Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

Random Variables, Distribution Functions, and Copulas – A Personal Look Backward and Forward

By A. Sklar
Illinois Institute of Technology

The author recalls his initial involvement with the basic notions of probability theory, which began in the late forties in the context of number theory, continued through his work with B. Schweizer on probabilistic metric spaces, and culminated in a correspondence with Fréchet that led to the identification and naming of copulas. The author speculates about possible future applications of the theory of distribution functions with given margins: In particular, there is the prospect of productive treatment of situations where, say, no common probability space can be found for a given set of "random variables," but such common probability spaces exist for arbitrary proper subsets of the given set.

This paper presents my recollections of, and outlook on, one phase of the development of our subject, up to the early sixties. It also indicates what I believe to be a promising direction for future work. I thank the organizers of the conference for the opportunity to do this, and I thank the referees for their very helpful comments.

My serious engagement with probability began in the late forties, in the context of number theory, where one meets statements such as: "almost all numbers are composite" [Hardy-Wright (1960), p. 8] and: "The probability that a number should be quadratfrei is $6/\pi^2$" [ibid, p. 267; "quadratfrei" is now usually anglicized to "squarefree"]. They refer to a function, often denoted by δ and called the "density," defined for certain sets of positive integers by

$$\delta(A) = \lim_{n\to\infty} \frac{1}{n} \#(A \cap \{1, 2, \cdots, n\})$$

whenever the limit exists, where $\#(S)$ denotes the number of elements in the (finite) set S. But δ is not a measure, since its domain is not closed under the binary Boolean operations: as pointed out in Niven (1951), p. 424 "sequences A and B can be constructed [Buck (1946), p. 571] so that $\delta(A)$ and $\delta(B)$ exist but $\delta(A \cup B)$ and $\delta(A \cap B)$ do not."

AMS 1991 Subject Classification: 60-03, 01A70, 60E05, 62H05.

Key words and phrases: Random variables, distribution functions, copulas.

Nevertheless, despite this apparent deficiency, by the late forties a large body of results involving δ, of strikingly probabilistic flavor, had been achieved. For a contemporary survey of the situation see Kac (1949); Billingsley (1969) provides an elegant exposition of an important special case; and a more recent, concise survey appears in Section 5 of Galambos (1982).

Elementary combinatorial arguments enabled me to contribute to this body of results the following straightforward extension of the quoted statement about squarefree numbers:

For any integer $n \geq 2$, the probability [i.e., density] of the set of positive integers that are n-free, i.e., not divisible by the nth power of any prime, is $1/\zeta(n)$, where $\zeta(n)$ is the value of the Riemann zeta-function at n (recall that $1/\zeta(2) = 6/\pi^2$).

As one consequence of this result, consider the function h defined on the positive integers by $h(1) = 0$ and

$$h(m) = \max(\alpha_1, \cdots, \alpha_k) \ \text{ for } \ m > 1,$$

where $m = k_1^{\alpha_1} \cdots k_k^{\alpha_k}$ is the standard representation of m as a product of prime-powers. Then for any positive integer n, the probability that h assumes the value n is equal to the probability of the set of integers that are $(n+1)$-free but not n-free, and so is equal to $1/\zeta(2)$ for $n = 1$, and $1/\zeta(n+1) - 1/\zeta(n)$ for $n \geq 2$. Since the sequence $1/\zeta(n)$ is strictly increasing and has the limit 1, and since the difference $1/\zeta(n+1) - 1/\zeta(n)$ behaves like 2^{-n-1} for n large, one can not only define a cumulative distribution function F for h by

$$F(x) = \delta\{n \mid h(n) \leq x\},$$

but show that this distribution function has finite moments of all orders. In particular, the mean value of h is

$$\frac{1}{\zeta(2)} + \sum_{n=2}^{\infty} n \left(\frac{1}{\zeta(n+1)} - \frac{1}{\zeta(n)} \right) = 1 + \sum_{n=2}^{\infty} \left(1 - \frac{1}{\zeta(n)} \right),$$

a number slightly larger than 1.7.

