Applied Mathematical Sciences, Vol. 4, 2010, no. 39, 1941 - 1953

Synthesis of Distributions through Sets: A Unitary Approach

Eugenio Melilli

Universitá "L. Bocconi"
Via Roentgen, 1 - Milano, Italy
eugenio.melilli@unibocconi.it

Abstract

This paper suggests a general framework for dealing in a unified way with the problem of the synthesis of a distribution through sets. Following this view, interesting methods and quantities of sinthesis generally discussed separately are seen as particular cases of the general concept of mean of a distribution here presented. Various examples of means, including some based on the use of probabilistic metrics, are discussed.

Mathematics Subject Classification: 62-01

Keywords: Mean of a distribution in a class; probability metric; modal set; center of a distribution

1 Introduction

The paper's aim is to provide a general framework in which to face the fundamental problem of the synthesis of a distribution. Here the synthesis of a distribution is given a very general meaning; to summarize a distribution means to replace it by a point, or a set of points, describing and representing it in the "best" way. For instance, one can summarize a distribution on $\Re$ through a mean in the usual sense, i.e. a single, appropriate real number. But it is possible, and in some situations it can be more interesting and useful, to replace the same distribution on $\Re$ by an interval, perhaps of fixed length, or more generally by a subset of $\Re$ belonging to some particular family. Similarly, if we consider a distribution on $\Re^2$, depending on the different situations and on the aims of the search, it may be interesting to look for a single point of the plane, or a line, or a circle with fixed radius, as mean of the distribution considered. For instance, in this way we can obtain the vector of the arithmetic means or medians in the first case, or the line of least squares in the second one. All these elements, found with respect to some reasonable criterion, can be useful to resume the distribution and, in our discussion, will be considered and denoted generally as means of the distribution itself.

So, problems usually discussed separately and by different tools can be dealt with in a unified way, referring to the here presented general concept of mean of a distribution μ, in a class $\mathcal{C}$, with respect to a function λ_μ. Through this concept, it is possible to re-obtain, and often to consider under a different view, many of the quantities usually discussed in the literature as synthesis of a distribution; and in some cases, this general setting will give rise to new kinds of means (according to our general meaning) of a distribution, some of which having interesting characteristics and properties.

Given a distribution μ on a set $\mathcal{X}$ and fixed the class $\mathcal{C}$ of subsets of $\mathcal{X}$ in which we want to search for the mean of μ, the fundamental step is to single out a criterion according to which we desire to make the choice. The most natural and general criterion is the following: first, to fix a function, say λ_μ, associating to each C in $\mathcal{C}$ a nonnegative number $\lambda_\mu(C)$; second, to select, as mean of μ, the element $\tilde{C}$ of $\mathcal{C}$ (if it exists) for which λ_μ is minimum. We can give λ_μ different meanings and each of such meanings can justify its choice. For instance, $\lambda_\mu(C)$ can represent a "distance" (of course, not in the sense of the theory of metric spaces even if , as we will seen, in some cases strictly connected with it), or a measure of the divergence, between μ and C; in other cases it can yield a measure of the (negative) consequences, or loss, due to the substitution of μ by C. Hence, $\tilde{C}$ can be interpreted as the element of $\mathcal{C}$ "nearest", or "most similar", to μ. Otherwise, if λ_μ represent a loss, $\tilde{C}$ becomes the element of $\mathcal{C}$ minimizing it, hence in some sense the element of $\mathcal{C}$ which summaries μ in the best way.

In fact, this point of view is largely present in the literature, even if it is often applied in more restricted framework and in particular situations. For instance, one can consider the well known notion of center of order k of a distribution on $\Re$, or those of linear and superficial centers of a distribution on the plane (see Pompilj, 1984, for these concepts). For distributions on the real line, this point of view is similar to that of Herzel (1961); indeed, our work can be considered an extension of Herzel's theory in two directions: from one hand it concerns distributions on arbitrary spaces, from the other hand it is not limited to "simgle-point mean" only.

The proposed concept of mean is applied to distributions; as distribution on a space $\mathcal{X}$ we mean a (formal) probability measure on the measurable space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{A}$ is a σ-field of subsets of $\mathcal{X}$. In general, μ can represent the (real or obtained through some fitting procedure) distribution of a characteristics in a statistical collective, or a distribution of quantity (see Benedetti, 1981), or, in some cases, a true probability distribution. The σ-field $\mathcal{A}$ can be completely arbitrary; however, in this paper, when $\mathcal{X}$ is a metric space, $\mathcal{A}$

will always be understood as the Borel σ-field of $\mathcal{X}$, i.e. the smallest σ-field containing all open subsets of $\mathcal{X}$.

