FRÉCHET CLASSES AND COMPATIBILITY OF DISTRIBUTION FUNCTIONS (*)

G. Dall'Aglio

In a paper published in 1951 [13] M. Fréchet gave for the first time a systematic study of the family of two-dimensional d.f.'s (distribution functions) with given marginals. For this reason we have called *Fréchet classes* this family and, more generally, the analogous ones arising in more than two dimensions.

Scope of this paper is to give an account (with some improvements) of results obtained in this field and in some related topics, particularly on the problem of *compatibility*. This problem arises naturally when one undertakes the study of Fréchet classes in more than two dimensions, if the given marginal d.f.'s are « overlapping »: take for instance the class of d.f.'s in R^3 having as marginal d.f.'s $F_{12}(x_1, x_2)$, $F_{13}(x_1, x_3)$, $F_{23}(x_2, x_3)$. In this case the first question to be answered is whether the class is not void, i.e. whether there exists at least a d.f. $F(x_1, x_2, x_3)$ whose marginal d.f.'s are F_{12}, F_{13}, F_{23}; we will say then that F_{12}, F_{13}, F_{23} are *compatible*.

Some applications will also be presented, with special regard to Gini's « indice di dissomiglianza » and measures of distances between two d.f.'s.

* * *

Some brief remarks about notations will be useful to avoid confusion. We will consider functions on R^n, $F(x) = F(x_1, ..., x_n)$.

We will denote by $F_A(x_A)$ functions on $R^{|A|}$, where A is a set of indices (a subset of $1, 2, ..., n$) and $|A|$ the number of elements in A.

A d.f. $F(x)$ will be a left continuous, non decreasing function on R^n, which tends to 1 when all variables tend to $+\infty$, and to 0 when at least one variable tends to $-\infty$.

(*) Conferenza tenuta il 17 marzo 1971.

The variation on an interval $[a, b)$, which by the « non decreasing » condition must be non negative, will be denoted by

$$\Delta_a^b F(x) = \Delta_{\substack{a_1 \\ x_1}}^{b_1} \dots \Delta_{\substack{a_n \\ x_n}}^{b_n} F(x_1, \dots, x_n)$$

where $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$ $(a_i < b_i)$.

When there is no danger of confusion, $F_A(x_A)$ will denote a marginal d.f. of $F(x)$:

$$F_A(x_A) = F(x_A, +\infty)$$

that is the d.f. obtained letting the variables $x_i (i \notin A)$ tend to $+\infty$.

1. Two-dimensional Fréchet classes.

Given two one-dimensional d.f.'s $F_1(x_1)$, $F_2(x_2)$, we call Fréchet class $\Gamma(F_1, F_2)$ the family of two-dimensional d.f.'s which have F_1, F_2 as marginal d.f.'s:

$$\Gamma(F_1, F_2) = \{F(x_1, x_2) : F \text{ is a d.f., } F(x_1, +\infty) = F_1(x_1), F(+\infty, x_2) = = F_2(x_2)\} .$$

This family of d.f.'s is introduced in [13], where its principal properties are studied. $\Gamma(F_1, F_2)$ is obviously not void, since it contains the « independence » d.f.

$$F^*(x_1, x_2) = F_1(x_1) F_2(x_2) .$$

The following results hold:

Theorem 1:

i) $F \in \Gamma(F_1, F_2)$ *iff* F *is a d.f. and*

$$F'(x_1, x_2) \leq F(x_1, x_2) \leq F''(x_1, x_2) \tag{1}$$

where

$$F'(x_1, x_2) = max \{F_1(x_1) + F_2(x_2) - 1, 0\}$$

$$F''(x_1, x_2) = min \{F_1(x_1), F_2(x_2)\}$$

are d.f.'s, and therefore the minimum and maximum elements of $\Gamma(F_1, F_2)$:

ii) $\Gamma(F_1, F_2)$ *is convex and closed (under usual convergence of d.f.'s)*

iii) $F^* = F' \Leftrightarrow F^* = F'' \Leftrightarrow \Gamma(F_1, F_2)$ *contanins only one element* $\Leftrightarrow F_1$ *and/or* F_2 *are degenerate d.f.'s.*

The proof of these results is very simple and for some of them will be given later for n dimensions. What is interesting is to point out the properties of extreme d.f.'s F' and F''. Let us first remark that their properties are exchangeable, since, if F' is the joint d.f. of the r.v.'s (random variables) X_1, X_2, F'' can be regarded as the d.f. of $(X_1, -X_2)$. Now if X_1, X_2 have the joint d.f. F'', each r.v. is functionally dependent on the other, in the sense that (except possibly for discontinuity points of F_1), the conditioned r.v.'s $X_2(x_1)$ are degenerate and $X_2(x_1)$ is not decreasing with x_1. In other words, the distribution is concentrated on a non decreasing curve of the plane (x_1, x_2). Thus X_1 and X_2 have the maximum positive dependence (or correlation); and, if $F_1 = F_2$, F'' corresponds to (a.s.) equal r.v.'s.

F', on the contrary, gives the maximum negative dependence, and the distribution is concentrated on a non increasing curve.

These remarks were earlier pointed out, for discrete distributions, by C. Gini [19] and by T. Salvemini [33]. The latter introduced the « tabelle di cograduazione e contrograduazione », which correspond to F'' and F', and gave a method to construct them, which is now known, in linear programming, as the N.W. corner rule.

As it will be seen later, remarks about maximum positive or negative correlation can be applied also to the value of correlation coefficient, when X_1 and X_2 possess finite second moments.

* * *

Among other studies about Fréchet classes in two dimensions, we recall here those by A. Nataf [29], from a geometric point of view, and by G. Letac [27], who, for atomic marginal distributions, gives a characterization of extremal functions of Γ.

