Dependence modeling using copulas

Régis LEBRUN

EADS IW SE-MA

CEMRACS 2013

Content

Uncertainty propagation

Given:

- A random vector X taking values into $\mathbb{R}^n$ (uncertainties)
- A measurable function $f : \mathbb{R}^n \to \mathbb{R}^p$

One want to gain information on the distribution of $Y = f(X)$ (hence the term of propagation):

- Some moments $\mathbb{E}[h(Y)]$ for various measurable functions h
- As a special case, the probability of some events $\mathbb{P}(Y \in B)$

Probabilistic modeling

The main objective is to build the distribution of X

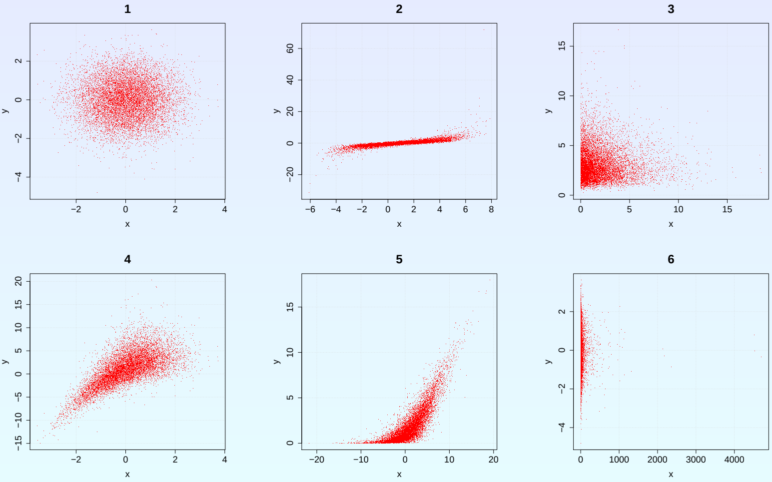
- from multivariate data
- or from univariate data only
- or from expert knowledge

From my personal experience:

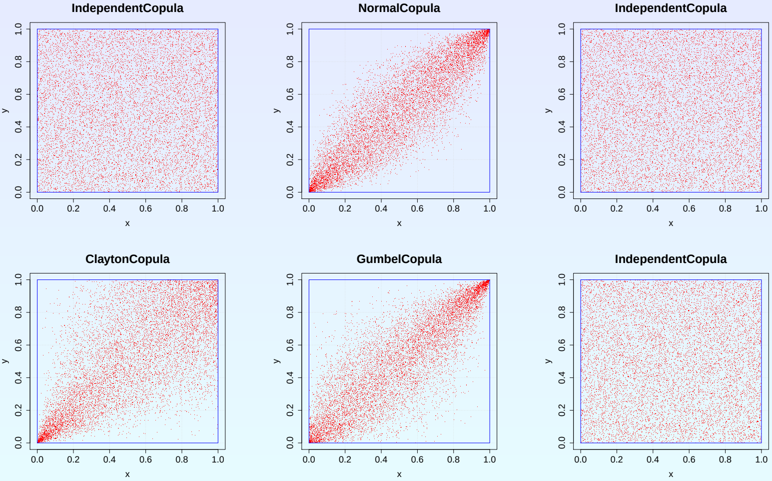
- in many applications, a rather good knowledge of the marginal distributions of X has been gained with time
- in contrast, the interaction between the components of X is rather unknown

dependence modeling is the description of this interaction, ie the description of the joint distribution function once the effect of the marginal distributions has been removed

Some bidimensional distributions. Which ones have independent components?



The same data, considering ranks



A short historical review on copulas and dependence modeling

1940 Hoeffding: measure of dependence, linear correlation multivariate distributions with uniform marginals on $[-1/2, 1/2]$.

1951 Fréchet: multivariate distributions with fixed marginal distributions.

1959 Sklar and Schweizer: probabilistic metric spaces, first occurence of the term copula.

1979 Deheuvels: independence tests, non parametric multivariate estimation.

1992 Darsow, Nguyen and Olsen: description of Markov processes in terms of copulas.

1999 Embrechts, Lindskog and McNeil: dissemination of copula methodology in financial and insurance applications.

2005 Mikosch: "Copulas: Tales and facts". Are copulas something else than a fashionable subject?

2009 Salmon: "Recipe for a disaster: the formula that killed Wall Street"

Copulas: a serious matter? [Mikosch]

Thomas Mikosch's analysis of the exponential growth of activity related to copulas:

2003 Google gives 10,000 responses to the word "copula"

2005 650,000 responses...

