Copulae and Their Uses

Carlo Sempi
Dipartimento di Matematica "Ennio De Giorgi"
Università di Lecce
I-73100 Lecce
Italy
carlo.sempi@unile.it

Abstract

This short survey of copulas presents their properties, tries to stress their relevance for statistics and their connection with Markov processes and conditional expectations.

1 What is a copula?

Copulæ were introduced by Sklar(1959). The reader is referred to Schweizer and Sklar (1983), Schweizer (1991), Sklar (1973), Sklar (1996), Nelsen (1999).

A copula is a function $C : [0,1] \times]0,1[\to]0,1[$ that satisfies the following properties:

- for all t in $[0,1]$, $C(t,0) = C(0,t) = 0$;
- for all t in $[0,1]$, $C(t,1) = C(1,t) = t$;
- if x, x', y, y' are in $[0,1]$ with $x \leq x'$ and $y \leq y'$, then

$$C(x',y') - C(x,y') - C(x',y) + C(x,y) \geq 0.$$

As a consequence of these properties it follows that (a) C satisfies the Lipschitz condition

$$|C(x,y) - C(x',y')| \leq |x - x'| + |y - y'|,$$

(b) that it is non–decreasing in each variable and (c) that it is absolutely continuous. In other words, a copula is a two–dimensional distribution function that concentrates all the probability mass on the unit square $[0,1] \times [0,1]$ and which has uniform marginals.

The importance of the concept of copula stems from the following

Theorem 1. (Sklar). *Let X and Y two random variables on the probability space $(\Omega, \mathcal{F}, P)$ having H as their joint distribution function and let F and G be the marginals of H,*

$$F(x) = H(x, +\infty), \qquad G(y) = H(+\infty, y).$$

Then there exists (at least) a copula C such that

$$H(x,y) = C\left(F(x), G(y)\right). \tag{1}$$

If both F and G are continuous, then the copula C is uniquely determined.

If either F or G, or both, is not continuous, then there may be more than one copula may satisfy (1); all of these coincide on the set $\mathrm{Ran}F \times \mathrm{Ran}G$. By a method of bilinear interpolation, which will always be adopted in the sequel, it is possible to choose a single copula that satisfies (1). For this method see Lemma 2.3.5 in Nelsen (1999).

The partial derivatives

$$D_1\, C(x,y) := \frac{\partial C(x,y)}{\partial x}, \qquad D_2\, C(x,y) := \frac{\partial C(x,y)}{\partial y}$$

exist almost everywhere and are almost everywhere non–decreasing.

2 Statistical Proterties

Copulæ are widely used in non–parametric statistics, especially in the study of dependence of random variables and in order to express the known measures of association between random variables, Kendall's tau, Spearman's rho, Gini's coefficient, see Nelsen (1999) and the bibliography quoted therein. A new measure of dependence was introduced by Schweizer and Wolff (1981) in terms of the copula of the two random variables involved.

Recently the work of some authors (Averous, Bassan, Dortet–Bernadet, and Spizzichino) has introduced the use of copulas in the study of multivariate aging. In a few cases, the multivariate aging function is not a copula but a quasi–copula (for which see Genet et al. (1999)).

3 Copulæ and Markov Processes

The connection between Copulas and Markov processes is established through an operation on the set $\mathcal{C}$ of all copulas, which was introduced by Darsow, Nguyen and Olsen (1992). Let A and B be copulæ; if x and y are in $[0,1]$ an operation on $\mathcal{C}$ is defined via

$$(A * B)(x,y) := \int_0^1 D_2A(x,t)\, D_1B(t,y)\, dt. \tag{2}$$

Then $A * B$ is a copula and if $\{A_n\}$ converges to A, then one has both

$$A_n * B \longrightarrow A * B$$

and

$$B * A_n \longrightarrow B * A;$$

moreover, the operation $*$ is associative but not jointly continuous.

Theorem 2. (Darsow, Nguyen, Olsen). *Let $\{X_t : t \in T\}$ be a real–valued stochastic process and let C_{st} be the copula of the random variables X_s and X_t. If, for s and t in T with $s < t$ and for a Borel set A, one sets*

$$P(s,x,t,A) := P\left(X_t \in A \mid X_s = x\right),$$

then the following are equivalent:

(a) The Chapman–Kolmogorov equations

$$P(s,x,t,A) = \int_{\boldsymbol{R}} P(u,\xi,t,A)\, P(s,x,u,d\xi) \quad (u \in \,]s,t[\, \cap T)$$

hold for almost all $x \in \boldsymbol{R}$.