2. General definition of Fréchet classes. The case of one-dimensional marginal distribution functions.

Let N denote the set of the first n natural numbers, and $\mathcal{A}$ a family of subsets of N. Given the set of d.f.'s $\{F_A(x_A),\ A \in \mathcal{A}\}$, we define

Fréchet class $\Gamma(F_A, A \in \mathcal{A})$ as the class of n-dimensional d.f.'s having the given F_A as marginal d.f.'s:

$$\Gamma(F_A, A \in \mathcal{A}) = \{F(x_1, \dots, x_n) : F \text{ is a d.f., } F(x_A, +\infty) = F_A(x_A) \text{ for } A \in \mathcal{A}\}.$$

We will assume that $\mathcal{A}$ does not contain N (since then F_N would obviously be the only d.f. in Γ) and that:

$$\bigcup_{A \in \mathcal{A}} A = N ,$$

since otherwise the study could be conducted, with same results, on an R^{N_1} with a smaller number of dimensions than N.

* * *

The most immediate generalization of the two-dimensional case is when $|A| = 1$ for every A in $\mathcal{A}$, i.e. the given marginal d.f.'s are all one-dimensional. The Fréchet class $\Gamma(F_1, \dots, F_n)$ is again obviously not void, since it contains the independence d.f. $F^* = \prod_i F_i$. Properties of $\Gamma(F_1, \dots, F_n)$ have been studied in [11] for $n = 3$ and in [32] for $n = 4$, but it is not difficult to extend the results obtained. It can be seen that $\Gamma(F_1, \dots, F_n)$ has all the properties as for $n = 2$, with only one major exception:

THEOREM 2:

i) *$F(x_1, \dots, x_n)$ belongs to $\Gamma(F_1, \dots, F_n)$ iff it is a d.f. and*

$$F'(x_1, \dots, x_n) \leq F(x_1, \dots, x_n) \leq F''(x_1, \dots, x_n) \tag{4}$$

where

$$F''(x_1, \dots, x_n) = \min\{F_1(x_1), \dots, F_n(x_n)\} \tag{5}$$

belongs to Γ, while

$$F'(x_1, \dots, x_n) = \max\{F_1(x_1) + \dots + F_n(x_n) - n + 1, 0\} \tag{6}$$

for $n > 2$ is not, in general, a d.f.

ii) *$\Gamma(F_1, \dots, F_n)$ is convex and closed*

iii) *$F^* = F' \Leftrightarrow F^* = F'' \Leftrightarrow \Gamma(F_1, \dots, F_n)$ contains only one element $\Leftrightarrow$ at least $n-1$ among the n d.f. F_i are degenerate.*

Some of the statements in the theorem are almost obvious. The left inequality in (4) will be proved as a separate lemma which will be useful in the sequel. The fact that F' is not always a d.f. will be investigated later. We prove now:

a) F'' is a d.f. Since other conditions are immediately verified, we have to prove that F'' is not decreasing, i.e. that it has non-negative variation on every interval $[x', x'')$ with $x' < x''$. We assume, for simplicity of notations, that $F_1(x_1') \leq F_2(x_2') \leq \dots \leq F_n(x_n')$. Then

$$\Delta_{x'}^{x''} F'' = \Delta_{x_1'}^{x_1''} \Delta_{x_2'}^{x_2''} \dots \Delta_{x_n'}^{x_n''} \min \{F_1(x_1), \dots, F_n(x_n)\} =$$

$$= \Delta_{x_2'}^{x_2''} \dots \Delta_{x_n'}^{x_n''} \min \{F_1(x_1''), F_2(x_2), \dots, F_n(x_n)\} -$$

$$- \Delta_{x_2'}^{x_2''} \dots \Delta_{x_n'}^{x_n''} \min \{F_1(x_1'), F_2(x_2), \dots, F_n(x_n)\}.$$

By assumption,

$$F_1(x_1') \leq F_i(x_i') \leq F_i(x_i'') \text{ for } i > 1$$

so that all the terms in the last variation written above are equal to $F_1(x_1')$, and the variation is zero. Now:

$$\Delta_{x'}^{x''} F'' = \Delta_{x_2'}^{x_2''} \dots \Delta_{x_n'}^{x_n''} \min \{F_1(x_i''), F_2(x_2), \dots, F_n(x_n)\} =$$

$$= \Delta_{x_3'}^{x_3''} \dots \Delta_{x_n'}^{x_n''} \min \{F_1(x_1''), F_2(x_2''), F_3(x_3), \dots, F_n(x_n)\} -$$

$$- \Delta_{x_3'}^{x_3''} \dots \Delta_{x_n'}^{x_n''} \min \{F_1(x_1''), F_2(x_2'), F_3(x_3), \dots, F_n(x_n)\}$$

and the last variation is again zero, since all the terms are equal to $\min \{F_1(x_1''), F_2(x_2')\}$.
Continuing, we obtain

$$\Delta_{x'}^{x''} F'' = \Delta_{x_n'}^{x_n''} \min \{F_1(x_1''), \dots, F_{n-1}(x_{n-1}''), F_n(x_n)\}$$

which is clearly non-negative

b) $F^* = F' \Rightarrow$ at least $n-1$ among the n d.f. F_i are degenerate. Suppose that there are two marginal d.f.'s, say F_1 and F_2, which are not degenerate. Then there exist two values x_1', x_2', such that $0 < F_1(x_1') < 1$, $0 < F_2(x_2') < 1$. If $F^* = F'$, letting $x_i \to +\infty$ for $i > 2$, we obtain

$$\max \{F_1(x_1) + F_2(x_2') - 1, 0\} = F_1(x_1')F_2(x_2').$$

Since the right member cannot vanish, we have

$$F_1(x_1') + F_2(x_2') - 1 = F_1(x_1')F_2(x_2')$$

and therefore:

$$[1-F_1(x_1')][1-F_2(x_2')]=0$$

which proves the inconsistency of the assumptions.

The same argument applies if we start from $F^* = F''$.