The plan of the work is the following. Section 2 will yield the general and precise definition of mean of a distribution μ, in a class $\mathcal{C}$, with respect to a function λ_μ. Then, in this section we will present a kind of mean, which we will call center, obtained by considering a particular form for the function λ_μ; some examples of means will conclude the section. In section 3 we will consider means, according to the definitions given, which comes out from particular probabilistic metrics. Section 4 will yield the concept of modal set; we will see as this concept extends in more directions the classical notion of mode of a distribution, giving it a new and interesting meaning.

2 The definition of mean of a distribution

The present section yields a precise and general definition of the concept of mean of a distribution, according to the point of view described in the introduction. Let μ be a distribution on $(\mathcal{X}, \mathcal{A})$, where $\mathcal{X}$ is an arbitrary set and $\mathcal{A}$ is a σ-field of subsets of $\mathcal{X}$. Let us consider a class $\mathcal{C}$ of subsets of $\mathcal{X}$ belonging to $\mathcal{A}$ and a function $\lambda_\mu : \mathcal{C} \to [0, +\infty)$.

Definition 1. A set $\tilde{C}$ in $\mathcal{C}$ such that

$$\inf_{C \in \mathcal{C}} \lambda_\mu(C) = \lambda_\mu(\tilde{C}) \tag{1}$$

is called *mean of μ, in $\mathcal{C}$, with respect to λ_μ.*

How to interpret such a quantity and the different elements with respect to which it is defined? We can consider $\tilde{C}$ as the element of $\mathcal{C}$ nearest to μ (or, more generally, one of the elements of $\mathcal{C}$ nearest to μ). Of course, when we say "nearest" we use an informal language; this notion is made precise by the function λ_μ. So, λ_μ can be interpreted as a distance between the elements of the class $\mathcal{C}$ and the distribution μ. More precisely, for C in $\mathcal{C}$, $\lambda_\mu(C)$ can be considered a measure of the distance, or difference, between C and μ, or, in some cases, a measure of the (negative) consequences of the substitution of C to the distribution μ.

The mean of the distribution μ, according to Definition 1, depends on a second element: the class $\mathcal{C}$. It represents the family of subsets of $\mathcal{X}$ in which we desire to search for that mean. In other words, the choice of $\mathcal{C}$ is related to the kind of set through which we want to summarize the distribution μ. For instance, if we are searching for a mean of μ according to the usual meaning (i.e., a point of $\mathcal{X}$), then we can take $\mathcal{C}$ as the class of all (or some) singletons of $\mathcal{X}$; indeed, it is sufficient to this aim to identify each singleton with its unique element. If, as a second example, our aim is to search for a line that summarize a distribution μ on the plane, then we take $\mathcal{C}$ to be the class of the lines in the plane; as we will observe later, if λ_μ is chosen appropriately, then in this case the mean of μ is the well known line of least squares. In other cases it may be interesting to summarize a distribution μ on a metric space through a sphere of fixed radius (for instance, an interval of fixed length in $\Re$). Then, we search for a mean of μ, with respect to some function λ_μ, in the class $\mathcal{C}$ of all such spheres.

Hence, different choices of the class $\mathcal{C}$ and of the function λ_μ yield different means, in the sense of Definition 1, of the distribution μ. So, in this way one can obtain synthesis of μ with various meanings and with respect to very different aims.

We can derive interesting and large classes of means, according to Definition 1, by selecting functions λ_μ having a particular structure, as precised by the following definition.

Definition 2. If ρ is a nonnegative function on $\mathcal{X} \times \mathcal{C}$ such that $\rho(\cdot, C)$ is μ-integrable for every C in $\mathcal{C}$, then a mean of μ, in the sense of Definition 1, with respect to the function

$$\lambda_\mu(C) = \int_{\mathcal{X}} \rho(x, C)\mu(dx), \tag{2}$$

is called a *center of μ, in $\mathcal{C}$, with respect to ρ.*

The "distance" $\lambda_\mu(C)$ between the element C of $\mathcal{C}$ and the distribution μ depends, in this case, on another function $\rho(x, C)$, whose task is to fix, in quantitative terms, the "divergence", or "difference", between C and the point x in $\mathcal{X}$. The function $\rho(x, C)$ is, except that for the condition of integrability, completely arbitrary, so that different choices of it give rise to many and different centers. Note that, in (2), the function $\lambda_\mu(C)$ is obtained by integrating $\rho(x, C)$ with respect to μ; i.e., we have computed the expected value of $\rho(X, C)$, where X is a random element having distribution μ. It is clear that it is possible to consider different ways to summarize $\rho(x, C)$ in order to produce a "distance" function λ_μ; we will briefly speak about this possible extension of the notion of center at the end of the present section.