2013 2,010,000 responses...

- "My main concern is that this very simple concept might be something like the emperor's new clothes because it promises to solve all problems of stochastic dependence but it falls short in achieving the goal."
- "I also observed that my students are likely to be attracted to copulas than to stochastic processes. A possible reason is that one needs less than 10 minutes to understand the fundamentals of copulas, but many years of studies in order to get an idea of a genuine stochastic process."

So we will have about 40 minute left for questions...

...and we will be on time for diner!

Copulas for dependence modeling I

Définition

A n-dimensional copula is the restriction to the unit cube $[0,1]^n$ of a multivariate distribution function with uniform univariate marginals on $[0,1]$.

Theorem

Let C be a n-dimensional copula, then $\forall u, v \in [0,1]^n, |C(u) - C(v)| \leq \sum_{i=1}^{n} |u_i - v_i|$

Theorem ([Sklar])

Let F be a n-dimensional distribution function whose marginal distribution functions are $F_1, \dots, F_n$. There exists a copula C of dimension n such that for $x \in \overline{\mathbb{R}}^n$, we have:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \tag{1}$$

In the case of continuous marginal distributions, for all $u \in [0,1]^n$, we have:

$$C(u) = F(F_1{}^{(-1)}(u_1), \dots, F_n{}^{(-1)}(u_n)) \tag{2}$$

Copulas for dependence modeling II

Proof.

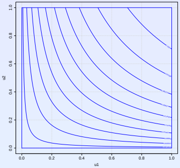
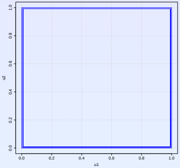
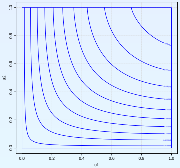
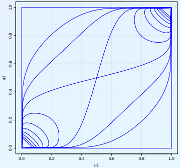
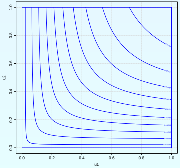
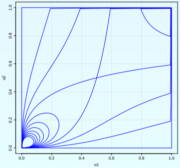
- Let X be a n-dimensional random vector with distribution function F.
- Let V be a random variable uniformly distributed over $[0,1]$ and independent from X
- For $k = 1, \ldots, n$, let U_k be defined by $U_k = F_k(X_k)$ if F_k is continuous, and by $U_k = F_k(X_k-) + V \sum_{v \in \Delta_k} \mathbb{P}(X_k = v)$ where Δ_k is the set of discontinuity points of F_k if F_k is not continuous.
- The random variables $U_1, \ldots, U_n$ are uniformly distributed on $[0,1]$ and $\forall x_k \in \mathbb{R}, \{X_k \leq x_k\} = \{U_k \leq F_k(x_k)\}$ a.s.
- Let C be the distribution function of $(U_1, \ldots, U_n)$. Then C is a copula and $\forall x \in \mathbb{R}^n$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(U_1 \leq F_1(x_1), \ldots, U_n \leq F_n(x_n)) = C(F_1(x_1), \ldots, F_n(x_n)) \quad (3)$$

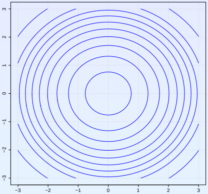
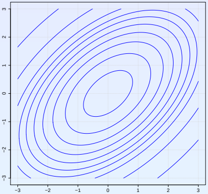
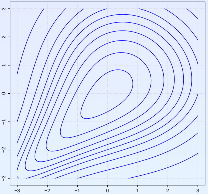
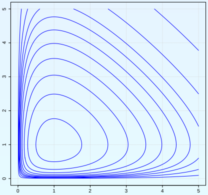
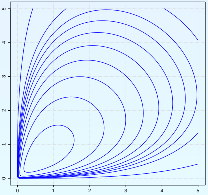
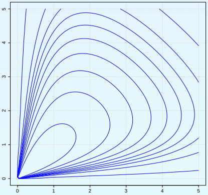
- If all the F_k are continuous, then their image include $(0,1)$ and by continuity of the copulas, there exists a unique C satisfying (1).