At this time I was essentially ignorant of measure theory and the measure-theoretic foundations of probability theory as codified in Kolmogorov (1933). (Having mentioned this classic book, I cannot resist the temptation, in a conference devoted to work ultimately based on that of Fréchet's, to note that in Chapter 1, Section 5, Kolmogorov says: "Random variables from a mathematical point of view represent merely functions measurable with respect to $P(A)$, while their mathematical expectations are abstract Lebesgue integrals. This analogy was explained fully for the first time in the work of Fréchet.") But as I only found out later, I was not exactly alone in this. Even if not afflicted by my lamentable ignorance, a good part of the mathematical community was apparently beset by inertia. To this day, I find it remarkable that 21 years after the appearance of his book, Kolmogorov himself, speaking to an audience of colleagues and specialists, still felt it desirable to repeat his basic definitions and axioms [Kolmogoroff-Prochorow (1956), pp. 113–114]. As for the general situation in the late forties, there is a remark by Eduard Marczewski. He had edited and published some papers of his teacher Stefan Mazurkiewicz, who had "died immediately after the war, as a result of wartime difficulties" [Zygmund (1951), p. 8]. In a note at the end of Mazurkiewicz (1949), Marczewski says:

"Il est cependant á remarquer que, pour les probabilistes, les distributrices des variables sont préférables aux variables mêmes et le jeu de pile et face est préférable à l'espace des fonctions mesurables."

In a footnote to this note, Marczewski referred to Paul Lévy and J. L. Doob, whose contrasting practices were illustrated in Lévy (1937) and Doob (1947). As for Lévy, we read in Taylor (1975), p. 308 that:

"He never learnt to use the machinery of modern measure theory so that when he studied sample paths he did not speak of points ω in an underlying probability space Ω and σ-algebras of sets which determined the correct conditioning fields for a proper definition of the Markov or the strong Markov property. His intuition was almost infallible, and this is all the more surprising because many of the truths he discovered go counter to the normal intuition of an analyst. One can only explain his success with the belief that he was using subconsciously and informally the points of a probability space to which he never referred."

Doob, of course, expressed the opinion (most explicitly and emphatically in Doob (1953)) that probability theory was simply a branch of measure theory, a point of view gently but firmly rebutted by M. Loève [(1977), pp. 172–174, unchanged from previous editions]. In any case, by the middle fifties, most, if not all, probabilists would have defined a random variable as a measurable function on a probability space. Statisticians, on the other hand, would most likely have behaved differently, and defined a random variable, if at all, as (a rather vague) "anything to which a distribution function can be attached." This contrast was forcefully pointed out by Karl Menger in his talk at the Third Berkeley Symposium on Mathematical Statistics and Probability, which he begins by saying:

"In his great book *Sequential Analysis*, Wald defines a random variable as a variable x such that 'for any given number c a definite probability can be ascribed to the event that x will take a value less than c' $\cdots$. In 1947, I submitted to Wald the following two observations: (1) the concept 'variable' on which the notion of random variable is based does not appear to be that of a numerical variable, the only one then clearly defined; (2) the statement and example on page 11 seem to be at variance with the definition of random variables on page 5.

"I believe that I carry out Wald's intentions by saying that he fully agreed with both remarks and expressed the hope to clarify the statistical concept of random variables at a later occasion. His untimely death in 1950, after the completion of his fundamental book on statistical decision functions (in which he essentially retained the treatment of random variables of *Sequential Analysis*) prevented him from carrying out this plan." [Menger (1956), p. 215.]

I mention Menger and Wald because each had earlier written about "statistical metric spaces" [Menger (1942), Wald (1943)] in which distances are statistical objects rather than fixed numbers. Menger defined these spaces in terms of distribution functions rather than random variables; and Wald, while retaining Menger's definition, in effect treated distances as independent random variables. Now when I arrived at the Illinois Institute of Technology in 1956, I met Berthold Schweizer; and Bert, having some time back been struck by a remark in Menger (1949), had come to IIT in order to work on statistical metric spaces under Menger's direction. Bert introduced me to the subject and persuaded me to work together with him on it; so began our long collaboration.