Now we discuss some examples of means obtained according to Definition 2. These means are well known and largely used in statistics, so that it is of some interest to recognize that they may be obtained and reconsidered in this general setting.

As a first example, let μ be a distribution on $\Re$ having a finite moment of order k, where k is a positive real number. If we search for a center of μ, in $\mathcal{C} = \{\{y\}, y \in \Re\}$, with respect to the function $\rho(x, \{y\}) = |x - y|^k$, we are led to look for a point $\tilde{y}$ in $\Re$ for which

$$\inf_{y\in\Re}\int_{\Re}|x-y|^k\mu(dx)=\int_{\Re}|x-\tilde{y}|^k\mu(dx).$$

It is well known that such a point $\tilde{y}$ exists for each $k>0$, and it is called center (in the classical sense) of order k of μ. In particular, if $k=1$ or $k=2$, we obtain the median and the (arithmetic) mean respectively.

Now suppose $\mathcal{X}=\{x_1,x_2,\ldots,x_n\}$ and consider a distribution μ on $\mathcal{X}$, determined by the point masses $p_1,p_2,\ldots,p_n$, where $p_i=\mu(\{x_i\})$. As in the previous example, let $\mathcal{C}$ be the class of the singletons of $\mathcal{X}$. We search for a center of μ, in $\mathcal{C}$, with respect to the function

$$\rho(x_i,\{x_j\})=\begin{cases}1 & i\neq j\\ 0 & i=j.\end{cases}$$

Since

$$\int_{\mathcal{X}}\rho(x,\{y\})\mu(dx)=1-\mu(\{y\}),$$

it is clear that x_i is a center of μ, in $\mathcal{C}$, with respect to ρ if and only if x_i is a mode of μ according to the usual, classical meaning. We will deal in more detail with the notion of mode of a distribution and more generally with the concept of modal set in Section 4.

Let now μ be a distribution on $\Re^k$ having a finite variances and covariances matrix and consider the class $\mathcal{C}=\left\{\{y\},y\in\Re^k\right\}$ and the function

$$\rho(x,\{y\})=\sum_{i=1}^k(x_i-y_i)^2=||x-y||^2.$$

According to Definition 2, a center of μ, in $\mathcal{C}$, with respect to ρ is a vector $\tilde{y}$ in $\Re^k$ such that

$$\inf_{y\in\Re^k}\sum_{i=1}^k\int_{\Re^k}(y_i-x_i)^2\mu(dx_1\ldots dx_k)=\sum_{i=1}^k\inf_{y_i\in\Re}\int_{\Re}(y_i-x_i)^2\mu(dx_i)$$

$$=\sum_{i=1}^k\int_{\Re^k}(\tilde{y}_i-x_i)^2\mu(dx_1\ldots dx_k);$$

hence, $\tilde{y}$ is the vector of the arithmetic means of the marginal distributions of μ. If, in place of $||x-y||^2$, we consider

$$\rho(x,\{y\})=\sum_{i=1}^k|x_i-y_i|,$$

i.e. the so called *metropolitan distance* between x ed y (see, for instance, Benedetti, 1981), we obtain, as a center of μ, in $\mathcal{C}$, with respect to ρ, a vector whose components are medians of the marginal distributions of μ.