The proof of the theorem will terminate with the following:

LEMMA 1: *If $F \in \Gamma(F_1, \dots, F_n)$, then for ever $x = (x_1, \dots, x_n)$*

$$F(x_1, \dots, x_n) \geqslant \Sigma_i F_i(x_i) - n + 1 \,. \tag{7}$$

Equality in (7) holds iff, for every pair i, j

$$F_{ij}(x_i, x_j) = F_i(x_i) + F_j(x_j) - 1 \,. \tag{8}$$

Considering n r.v.'s $X_1, \dots X_n$ whose joint d.f. is F we have:

$$F(x_1, \dots, x_n) = P\{\cap_i (X_i < x_i)\} = 1 - P\{\cup_i (X_i \geqslant x_i)\} \geqslant$$
$$\geqslant 1 - \Sigma_i P\{X_i \geqslant x_i\} = 1 - n + \Sigma_i F_i(x_i)$$

which proves (7). Condition (8) follows from the fact that, in the inequality above, equality holds iff, for every pair i, j, $P\{X_i \geqslant x_i, X_j \geqslant x_j\} = 0$.

* * *

When sets A in $\mathcal{A}$ are pairwise disjoint, the existence of the independence d.f. $F^*(x) = \prod_{A \in \mathcal{A}} F_A(x_A)$ assures again that $\Gamma(F_A, A \in \mathcal{A})$ is not void, also if sets A contain more than one element. But in this case Γ will not have, in general, a maximum d.f.. This can be seen, for instance, in R^3, if F_1 and F_{23} are given. Then a d.f. $F(x_1, x_2, x_3)$ belongs to $\Gamma(F_1, F_{23})$ iff:

$$max\,\{F_1(x_1) + F_{23}(x_2, x_3) - 1, 0\} \leqslant F(x_1, x_2, x_3) \leqslant min\,\{F_1(x_1), F_{23}(x_2, x_3)\}$$

but the functions on the two sides may not be d.f. (see [31], 2.6.1).

3. Compatibility of distribution functions.

As soon as sets A of $\mathcal{A}$ are not pairwise disjoint, the question arises whether the class $\Gamma(F_A, A \in \mathcal{A})$ is not void.

We will say that d.f.'s. F_A, $A \in \mathcal{A}$ are *compatible* if $\Gamma(F_A, A \in \mathcal{A})$ is not void, i.e. if there exists a d.f. $F(x_1, \dots, x_n)$ such that $F_A(x_A, +\infty) = F_A(x_A)$ for every $A \in \mathcal{A}$.

A necessary condition for compatibility can be immediately established:

$$A_1, A_2 \in \mathcal{A};\ A \subset A_1 \cap A_2 \Rightarrow F_{A_1}(x_A, +\infty) = F_{A_2}(x_A, +\infty). \tag{9}$$

This condition is not always sufficient, as can be easily seen.

* * *

First studies about compatibility where developed in R^3, when d.f.'s $F_{12}(x_1, x_2)$, $F_{13}(x_1, x_3)$, $F_{23}(x_2, x_3)$ are given. Assuming finitess of second moments, J. Bass [1] gave a necessary and sufficient condition on the moments EX_1X_2, EX_1X_3, EX_2X_3. The class of d.f.'s F_{23} compatible with given F_{12}, F_{13} is studied in [7], through the class $\Gamma(F_{12}, F_{13})$. It is first shown (under assumption of validity of (9) for F_{12}, F_{13}, i.e. $F_{12}(x_1, +\infty) = F_{13}(x_1, +\infty)$) that all the d.f.'s of $\Gamma(F_{12}, F_{13})$ are given by

$$F(x_1, x_2, x_3) = \int_{-\infty}^{x_1} F_{23}(x_2, x_3 | x_1)\, dF_1(x_1) \tag{10}$$

where $F_{23}(x_2, x_3|x_1)$ belong to the Fréchet class $\Gamma(F_2(x_2|x_1), F_3(x_3|x_1))$ determined by the conditional d.f.s. $F_2(x_2|x_1)$, $F_3(x_3|x_1)$ of F_{12}, F_{13}. Then all the d.f.s F_{23} compatible with given F_{12}, F_{13} are obtained letting $x_1 \to +\infty$ in (10). They are thus given by

$$F_{23}(x_2, x_3) = \int_{-\infty}^{+\infty} F_{23}(x_2, x_3|x_1)\, dF_1(x_1)$$

and this result permits to obtain some properties of the class of d.f.'s F_{23} compatible with F_{12} and F_{13}. This class has a maximum function given by

$$\int_{-\infty}^{+\infty} min\{F_2(x_2|x_1), F_3(x_3|x_1)\} dF_1(x_1)$$

and a minimum one, given by

$$\int_{-\infty}^{+\infty} max\{F_2(x_2|x_1) + F_3(x_3|x_1) - 1, 0\} dF_1(x_1).$$

The class is not convex; it is easy (see [7]) to give counter-examples. Some extension to four dimensions of these results are given by V. Pagani in [30].

* * *

As the example seen above shows, condition (9) is sufficient in some cases to assure compatibility. This example can be extended to more general situations.

A comprehensive result has been given in [24] by H. Kellerer. He proves that, in order that condition (9) be sufficient for compatibility, it is necessary and sufficient that the family $\mathcal{A}$ possess the following structure: it is possible to order the sets A of $\mathcal{A}$ in a way $A_1, \ldots, A_l$ such that:

$$A_k \cap \Big(\bigcup_{j<k} A_j\Big) \in \bigcup_{j<k} \mathcal{T}(A_j) \qquad \text{for } 1 \leqslant k \leqslant l$$

where $\mathcal{T}(A_j)$ denotes the family of subsets of A_j.

* * *

In the same paper H. Kellerer solves the general problem of compatibility, proving the following result: a necessary and sufficient condition for compatibility of d.f.'s $\{F_A, A \in \mathcal{A}\}$ is that

$$\sum_{A\in\mathcal{A}} g_A(x_A) \geqslant 0 \Rightarrow \sum_{A\in\mathcal{A}} \int_{R^{|A|}} g_A(x_A)\, dF_A(x_A) \geqslant 0 \tag{11}$$

for every set $\{g_A(x_A),\ A \in \mathcal{A}\}$ of continuous bounded functions.