As we started to work together, I felt that I needed to learn much more about modern probability theory. I decided to do this by developing the subject, from the ground up, as far as possible in my own way. (I recommend this as a good way to learn virtually anything, provided that in thus "reinventing the wheel" one does not lose sight of the fact that one is redoing something that has been earlier, and most likely better, by others.) Individual (numerical) random variables were no problem: I defined them as functions on the unit interval. It was only later that I found out that I had been anticipated in this by at least 31 years. As Doob says in his *Appreciation of Khinchin*:

"In 1925 Khinchin and Kolmogorov initiated the systematic study of the convergence theory of infinite series whose terms are independent random variables $\cdots$. It is interesting historically to note that Khinchin considered it necessary to construct his random variables as functions on the interval $[0,1]$ with Lebesgue measure. It is of course no longer necessary to go through such construction procedures." [Doob (1961), p. 17.]

But when it came to two or more (nonindependent) random variables, I ran into difficulties. These largely stemmed from the fact that multi-dimensional distribution functions, unlike one-dimensional ones, are quite complicated objects: just how complicated, I realized when I saw Mann (1952). Now Henry Mann, who had proved the "$\alpha + \beta$ Theorem," one of the "Three Pearls of Number Theory" in Khinchin (1952) was one of my heroes; and it was a bit dismaying to see how hard and intricately he had to work to deal with stochastic processes in terms of joint distribution functions. Of course, as R. Fortet says in *Mathematical Reviews* (v. 14, 1953, p. 663) in the end Mann "réussit la gageure d'enclore en 44 pages un exposé d'ensemble sur les processus stochastiques, correct et dense $\cdots$."

In the meantime, Bert and I had been making progress in our work on statistical metric spaces, to the extent that Menger suggested it would be worthwhile for us to communicate our results to Fréchet. We did: Fréchet was interested, and asked us to write an announcement for the *Comptes Rendus* [Schweizer and Sklar (1958)]. This be an an exchange of letters with Fréchet, in the course of which he sent me several packets of reprints, mainly dealing with the work he and his colleagues were doing on distributions with given marginals. These reprints, among the later arrivals of which I particularly single out that of Dall'Aglio (1959) were important for much of our subsequent work. At the time, though, the most significant reprint for me was that of Féron (1956).

Féron, in studying three-dimensional distributions had introduced auxiliary functions, defined on the unit cube, that connected such distributions with their one-dimensional margins. I saw that similar functions could be defined on the unit n-cube for all $n \geq 2$ and would similarly serve to link n-dimensional distributions to their one-dimensional margins. Having worked out the basic properties of these functions, I wrote about them to Fréchet, in English. He asked me to write a note about them in French. While writing this, I decided I needed a name for these functions. Knowing the word "copula" as a grammatical term for a word or expression that links a subject and predicate, I felt that this would make an appropriate name for a function that links a multidimensional distribution to its one-dimensional margins, and used it as such. Fréchet received my note, corrected one mathematical statement, made some minor corrections to my French, and had the note published by the Statistical Institute of the University of Paris as Sklar (1959). Subsequent developments are summarized in Schweizer (1991).

Two footnotes to the preceding: The first is that, though neither Sklar (1959) nor the longer Sklar (1973) contain proofs, proofs of a combinatorial nature exist for all the basic statements about copulas, and are presented, with one important exception, in Chapter 5 of Schweizer and Sklar (1983). The missing proof is that for the extension theorem that is stated as Lemma 5 in Sklar (1973) and Theorem 6.2.6 in Schweizer and Sklar (1983). Note though that the two-dimensional case of the extension theorem is stated and proved as Lemma 5 in Schweizer and Sklar (1974), and that the general case is proved by probabilistic arguments, somewhat indirectly in Moore and Spruill (1975), and directly, along with other important results in Deheuvels (1978). Since people have expressed interest in a combinatorial proof of the general extension theorem, an outline of such a proof is given in Appendix 1.

The second footnote is that at the Fourth Berkeley Symposium on Mathematical Statistics and Probability in 1960 I was privileged to meet and talk at length with Alfréd Rényi. At one point Rényi, who was of course completely familiar with the literature, remarked that in defining statistical metric spaces, Menger should have started with random variables instead of distribution functions, since then dependence properties would come in automatically. I don't remember what, if anything, I said to this. Much later, I could have said that, as is shown for example by the results in Schweizer and Sklar (1974), Frank (1975), and the paper by R. Nelsen et al. in this volume, it was very fortunate that Menger proceeded as he did.

