcal{P}$ in $\mathcal{T}_V$. It is easy to prove that there exists ε such that

$$U_V \supseteq \{P \in \mathcal{P} : \varphi_P^{P_0}(x) \geq \max\{0, x - \varepsilon\},\ \forall x \in [0,1]\}.$$

Let $\alpha > 1$ be such that $x^\alpha \geq \max\{0, x - \varepsilon\}$ for any $x \in [0,1]$, then it follows that

$$U_V \supseteq K_{g_\alpha}(P_0)$$

so that U_V is a neighbourhood of P_0 in the topology $\mathcal{T}_\alpha$. Hence $\mathcal{T}_\alpha$ is finer than $\mathcal{T}_V$. It follows that every continuous functional on $(\mathcal{P}, \mathcal{T}_V)$ is a continuous functional on $(\mathcal{P}, \mathcal{T}_\alpha)$.

As pointed out by Cifarelli and Regazzini (1987), the concentration function w.r.t. a fixed measure P_0 can be used to introduce a partial ordering in the space $\mathcal{P}$ of all the probability measures on $(\Theta, \mathcal{F})$.

Definition 3. The probability measure P_1 is to be said not less concentrated than P_2 w.r.t. P_0 if and only if $\varphi_{P_1}(x) \le \varphi_{P_2}(x)$, for any $x \in [0,1]$. We will denote it as $P_2 \preceq P_1$.

If there exists a probability measure, say $\bar{P}$, whose c.f. coincides with g, then the definition of g-neighbourhood can be reformulated as:

$$K_g = \{P \in \mathcal{P} : P \preceq \bar{P}\}.$$

Such a $\bar{P}$ exists, provided that g is compatible with P_0, according to the following definition:

Definition 4. A function $g : [0,1] \to [0,1]$ is said to be compatible with P_0 if g is a monotone nondecreasing, continuous, convex function, with $g(0) = 0$, and there exists a correspondence $\chi : S = \{\theta \in \Theta : P_0(\{\theta\}) \neq 0\} \to [0,1]$ such that, if $\theta_1 \neq \theta_2$,

$$(\chi(\theta_1), \chi(\theta_1) + P_0(\{\theta_1\})) \cap (\chi(\theta_2), \chi(\theta_2) + P_0(\{\theta_2\})) = \emptyset$$

and for any $\theta \in S$, it follows that there exist c_θ and d_θ such that

$$g_{|(\chi(\theta),\chi(\theta)+P_0(\{\theta\}))}(x) = c_\theta x + d_\theta,$$

where $g_{|(a,b)}(x)$ is the restriction of $g(x)$ to the interval (a,b).

Observe that if P_0 is nonatomic, then every monotone nondecreasing, continuous, convex function g such that $g(0) = 0$ is compatible with P_0.

Lemma 2. *Let Θ be a Polish space and $\mathcal{F}$ the σ-algebra of Borel sets. Then there exists a total ordering on Θ, denoted by $<$, such that, for any $\bar{\theta}$, the set $\{\theta \in \Theta : \theta < \bar{\theta}\}$ belongs to $\mathcal{F}$.*

Proof. Let $\{\theta_n\}_{n\ge1}$ be a dense subset of Θ. Introduce in Θ the following ordering: $\theta \le \bar{\theta}$ if and only if

$$\left(d(\theta_1,\theta) < d(\theta_1,\bar{\theta})\right) \bigvee \left(d(\theta_n,\theta) = d(\theta_n,\bar{\theta}), n = 1,2,\ldots\right)$$

$$\bigvee \left(\bigvee_{n=1}^{\infty} \left(d(\theta_i,\theta) = d(\theta_i,\bar{\theta}),\ i = 1,2,\ldots,n, d(\theta_{n+1},\theta) < d(\theta_{n+1},\bar{\theta})\right) \right).$$