The necessity of (11) is immediate, since the existence of an n-dimensional d.f. F with the given marginals F_A implies

$$\sum_{A\in\mathcal{A}} \int_{R^{|A|}} g_A(x_A)\, dF_A(x_A) = \int_{R^n} \Big(\sum_{A\in\mathcal{A}} g_A(x_A)\Big)\, dF(x)\,.$$

The sufficiency is proved on the basis of a previous result obtained by the same author for general measures [23]. This result is first proved in the finite case (i.e. for given marginal distributions concentrated on a finite set of points): the set of given data of the problem is represented as a point of a space R^m; it is shown that the set of points of R^m for which the problem is solvable forms a convex cone, and then a necessary and sufficient condition is given in order that a point belong to this cone. Then the general condition is obtained by limit on suitable sequence of « finite case » given marginal measures.

Kellerer investigates also whether it is possible to restrict condition (11) to functions g assuming only values $-1, 0, 1$, and obtains that for a particular case, which will be seen later. In general he proves that if given marginal measures are concentrated on a finite number of points, (11) is sufficient for compatibility if functions g_A are integer-valued, with $|g_A| \leq \gamma$, where γ is determined from the given marginal measures. The same problem is investigated by G. Bernardini [2] for compatibility of F_{12}, F_{13}, F_{23} in R^3, obtaining only partial results: he proves that, if one-dimensional distributions F_1, F_2, F_3 are concentrated on at most three points, it is sufficient for compatibility that

$$g_{23}(x_2, x_3) \leq g_{12}(x_1, x_2) + g_{13}(x_1, x_3) \Rightarrow \iint g_{23}\, dF_{23} \leq \iint g_{12}\, dF_{12} + \iint g_{13}\, dF_{13}$$

for g_{12}, g_{13}, g_{23} assuming only values 0 and 1, i.e. for set indicators.

* * *

Kellerer extends also his results to infinite index sets I, giving a slight generalization of Kolmogorov theorem, in the sense that joint distributions are not required for every finite set of indices, but for a slightly smaller class of sets.

Such results are extended by V. Strassen [36] to more general spaces. He finds also a condition for existence of martingales with given marginals: given a sequence $\{\mu_n\}$ of probability distribution in R^k, a k-dimensional martingale with marginals μ_n exists iff distributions μ_n possess finite expectations and for any concave function f on R^k the sequence $\int f d\mu_n$ is non increasing.

4. Compatibility of minimum distribution functions.

It is interesting to examinate when two-dimensional d.f.'s minimum in their Fréchet classes are compatible, that is under what conditions, given n one-dimensional d.f.'s. $F_1, \ldots, F_n$, the $\binom{n}{2}$ d.f.'s. $F'_{ij}(x_i, x_j) = \max\{F_i(x_i) + F_j(x_j) - 1, 0\}$ are compatible.

The question arises since if two r.v.'s. X_2, X_3 have maximum negative correlation to a third r.v. X_1, one should expect that X_2 and X_3 are positively correlated. As a matter of fact in [7] it is shown that, given $F_{1i} = \max\{F_1 + F_i - 1, 0\}$ $(i = 2, 3)$, under mild conditions (e.g. F_1 continuous) the only d.f. $F_{23}(x_2, x_3)$ compatible with them is $F_{23} = \min\{F_2, F_3\}$. This result is easily achieved using the representation of section 3 ((10) and following remarks).

It is therefore clear that compatibility of minimum d.f.'s. brings severe restrictions on the d.f.'s. F_i. The problem is studied in [8] in the three-dimensional case; the results there obtained can be extended to R^n as it is shown in following Theorem 3.

First a lemma will be given, which is useful for the proof of Theorem 3, and clarifies the structure of n-dimensional distributions when the two-dimensional marginal d.f.'s. are minimum.

LEMMA 2: *If in* R^n *the* $\binom{n}{2}$ *d.f.'s* $F'_{ij} = max\{F_i + F_j - 1, 0\}$ $(i, j = 1, 2, \dots, n;\ i \neq j)$ *are compatible, and at least three among the* F_i *are not degenerate, then the corresponding n-dimensional distribution is concentrated on a set formed by n ortogonal half-lines parallel to the axes, starting from the same point either all in the positive direction or all in the negative one.*

The essential idea of the proof is that the projection of the support of the n-dimensional distribution on every coordinate plane must belong to a non-increasing curve, and the only sets which have this property are the two sets described by the lemma. In order not to cumber excessively the calculations, the proof will be given for atomic distributions.

Let us consider two non-null (i.e. of positive probability) points $(x'_1, \dots x'_n)$, $(x''_1, \dots, x''_n)$ with $x'_i \neq x''_i$ for $i = 1, 2$. Since the two-dimensional d.f.'s are minimum, by remarks after Theorem 1 points (x'_1, x'_2) and (x''_1, x''_2) must be on a non-increasing curve, say $x'_1 < x''_1$, $x'_2 > x''_2$. Now on the (x_1, x_i) plane $(i \neq 1, 2)$ for the same reason it must be $x'_i \geqslant x''_i$, and on the (x_2, x_i) plane $x'_i \leqslant x''_i$, so that $x'_i = x''_i$, and the considered points belong to a plane parallel to the coordinate plane (x_1, x_2).

Let us now assume that at least three of d.f.'s. F_i don't degenerate, say F_1, F_2, F_3. Then there must be at least two non-null points $(x'_1, \dots, x'_n)$ and $(x''_1, \dots, x''_n)$, with the two first coordinates different, say $x'_1 < x''_1$, $x'_2 > x''_2$, and, as it has been shown, they must belong to a plane of equations, say, $x_i = y_i$ $(i = 3, 4, \dots, n)$. The non-null points outside this plane must have the first two coordinates either all equal to (x'_1, x''_2) or all equal to (x''_1, x'_2), and at least one such point exists, since F_3 do not degenerate.

If the first case applies, putting $x'_1 = y_1$, $x''_2 = y_2$, we may affirm that the distribution is concentrated on a set formed by the two positive half-lines of plane $x_i = y_i (i = 3, 4, \dots, n)$, parallel to the axes, starting from (y_1, y_2) with positive direction, and the $(n-2)$-dimensional space orthogonal to that plane and passing through the point $(y_1, \dots, y_n)$. By same argument applied to other coordinate planes, we may conclude that the distribution is concentrated on the n positive half-lines parallel to the axes, issued from $(y_1, \dots, y_n)$. If on the contrary the first two coordinates are equal to (x_1'', x_2'), we obtain the same result except that all the n orthogonal half-lines have negative direction.