All of the preceding indicates to me that our view of what constitutes a proper theory of probability has to be broadened. This thought is by no means original: others have not only had this thought but have acted on it. I need only mention here Fréchet's treatment of "Zufallselemente" in separable metric spaces in Fréchet (1956), Rényi's well known "conditional probability spaces," and the generalized probability spaces, motivated by physics, foreshadowed in Suppes (1969) and defined in Gudder (1969), (1984).

Looking ahead, I think we are just now beginning to be able to deal with the following type of situation, which will increasingly be seen to be ubiquitous: For a given set S and a given integer $n \geq 2$, it is possible to associate a classical probability space with each n-element subset of S, but no such classical probability space exists for any subset of S with more than n elements. (Thus, if S has $n+1$ elements, there is a classical probability space for every proper subset of S, but no such space for S itself.) The existence of such situations in the case $n = 2$ has been implicitly, if not explicitly, known for a long time, and the possibility of such existence in the case $n > 2$ is at least implicit in some (apparently unpublished) work by R. M. Dudley, referred to in Schweizer and Sklar (1983), p. 161. The existence of such situations for all $n \geq 2$ is shown, by an explicit example, in Appendix 2. I feel that the investigation of these and related situations offers great promise for the future.

Appendix 1
Outline of a Combinatorial Proof of the Extension Theorem for Copulas

To make this appendix reasonably self-contained, I begin by recalling some basic definitions.

For a positive integer n, an *n-box* B is the Cartesian product of n closed real intervals. If the intervals are $[a_k, b_k]$, $k = 1, 2, \cdots, n$, then B is *n-small* if $a_k = b_k$ for at least one k. Otherwise, B is *n-big* and the *vertices* of B are the 2^n n-tuples $(c_1, c_2, \cdots, c_n)$ where each c_k is either a_k or b_k. The *sign*, sgn $\boldsymbol{v}$, of the vertex $\boldsymbol{v} = (c_1, c_2, \cdots, c_n)$ is $(-1)^{\sigma(\boldsymbol{v})}$, where $\sigma(\boldsymbol{v})$ is the number of k's for which $c_k = a_k$. Thus $\operatorname{sgn}(b_1, b_2, \cdots, b_n) = 1$, $\operatorname{sgn}(a_1, b_2, \cdots, b_n) = -1$, $\operatorname{sgn}(a_1, a_2, \cdots, a_n) = (-1)^n$, etc. We need not define the sgn function for the fewer than 2^n vertices of an n-small n-box.

If F is a real-valued function whose domain is a Cartesian product of n real sets, then the *F-volume*, $V_F(B)$ of an n-box B whose vertices are in the domain of F is 0 if B is n-small; otherwise, summing over the vertices of B,

$$V_F(B) = \sum_{\boldsymbol{v}} (\operatorname{sgn} \boldsymbol{v}) F(\boldsymbol{v}).$$

Such a function F is *n-increasing* if $V_F(B) \geq 0$ for all n-boxes B whose vertices are in the domain of F.

An *n-subcopula* is a real-valued function C' such that:

i. The domain of C' is a Cartesian product of n subsets of the closed unit interval $I = [0, 1]$, each subset containing at least the points 0 and 1.

ii. C' is n-increasing.

iii. If $(x_1, x_2, \cdots, x_n)$ is in the domain of C' and at least one of the x's is 0, then

$$C'(x_1, x_2, \cdots, x_n) = 0.$$

If all the x's, with the possible exception of x_a are 1, then $C'(x_1, \cdots, x_n) = x_k$.

An *n-copula* is an n-subcopula whose domain is the entire n-cube I^n.