If $\theta \le \bar{\theta}$ and $\bar{\theta} \le \theta$, then $d(\theta_n,\theta) = d(\theta_n,\bar{\theta})$, for any $n = 1,2,\ldots$. Since the topology is Hausdorff and $\{\theta_n\}_{n\ge1}$ is dense, it follows that $\theta = \bar{\theta}$. Moreover the set $\{\theta \in \Theta : \theta < \bar{\theta}\}$ is a denumerable union of measurable sets and therefore it is measurable. ▱

Theorem 2. *Given a function $g : [0,1] \to [0,1]$, there exists at least one measure P such that g is the c.f. of P w.r.t. P_0 if and only if g is compatible with P_0.*

Proof. The necessity of the condition is trivial. Let us prove the sufficiency. From the Lemma 2, Θ can be endowed with a total ordering, which will be denoted by $<$, such that, for any $\bar{\theta}$, the set $\{\theta \in \Theta : \theta < \bar{\theta}\}$ is measurable. Let P_s be a measure on Θ, singular w.r.t. P_0 and such that $P_s(\Theta) = 1 - g(1)$. Let χ and S be as in Definition 4. Define $T = [0,1] \setminus \cup_{\theta \in S}(\chi(\theta), \chi(\theta) + P_0(\{\theta\})$. The monotone function $V(x) = x - \sum_{\{\theta \in S; \chi(\theta) < x\}} P_0(\theta)$ determines a one to one correspondence between T and $[0,t]$, where $t = P_0(S^C)$. Hence there exists a measurable monotone function $W : S^C \to T$ such that

$$P_0(\{\theta \in S^C : \theta < \bar{\theta}\}) = V(W(\bar{\theta})). \qquad \ldots (4)$$

From (4) the measure μ defined on every Borel subset A of T as $\mu(A) = P_0(W^{-1}(A))$ is the restriction of the Lebesgue measure. Since g is a convex function, then it is differentiable almost everywhere. Hence, if A_x denotes the set where g is not differentiable and $A = W^{-1}(A_x)$, then $P_0(A) = \mu(A_x) = 0$. Consider the function

$$h(\theta) = \begin{cases} c_\theta & \theta \in S \\ g'_x(W(\theta)) & \theta \in \Theta \setminus (A \cup S) \\ 0 & \theta \in A. \end{cases}$$

Observe that the restriction of h to $\Theta \setminus (A \cup S)$ is monotone nondecreasing. Moreover, since W is measurable, then h is measurable. Let P_a be the measure whose Radon-Nikodym derivative w.r.t. P_0 is h and let $P = P_s + P_a$. Then, since h is monotone nondecreasing on $\Theta \setminus (A \cup S)$ then, if $x = H(c(x)^-)$,

$$\int_{h(\theta) < c_x} h(\theta) P_0(d\theta) = \sum_{\theta \in S; \chi(\theta) < x} c_\theta P_0(\{\theta\}) + \int_{T \cap [0,x]} g'(t) \mu(dt) = g(x).$$

Hence it is easily seen that P is a probability measure and that g is the c.f. of P w.r.t. P_0. $\square$

Observe that, as it results from the proof, the measure, whose concentration function is g, is generally not unique, since it depends on P_s and the arbitrary ordering on Θ. Furthermore, the singular component P_s is present if and only if $g(1) < 1$.

4. Representation Theorem

Let P_0 be a nonatomic probability measure on θ and let $g : [0,1] \to [0,1]$ be a function compatible with P_0. It will be proved now that all the probability measures in a g-neighbourhood K_g of P_0 are mixtures of the extremal ones in E_g, where $E_g = \{P \in \mathcal{P} : \phi_P(x) = g(x)\}$. Different proofs lead to such a result when either

$g(1) = 1$ or $g(1) < 1$; the former case admits just probability measures absolutely continuous w.r.t. P_0 while singularities are admitted in the latter one.

Consider the space $\mathcal{P}$ of all probability measures on Θ endowed with the weak topology; then $\mathcal{P}$ can be metrized as a complete separable metric space (see Parthasarathy, 1967, pp. 43–46).