The lemma is thus proved.

THEOREM 3: *For n one-dimensional d.f.'s F_i, following statements are equivalent*:

i) $max\{F_1 + \dots + F_n - n + 1, 0\}$ *is a d.f.*

ii) $\Gamma(F_1, \dots, F_n)$ *has a minimum d.f.*

iii) *The* $\binom{n}{2}$ *d.f.'s* $F'_{ij} = max\{F_i + F_j - 1, 0\}$ *are compatible*

iv) *If at least three among F_i are not degenerate, then putting:*

$$z_i' = inf\{x_i : F_i(x_i) > 0\}, \; z_i'' = sup\{x_i : F_i(x_i) < 1\}$$

either

$$F_1(z_1' +) + \dots + F_n(z_n' +) \geq n - 1 \tag{12}$$

or

$$F_1(z_1'') + \dots + F_n(z_n'') \leq 1 \tag{13}$$

Obviously i) implies ii) and iii). We prove now the other implications

a) ii) *implies* iii). If $\Gamma(F_1, \dots, F_n)$ has a minimum d.f. $G(x_1, \dots, x_n)$, then for every pair i, j it is:

$$G(x_1, \dots, x_n) \leq max\{F_i(x_i) + F_j(x_j) - 1, 0\} \prod_{s \neq i,j} F_s(x_s)$$

since the function on the right belongs to $\Gamma(F_1, \dots, F_n)$; hence, letting $x_s \to +\infty$ for $s \neq i, j$,

$$G_{ij}(x_i, x_j) = G(x_i, x_j, +\infty) \leq max\{F_i(x_i) + F_j(x_j) - 1, 0\}.$$

But also the inverse inequality holds, since $G_{ij} \in \Gamma(F_i, F_j)$, so that the d.f.'s $max\{F_i + F_j - 1, 0\}$ are marginals of G, and they are compatible.

b) iii) *implies* iv). If iii) holds, and there are at least three non degenerate d.f.'s. F_i, Lemma 2 applies. We assume that the first case hold, i.e. the n-dimensional underlying distribution is concentrated on n half lines parallel to the axes, starting from $(y_1, \dots, y_n)$ in the positive direction.

Now it is clearly, for every i, $F_i(y_i) = 0$, $F_i(y_i + \varepsilon) > 0$ for $\varepsilon > 0$, so that $y_i = inf\{x_i : F_i(x_i) > 0\} = z_i'$. Moreover, if $x_i > y_i$ for every i, it follows from the structure of the support that, for every pair i, j:

$$1 - F_i(x_i) - F_j(x_j) + F_{ij}(x_i, x_j) = Pr\{X_i > x_i, X_j > x_j\} = 0$$

so that, by Lemma 1,

$$F_1(x_1) + \dots + F_n(x_n) - n + 1 = F(x_1, \dots, x_n) \geqslant 0$$

and (12) holds.

By same argument, if the half-lines of Lemma 2 have negative direction, (13) follows. The assertion is thus proved.

c) iv) *implies* i). We have to prove that the variation of $F = max\{F_1 + \dots + F_n - n + 1, 0\}$ is non-negative. This is easily seen if at most two among the F_i are not degenerate.

Let us now assume that (12) holds. If $F_i(x_i) = 0$ for at least one i, $F_1(x_1) + \dots + F_n(x_n) \leqslant n - 1$ and $max\{F_1(x_1) + \dots + F_n(x_n) - n + 1, 0\} = 0$. If on the contrary $F_i(x_i) > 0$ for all i, by (12) $max\{F_1(x_1) + \dots + F_n(x_n) - n + 1, 0\} = F_1(x_1) + \dots + F_n(x_n) + n - 1$.

Consider now the variation $\Delta_{x'}^{x''} F'$, with $x' = (x_1', \dots, x_n')$; $x'' = (x_1'', \dots, x_n'')$; $x_i' < x_i''$. If $x_i'' \leqslant z_i'$ for at least one i, all terms vanish and $\Delta_{x'}^{x''} F' = 0$. Assume then $x_i'' > z_i'$ for every i. If $x_i' \leqslant z_i'$ for every i, $\Delta_{x'}^{x''} F' = max\{F_1(x_1'') + \dots + F_n(x_n'') - n + 1, 0\} \geqslant 0$. If $x_i' > z_i'$ for exactly $s (1 \leqslant s \leqslant n)$ indices i, say $i = 1, 2, \dots, s$, then

$$\Delta_{x'}^{x''} F' = \Delta_{x_1 x_1}^{x_1''} \dots \Delta_{x_s x_s}^{x_s''} [F_1(x_1) + \dots + F_s(x_s) + F_{s+1}(x_{s+1}'') + \\ + \dots + F_n(x_n'') - n + 1]$$

and it is easily seen that

$$\Delta_{x'}^{x''} F' = F_1(x_1'') - F_1(x_1') \geqslant 0$$

if $s = 1$, while $\Delta_{x'}^{x''} F' = 0$ if $s > 1$.

Finally let us assume that (13) holds.

Then if $x_i \leqslant z''$ for at least two variables, say x_1 and x_2, it follows from (13):

$$F_1(x_1) + \dots + F_n(x_n) - n + 1 \leqslant 1 + F_3(x_3) + \dots + F_n(x_n) - n + 1 \leqslant 0$$

Moreover the variation $\Delta_{x'}^{z''} F'$ vanishes if $x'_i > z''_i$ for at least one i, since (take $x'_1 > z''_1$):

$$\Delta_{x'}^{z''} F' = \Delta_{x_2}^{x''_2} \dots \Delta_{x_n}^{x''_n} \left[max \left\{ F_2(x_2) + \dots + F_n(x_n) - n + 2, 0 \right\} - max \left\{ F_2(x_2) + \dots + F_n(x_n) - n + 2, 0 \right\} \right] = 0 .$$

Therefore $\Delta_{x'}^{z''} F'$ can be different from zero only if $x'_i < z''_i$ for all i, and $x''_i < z''_i$ for at most one i. If $x''_i < z''_i$ only for $i = j$, the only non vanishing terms are when $x_i = x''_i$ for $i \neq j$, and then $\Delta_{x'}^{z''} F' = F_i(x''_j) - F_j(x'_j) \geq 0$. If $x''_i > z''_i$ for all i, the only non vanishing terms are when $x_i = x''_i$ for all i, or $x_i = x'_i$ for only one i, so that

$$\Delta_{x'}^{z''} F' = 1 - F_1(x'_1) - \dots - F_n(x'_n) \geq 0$$

The assertion is thus proved and the proof of the theorem is completed.