It is proved in Chapter 6 of Schweizer and Sklar (1983) that if C' is an n-subcopula and $(x_1, \cdots, x_n)$, $(y_1, \cdots, y_n)$ are in the domain of C', then

$$|C'(x_1, \cdots, x_n) - C'(y_1, \cdots, y_n)| \leq \sum_{i=1}^{n} |x_i - y_i|.$$

It follows that any n-subcopula is uniformly continuous on its domain. The extension theorem for copulas now states that:

Every n-subcopula can be extended to an n-copula, i.e., given any n-subcopula C' there is an n-copula C such that

$$C(x_1, x_2, \cdots, x_n) = C'(x_1, x_2, \cdots, x_n)$$

for all $(x_1, x_2, \cdots, x_n)$ in the domain of C'.

To prove this, we first use the uniform continuity of C' to extend C' to a subcopula C'' whose domain is the closure of that of C' (hence the domain of C'' is the Cartesian product of n closed subsets of I). If the domain of C'' is all of I^n, we are finished. If not, then we extend C'' to a function C defined on all of I^n as follows: Given any point $\boldsymbol{x}$ in I^n not in the domain of C'', then $\boldsymbol{x}$ lies in a unique n-box B whose vertices are in the domain of C'' and that contains no smaller such n-box. Then $\boldsymbol{x}$ has a unique (vector) representation in the form

$$\boldsymbol{x} = \sum_{\boldsymbol{v}} \beta(\boldsymbol{v})\boldsymbol{v},$$

where the sumation is over the vertices of B, each $\beta(\boldsymbol{v})$ is a non-negative number, and the sum of all the $\beta(\boldsymbol{v})$'s is 1. Now we define $C(\boldsymbol{x})$ by:

$$C(\boldsymbol{x}) = \sum_{\boldsymbol{v}} \beta(\boldsymbol{v})C''(\boldsymbol{v}).$$

We then complete the definition of C by setting $C(\boldsymbol{x}) = C''(\boldsymbol{x})$ for all $\boldsymbol{x}$ in the domain of C''. It is not hard to show that the function C so defined is continuous.

To show that C is a copula, it suffices to show that C is n-increasing since the other copula properties follow almost immediately from the definition of C. Now consider an n-box B in I^n. Then B is the Cartesian product of n closed subintervals $[v_k, y_k]$ $(k = 1, 2, \cdots, n)$ of I. Call $[v_k, y_k]$ the kth *edge* of B. Since the domain of C'' is a Cartesian product of n closed subsets c_k of I, the edges of B can be divided into two classes as follows: $[v_k, y_k]$ is a *good* edge if both v_k and y_k are in c_k; otherwise, $[v_k, y_k]$ is a *bad* edge. We proceed to do an induction on the number b of bad edges of B.

If $b = 0$, then every vertex of B is in the domain of C'', whence

$$V_C(B) = V_{C''}(B) \geq 0.$$

Now suppose that it has been shown for some integer m between 0 and $n-1$ that $V_C(B) \geq 0$ for all boxes B with $b \leq m$. It will now be shown that $V_C(B) \geq 0$ if $b = m + 1$. Consider a box B with $m + 1$ bad edges. Without loss of generality, we may assume that the edge $[v_1, y_1]$ is bad. There are two cases to consider: either there is no point of c_1 strictly between v_1 and y_1, or there is at least one such point.

In the first case, let u_1 be the largest number in c_1 that is $\leq v_1$, and z_1 the smallest number in c_1 that is $\geq y_1$ (such numbers exist, since c_1 is closed). Then B is entirely contained in the box

$$B_0 = [u_1, z_1] \times [v_2, y_2] \times \cdots \times [v_n, y_n],$$

and B_0 has only m bad edges. It follows that $V_C(B_0) \geq 0$, and from this it is straightforward, though a bit tedious, to show that $V_C(B) \geq 0$. (In fact, if $v_1 = y_1$, then $V_C(B) = 0$ automatically, while if $v_1 < y_1$, then $u_1 < z_1$ and

$$V_C(B) = (\lambda_2 - \lambda_1)V_C(B_0),$$

where $\lambda_1 = (v_1 - u_1)/(z_1 - u_1)$ and $\lambda_2 = (y_1 - u_1)/(z_1 - u_1)$.)