(b) $C_{st} = C_{su} * C_{ut}$.

Of course $\{X_t\}$ may satisfy the Chapman–Kolmogorov equations without being a Markov process. A necessary and sufficient condition in terms of the copulas C_{st} is known.

The crucial step in the proof of Theorem 2 is the following equality

$$P(s,t,x,]-\infty,a]) = D_1C_{st}\left(F_s(x), F_t(a)\right) \qquad a.e..$$

4 Copulæ and Conditional Expectations

While proving Theorem 2 Darsow, Nguyen and Olsen (1992) proved that if the random variables X and Y have copula C, then one has almost surely

$$E\left(1_{\{X<x\}} \mid Y\right)(\omega) = D_2 C\left(F_X(x), F_Y(Y(\omega))\right),$$

$$E\left(1_{\{Y<y\}} \mid x\right)(\omega) = D_1 C\left(F_X(X(\omega)), F_Y(y)\right).$$

The last relationships point at a connection between Conditional Expectations and copulas. This connection is best established through Markov operators. Given a probability space $(\Omega, \mathcal{F}, P)$, a Markov operator is a linear operator $T : L^\infty \to L^\infty$ such that

(a) T is positive, $f \geq 0 \implies TF \geq 0$;

(b) $T1 = 1$;

(c) (expectation invariance) for every function f in L^∞, one has $E(Tf) = E(f)$.

Above, L^∞ may be replaced by L^1. Notice that if $\mathcal{G}$ is a sub–σ–field of $\mathcal{F}$, then the conditional expectation $E_{\mathcal{G}} := E(\cdot \mid \mathcal{G})$ is a Markov operator. An explicit one–to–one correspondence between Markov operators and copulas can be established; to a copula C there corresponds a unique Markov operator T_C and, conversely, to a Markov operator T there corresponds a unique copula C_T in such a way that $T_{A*B} = T_A \circ T_B$. A Markov operator T is a Conditional Expectation if, and only if, it is idempotent: $T^2 := T \circ T = T$. The Markov operator T_C corresponding to a copula C is idempotent if, and only if, the copula C is idempotent with respect to the operation $*$ introduced above, $C = C * C$. Thus, there exists a one–to–one correspondence between idempotent copulas and Conditional Expectations; for this see the forthcoming paper (Sempi, 2002).

Acknowledgements

The support of the research group "Processi Stocastici" of the Italian M.U.R.S.T. is gratefully acknowledged.

References

Darsow, W.F., Nguyen, B., Olsen, E.T. (1992) Copulas and Markov processes. *Illinois J. Math. 36*, 600–642.

Genest, C., Quesada–Molina, J.J., Rodríguez–Lallena, J.A. and Sempi, C. (1999) A characterization of quasi–copulas. *J. Multivariate Anal. 69*, 193–205.

Nelsen, R.B. (1999) *An introduction to copulas.* New York: Springer (Lecture Notes in Statistics 139).

Olsen, E.T., Darsow, W.F., Nguyen, B. (1996). Copulas and Markov operators. In L. Rüschendorf, B. Schweizer and M.D. Taylor (Eds.), *Distributions with fixed marginals and related topics*, Lecture Notes — Monograph Series Volume 28, pp. 244–259. Hayward CA: Institute of Mathematical Statistics

Schweizer, B. and Sklar, A. (1983). *Probabilistic metric spaces.* New York: North–Holland.

Schweizer, B. (1991). Thirty years of copulas. In G. Dall'Aglio, S. Katz and G. Salinetti (Eds.), *Probability distributions with given marginals*, pp. 13–50. Dordrecht: Kluwer.

Schweizer, B. and Wolff, E.F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist. 9*, 870–885.

Sempi, C. (2002). Conditional expectations and idempotent copulæ. In C. Cuadras, J. Fortuny and J.A. Rodríguez Lallena (Eds.), *Distributions with fixed marginals and applications*, Dordrecht: Kluwer.

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.

Sklar, A. (1996) Random variables, distribution functions, and copulas — A personal look backward and forward. In L. Rüschendorf, B. Schweizer and M.D. Taylor (Eds.),*Distributions with fixed marginals and related topics*, Lecture Notes — Monograph Series Volume 28, pp. 1–14. Hayward CA: Institute of Mathematical Statistics