As a final example, let us consider a distribution μ on $\Re^2$ with finite variances and covariance and the class $\mathcal{C}$ of all the lines of the plane not parallel to the y-axis. If we denote by $x = (x_1, x_2)$ a point of $\Re^2$ and we identify the line in $\mathcal{C}$ having equation $x_2 = a_0 + a_1 x_1$ with the pair (a_0, a_1), we obtain, as center of μ, in $\mathcal{C}$, with respect to the function

$$\rho\left((x_1, x_2), (a_0, a_1)\right) = [x_2 - a_0 - a_1 x_1]^2, \tag{3}$$

the line whose associated pair (a_0, a_1) minimizes $\int_{\Re^2} [x_2 - a_0 - a_1 x_1]^2 \mu(dx_1 dx_2)$; that is, the well known least squares line, or regression line, in the direction of the y-axis. If we consider, instead of the function (3) (representing the distance, in the direction of x_2, between a point and a line), as function ρ the distance between a point and a line in a different direction, we obtain the least squares line in that direction as center of μ in the class of the lines not parallel to that direction; see, for the determination and some properties of this kind of mean, Pompilj (1984). More generally, for a distribution μ on $\Re^k$, one can obtain, with proper choices of $\mathcal{C}$ and ρ, the hyperplanes of least squares as centers of μ in the sense of Definition 2.

We conclude the section with a brief mention to means whose definition is similar to Definition 2, except for the fact that the mean value (expected value) is replaced by other quantities, for instance the median. More precisely, given a distribution μ on $(\mathcal{X}, \mathcal{A})$ and a class $\mathcal{C}$ of measurable subsets of $\mathcal{X}$, consider a function $\rho : \mathcal{X} \times \mathcal{C} \to [0, +\infty)$. The quantity $\lambda_\mu(C) = \int_{\mathcal{X}} \rho(x, C)\mu(dx)$, considered in (2), can be viewed as a mean distance between $\mathcal{C}$ and the points of $\mathcal{X}$ distributed according to μ. Instead of the mean distance, we can consider, for instance, the median of this distance; in this case it is also possible to drop the hypothesis of integrability of ρ. Hence, if we put

$$\lambda_\mu(C) = \inf\{t \in \Re : \mu(\{x \in \mathcal{X} : \rho(x, C) \le t\}) \ge 1/2\}, \tag{4}$$

(i.e., $\lambda_\mu(C)$ is the median of the distribution of $\rho(X, \mathcal{C})$, being X a random element with distribution μ), then the mean of μ, in $\mathcal{C}$, with respect to λ_μ is a set $\tilde{C}$ in $\mathcal{C}$ for which such λ_μ is minimum. As an example, consider $\mathcal{X} = \Re$, $\mathcal{C} = \{\{y\}, y \in \Re\}$ and $\rho(x, C) = \rho(x, \{y\}) = |x - y|$. Denoting by $h(y)$ the quantity $\lambda_\mu(\{y\})$, we have

$$h(y) = \inf\{t \in \Re : \mu([y - t, y + t]) \ge 1/2\}.$$

The mean of μ, in $\mathcal{C}$, with respect to λ_μ is the middle point $\tilde{y}$ of the interval of minimum length between those with mass larger than $1/2$. If, for a particular case, μ is the negative exponential distribution with mean $1/\theta$, then $\tilde{y}$ is $(2\theta)^{-1} \log 2$. If μ is gaussian with mean m and variance σ^2, we have $\tilde{y} = m$.

3 Means based on probabilistic metrics

We said in the previous section that the function λ_μ which, given the distribution μ and the class $\mathcal{C}$, determines the mean of μ in $\mathcal{C}$ can be interpreted as a distance between an element C of $\mathcal{C}$ and μ. Clearly, the word distance has to be taken in an informal way, without any reference to the theory of metric spaces. In fact, interesting and in some cases well known means derive from functions λ_μ strictly connected with metrics defined on spaces of (formal) probability measures. In the present section we will discuss some examples of means obtained in this way, fixing our attention on classes $\mathcal{C}$ whose elements are singletons. Let us consider a metric space $\mathcal{X}$ with distance d, a distribution μ on $(\mathcal{X}, \mathcal{A})$ and the class $\mathcal{C} = \{\{y\}, y \in \mathcal{X}\}$. We are interested in means of μ in $\mathcal{C}$, in the sense of Definition 1, with respect to functions of the following kind:

$$\lambda_\mu(\{y\}) = \sigma(\mu, \delta_y), \tag{5}$$

where δ_y denotes the distribution having all its mass in y and σ is a metric on a space $\mathcal{P}$ of probability measures on $(\mathcal{X}, \mathcal{A})$ containing μ and all the degenerate distributions. According to Definition 1, a mean of μ in $\mathcal{C}$ is a point $\tilde{y}$ in $\mathcal{X}$ such that $\delta_{\tilde{y}}$ is the degenerate distribution nearest to μ (with respect to the metric σ). Clearly, for each choice of the metric σ there is a particular function λ_μ and hence, in general, a different mean of μ in $\mathcal{C}$.