* * *

Lemma 2 makes clear the nature of condition (12): the d.f.'s $F_i(x_i)$ must start their increase by a jump, and the sum of the jumps must be large enough in order to obtain a distribution of the type described by the lemma.

Condition (13) is clearly symmetric of (12), in the sense that is obtained from it changing sign to all the variables. To symmetry of conditions on the F_i's corresponds the symmetry of the n-dimensional distributions.

From Lemma 2, and from the proof of *c*) of Theorem 3, which confirm it, one deduces as corollary that, if at least three of the d.f.'s F_i are not degenerate, the class $\Gamma(max\{F_i + F_j - 1, 0\}; i, j = 1, 2, \dots, n; i \neq j)$ contains only one function. This shows that, when the given marginal d.f.'s are not one-dimensional, Γ may contain only one function even if one-dimensional d.f.'s are not degenerate. In the case considered, of course, the given two-dimensional marginal d.f.'s are in some way degenerate in the sense that, for each one of them, distribution is concentrated on a curve.

5. Fréchet classes with bounds.

In connection with some problems in linear programming, M. Fréchet posed the following question [14]. Given $r + s + rs$ non negative numbers N_i, N'_j and m_{ij} ($i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$), with $\Sigma_i N_i = \Sigma_j N'_j = N$, is it possible to find rs non negative numbers n_{ij} such that

$$\Sigma_i n_{ij} = N'_j, \quad \Sigma_j n_{ij} = N_i; \; n_{ij} \leqslant m_{ij}?$$

If $\{N_i\}$ and $\{N'_j\}$ are regarded as marginal frequency or probability distributions, the question corresponds to ask whether, given two one-dimensional d.f.'s $F_1(x_1)$, $F_2(x_2)$, each one concentrated on a finite number of values $\{x_1^{(i)}\}$, $\{x_2^{(j)}\}$, and non negative numbers m_{ij}, there exists in $\Gamma(F_1, F_2)$ a d.f. under which $P\{X_1 = x_1^{(i)}, X_2 = x_2^{(j)}\} \leqslant m_{ij}$.

Fréchet investigated the conditions for existence, finding the necessary condition

$$\sum_{i\in I} \sum_{j\in J} m_{ij} - \sum_{i\in I} N_i - \sum_{j\in J} N'_j + N \geqslant 0 \tag{14}$$

for every I and J, where $I(J)$ is a subset of the set of the first $r(s)$ natural numbers.

Fréchet proved also the sufficiency of condition (14) for particular cases, studying also, in these cases, the class of solutions ([14], [17], [18]).

The sufficiency, in general, of condition (14) was proved in [9], where a method to build a solution was presented. The proof is by induction on the number of lines of the matrix, and it shown that if N_i, N'_j, m_{ij} are integer, also the particular solution constructed is integer-valued.

A simpler proof was given later by F. Stivali [37], which, on a hint by C. Berge, makes use of results in graph theory.

The proof of sufficiency of (14) is also given by H. Kellerer in [22], where, in a measure-theoretic framework, more general results are obtained. The main theorem, established for more general spaces, in our framework states that, given two one-dimensional d.f.'s F_1, F_2, and a (non negative) finite measure $\varrho(A)$ on R^2, in order that there exists a d.f. $F \in \Gamma(F_1, F_2)$ with

$$\iint_A dF(x_1, x_2) \leqslant \varrho(A)$$

for every Borel set A in R^2, it is necessary and sufficient that

$$\varrho(A_1 \times A_2) - \int_{A_1} dF_1(x_1) - \int_{A_2} dF_2(x_2) + 1 \geqslant 0 \tag{15}$$

for every pair A_1, A_2 of Borel sets in R.

This condition is the direct extension of (14). The proof of sufficiency is given starting from the result for the « finite case » i.e. from (14), and extending it, by limit on suitably chosen sequences of measures, first to the « denumerable » case and then to the general one.

* * *

It is easy to see (for the « finite case » it is shown in [9], and the extension to the general case is immediate) that the problem of Fréchet classes with bounds can be stated as a problem of compatibility of three two-dimensional d.f.'s, where one of the one-dimensional distributions is concentrated in two points. In this framework (15) appears as condition (11) restricted a particular kind of functions g_A which assume only values $-1, 0, 1$; more precisely, to functions $g_{12}, g_{13}, -g_{23}$, where g_{ij} are indicators of intervals in R^2.

6. Applications: distance between distributions.

In this section we will discuss the application of Fréchet classes to distance between distributions. First we recall some other applications.

Connections with linear programming, particularly the transport problem, have been already hinted. They have been studied by M. Fréchet [15], A. Herzel [20] and R. Feron [12]. Condition have been found under which the « N. W. corner rule », corresponding to the maximum function in the Fréchet class, gives the solution of the linear programming problem; and ways have been suggested to restrict the class of admissible solutions among which the optimal solution must be searched for.

Fréchet classes have been also applied in [10] to the study of relations between convergence in distribution and in probability. It is easily proved that the sequence of r.v.'s X_n, with d.f.'s F_n, tends in probability to X, with d.f. F, iff

$$G_n(x, y) = P\{X < x, X_n < y\} \to \min\{F(x), F(y)\} .$$

It is thus clarified the intuitive idea that X_n converges in probability to X if it converges in distribution and the correlation between X_n and X tends to be maximum. In addition, the new form of the convergence permits to obtain immediately various known results.