Lemma 3. *If $g(1) = 1$, then K_g is a convex, compact subset of $\mathcal{P}$ in the weak topology.*

Proof. K_g is convex. Let $P_1, P_2 \in K_g, 0 < \alpha < 1$ and $P = \alpha P_1 + (1-\alpha)P_2$. If $A \in \mathcal{F}$ is such that $P_0(A) \geq x$, then it follows, from the definition of K_g and Theorem 1, that

$$P(A) \geq \inf_{P_0(B)>x} P(B) \geq \alpha \inf_{P_0(B)>x} P_1(B) + (1-\alpha) \inf_{P_0(B)\geq x} P_2(B) \geq g(x).$$

K_g is closed. Let $\{P_n\}_{n\geq 1}$ be a sequence of probability measures in K_g converging weakly to a probability measure P. For closed sets C we have $P(C) \geq \lim_{n\to\infty} P_n(C)$. Hence for any $\varepsilon > 0$ there exists $n_0(C)$ such that, for any $n(C) > n_0(C), P_{n(C)}(C) \leq P(C) + \varepsilon$. Furthermore, let $P_0(C) = x_0 - \delta$, then it follows that $g(x_0 - \delta) - \varepsilon \leq P(C)$. Let $S \in \mathcal{F}$ be such that $P_0(S) = x_0$. Then it follows that

$$P(S) = \sup_{\{C \text{ closed } : C\subset S\}} P(C) \geq \sup_{\delta>0} g(x_0 - \delta) - \varepsilon \geq g(x_0) - \varepsilon.$$

Hence $\varphi_P(x) \geq g(x)$.

K_g is compact. Since K_g is closed, it is sufficient to show that for any $\varepsilon > 0$ there exists a compact set $R \in \mathcal{F}$ such that for every $P \in K_g, P(R) \geq 1 - \varepsilon$ (see Parthasarathy, 1967, p. 47). Let x_ε be such that $g(x_\varepsilon) = 1 - \varepsilon$. Since Θ is a Polish space, it follows (see Parthasarathy, 1967, p. 29) that P_0 is tight, i.e. there exists a compact set $R \in \mathcal{F}$ such that $P_0(R) > x_\varepsilon$. Then, applying Theorem 1, it follows that $P(R) > 1 - \varepsilon$, for any $P \in K_g$. ▱

Let P_0 be a nonatomic probability measure. Consider the set of extremal points of K_g, that is the probability measures $P \in K_g$ such that

$$P = \alpha P_1 + (1-\alpha)P_2, P_1 \in K_g, P_2 \in K_g, 0 < \alpha < 1 \Longleftrightarrow P = P_1 = P_2.$$

Proposition 3. *The set of all the extremal points of K_g is contained in E_g. If $g(1) = 1$, then it coincides with E_g.*

Proof. Suppose $g(1) = 1$ and let $P \in E_g$. If $P = \alpha P_1 + (1-\alpha)P_2, P_1, P_2 \in K_g$, then P_1 and P_2 belong to E_g. In fact, suppose that $P_1 \notin E_g$, so that there exists

$x \in [0,1]$ such that $\inf_{P_0(A)\geq x} P_1(A) > g(x)$ and then

$$\inf_{P_0(A)\geq x} P(A) \geq \alpha \inf_{P_0(A)\geq x} P_1(A) + (1-\alpha) \inf_{P_0(A)\geq x} P_2(A) > g(x),$$

because $P_2 \in K_g$ implies that $\inf_{P_0(A)\geq x} P_2(A) \geq g(x)$. Therefore, it follows that $\varphi_{P_1}(x) = \varphi_{P_2}(x) = g(x)$ and $c_{P_1}(x) = c_{P_2}(x) = c_P(x)$ almost everywhere.