We now deal with some of this means, obtained from particular and well known probabilistic metrics. As a first example, let $\mathcal{X} = \Re$ and consider the Gini-Kantorovich metric, defined by

$$\sigma(\mu, \nu) = \int_{-\infty}^{+\infty} |F(t) - G(t)| dt,$$

for probability measures μ and ν having finite first moment; F and G denote the distribution functions corresponding to μ and ν respectively. Given μ on $(\Re, \mathcal{B})$, we are then interested in the quantity

$$\lambda_\mu(\{y\}) = \sigma(\mu, \delta_y) = \int_{-\infty}^{y} F(t) dt + \int_{y}^{+\infty} [1 - F(t)] dt,$$

where y is a real number. The following proposition holds.

PROPOSITION 1. *We have*

$$\inf_{y \in \Re} \sigma(\mu, \delta_y) = \sigma(\mu, \delta_{\tilde{y}})$$

if, and only if, $\tilde{y}$ *is a median of* μ.

Proof. For y in $\Re$, let us put $g(y) = \sigma(\mu, \delta_y)$, $a = \inf\{t \in \Re : F(t) \geq 1/2\}$ and $b = \sup\{t \in \Re : F(t) \leq 1/2\}$; we have $a \leq b$ and $[a, b] = \{t \in \Re :$

$\lim_{h\to 0^+} F(t-h) \le 1/2 \le F(t)\}$ is the set of all medians of F (or of μ). Let $y < a$. Since $F(t) < 1/2$ for each t in (y,a), then

$$\int_y^a [1-F(t)]dt > \int_y^a F(t)dt,$$

from which follows

$$g(y) = \int_{-\infty}^y F(t)dt + \int_y^a [1-F(t)]dt + \int_a^{+\infty} [1-F(t)]dt$$

$$> \int_{-\infty}^y F(t)dt + \int_y^a F(t)dt + \int_a^{+\infty} [1-F(t)]dt = g(a).$$

In the same way, if $y > b$, we have, for each t in (b,y), $F(t) > 1/2$, so that

$$\int_b^y F(t)dt > \int_b^y [1-F(t)]dt,$$

which implies

$$g(y) = \int_{-\infty}^b F(t)dt + \int_b^y F(t)dt + \int_y^{+\infty} [1-F(t)]dt$$

$$> \int_{-\infty}^b F(t)dt + \int_b^y [1-F(t)]dt + \int_y^{+\infty} [1-F(t)]dt = g(b).$$

Hence, if y does not belong to $[a,b]$, $g(y) > \max\{g(a), g(b)\}$. Let us now verify that g is constant on the interval $[a,b]$. Let y be a point in $[a,b]$; we have

$$g(y) = \int_{-\infty}^a F(t)dt + \int_a^y F(t)dt + \int_y^b [1-F(t)]dt + \int_b^{+\infty} [1-F(t)]dt.$$

But, if $a < t < b$, $\lim_{h\to 0^+} F(t-h) \le 1/2 \le F(t) \le 1/2$, that is $F(t) = 1/2$, so that

$$g(y) = \int_{-\infty}^a F(t)dt + \frac{1}{2}(b-a) + \int_b^{+\infty} [1-F(t)]dt$$

does not depend on y. So, $\tilde{y}$ is a point of minimum for g if and only if it belongs to the interval $[a,b]$; this concludes the proof.

By Proposition 1, a median of μ can be seen as a point whose associated degenerate distribution is the nearest one, with respect to Gini-Kantorovich's distance, to μ. This fact is more evident and its proof straightforward if we observe that Gini-Kantorovich's distance is a particular Wasserstein's distance; we will discuss below this class of metrics. We recall that a median can be obtained as a mean, according to our point of view, also in a different way (see one of the examples in the previous section).