* * *

Application of Fréchet classes to distance between distributions starts from the study of « dissimilarity index » firstly introduced by

C. Gini in [19] for discrete frequency distributions and investigated, among others, by T. Salvemini in [32], [33].

The dissimilarity index is defined as the minimum of

$$\Sigma_{i,j}|x_i - x_j| f_{i,j}$$

with respect to the joint frequency distribution, given by $f_{i,j}$, when the marginal frequencies f_i, f'_j are fixed. It is shown that the minimum is attained for the joint frequency distribution corresponding to maximum association, i.e. for the « tabella di cograduazione », which corresponds to the maximum d.f. in the Fréchet class.

In the same way, starting from the square of the differences, the « quadratic dissimilarity index » is defined.

This study was continued later by G. Landenna ([25], [26]) who, utilizing Fréchet classes, gave a systematic treatment of the study of the index, extending it to general distributions and pointing out its distance properties.

The extension is based on the following results, obtained in [6].

Given two one-dimensional r.v.'s X, Y, with d.f.'s F_1, F_2, and joint d.f. F, we denote by

$$d_\alpha(X, Y) = d_\alpha(F) = E_F|X - Y|^\alpha = \iint_{-\infty}^{+\infty} |x - y|^\alpha \, dF(x, y) \tag{16}$$

the moment of order α with respect to the line $x = y$. It is assumed that X^α and Y^α are integrable.

For $\alpha = 1$, splitting the integral in the sum of the integrals for $x > y$ and for $x < y$, putting (for $x > y$)

$$x - y = \int_y^x dt$$

and changing the order of integration, it is easily obtained that:

$$d_1(X, Y) = d_1(F) = \int_{-\infty}^{+\infty} [F_1(z) + F_2(z) - 2F(z, z)] \, dz . \tag{17}$$

For $\alpha > 1$, in a slightly more complicate way, one obtains:

$$d_\alpha(X, Y) = d_\alpha(F) = \alpha(\alpha - 1) \iint_{u>v} [F_2(v) - F(u, v)](u - v)^{\alpha-2} \, du \, dv + \alpha(\alpha - 1) \iint_{u<v} [F_1(u) - F(u, v)](v - u)^{\alpha-2} \, du \, dv . \tag{18}$$

The difference between case $\alpha = 1$, in which only values of F on the line $x = y$ give a contribution, and case $\alpha > 1$, in which all values are involved, accounts for difference in extremal. For $\alpha > 1$, (18) shows immediately that the minimum of $d_\alpha(F)$, when $F \in \Gamma(F_1, F_2)$, is obtained for the maximum function of $\Gamma(F_1, F_2)$, i.e. when $F = F'' = min\{F_1(x), F_2(y)\}$. The maximum of $d(F)$ is given by $F = F'$ but this will not insisted on, since, as it has been already remarked, we can pass from maximum to minimum d.f.'s changing sign one variable.

For $\alpha = 2$, (18) was already obtained by W. Hoeffding [21], and since $d_2(F) = EX^2 + EY^2 - 2EXY$, it shows that the maximum of correlation coefficient, for given marginal distributions, is obtained when joint d.f. is F''.

For $\alpha = 1$ the minimum is obtained when $F(x, y)$ is maximum for $x = y$, and there is a class $\Gamma^*(F_1, F_2) \subset \Gamma(F_1, F_2)$ of d.f.'s satisfying this condition. $\Gamma^*(F_1 F_2)$ has obviously F'' as maximum d.f.; it has also a minimum d.f. F^{**} given by

$$F^{**}(x, y) = \begin{cases} F_1(x) - max\left\{ \inf\limits_{x \leqslant z \leqslant y} [F_1(z) - F_2(z)],\ 0\right\} & \text{if } x \leqslant y, \\ F_2(y) - max\left\{ \inf\limits_{y \leqslant z \leqslant x} [F_2(z) - F_1(z)],\ 0\right\} & \text{if } x > y. \end{cases}$$

It is easily seen that for $F \in \Gamma^*$, i.e. if $F(z, z) = min\{F_1(z), F_2(z)\}$, the value of $d_1(F)$ is

$$\min_{F \in \Gamma} d_1(F) = \int_{-\infty}^{+\infty} |F_1(z) - F_2(z)|\, dz\,.$$

According to discussion above, this is the dissimilarity index between X and Y, and it appears as a reasonable definition of a distance between F_1 and F_2.

For $\alpha > 1$ the minimum has not a simple form; some expressions for it are given by M. Fréchet in [16].

A more general definition of distance between distributions, along the same lines, is given by P. Levy in [28]. He proves that, given a distance $d(X, Y) = d(F)$ between r.v.'s, the minimum of $d(X, Y)$ when all the joint distributions of X, Y are considered (i.e. the minimum of $d(F)$ when $F \in \Gamma(F_1, F_2)$) has again the properties of a distance, and thus can be considered as a distance between d.f.'s.

M. Fréchet [16] develops this idea defining a class of distances between r.v.'s as

$$d(X, Y) = d(F) = Ef(|X - Y|) \tag{19}$$

where $f(z)$, for $z \geq 0$, is an increasing sub-additive function with $f(0) = 0$; a corresponding class of distance between d.f.'s remains defined as said above. He suggests also, according to the results above, that a distance between d.f.'s can be defined, starting from a distance between r.v.'s, and choosing F'' as their joint d.f.'s.

Further results are obtained by S. Bertino in [11] (see also [31], [8]). He considers a distance between r.v.'s defined as above, with function f possessing first and second derivatives, and by a transformation which generalizes (17) and (18) shows that if the second derivative f'' of f is non-negative, the minimum of the distance is obtained by $F = F''$. Moreover, in a paper yet to appear, he proves (under some limitations) that, if f'' is non-positive, the minimum is obtained by the d.f. F^{**} given above.

An interesting property of d.f. F^{**} is that the underling distribution has the maximum concentration on the line. That means, for discrete distributions, that for every z:

$$P\{X = z,\ Y = z\} = min\,[P\{X = z\},\, P\{Y = z\}]$$

and a similar property holds for other kind of distributions. Such property is not possessed by F'', for which, as for all d.f.'s of Γ^*, it is for every z:

$$P\{X < z,\ Y < z\} = min\,[P\{X < z\},\, P\{Y < z\}]\,.$$

The results above shows the important role played by the class $\Gamma^*(F_1, F_2)$ in minimizing $d(F)$ in (19). If $f'' \geq 0$ (but not identically zero), the only minimizing d.f. is the maximum function F'' of Γ^*; if f'' is identically zero all the d.f.'s of Γ^* give the minimum value of $d(F)$; if $f'' \leq 0$ the only minimizing d.f. is the minimum, F^{**}, of Γ^*.