Consider $L_P(x)$ for any $x \in [0,1]$ such that $x = H_P(c_P(x))$. Since $P, P_1, P_2 \in E_g$, then $P(L_P(x)) = \alpha P_1(L_P(x)) + (1-\alpha)P_2(L_P(x))$ implies that for every $x \in [0,1]$

$$P_1(L_P(x)) = P_2(L_P(x)) = P(L_P(x)) = g(x). \qquad \ldots (5)$$

Analogously, we have $P_1(L_P^-(x)) = P_2(L_P^-(x)) = P(L_P(x))$ if $x = H_P(c_P(x)^-)$. If $x = H_P(c_P(x))$, then, for any $\delta > 0, c_P(x+\delta) > c_P(x)$ which implies $c_{P_1}(x+\delta) > c_{P_1}(x)$ and $c_{P_2}(x+\delta) > c_{P_2}(x)$. Hence $x = H_{P_1}(c_{P_1}(x)) = H_{P_2}(c_{P_2}(x))$ so that $P_1(L_{P_1}(x)) = P_2(L_{P_2}(x)) = g(x)$. From this and (5) it follows that, for any $y \geq 0$,

$$\{\theta \in \Theta : h_P(\theta) \leq y\} = \{\theta \in \Theta : h_{P_1}(\theta) \leq y\} = \{\theta \in \Theta : h_{P_2}(\theta) \leq y\} \text{ a.s.} - P_0.$$

Analogously,

$$\{\theta \in \Theta : h_P(\theta) < y\} = \{\theta \in \Theta : h_{P_1}(\theta) < y\} = \{\theta \in \Theta : h_{P_2}(\theta) < y\} \text{ a.s.} - P_0.$$

Hence $P_1 = P_2 = P$ and every probability measure in E_g is an extremal point for K_g.

Let now P be a probability measure not belonging to E_g. Then there exists $\dot{x} \in [0,1]$ such that $\varphi_P(x) > g(x)$. Since φ_p and g are continuous, there exists a neighbourhood $\mathcal{U}$ of x such that $\varphi_P(x) > g(x)$ for every x in $\mathcal{U}$.

Let (x_1, x_2) be the largest interval, eventually $(0,1)$, such that $x \in (x_1, x_2)$ implies $\varphi_P(x) > g(x)$. Since φ_P and g are continuous, $\varphi_P(x_1) = g(x_1)$ and $\varphi_P(x_2) = g(x_2)$.

Let $\bar{x} \in (x_1, x_2)$ and let $a = (\varphi_P)'_+(\bar{x})$. Then $g'_-(x_1) < a < g'_+(x_2)$. It is easily seen that $a_1, a_2, c_{a_1}, c_{a_2}$ can be chosen such that $g'_-(x_1) < a_1 < a < a_2 < g'_+(x_2)$ and

$$\bar{\varphi}(x) = \begin{cases} \varphi_P(x) & (\varphi_P)'_+(x) \leq a_1 \text{ or } (\varphi_P)'_+(x_2) > a_2 \\ a_1 x + c_{a_1} & a_1 < (\varphi_P)'_-(x) \leq (\varphi_P)'_+(x) \leq a \\ a_2 x + c_{a_2} & a < (\varphi_P)'_-(x) \leq (\varphi_P)'_+(x) \leq a_2 \end{cases}$$

is a continuous function with the property $g(x) \leq \bar{\varphi}(x) \leq \varphi_P(x)$ for any $x \in [0,1]$. Let $\epsilon = \min(a_2 - a, a - a_1)$. Since P_0 is nonatomic, the set $\{\theta \in \Theta : h_P(\theta) \in [a-\epsilon/3, a+\epsilon/3]\}$ contains more than one point. Hence there exists a non-constant function δ defined on Θ such that $\delta(\theta) = 0$ if $h_P(\theta) \notin [a-\epsilon/3,\ a+\epsilon/3], |\delta(\theta)| < \epsilon/3$ and $\int_\Theta \delta(\theta)P_0(d\theta) = 0$. Define $h_1(\theta) = h_P(\theta) + \delta(\theta), h_2(\theta) = h_P(\theta) - \delta(\theta)$. Hence $h_i(\theta) = h_P(\theta)$ if $h_P(\theta) \notin [a_1, a_2], a - \epsilon < h_i(\theta) < a + \epsilon$ if $a - \epsilon < h_P(\theta) < a + \epsilon$

and $\int_{\{h_i(\theta)\le a-\epsilon\}} h_i(\theta)P_0(d\theta) = \int_{\{h_P(\theta)\le a-\epsilon\}} h_P(\theta)P_0(d\theta), i = 1,2$. Define P_i as the probability measure on Θ whose Radon-Nikodym derivative with respect to P_0 is $h_i, i = 1,2$. It follows that φ_{P_1} and φ_{P_2} satisfy the condition

$$\varphi(x) = \varphi_P(x) \text{ for } (\varphi_P)'_+(x) < a_1 \text{ and } (\varphi_P)'_+(x) > a_2. \qquad \ldots (6)$$