□

Examples

Copula	CDF	PDF
Independent $u_1 u_2$		
Normal $\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{\exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right)}{2\pi\sqrt{1-\rho^2}} \, \mathrm{d}s \, \mathrm{d}t,$ $\lvert\rho < 1\rvert$		
Clayton $\left(u_1^\theta + u_2^\theta - 1\right)^{1/\theta}, \theta > 0$		

Examples of composed distributions

Copula	Independent	Normal	Clayton
$\mathcal{N}(0,1)$ marginals			
$\Gamma(2,1)$ marginals			

Sampling of composed distributions

Let X be a random vector with marginal distribution functions $F_1, \ldots, F_n$ and copula C. One can sample X by the following two-steps procedure:

1. Generate $u \sim C$;
2. A realization x of X is given by:

$$x = \left(F_1^{(-1)}(u_1), \ldots, F_N^{(-1)}(u_n)\right) \tag{4}$$

The key point is to be able to sample C.

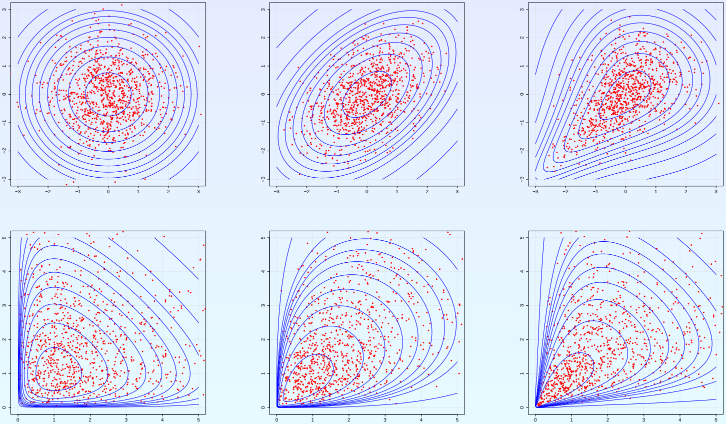
Define $C_k(u_1, \ldots, u_k) = C(u_1, \ldots, u_k, 1, \ldots, 1)$
and $C_k(u_k | u_1, \ldots, u_{k-1}) = \dfrac{\partial^{k-1} C_k(u_1, \ldots, u_k)}{\partial u_1 \ldots u_{k-1}} / \dfrac{\partial^{k-1} C_{k-1}(u_1, \ldots, u_{k-1})}{\partial u_1 \ldots u_{k-1}}$

1. Generate $u_1 \sim \mathcal{U}(0, 1)$;
2. For $k \in \{2, \ldots, n\}$, generate $u_k \sim C_{k|1,\ldots,k-1}(u_1, \ldots, u_{k-1})$.
3. The resulting point $(u_1, \ldots, u_n)$ is a realization of C.

Remark: for many copulas, more efficient specialized algorithms exist

Sampling of composed distributions

Copula	Independent	Normal	Clayton
$\mathcal{N}(0,1)$ marginals			
$\Gamma(2,1)$ marginals			



More modeling tools: composed copulas

Let $C_1, \ldots, C_k$ be k copulas of dimensions $n_1, \ldots, n_k$ and $N = \sum_{i=1}^{k} n_i$. The function C defined on $[0,1]^N$ by:

$$C(u_1, \ldots, u_N) = C_1(u_1, \ldots, u_{n_1}) \times \cdots \times C_k(u_{N-n_k+1}, \ldots, u_N) \tag{5}$$

is a copula of dimension N.

It is a sparse block-diagonal dependence structure based on several low dimensional dense dependence structures.

More modeling tools: copula tree [Kurowicka] I

Définition

$\mathcal{T} = (N, E)$ is a tree with nodes $N = \{1, \ldots, n\}$ and edges E, where E is a subset of unordered pairs of N with no cycle; that is, there does not exists a sequence $a_1, \ldots, a_k$ $(k > 2)$ of elements of N such that:

$$\{a_1, a_2\} \in E, \ldots, \{a_{k-1}, a_k\} \in E, \{a_k, a_1\} \in E \tag{6}$$

The degree of node $a_i \in N$ is $\#\{a_j \in N \,|\, \{a_i, a_j\} \in E\}$, ie the number of edges attached to a_i.

Remark: this definition allows for non-connected trees (forest).

Définition

$(\mathcal{T}, B)$ is a copula-tree specification if:

1. $\mathcal{T}$ is a tree on n elements with nodes $N = \{1, \ldots, n\}$ and edges E.
2. $B = \{C_{ij} \,|\, \{i, j\} \in E$ and C_{ij} is a bidimensional copula$\}$

More modeling tools: copula tree [Kurowicka] II

Définition

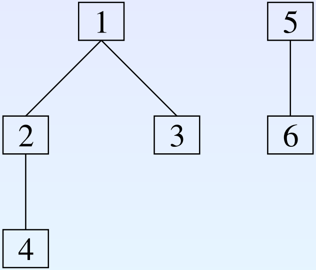
$(\mathcal{T}, B)$ is a Markov tree dependence whenever disjunct subsets a and b of variables are separated by subset c of variables in $\mathcal{T}$ (i.e. every path from a to b intersect c), in which case the variables in a and b are conditionally independent given the variables in c.