In the second case, define u_1 and z_1 as in the first case, and in addition define w_1 to be the smallest number in c_1 that is $\geq v_1$, and x_1 to be the largest number in c_1 that is $\leq y_1$. It follows that

$$u_1 \leq v_1 \leq w_1 \leq x_1 \leq y_1 \leq z_1$$

with some of the inequalities being strict. Now split B into the three boxes

$$\begin{aligned} B_1' &= [v_1, w_1] \times [v_2, y_2] \times \cdots \times [v_n, y_n], \\ B_2 &= [w_1, x_1] \times [v_2, y_2] \times \cdots \times [v_n, y_n], \\ B_3' &= [x_1, y_1] \times [v_2, y_2] \times \cdots \times [v_n, y_n]. \end{aligned}$$

Since B_1' is entirely contained in the box

$$B_1 = [u_1, w_1] \times [v_2, y_2] \times \cdots \times [v_n, y_n]$$

and B_3' in the box

$$B_3 = [x_1, z_1] \times [v_2, y_2] \times \cdots \times [v_n, y_n];$$

and since B_1, B_2, B_3 each has only m bad edges, it follows as before that

$$\begin{aligned} V_C(B_1) &\geq V_C(B_1') \geq 0, \\ V_C(B_2) &\geq 0, \\ \text{and } V_C(B_3) &\geq V_C(B_3') \geq 0. \end{aligned}$$

Therefore $V_C(B) = V_C(B_1') + V_C(B_2) + V_C(B_3') \geq 0$, and the induction is complete. Hence $V_C(B) \geq 0$ for all n-boxes B in I^n, so C is n-increasing on its domain and therefore is a copula.

Appendix 2
Common Probability Spaces for Arbitrary n-Element Subsets of a Set, but None for Larger Subsets

Let n be an integer ≥ 2. Define a function C_{nn} on the unit n-cube by

$$C_{nn}(x_1, \cdots, x_n) = (\max(x_1^{1/(n-1)} + \cdots + x_n^{1/(n-1)} - n + 1, 0))^{n-1}.$$

It is not hard to show that C_{nn} is an n-copula (not absolutely continuous, but that plays no role in what follows). For $m < n$, each m-dimensional margin of C_{nn}, i.e., each function obtained by fixing $n - m$ of the n arguments of C_{nn} to be 1, is the same m-copula C_{mn}, given by

$$C_{mn}(x_1, \cdots, x_m) = (\max(x_1^{1/(n-1)} + \cdots + x_m^{1/(n-1)} - m + 1, 0))^{n-1}.$$

For $n = 2$, $C_{22}(x_1, x_2) = \max(x_1 + x_2 - 1, 0)$. For $n > 2$, C_{nn} is the $(n-1)$st *serial iterate* of C_{2n}, i.e.,

$$C_{nn}(x_1, x_2, x_3, \cdots, x_n) = C_{2n}(\cdots C_{2n}(C_{2n}(x_1, x_2)x_3) \cdots, x_n).$$

For n increasing without bound, the sequence C_{2n} has the limit Π, the two-copula given by $\Pi(x, y) = xy$.

THEOREM. *For any $n \geq 2$, there is no $(n+1)$-copula each of whose n-dimensional margins is the n-copula C_{nn}.*

PROOF. Assume otherwise, i.e., assume that there is an $(n+1)$-copula C, each of whose n-margins is C_{nn}. Set $x_0 = ((n-1)/n)^{n-1}$. Then

$$C_{nn}(x_0, \cdots, x_0) = 0,$$

while for $m < n$, $C_{mn}(x_0, \cdots, x_0) = (n-m)^{n-1}/n^{n-1}$. Since C is a copula,

$$0 \leq C(x_0, x_0, \cdots, x_0) \leq C(1, x_0, \cdots, x_0) = C_{nn}(x_0, \cdots, x_0) = 0,$$

so $C(x_0, x_0, \cdots, x_0) = 0$.