Moreover, since $\bar{\phi}$ is the most concentrated curve among those satisfying (6),

$$\varphi_{P_i}(x) \ge \bar{\varphi}(x) > g(x). \qquad \ldots (7)$$

Since $P = (P_1 + P_2)/2$ and (7) holds, then P is not an extremal point of K_g.

Suppose now $g(1) < 1$; it can be proved, as before, that E_g is an extremal subset, i.e. that $P \in E_g$ is a convex combination just of $P_1, P_2 \in E_g$. No probability measure $P \notin E_g$ is an extremal one, because it can be expressed as a combination of two other measures, as before, whose c.f.'s are above g. The extremal points, if any, are thus in E_g. ▱

Every probability measure whose c.f. is greater than g can be represented as a mixture of probability measures having g as c.f.

Theorem 3. *Let the function $g : [0,1] \to [0,1]$ be compatible with a nonatomic probability measure P_0. For any probability measure $P \in K_g$, there exists a probability measure $\mu_{\bar{P}}$ on $\mathcal{P}$ such that $\mu_{\bar{P}}(E_g) = 1$ and $\bar{P} = \int_{\mathcal{P}} P\mu_{\bar{P}}(dP)$.*

Proof. Suppose $g(1) = 1$ and let K_g be the set of the probability measures P such that $\varphi_P(x) \ge g(x)$. It was proved in Lemma 3 that K_g is convex and compact in $\mathcal{P}$. Moreover in Proposition 3 it was proved that E_g is the extremal set of K_g.

Consider the topological vector space $C(\Theta)$ of all bounded continuous functions on Θ, endowed with the supremum topology, and let C' be its dual space; it can be easily seen that $\mathcal{P} \subset C'$.

The $C(\Theta)$ -topology of C' is a locally convex vector topology on C' (see Rudin, 1991, p. 68). Since K_g is metrizable, because $\mathcal{P}$ is, and it is also convex and compact in C', then the Choquet's theorem (Phelps, 1966, p. 19) implies the result.

Suppose now $g(1) = 1 - \varepsilon < 1$. Let $\bar{P}_a(\Theta) = 1 - \eta$, where, for any $P, \bar{P}_a$ denote the absolutely continuous part of P with respect to P_0. If $\eta = \varepsilon$, then the previous proof is applied to the set $\mathcal{P}_{1-\varepsilon}$ of the measures P on Θ such that $P(\Theta) = 1 - \varepsilon$, proving that there exists a probability measure μ on $\mathcal{P}_{1-\varepsilon}$ such that $\bar{P}_a = \int_{\{P\in\mathcal{P}_{1-\varepsilon},\varphi_P=g\}} P_\mu(dP)$. Hence

$$\bar{P} = \bar{P}_s + \int_{\{P\in\mathcal{P}_{1-\varepsilon},\varphi_P=g\}} P_\mu(dP) = \int_{\{P\in\mathcal{P},\varphi_P=g\}} P_{\bar{\mu}}(dP).$$

Consider the case $\eta < \varepsilon$. Let $\delta > 0$ be such that $\varphi_{\bar{P}}(x) > g(x)$ for every $x \in (1-\delta, 1]$. Let $\bar{g} : [0,1] \to [0, 1-\eta]$ be a continuous convex monotone nondecreasing function such that $\tilde{g}(1) = 1 - \eta$ and

$$\begin{cases} \tilde{g}(x) = g(x) & 0 \le x \le 1-\delta \\ g(x) \le \tilde{g}(x) \le \varphi_P(x) & 1-\delta \le x \le 1 \end{cases}$$