Another distance largely used in probability theory is Prohorov's metric, which can be defined for distributions on arbitrary metric spaces. If μ and ν are probability measures on $(\mathcal{X}, \mathcal{A})$, where $\mathcal{X}$ is a metric space with metric d, the Prohorov distance between μ and ν is defined as follows:

$$\sigma(\mu, \nu) = \inf\{t > 0 : \mu(A^t) + t \geq \nu(A), \forall A \in \mathcal{A}\}, \tag{6}$$

where $A^t = \{z \in \mathcal{X} : d(z, A) < t\}$ and $d(z, A) = \inf_{x \in A} d(x, z)$. We are interested in the quantity $\lambda_\mu(\{y\}) = \sigma(\mu, \delta_y)$; it is simple to see that $\sigma(\mu, \delta_y) = \inf\{t > 0 : \mu(U_y(t)) \geq 1 - t\}$, where $U_y(t)$ denotes the open sphere with center y and radius t. If we denote by $\mathcal{U}$ the class of all open spheres $U_y(t)$ of $\mathcal{X}$ with $t > 0$, y in $\mathcal{X}$ and such that $\mu(U_y(t)) \geq 1 - t$, then a point y in $\mathcal{X}$ minimizes $\lambda_\mu(\{y\})$, and hence is a mean of μ, in $\mathcal{C}$, with respect to λ_μ, if it is the center of a sphere in $\mathcal{U}$ having minimum radius. Of course, a point with these characteristics does not necessarily exist and, if it does, needs not be unique; in the particular case in which μ is diffuse (i.e., it does not have concentrated masses), such a point does exist and it is unique. This mean, having an intuitive and interesting meaning, is in some way connected with the classical concept of mode of a distribution or, more directly and generally, with that of modal set which we will discuss in section 4. For a more detailed analysis of the mean obtained through Prohorov's distance and for the discussion of some related examples, we refer to Melilli (1995), where the function $\sigma(\mu, \delta_y)$ defined in (6) is dealt with as a measure of the dispersion of μ around the point y.

Let us now consider the distance, for distributions μ and ν on an arbitrary measurable space $(\mathcal{X}, \mathcal{A})$,

$$\sigma(\mu, \nu) = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)|.$$

We have

$$\lambda_\mu(\{y\}) = \sigma(\mu, \delta_y) = 1 - \mu(\{y\}),$$

so that

$$\lambda_\mu(\{\tilde{y}\}) = \inf_{y \in \mathcal{X}} \lambda_\mu(\{y\}) = 1 - \sup_{y \in \mathcal{X}} \mu(\{y\})$$

if and only if

$$\mu(\{\tilde{y}\}) \geq \mu(\{y\}) \qquad \forall y \in \mathcal{X}.$$

In other words, the point $\tilde{y}$ minimizes λ_μ if and only if it is a mode of μ. Clearly, this quantity has interest for discrete distributions only.

For $\mathcal{X} = \Re$, we can also consider the uniform metric

$$\sigma(\mu, \nu) = \sup_{t \in \Re} |F(t) - G(t)|,$$

F and G being the distribution functions corresponding to μ and ν respectively. We have

$$\lambda_\mu(\{y\}) = \sigma(\mu, \delta_y) = \max\{F(y), 1 - F(y)\},$$

and, as usual, a point $\tilde{y}$ of $\Re$ is a mean of μ, with respect to λ_μ, if and only if it minimizes this latter. If there exists y such that $F(y) = 0.5$, then $\tilde{y} = y$ (in this case, of course, $\tilde{y}$ is a median of μ). If such a point y does not exist, then consider the median y_0 (unique, in this case) of μ, for which $F(y_0) > 0.5$ and $\lim_{h\to 0^+} F(y_0 - h) \leq 0.5$. If $F(y_0) \leq 1 - \lim_{h\to 0^+} F(y_0 - h)$, then $\tilde{y} = y_0$, otherwise $\tilde{y}$ does not exist.

Let us now analyze the means derived from functions λ_μ of type (5), when we use as metric σ a Wasserstein distance. This metric is defined, for probability measures on a space with arbitrary metric d, in the following way:

$$\sigma(\mu, \nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\mathcal{X}^2} d(x, z)\gamma(dxdz), \tag{7}$$