Testo pervenuto il 26 maggio 1971.
Bozze licenziate il 17 aprile 1972.

REFERENCES

[1] J. Bass, *Sur la compatibilité des fonctions de répartition*, C. R. Academie de Sciences, 240 (1955), 839-841.
[2] G. Bernardini: Unpublished Thesis (1969).
[3] S. Bertino, *Una generalizzazione della dissomiglianza di C. Gini tra variabili casuali semplici*, Giorn. Ist. It. Attuari, 29 (1966), 155-178.

[4] S. Bertino, *Su di una sottoclasse della classe di Fréchet*, Statistica, 28 (1968), 511-542.

[5] S. Bertino, *Sulla distanza tra distribuzioni*, Pubbl. Ist. Calc. Prob. Univ. Roma, n. 82 (1968).

[6] G. Dall'Aglio, *Sugli estremi dei momenti delle funzioni di ripartizione doppia*, Ann. Sc. Norm. Sup., 10 (1956), 35-74.

[7] G. Dall'Aglio, *Sulla compatibilità delle funzioni di ripartizione doppia*, Rendiconti di Mat., 18 (1959), 385-413.

[8] G. Dall'Aglio, *Les fonctions extrêmes de la classe de Fréchet à trois dimensions*, Publ. de l'Inst. de Stat. de l'Univ. de Paris, 9 (1960), 175-188.

[9] G. Dall'Aglio, *Sulle distribuzioni doppie con margini assegnati soggette a delle limitazioni*, Giorn. Ist. Ital. Attuari, 24 (1961), 94-108.

[10] G. Dall'Aglio, *Osservazioni sulla convergenza in distribuzione e in probabilità*, Giorn. Ist. Ital. Attuari, 29 (1966), 255-269.

[11] R. Feron, *Sur les tableaux de corrélation dont les marges sont données, cas de l'espace à trois dimensions*, Publ. Inst. Stat. Univ. Paris, 5 (1956), 3-12.

[12] R. Feron, *Tableaux de corrélation dont les marges sont données et programmes linéaires*, Publ. Inst. Stat. Univ. Paris, 12 (1963), 103-116.

[13] M. Fréchet, *Sur les tableaux de corrélation dont les marges sont données*, Annales de l'Université de Lyon, Sciences, 4 (1951), 13-84.

[14] M. Fréchet, *Les tableaux de corrélation dont les marges et des bornes sont données*, Annales Univ. Lyon, Sciences, 20 (1957), 13-31.

[15] M. Fréchet, *Les tableaux de corrélation et les programmes linéaires*, Revue Inst. Int. Stat., 25 (1957), 23-40.

[16] M. Fréchet, *Sur la distance de deux lois de probabilité*, Publ. Inst. Stat. Univ. Paris, 6 (1957), 185-198.

[17] M. Fréchet, *Sur les tableaux de corrélation dont les marges et des bornes sont données*, Annales Univ. Lyon, Sciences, 21 (1958), 19-32.

[18] M. Fréchet, *Sur les tableaux de corrélation dont les marges et des bornes sont données*, Revue Inst. Int. de Statistique, 28 (1960), 10-32.

[19] C. Gini, *Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche*, Att. R. Ist. Veneto Sc. Lettere, Art, 74 (1914), 185-213.

[20] A. Herzel, *Le tabelle di co- e contrograduazione e la programmazione lineare*, Metron, 21 (1961), 186-200.

[21] W. Hoeffding, *Massstabinvariante Korrelationtheorie*, Schriften Math. Inst. Univ. Berlin, 5 (1940), 181-233.

[22] H. G. Kellerer, *Funktionen auf Produkträumen mit vorgegebenen Marginal-Funktionen*, Math. Annalen, 144 (1961), 323-344.

[23] H. G. Kellerer, *Masstheoretische Marginal probleme*, Math. Annalen, 153 (1964), 168-198.

[24] H. G. Kellerer, *Verteilungsfunktionen mit gegebenen Marginalverteilungen*, Z. Wahrscheinlickkeitstheorie, 3 (1964), 247-270.

[25] G. Landenna, *La dissomiglianza*, Statistica, 16 (1956), 21-57.

[26] G. Landenna, *Osservazioni sulla connessione*, Statistica, 17 (1957), 351-392.

[27] G. Letac, *Répresentation des mesures de probabilité sur le produit de deux espaces denombrambles de marges données*, Illinois J. Math., 10 (1966), 497-507.

[28] P. Levy, *Distance de deux variables aléatoires et distance de deux lois de probabilité*, in Fréchet, *Généralités sur les probabilités. Eléments aléatoires*, Traité de probabilité dirigé par E. Borel, Gauthier-Villars, Paris (1950).

[29] A. Nataf, *Etude graphique de détermination de distributions de probabilité planes dont les marges sont données*, Publ. Inst. Stat. Univ. Paris, 11 (1962), 257-260.

[30] V. Pagani, Unpublished Thesis (1963).
[31] G. Pompilj, *Le variabili casuali*, fasc. II, Veschi, Roma (1970).
[32] A. Rizzi, *Osservazioni sulle classi di Fréchet a più variabili*, Boll. Unione Mat. Ital., 12 (1957), 263-277.
[33] T. Salvemini, *Sugli indici di omofilia*, Atti I Riunione della Società Italiana di Statistica (1939), 105-115.
[34] T. Salvemini, *Nuovi procedimenti per il calcolo degli indici di dissomiglianza e di connessione*, Statistica (1949), 3-26.
[35] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Stat. Univ. Paris, 8 (1959), 229-231.
[36] V. Strassen, *The existence of probability measures with given marginals*, Annals Math. Stat., 36 (1965), 423-439.
[37] F. Stivali, Unpublished Thesis (1969).