Theorem

Let $(\mathcal{T}, B)$ be an n-dimensional bivariate copula-tree specification with absolutely continuous bivariate copulas C_{ij} with density c_{ij}. Then, there is a unique n-dimensional absolutely continuous copula C with density c such that C has Markov tree dependence for $\mathcal{T}$. Its density c is given by:

$$c(u_1, \ldots, u_n) = \prod_{\{i,j\} \in E} c_{ij}(u_i, u_j) \tag{7}$$

More modeling tools: copula tree [Kurowicka] III



$$c(u_1, \ldots, u_6) = c_{12}(u_1, u_2)c_{13}(u_1, u_3)c_{24}(u_2, u_4)c_{56}(u_5, u_6)$$

To simulate realizations of such a copula, draw u_1 and u_5 independently, then draw u_2 and u_3 independently conditional on the value of u_1, and u_6 independently conditional on the value of u_5. u_4 is drawn independently of the others conditional on the value of u_2.

More modeling tools: vines copulas [Kurowicka] I

Définition

$\mathcal{V}$ is a vine on n elements if

1. $\mathcal{V} = (\mathcal{T}_1, \dots, \mathcal{T}_{n-1})$.
2. $\mathcal{T}_1$ is a connected tree with nodes $N_1 = \{1, \dots, n\}$ and edges E_1; for $i \in \{2, \dots, n-1\}$, $\mathcal{T}_i$ is a connected tree with nodes $N_i = E_{i-1}$.

$\mathcal{V}$ is a regular vine on n elements if additionally:

3. For $i \in \{2, \dots, n-1\}$, if $\{a, b\} \in E_i$, then $\#a \triangle b = 2$, where $\triangle$ denotes the symmetric difference. In other words, if a and b are nodes of $\mathcal{T}_i$ connected by an edge in $\mathcal{T}_i$, where $a = \{a_1, a_2\}$, $b = \{b_1, b_2\}$, then exactly one of the a_j equals one of the b_j.

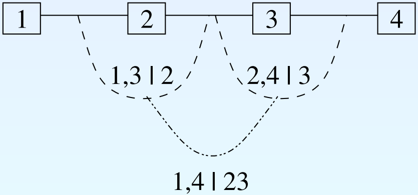
The edges E_1 express the unconditioned pairwise dependence, the edges E_2 express the pairwise dependence conditional on the value of nodes in N_1, the edges E_3 express the pairwise dependence conditional on the value of nodes in N_2 and so one.

More modeling tools: vines copulas [Kurowicka] II

Définition

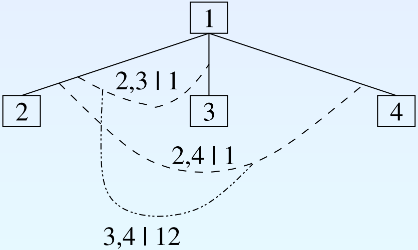
A regular vine is called a

- D-vine if each node in T_1 has a degree of at most 2
- Canonical or C-vine if each tree $\mathcal{T}_i$ has a unique node of degree $n-i$. The node with maximal degree in $\mathcal{T}_1$ is the root.



D-vine

$$c(u_1,\ldots,u_4) = c_{12}(u_1,u_2)c_{23}(u_2,u_3)\times c_{34}(u_3,u_4)c_{13|2}(u_1,u_3|u_2)\times c_{24|3}(u_2,u_4|u_3)c_{14|23}(u_1,u_4|u_2,u_3)$$



C-vine

$$c(u_1,\ldots,u_4) = c_{12}(u_1,u_2)c_{13}(u_1,u_3)\times c_{14}(u_1,u_4)c_{24|1}(u_2,u_4|u_1)\times c_{34|12}(u_3,u_4|u_1,u_2)$$

Additional properties of copulas

Theorem (Fréchet-Hoeffding bounds)

Let C be a n-dimensional copula. Then for all $u \in [0,1]^n$ we have:

$$W_n(u) := \max(0, u_1 + \ldots + u_n) \leq C(u) \leq M_n(u) := \min(u_1, \ldots, u_n) \quad (8)$$

The copula M_n is called the Min copula and corresponds to random vectors X for which all the components are almost surely strictly increasing functions of a common random variable: $X = (f_1(U), \ldots, f_n(U))$.

