Fonctions de répartition à N dimensions et leurs marges

M. Sklar

FONCTIONS DE RÉPARTITION A N DIMENSIONS ET LEURS MARGES

M. SKLAR

M. M. Fréchet [1] a étudié le problème de la détermination des fonctions de répartition à n dimensions dont les marges sont données. En ce qui suit nous considérons ce problème d'un point de vue nouveau.

Pour les définitions des fonctions de répartition, des marges, etc., voir p. 67 dans le premier ouvrage cité en (1). Les définitions qu'on y trouve sont exprimées pour n = 2, mais elles peuvent être étendues à n quelconque.

THEOREME 1 -

Soit G_n une fonction de répartition à n dimensions avec les marges F_1, F_2, ..., F_n. Soit R_k l'ensemble des valeurs de F_k, pour $k = 1, 2, \ldots, n$. Alors il existe une fonction unique H_n, dont l'ensemble de définition est le produit Cartésien $R_1 \times R_2 \times \ldots \times R_n$ et telle que :

$$G_n(x_1, \ldots, x_n) = H_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

<u>Définition 1</u> :

Nous appellerons <u>copule</u> (à n dimensions) toute fonction C_n continue et non-décroissante [2] définie sur le produit Cartésien de n intervalles fermés [0, 1] et satisfaisant aux conditions :

----

(1) M. Fréchet : sur les tableaux de corrélation dont les marges sont données, Ann. Univ. Lyon, Sect. A (3), 14, 1951, 53-77. Voir aussi M. Fréchet, Comptes Rendus Acad. Sci., Paris, 242, 1956, 2426-2428 ; E. J. Gumbel, Ibid., 246, 1958, 2712-2719 ; R. Féron, Publ. Inst. Statistique Univ. Paris, 5, 1956, 3-12.

(2) "Non-décroissante" au sens employé pour une fonction de répartition à n dimensions.

$$C_n(0, 0, \ldots, 0) = 0, \; C_n(1, \ldots, 1, \alpha, 1, \ldots, 1) = \alpha^{(1)}.$$

THEOREME 2 -

La fonction H_n de Théorème 1 peut être prolongée (en général de plusieurs façons) à une copule C_n. Etant un prolongement de H_n, la copule C_n satisfait à la condition :

$$G_n(x_1, \ldots, x_n) = C_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$$

THEOREME 3 -

Soient données des fonctions de répartition $F_1, F_2, \ldots, F_n$ à une dimension. Soit C_n une copule quelconque à n dimensions. Alors la fonction G_n définie par :

$$G_n(x_1, \ldots, x_n) = C_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$$

est une fonction de répartition à n dimensions avec les marges $F_1, F_2, \ldots, F_n$.

Les Théorèmes 1 - 3 réduisent le problème mentionné de M. Fréchet au problème de la caractérisation des couples à n dimensions. Il n'y a qu'une seule copule à une dimension, notamment P_1, qui satisfait la condition $P_1(x) = x$ pour tout x entre 0 et 1. Pour $n > 1$, il y a une infinité de copules de dimension n, limitées par deux fonctions particulières.

THEOREME 4 -

Une copule quelconque C_n à n dimensions satisfait aux inégalités :

$$L_n \leqslant C_n \leqslant M_n,$$

où :

$$L_n(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathrm{Max}(\alpha_1 + \alpha_2 + \ldots + \alpha_n - n + 1, 0)$$

$$M_n(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathrm{Min}_n(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Pour $n \geqslant 1$, les fonctions P_n, définies par :

$$P_n(\alpha_1, \alpha_2, \ldots, \alpha_n) = \alpha_1 \alpha_2 \ldots \alpha_n \; (0 \leqslant \alpha_k \leqslant 1, \; k = 1, 2, \ldots, n)$$

sont des copules à n dimensions. Elles déterminent le tableau de corrélation de n variables aléatoires indépendantes.

---

(1) Des cas spéciaux de telles fonctions ont été considérés par M. Féron, op. cit.

Or, les copules sont en général d'une structure plus simple que les fonctions de répartition. On peut utiliser cette simplicité en introduisant la notion du "quasi-inverse" d'une fonction monotone.

Soit f une fonction monotone d'une variable réelle. Ajoutons à l'ensemble des points (x, f(x)) tous les segments fermés de la forme [(x, f(x+)), (x, f(x-))]. Réfléchissons l'ensemble résultant dans la ligne y = x. Enlevons de l'ensemble réfléchi tous les segments verticaux à l'exception d'un seul point sur chacun d'eux. Tout ensemble qu'on peut obtenir de cette façon représente une fonction, que nous nommerons une quasi-inverse de la fonction f [1].

Selon cette définition, chaque fonction de répartition F à une dimension possède au moins un quasi-inverse F^*, défini et non-décroissant sur l'intervalle fermé [0, 1] et prenant des valeurs finies sur l'intervalle ouvert (0,1). Parmi ces quasi-inverses, il y en a un qui est continu à gauche sur (0, 1) et continu à 0 et à 1.

THEOREME 5 -

Soit Q_n une fonction de n variables, intégrable dans l'espace euclidien E^n à n dimensions par rapport à une fonction de répartition G_n à n dimensions. Soit C_n une copule telle que l'on ait :

$$G_n(x_1, x_2, \ldots, x_n) = C_n(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$$

où $F_1, F_2, \ldots, F_n$ sont les marges de G_n. Alors :

$$(1) \quad \int_{E^n}\cdots\int Q_n(x_1, x_2, \ldots, x_n)\, d\, G_n(x_1, x_2, \ldots x_n)$$

$$= \int_{U^n}\cdots\int Q_n(F_1^*(\alpha_1), F_2^*(\alpha_2), \ldots, F_n^*(\alpha_n))\, d\, C_n(\alpha_1, \alpha_2, \ldots, \alpha_n),$$

où U^n est le produit Cartésien de n intervalles [0, 1].

L'expression au côté droit de la formule (1) peut être plus simple que celle au côté gauche. C'est le cas, par exemple, si C_n est absolument continue, comme au cas $C_n = P_n$; ou n = 2 et $C_2 = L_2$ ou M_2 ; ou si n = 3 et $C_3 = M_3$.

----

(1) Des formes diverses de telles fonctions quasi-inverses ont été considérées par plusieurs auteurs.