Let $\tilde{K}_g = \{P : P(\Theta) = 1-\eta,\ \tilde{g}(x) \le \varphi_P(x) \le (1-\eta)x\}$. Then it can be proved as in Lemma 3 that $\tilde{K}_g$ is a convex compact subset of the space $\mathcal{P}_{(1-\eta)}$ of the measures P on Θ such that $P(\Theta) = 1-\eta$ endowed with the weak topology. Moreover $P \in \tilde{K}_g$ is an extremal point if and only if $\varphi_P(x) = \tilde{g}(x)$ (see Propositon 3). Since $\bar{P}_a \in \tilde{K}_g$, there exists a probability measure μ on $\mathcal{P}_{1-\eta}$ such that

$$\bar{P}_a = \int_{\{P(\Theta)=1-\eta,\ \varphi_P(x)=\tilde{g}(x)\}} P_\mu(dP).$$

Let P be such that $P(\Theta) = 1-\eta$ and $\varphi_P(x) = \tilde{g}$ and let h be its Radon-Nikodym derivative with respect to P_0. It is easy to see that there exist δ and h_g such that $h_g(\theta) = h(\theta)$ for $\theta \in L_P(1-\delta), h_g(\theta) < h(\theta)$ for $\theta \in \Theta \backslash L_P(1-\delta), P_g(\Theta) = 1-\varepsilon$ and $\varphi_{P_g}(x) = g(x)$ for any $x \in [0,1]$, where h_g is the Radon-Nikodym derivative of P_g with respect to P_0.

Since $P - P_g$ is a positive measure on Θ, it is well-known that there exists a probability measure Q on Θ such that $P - P_g = (\varepsilon - \eta)\int_\Theta \delta_\theta Q(d\theta)$ where δ_θ is the Dirac measure on θ. Hence $P = P_g + \int_\Theta (\varepsilon-\eta)\delta_\theta Q(d\theta)$ and

$$\begin{aligned} \bar{P}_a &= \int_{\{(P_g,Q)\}} \int_\Theta [P_g + (\varepsilon-\eta)\delta_\theta] Q(d\theta)\mu'(dP_g, dQ) \\ &= \int_{\{P(\theta)=1-\eta,\ \varphi_p=g\}} P_{\mu_P}(dP). \end{aligned}$$

Hence $\bar{P} = \int_{\{P(\Theta)=1, \varphi_P=g\}} P_{\mu_P}(DP)$. ▱

Observe that if P_0 is nonatomic, then Thoerem 3 does not hold in general.

Example 4. Suppose $\Theta = \{\theta_1, \theta_2, \theta_3\}$; let $P_0(\theta_1) = 0, P_0(\theta_2) = P_0(\theta_3) = 1/2, P(\theta_1) = 0, P(\theta_2) = 1/3, P(\theta_3) = 2/3$ and $g(x) = 2x/3$. Let μ be a probability measure on P such that $\mu(E_g) = 1$ and let

$$P_\mu = \int_{\mathcal{P}} P_\mu(dP).$$

Since $P(\theta_1) = 1/3$ for any $P \in E_g$, then P_1 can not be represented as a mixture of measures belonging to E_g. ▱

5. Bounds on functionals

Theorem 3 can be applied to show that the supremum (or the infimum) over K_g of a large class of functionals is equal to the supremum (or the infimum) over E_g. The proof can be obtained by slightly modifying the one in Sivaganesan and Berger (1989).

Theorem 4. *Let f and g be real-valued functions on Θ such that $\int_{\Theta} |f(\theta)|P(d\theta) < \infty$ and $0 < \int_\Theta g(\theta)P(d\theta) < \infty$ for any $P \in K_g$. Then*

$$\sup_{P\in K_g} \frac{\int_\Theta f(\theta)P(d\theta)}{\int_\Theta g(\theta)P(d\theta)} = \sup_{P\in E_g} \frac{\int_\Theta f(\theta)P(d\theta)}{\int_\Theta g(\theta)P(d\theta)}. \qquad \ldots (8)$$

The same result holds with "sup" replaced by "inf".

