n-Dimensional Distribution Functions And Their Marginals

M. Sklar
translated by Gheis Hamati

December 8, 2004

Abstract

This is a translation of the original article by M. Sklar in the 1959 issue of *Comptes Rendus de L'académie des Sciences de Paris*, entitled "Fonctions de Répartition a N Dimensions et Leurs Marges." This is the first work to coin the word *copula*.

Mr. M. Fréchet has studied the problem of determining distribution function whose marginal distributions are given. In what follows, we consider this problem from a new point of view.

Theorem 1 *Let G_n be an n-dimensional distribution function, having marginal distributions $F_1, F_2, ..., F_n$. Let $\mathbb{R}_k$ be the set of values of $F_k, k = 1, 2, \ldots, n$. Then, there exists a unique function H_n, defined on the Cartesian product $\mathbb{R}_1 \times \mathbb{R}_2 \times \ldots \mathbb{R}_n$, such that*

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

Definition 1 *We will call* copula *(having n dimensions) any continuous and non-decreasing function C_n, defined on $[0,1]^n$, satisfying the following conditions:*

(i) $C_n(0, \ldots, 0) = 0$*, and*

(ii) $C_n(1, \ldots, 1, \alpha, 1, \ldots, 1) = \alpha$.

Theorem 2 *The function H_n of Theorem 1 can be extended (in general, non-uniquely) to a copula C_n. Being an extension of H_n, the copula C_n satisfies the condition*

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

Theorem 3 *Let one-dimensional distribution functions $F_1, \ldots, F_n$ be given. Let C_n be any n-dimensional copula. Then, the function*

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

is an n-dimensional distribution function having marginals $F_1, F_2, \ldots, F_n$.

Theorems 1-3 reduce Fréchet's problem to the problem of characterizing copulas with n dimensions. There is but one copula with 1 dimension, P_1, which satisfies the condition $P_1(x) = x$ for any $x \in [0,1]$. For $n > 1$, the set of n-dimensional copulas is infinite, limited by two particular functions.

Theorem 4 *Any n-dimensional copula C_n satisfies the inequality*

$$L_n \leq C_n \leq M_n,$$

where the bounds L_n, M_n are defined as

$$L_n(a_1, \ldots a_n) = \max\{0, \textstyle\sum_{k=1}^{m} \alpha_k - n + 1\}$$
$$M_n(a_1, \ldots a_n) = \min\{\alpha_1, \ldots, \alpha_n\}.$$

For $n \geq 1$, the functions $P_n = \prod_{k=1}^{n} \alpha_k$ defined on $[0,1]^n$ are n-dimensional copulas. They determine the correlation scheme of n independent random variables.

But copulas have structure generally simpler than distribution functions. We can use this simplicity by introducing the notion of "quasi-inverse" of a monotone function.

Definition 2 *Let f be a monotone function of a real variable. Add to the set of points $(x.f(x))$ all closed segments of the form $[(x, f(x+), (x, f(x-)]$. Reflect the resulting set about the line $y = x$. Remove from the reflected set all vertical segments except for a point in each. Any set thus obtained is a graph of a function, which we call a* quasi-inverse *of f.*

According to this definition, any one-dimensional distribution function F possesses at least one quasi-inverse function F^* defined and non-decreasing on the closed interval $[0,1]$ and taking finite values in $(0,1)$. Among such values, there is one which will make F^* left-continuous on $(0,1)$ and continuous at 0 and 1.

Theorem 5 *Let Q_n be a function of n variables, integrable in $\mathbb{R}^n$ with respect to a distribution function G_n. Let C_n be a copula such that*

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

where $F_1, \ldots, F_n$ are the marginal distributions of G_n. Then,

$$\begin{aligned} &\int_{\mathbb{R}^n} Q_n(x_1, \ldots, x_n)\, dG_n(x_1, \ldots, x_n) \\ &\qquad = \int_{U^n} Q_n(F_1^*(\alpha_1), \ldots, F_n^*(\alpha_n))\, dC_n(\alpha_1, \ldots, \alpha_n). \end{aligned} \tag{1}$$

The expression on the right-hand side of (1) may be simpler than that of the left-hand side. This is the case, for example, if C_n is absolutely continuous, as is the case $C_n = P_n$; or $n = 2$ and $C_2 = L_2$ or $C_2 = M_2$; or if $n = 3$ and $C_3 = M_3$.

Translator's Note. This is a partial answer to a question by Zhengao Huang.