Now consider the $(n+1)$-cube $[x_0, 1]^{n+1}$. Its C-volume $V_C[x_0, 1]^{n+1}$ is

$$\begin{aligned} 1 - (n+1)x_0 + \binom{n+1}{2} C_{2n}(x_0, x_0) + \cdots + (-1)^n \binom{n+1}{n} C_{nn}(x_0, \cdots x_0) \\ + (-1)^{n+1} C(x_0, x_0, \cdots, x_0) \\ = \frac{1}{n^{n+1}} \sum_{m=0}^{n-1} (-1)^m \binom{n+1}{m} (n-m)^{n-1}. \end{aligned}$$

Now, $\binom{n+1}{m} = \sum_{\ell=0}^{m} \binom{n-\ell}{n-m}$, so

$$\begin{aligned}
V_C[x_0,1]^{n+1} &= \frac{1}{n^{n-1}} \sum_{m=0}^{n-1} \sum_{\ell=0}^{m} (-1)^m \binom{n-\ell}{n-m} (n-m)^{n-1} \\
&= \frac{1}{n^{n-1}} \sum_{\ell=0}^{n-1} \sum_{m=\ell}^{n-1} (-1)^m \binom{n-\ell}{n-m} (n-m)^{n-1} \\
&= \frac{1}{n^{n-1}} \sum_{\ell=0}^{n-1} (-1)^\ell \sum_{m=\ell}^{n-1} (-1)^{m-\ell} \binom{n-\ell}{n-m} (n-m)^{n-1} \\
&= \frac{1}{n^{n-1}} \sum_{\ell=0}^{n-1} (-1)^\ell (n-\ell)! S(n-1, n-\ell),
\end{aligned}$$

where $S(n-1, n-\ell)$ is a Stirling number of the second kind. But the sum in the last expression can be written in the form

$$-\sum_{\ell=0}^{n-1} (-1)^{(n-1)-(n-\ell)} (n-\ell)! S(n-1, n-\ell)$$

which is equal to -1 by a standard identity for Stirling numbers (see, e.g. p. 825 in Abramowitz and Stegun (1964). So $V_C[x_0,1]^{n+1} = -n^{-n+1} < 0$, which is impossible for a copula. Therefore the $(n+1)$-copula C does not exist, and this proves the theorem.

It is worth noting that the theorem applies to many copulas in addition to the C_{nn}'s. For example, for each α in the open interval $(1, \log 3/\log 2)$, there is no three-copula each of whose two-dimensional margins is the two-copula C_α given by

$$C_\alpha(x,y) = (\max(x^{1/\alpha} + y^{1/\alpha} - 1, 0))^\alpha.$$

Moreover, unlike the two-copula C_{22}, each such C_α (so in particular $C_{3/2}$ and $C_{\pi/2}$) is absolutely continuous.

Now let E be a set with more than n elements, and suppose that each element of E can be regarded as a (continuous) random variable in the loose "statistical" sense mentioned in the text, i.e., as something which has a continuous distribution function. To each n-element subset of E assign the n-copula C_{nn}. Then for each n-element subset of E a standard construction yields a classical probability space with respect to which the elements of the subset are random variables in the strict sense (and for large n, any two elements of the subset are close to being independent). But for any subset of E with more than n elements no such classical probability space exists.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A., eds. (1964). *Handbook of Mathematical Functions.* National Bureau of Standards Applied Mathematics Series 55. U.S. Government Printing Office.

BILLINGSLEY, PATRICK (1969). On the central limit theorem for the prime divisor function *Amer. Math. Monthly* **76**, 132–139.

BUCK, R. C. (1946). The measure theoretic approach to density. *Amer. J. Math.* **68**, 560–580.

DALL'AGLIO, G. (1959). Sulla compatibilità delle funzioni di ripartizione doppia. *Rend. Mat.* **18**, 385–413.

DEHEUVELS, PAUL (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publ. Inst. Statist. Univ. Paris* **23**, 1–37.

DOOB, J. L. (1947). Probability in function space. *Bull. Amer. Math. Soc.* **53**, 17–30.

DOOB, J. L. (1953). *Stochastic Processes.* John Wiley & Sons, New York.

DOOB, J. L. (1961). Appreciation of Khinchin, pp. 17–20 in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. II, University of California Press.

FÉRON, R. (1956). Sur les tableaux de corrélation dont les marges sont donées: cas de l'espace a trois dimensions. *Publ. Inst. Statist. Univ. Paris* **5**, 3–12.