$\Gamma(\mu, \nu)$ being the Fréchet class with marginals μ e ν, i.e. the class of all distributions on $(\mathcal{X}^2, \mathcal{A}^2)$ having marginals μ and ν respectively. We are so led to consider the quantity

$$\lambda_\mu(\{y\}) = \sigma(\mu, \delta_y) = \int_{\mathcal{X}} d(x, y)\mu(dx),$$

since the Fréchet class for which a marginal is degenerate has a unique element. The search for a mean of μ, in $\mathcal{C}$, with respect to such function λ_μ is then equivalent to the determination of the point $\tilde{y}$ for which is minimum $\int_{\mathcal{X}} d(x, y)\mu(dx)$; so we are led to find particular means in the sense of Definition 2, i.e. particular centers of μ, in $\mathcal{C} = \{\{y\}, y \in \mathcal{X}\}$, with respect to the function ρ which, in this case, coincides with the metric fixed on the space $\mathcal{X}$. Hence, we can obtain, following this different direction, all the (point) means discussed in the examples of the previous section. Let us observe that, if $\mathcal{X} = \Re$ and d is the euclidean distance, the Wasserstein metric defined in (7) coincides with Gini-Kantorovich's distance (see, for a proof, dall'Aglio, 1956). Hence, the proof of Proposition 1 follows directly from the fact that the medians of a distribution are the centers (according to the classical meaning of the term) of μ of order one.

4 Modal sets

Here we consider a particularly interesting class of means, defined according to the point of view adopted in this paper. These means will be called modal sets, since, as we are going to see, they are in some way generalizations of the well known mode of a distribution.

With reference to Definition 2, let $(\mathcal{X}, \mathcal{A})$ be an arbitrary measurable space and $\mathcal{C}$ a class of elements of $\mathcal{A}$. Define

$$\rho(x, C) = I_{C^c}(x) = \begin{cases} 1 & x \notin C \\ 0 & x \in C. \end{cases} \tag{8}$$

Let us notice that, if $\mathcal{X}$ is a metric space with metric $d(x, y) = I_{\{y\}^c}(x)$ (i.e., a discrete metric space), then $\rho(x, C)$ is the usual distance between a point and a set; indeed,

$$\inf_{y\in C} d(x, y) = \inf_{y\in C} I_{\{y\}^c}(x) = I_{C^c}(x) = \rho(x, C).$$

According to Definition 2, a mean in $\mathcal{C}$ of a distribution μ on $(\mathcal{X}, \mathcal{A})$, with respect to ρ, is any element $\tilde{C}$ in $\mathcal{C}$ such that

$$\int_{\mathcal{X}} \rho(x, \tilde{C})\mu(dx) = 1 - \mu(\tilde{C}) = \inf_{C\in\mathcal{C}}\{1 - \mu(C)\},$$

or, equivalently, such that

$$\mu(\tilde{C}) = \sup_{C\in\mathcal{C}} \mu(C). \tag{9}$$

A set $\tilde{C}$ which satisfies (9) is called *modal set of* μ *in* $\mathcal{C}$, or, more briefly, *mode of* μ *in* $\mathcal{C}$. In short, $\tilde{C}$ is the element of $\mathcal{C}$ which has the largest μ-mass. Clearly, such a mean needs not to exist or to be unique. It is simple to show how the well known mode of a (discrete) distribution can be seen as a particular case of the now defined modal set. Indeed, if μ has support $S = \{x_1, x_2, \ldots, \}$, with $\mu(\{x_i\}) = p_i$, and we consider the class $\mathcal{C} = \{\{y\}, y \in \mathcal{X}\}$ (and, as usual, we identify the set $\{y\}$ with its element y), then a mode of μ in $\mathcal{C}$, according to our definition, is a point x_i of S such that $p_i \geq p_j$ for all $j = 1, 2, \ldots$; in other words, x_i is a mode of μ in the usual, classical meaning.

Dealing with modal sets, it turns out to be particularly interesting the following situation. We assume, with reference to our definition of modal set, $\mathcal{X}$ metric space with distance d and $\mathcal{C} = \mathcal{C}_t = \{U_x, x \in \mathcal{X}\}$, where $U_x = \{y \in \mathcal{X} : d(x, y) \leq t\}$ and $t > 0$; that is, we consider the class of all closed spheres of $\mathcal{X}$ having fixed radius t. Then, $\tilde{C} = \tilde{U}_x = U_{\tilde{x}}$ is the sphere with radius t to which μ assignes the largest mass; hence, $\tilde{x}$ is a point (not necessarily the unique point) which is the center of the closed sphere of radius t on which is concentrated the largest mass (of μ). We have already noticed how a modal set, in general, needs not to exist; in this case, we can prove its existence. Indeed, the following proposition holds.