Theorem ([Nelsen])

Let X be a n-dimensional random vector with copula C and $\alpha_1, \ldots, \alpha_n$ be n strictly increasing functions from $\mathbb{R}$ to $\mathbb{R}$, then C is also a copula for the random vector $(\alpha_1(X_1), \ldots, \alpha_n(X_n))$.

Measures of association [Joe], [Nelsen]

Définition

A measure of association r between two random variables X_1 and X_2 is a scalar function of X_1 and X_2 such that:

1. r is defined for all pair (X_1, X_2).
2. $r(X_1, X_2) \in [-1, 1], r(X_1, X_1) = 1, r(X_1, -X_1) = -1$.
3. If X_1 and X_2 are independent, $r(X_1, X_2) = 0$.
4. If g and h are two strictly increasing functions, $r(X_1, X_2) = r(g(X_1), h(X_2))$.

Such a r is a function of the copula of (X_1, X_2) only. The objective of such a measure is to provide a scalar summary of the intensity of the dependence between X_1 and X_2.

Bad properties of linear correlation I

Définition

The linear correlation ρ between two random variables X_1 and X_2 such that $\mathrm{Var}(X_1) = \sigma_1^2 < \infty$ and $\mathrm{Var}(X_2) = \sigma_2^2 < \infty$ is defined by:

$$\begin{aligned}\rho(X_1, X_2) &= \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} \\ &= \frac{1}{\sigma_1\sigma_2}\iint_{\mathbb{R}^2} F_{12}(x_1, x_2) - F_1(x_1)F_2(x_2)\,\mathrm{d}x_1\mathrm{d}x_2 \end{aligned} \tag{9}$$

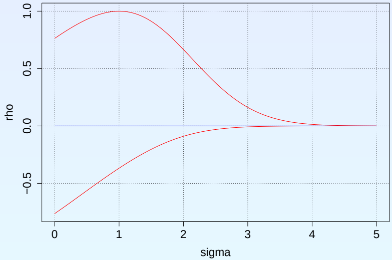
It is not a measure of association because it is defined for finite second moment random variables only, and is not invariant by increasing transformation.

Theorem (Fréchet)

Let (X_1, X_2) be a random vector with given marginal distribution functions F_1, F_2. The possible values of the linear correlation $\rho(X_1, X_2)$, if defined, form an interval $[\rho_{min}, \rho_{max}]$ that is in general a strict subset of $[-1, 1]$.

Example

If $X_1 \hookrightarrow \mathcal{LN}(0,1)$ and $X_2 \hookrightarrow \mathcal{LN}(0,\sigma^2)$ are two log-normal random variables, then
$$\rho(X_1,X_2) \in \left[\rho_{min} = \frac{e^{-\sigma}-1}{\sqrt{e-1}\sqrt{e^{\sigma^2}-1}}, \rho_{max} = \frac{e^{\sigma}-1}{\sqrt{e-1}\sqrt{e^{\sigma^2}-1}}\right] \subsetneq [-1,1]$$



We see that $\lim_{\sigma\to\infty} \rho_{min} = \lim_{\sigma\to\infty} \rho_{max} = 0$. For $\sigma = 5, \rho \in [-3\,10^{-6}, 4\,10^{-4}]$! As a consequence, from a modeling perspective, the value of $\rho(X_1,X_2)$ cannot be specified independently from F_1 and F_2.

Spearman'rho and Kendall's tau

Let X_1 and X_2 be two random variables.

Définition

Spearman's rho $\rho_S(X_1, X_2)$ is defined by:

$$\rho_s(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) = 12 \iint_{[0,1]^2} C(u, v)\, \mathrm{d}u\, \mathrm{d}v - 3 \qquad (10)$$

where C is the copula of (X_1, X_2).