FRANK, M. J. (1975). Associativity in a class of operations on spaces of distribution functions. *Aequat. Math.* **12**, 121–144.

FRÉCHET, MAURICE (1956). Abstrakte Zufallselemente, pp. 23–28 in: *Bericht über die Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik in Berlin 1954*, VEB Deutsch. Verl. Wissen.

GALAMBOS, JANOS (1982). The role of functional equations in stochastic model building. *Aequat. Math.* **25**, 21–41.

GUDDER, STANLEY P. (1969). Quantum probability spaces. *Proc. Amer. Math. Soc.* **21**, 296–302.

GUDDER, STANLEY P. (1984). An extension of classical measure theory. *SIAM Rev.* **26**, 71–89.

HARDY, G. H. and WRIGHT, F. M. (1960). *An Introduction to the Theory of Numbers.* Oxford University Press (4th Edition).

KAC, M. (1949). Probability methods in some problems of analysis and number theory. *Bull. Amer. Math. Soc.* **55**, 641–665.

KHINCHIN, A. YA. (1952). *Three Pearls of Number Theory.* Graylock Press.

KOLMOGOROV, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer-Verlag.

KOLMOGOROFF, ANDREJ und PROCHOROW, JURI (1956). Zufällige Funktionen und Grenzverteilungssatze, pp. 113–126 in: *Bericht über die Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik in Berlin 1954.* VEB Deutsch. Verl. Wissen.

LÉVY, PAUL (1937). *Théorie de l'addition de variables aléatoires.* Gauthier-Villars, Paris.

LOÈVE, M. (1977). *Probability Theory.* Springer-Verlag (4th Edition).

MANN, HENRY B. (1952). Introduction to the theory of stochastic processes depending on a continuous parameter. NBS Appl. Math. Series No. 24, U.S. Government Printing Office.

MAZURKIEWICZ, STEFAN (1949). Sur les espaces de variables aléatoires. *Fund. Math.* **36**, 288–302.

MENGER, KARL (1942). Statistical metrics. *Proc. Nat. Acad. Sci.* **28**, 535–537.

MENGER, KARL (1949). The theory of relativity and geometry, pp. 459–474 in: *Albert Einstein, Philosopher–Scientist*, R. A. Schilpp, ed. Tudor Publishing Co.

MENGER, KARL (1956). Random variables from the point of view of a general theory of variables, pp. 215–229 in: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. II. University of California Press.

MOORE, D. S. and SPRUILL, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3**, 599–616.

NIVEN, IVAN (1951). The asymptotic density of sequences. *Bull. Amer. Math. Soc.* **57**, 420–434.

SCHWEIZER, BERTHOLD (1991). Thirty years of copulas, pp. 13–50 in: G. Dall'Aglio, S. Katz and G. Salinetti, eds. *Advances in Probability Distributions with Given Marginals.* Kluwer, Dordrecht.

SCHWEIZER, BERTHOLD and SKLAR, ABE (1958). Espaces métriques aléatoires. *C. R. Acad. Sci. Paris* **247**, 2092–2094.

SCHWEIZER, BERTHOLD and SKLAR, ABE (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Math.* **52**, 43–52.

SCHWEIZER, BERTHOLD and SKLAR, ABE (1983). *Probabilistic Metric Spaces.* North-Holland, New York.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

SKLAR, A. (1973). Random variables, joint distribution functions and copulas. *Kybernetika* **9**, 449–460.

SUPPES, PATRICK (1969). The probabilistic argument for a non-classical logic of quantum mechanics. *Philos. Sci.* **33**, 14–21.

TAYLOR, S. J. (1975). Paul Lévy. *Bull. London Math. Soc.* **7**, 300–320.

WALD, ABRAHAM (1943). On a statistical generalization of metric spaces. *Proc. Nat. Acad. Sci. USA* **29**, 196–197.

ZYGMUND, ANTONI (1951). Polish mathematics between the two wars, pp. 3–9 in: *Proceedings of the Second Canadian Mathematical Congress*, University of Toronto Press.

DEPARTMENT OF MATHEMATICS
5044 MARINE DRIVE
CHICAGO, IL 60640