PROPOSITION 2. *For any* $t > 0$, *there exists at least one solution for the equation (in* y*)*

$$\mu\left(\{x \in \mathcal{X} : d(x, y) \leq t\}\right) = \sup_{z\in\mathcal{X}} \mu\left(\{x \in \mathcal{X} : d(x, z) \leq t\}\right). \tag{10}$$

Proof. Let $c = \sup_{z\in\mathcal{X}} \mu\left(\{x \in \mathcal{X} : d(x, z) \leq t\}\right)$ and consider the following sets: $A = \{y \in \mathcal{X} : \mu\left(\{x \in \mathcal{X} : d(x, y) \leq t\}\right) = c\}$, $A_\varepsilon = \{y \in \mathcal{X} : \mu(\{x \in \mathcal{X} :$

$d(x,y) \leq t\}) \geq c-\varepsilon\}$, where ε is a positive number, and $B = \cap_{\varepsilon>0} A_\varepsilon$. We show that $A = B$. Of course, $A \subseteq B$. To show the converse, let us consider a point y in B, so that $\mu(\{x \in \mathcal{X} : d(x,y) \leq t\}) \geq c - \varepsilon$ for all $\varepsilon > 0$, or equivalently $\mu(\{x \in \mathcal{X} : d(x,y) \leq t\}) \geq c$. From the definition of c follows $\mu(\{x \in \mathcal{X} : d(x,y) \leq t\})$
$= c$, so that y belongs to A and hence $B \subseteq A$. The proof is complete if we show that B is non-empty. Observe that A_ε is, for any $\varepsilon > 0$, a compact set. Indeed, let $\{y_n\}_{n\geq 1}$ a sequence of points in A_ε converging to a point y. Then, for every $\eta > 0$ and n large enough, we have $d(y_n, y) \leq \eta$. Hence

$$c - \varepsilon \leq \mu\ (\{x \in \mathcal{X} : d(x, y_n) \leq t\}) \leq \mu(\{x \in \mathcal{X} : d(x,y) \leq t + \eta\}), \qquad (11)$$

where the first inequality holds since y_n belongs to A_ε and the second one is true since, for n large enough, $d(y,x) \leq d(y,y_n) + d(y_n,x) \leq \eta + t$. Now, as we have $\mu(\{x \in \mathcal{X} : d(x,y) \leq t + \eta\}) \geq c - \varepsilon$ for all $\eta > 0$, it follows $\mu(\{x \in \mathcal{X} : d(x,y) \leq t\})$
$\geq c - \varepsilon$ and hence y belongs to A_ε, so that A_ε is compact. Since A_ε is, for all $\varepsilon > 0$, a compact, non-empty set and $A_{\varepsilon_1} \subseteq A_{\varepsilon_2}$ if $\varepsilon_1 < \varepsilon_2$, it follows $B \neq \emptyset$ and, as $A = B$, A is non-empty; in other words, equation (10) has at least one solution.

We could consider the open spheres of $\mathcal{X}$ in place of the closed ones; i.e., we could be interested in the open sphere with radius $t > 0$ which is modal for a distribution μ. In this case, if μ is not diffuse, such a mode needs not to exist. Note that, when $\mathcal{X} = \Re$, a center of μ, in $\mathcal{C}_t$, with respect to the function ρ defined in (8), is a modal interval of length $2t$, a certainly interesting quantity as a synthesis of a distribution.

To conclude the section, we point out the existence of a strict link between the modal spheres defined above and the well known Lévy concentration function; we refer to the paper by Hengartner and Theodorescu (1973) for its definition and the proof of some results that clarify these connections.

References

[1] D. Dabrowska, Descriptive Parameters of Location, Dispersion and Stochastic Dependence, *Statistics,* **16** (1985), 1563 - 1572.

[2] G. Dall'Aglio, Sugli Estremi dei Momenti delle Funzioni di Ripartizione Doppia, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.,* **3** (1956), 33-74.

[3] W. Hengartner, R. Theodorescu, *Concentration Functions,* Academic Press, New York, 1973.

[4] A. Herzel, Sulla definizione dei concetti di Media e di Interpolazione , *Metron,* **XXI** (1961), 126-138.

[5] E. Melilli, A Measure of Dispersion based on Prohorov's Distance, *Metron,* **LIII** (1995), 115-134.

[6] G. Pompilj, *Le Variabili Casuali,* Istituto di Calcolo delle Probabilitá dell'Universitá di Roma, Roma, 1984.

[7] Rachev, S.T., *Probability Metrics,* Wiley, London, 1991.

Received: November, 2010