Définition

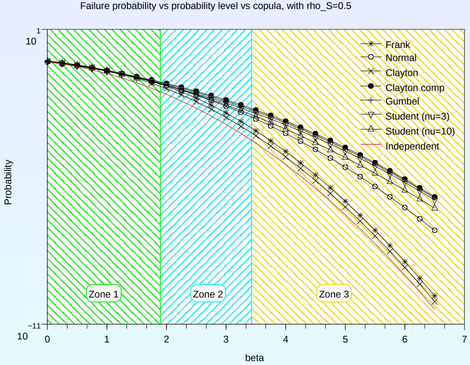
Kendall's tau $\tau(X_1, X_2)$ is defined by:

$$\begin{aligned}\tau(X_1, X_2) =& \mathbb{P}[(\hat{X}_1 - \tilde{X}_1)(\hat{X}_2 - \tilde{X}_2) > 0] - \mathbb{P}[(\hat{X}_1 - \tilde{X}_1)(\hat{X}_2 - \tilde{X}_2) < 0] \\ =& 4 \iint_{[0,1]^2} C(u, v)\, \mathrm{d}C(u, v) - 1\end{aligned}$$

where $(\hat{X}_1, \hat{X}_2)$ and $(\tilde{X}_1, \tilde{X}_2)$ are iid copies of (X_1, X_2).

Is a measure of association enough to quantify the dependence?

$\mathbb{P}\left(X_1 + X_2 \geq \beta\sqrt{2}\right)$ for $X_1, X_2 \sim \mathcal{N}(0,1)$ and various copulas C such that $\rho_S(X_1, X_2) = 1/2$.



β	$P_{min}(\beta)$	$P_{max}(\beta)$	ratio
1.89	$6.5\,10^{-2}$	$8.7\,10^{-2}$	1.5
3.41	$1.1\,10^{-3}$	$8.6\,10^{-3}$	10.0
6.5	$8.3\,10^{-11}$	$1.9\,10^{-6}$	$2.3\,10^{4}$

Parametric estimation of copulas

Estimation of an elementary copula:

- Based on estimators of measures of association
- Inversion of the relation between the parameters of the copula and the measures of association:
  - Normal copula C_R: $R_{ij} = 2\sin\left(\frac{\pi}{6}\rho_{S\,ij}\right) = \sin\left(\frac{\pi}{2}\tau_{ij}\right)$
  - Clayton's copula C_θ: $\theta = \dfrac{2\tau}{1-\tau}$
  - etc.

Estimation of vines copulas or of copula trees:

- Semi-heuristic estimation of the structure
- Based on partial rank correlation
- see [Kurowicka] for the details

Rank and order statistics

The notion of rank plays a key role in the estimation of measures of association.

Définition

Let $(X^k)_{k=1,\dots,N}$ be a sample of size N of the random variable X and $\sigma \in \mathfrak{S}_N$ a random permutation such that $X_{\sigma(1)} \leq \dots X_{\sigma(N)}$ a.s. (such a permutation is almost surely unique if X is continuous). The rank of X^k is defined by:

$$\text{rank}(X^k) = \sigma^{-1}(k)$$

It is the random position of X^k in the order statistics $X_{1:N} = X_{\sigma(1)}, \dots, X_{N:N} = X_{\sigma(N)}$.

Estimation of measures of association I

Définition

Let $\left((X_1^k, X_2^k)\right)_{k=1,\dots,N}$ be a sample of size N of the random vector $X = (X_1, X_2)$. The Spearman rho estimator $\hat{\rho}_{S,N}(X)$ is given by:

$$\hat{\rho}_{S,N}(X) = \frac{\sum_{k=1}^{N} \left(\text{rank}(X_1^k) - \overline{\text{rank}}(X_1)\right)\left(\text{rank}(X_2^k) - \overline{\text{rank}}(X_2)\right)}{\sqrt{\sum_{k=1}^{N} \left(\text{rank}(X_1^k) - \overline{\text{rank}}(X_1)\right)^2 \sum_{k=1}^{N} \left(\text{rank}(X_2^k) - \overline{\text{rank}}(X_2)\right)^2}} \tag{11}$$

where $\overline{\text{rank}}(X_1) = \frac{1}{N}\sum_{k=1}^{N} \text{rank}(X_1^k)$ and $\overline{\text{rank}}(X_2) = \frac{1}{N}\sum_{k=1}^{N} \text{rank}(X_2^k)$.

Theorem

Let X be a bi-dimensional continuous random vector. Then:

$$\hat{\rho}_{S,N}(X) \xrightarrow{a.s} \rho_S(X) \text{ when } N \to \infty$$

$$\sqrt{N}\left(\hat{\rho}_{S,N}(X) - \rho_S(X)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{\rho_S}) \text{ when } N \to \infty$$

Estimation of measures of association II

where the asymptotic variance $\sigma^2_{\rho_S}$ is given by:

$$\sigma^2_{\rho_S} = \left(1 + \frac{\rho_S(X)^2}{2}\right) \frac{4(5+192\eta_{10})}{3(4\eta_{00}-1)^2} + \frac{\rho_S(X)^2}{4}\left(\frac{342}{125} - \frac{12}{5}\left(\frac{24(\eta_{20}+\eta_{02})-1}{4\eta_{00}-1}\right)\right)$$

where $\eta_{k\ell} = \iint_{[0,1]^2} \left(u_1 - \frac{1}{2}\right)^k \left(u_2 - \frac{1}{2}\right)^\ell C(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2$ and C is the copula of X.