Observe that if $f(\theta) = A + f_1(\theta)$ and $g(\theta) = B + g_1(\theta)$, then (8) becomes

$$\sup_{P\in K_g} \frac{A + \int_\Theta f_1(\theta)P(d\theta)}{B + \int_\Theta g_1(\theta)P(d\theta)} = \sup_{P\in E_g} \frac{A + \int_\Theta f(\theta)P(d\theta)}{B + \int_\Theta g(\theta)P(d\theta)} \quad . \qquad \ldots (9)$$

If $g(\theta) \equiv c$ for some $c > 0$, then the computation of

$$\sup_{P\in K_g} \frac{\int_\Theta f(\theta)P(d\theta)}{\int_\Theta g(\theta)P(d\theta)} = \frac{1}{c} \sup_{P\in E_g} \int_\Theta f(\theta)P(d\theta)$$

becomes easy.

Theorem 5. *Let $H_f(y) = P_0(\{\theta \in \Theta : f(\theta) \le y\})$, $c_f(x) = inf\{y : H_f(y) \ge x\}$. Then $\sup_{P\in K_g} \int_\Theta f(\theta)P(d\theta) = \int_0^1 c_f(x)c(x)dx$, where $c(x) = g'(x)$ a.e.*

Proof. Observe that $\sup_{P\in K_g} \int_\Theta f(\theta)P(d\theta) = \sup_{P\in K_g} \int_\Theta f(\theta)h_P(\theta)P_0(d\theta)$. Since $c_f(x)$ and $c(x)$ are rearrangements (see Hardy-Littlewood-Pólya, 1988, pp. 276-278) of $f(\theta)$ and $h(\theta)$, then $\int_\Theta f(\theta)P(d\theta) \le \int_0^1 c_f(x)c(x)dx$. On the other hand there exists $P \in E_g$ such that $f(\theta_1) \le f(\theta_2)$ implies $h_P(\theta_1) \le h_P(\theta_2)$. For such a P, $\int_\Theta f(\theta)P(d\theta) = \int_0^1 c_f(x)c(x)dx$, which proves the Theorem. ▱

Since $\inf_{P\in K_g} \int_\Theta -f(\theta)P(d\theta) = -\sup_{P\in K_g} \int_\Theta f(\theta)P(d\theta)$, the following corollary to Theorem 5 holds.

Corollary 1. $\inf_{P\in K_g} \int_\Theta f(\theta)P(d\theta) = -\int_0^1 c_{-f}(x)c(x)dx.$

6. Discussion

In this paper we have defined neighbourhoods of a probability measure P_0 and then, by means of a representation theorem, we have been able to provide results on the bounds of functionals defined in such neighbourhoods. The results presented can be used both in economics and in statistics, and not only in the problems in which the concentration curve has been widely used.

We just mention that the Lorenz curve, as well as the related Gini's index, has been sometimes used as a tool to examine how far actual situations are from ideal ones (e.g. to check the fairness in allocating seats in a U.S. legislature, so that representatives are elected by equivalent numbers of voters). In such problems, the function g could be considered as the maximum allowed distance from a uniform probability measure, expressed by means of c.f.'s, obviously.

Fortini and Ruggeri (1994) already applied the results in this paper to the robust Bayesian analysis, where the authors faced the problems of building a class of prior measures in a neighbourhood of a given one and checking if inferences lead to either posterior measures close to a base one or posterior functionals quite insensitive to the changes in the prior.

References

Bourbaki, N. (1989). *Elements of Mathematics, General Topology.* (2nd Edition), Springer Verlag : Berlin.

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

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

Hardy, G.H., Littlewood, J.E., Pólya, G. (1988). *Inequalities*, Cambridge University Press: Cambridge.

Huber, P.J. (1981). *Robust Statistics*, Wiley : New York.

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

Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces,* Academic Press: San Diego.

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

Rudin, W. (1991). *Functional Analysis* (2nd Edition), McGraw-Hill: New York.

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

CNR-IAMI
Via A.M. Ampère 56
I-20131 Milano
Italy