Définition

Let $\left((X_1^k, X_2^k)\right)_{k=1,\ldots,N}$ be a sample of size N of the random vector $X = (X_1, X_2)$. The sampling Kendall tau $\hat{\tau}_N(X_1, X_2)$ is given by

$$\hat{\tau}_N(X) = \frac{2}{N(N-1)} \sum_{1\leq i<j\leq N} \mathrm{sgn}(X_1^i - X_1^j)\,\mathrm{sgn}(X_2^i - X_2^j) \quad (12)$$

Estimation of measures of association III

Theorem

Let X be a bi-dimensional random vector. Then:

$$\hat{\tau}_N(X) \xrightarrow{a.s} \tau(X) \text{ when } N \to \infty$$

$$\sqrt{N}\left(\hat{\tau}_N(X) - \tau(X)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\tau^2) \text{ when } N \to \infty$$

where the asymptotic variance σ_τ^2 is given by:

$$\sigma_\tau^2 = 4\,\mathbf{Var}\left[\operatorname{sgn}(X_1 - X_1')\operatorname{sgn}(X_2 - X_2') \,|\, X_1, X_2\right]$$

where $X' = (X_1', X_2')$ is an independent copy of X.

Fitting test

- An active research area with relatively few results
- See [Genest2], [Berg] and [Fermanian]
- Good news: in dimension 2, the tests are powerful enough to discriminate rather close hypotheses for sample sizes N as small as $N = 150$.

Order statistics and copulas I

Let X be an n-dimensional random vector with known univariate marginal distribution functions $F_1, \ldots, F_n$. We look for the set of copulas $\mathcal{C}$ such that the resulting distribution function satisfies:

$$X_1 \leq \ldots \leq X_n \text{ a.s.} \tag{13}$$

Theorem

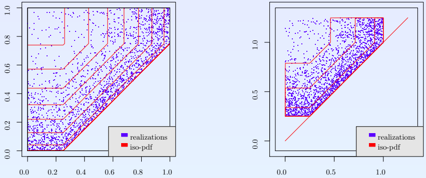
1. *$\mathcal{C} \neq \emptyset$ if and only if $\forall x \in \mathbb{R}, F_n(x) \leq \ldots F_1(x)$;*
2. *If $F_1,\ldots,F_n$ verify (1 and are continuous, then $C \in \mathcal{C}$ if and only if the support of C is included in $\{u \in [0,1]^n \,|\, F_1^{\leftarrow}(u_1) \leq \ldots \leq F_n^{\leftarrow}(u_n)\}$*

where $F^{\leftarrow}$ is the generalized inverse of F:

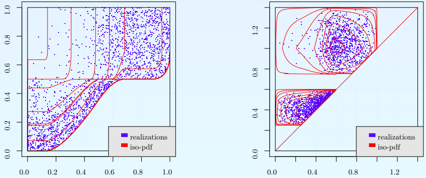
$$F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} \,|\, F(x) \geq q\} \tag{14}$$

Order statistics and copulas II

$F_1 \sim \mathcal{U}(0,1),\ F_2 \sim \mathcal{U}(\frac{1}{4},\frac{5}{4})$



$F_1 \sim \mathcal{T}(0,\frac{1}{2},1),\ F_2 \sim \frac{1}{2}\mathcal{T}(\frac{1}{4},\frac{1}{2},\frac{3}{5}) + \frac{1}{2}\mathcal{T}(\frac{3}{4},1,\frac{7}{5})$



Example of compatible absolutely continuous copulas for order statistics.

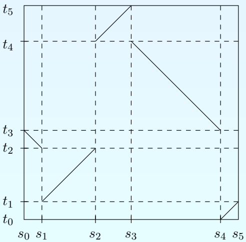
Are perfect dependence and independence so different? I

Définition

Let $X_1, \ldots, X_n$ be n random variables. They are said to be perfectly dependent if there exist a random variable U and n almost surely bijective functions $f_1, \ldots, f_n$ such that $X_1 = f_1(U), \ldots, X_n = f_n(U)$.

Définition

A copula C of dimension n is a shuffle of min if and only if there is a positive integer N, n partitions $(0 = s_0^k < s_1 < \ldots < s_n = 1)_{k=1,\ldots,n}$ of $[0,1]$, and $n-1$ permutations σ^k on $\{1,\ldots,n\}$ such that each $[s_{i-1}, s_i] \times \ldots \times [s^n_{\sigma^{n-1}(i-1)}, s^{n-1}_{\sigma^{n-1}(i)}]$ is a hypercube in which C deposits a mass of size $s_i - s_{i-1}$ spread uniformly along one of the diagonals.



$$s = (0, 1/12, 1/3, 1/2, 11/12, 1)$$
$$t = (0, 1/12, 1/3, 5/12, 5/6, 1)$$
$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 4 & 1 \end{pmatrix}$$

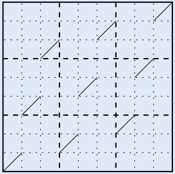
Are perfect dependence and independence so different? II

Theorem

Shuffles of Min are dense in the set of copulas endowed with the sup norm.

We give the demonstration for target compula Π_n, the n-dimensional independent copula.

- $\epsilon > 0$ given, m an integer such that $m \geq 1/\epsilon$
- Take $M = m^n$ and build C_ϵ a shuffle of Min associated with the n uniform partitions of $[0,1]$ into M sub-intervals of equal width and the permutations $\sigma^k(m^k(j-1)+i) = m^k(i-1)+j$ for $i, j = 1, \ldots, m$, $k = 1, \ldots, m-1$.
- C_ϵ distributes a mass of $1/M$ in each of the M sub-hypercubes of $[0,1]^n$, and $C_\epsilon(p_1/m, \ldots, p_n/m) = p_1 \times \ldots \times p_n/m$ for all $p_i = 0, \ldots, m$ so C_ϵ and Π_n are equal on these points. As both Π_n and C_ϵ are Lipschitz we have $\|C_\epsilon - \Pi_n\|_\infty \leq n\epsilon$.



□

MANY THANKS FOR YOUR ATTENTION!

Copulas and Markov processes I

Définition

1. Let A and B be two bidimensional copulas. We define the product $C = A * B$ of these copulas by:

$$C(u_1, u_2) = \int_0^1 A_{1|2}(u_1, t) B_{2|1}(t, u_2)\, dt \tag{15}$$

The null element is Π_2 (the bidimensional independent copula) and the neutral element is M_2.

2. Let A be a copula of dimension m and B a copula of dimension n. We define the product $C = A \star B$ of these two copulas by:

$$\begin{aligned} C(u_1, \ldots, u_{m+n-1}) &= \int_0^{u_m} A_{1,\ldots,m-1|m}(u_1, \ldots, u_{m-1}, t) \times \\ &\quad \times B_{2,\ldots,m|1}(t, u_{m+1}, \ldots, u_{m+n-1})\, dt \end{aligned} \tag{16}$$

We have the relation $A * B(u, v) = A \star B(u, 1, v)$

Copulas and Markov processes II

Properties

These products have the following properties (where $\bullet \in \{*, \star\}$*):*

- C *is a copula (of dimension 2 for* $*$*, of dimension* $m+n-1$ *for* $\star$*);*
- *These products are continuous with respect to* A *and* B*: if* $(A_n)_{n\in\mathbb{N}} \to A$ *and* $(B_n)_{n\in\mathbb{N}} \to B$*,* $A_n \bullet B \to A \bullet B$ *et* $A \bullet B_n \to A \bullet B$*;*
- *These products are associative:* $(A \bullet B) \bullet C = A \bullet (B \bullet C)$*;*
- *These products are left and right distributive with respect to convex combinations of copulas.*

Copulas and Markov processes III

Theorem

Let $X_t, t \in T$ be a real-valued stochastic process and for all $s, t \in T$, let C_{st} be the copula of the random vector (X_s, X_t). There is an equivalence between:

1. *The transition probabilities $\mathbb{P}(s, x, t, A) = \mathbb{P}(X_t \in A | X_s = x)$ of the process satisfy the Chapman-Kolmogorov equations:*

$$\mathbb{P}(s, x, t, A) = \int_{-\infty}^{\infty} \mathbb{P}(u, \xi, t, A) \mathbb{P}(s, x, u, d\xi) \tag{17}$$

   for all Borel set A, all $s < t$ in T, all $u \in]s, t[\cap T$ and almost all $x \in \mathbb{R}$;

2. *For all $s, u, t \in T$ such that $s < u < t$,*

$$C_{st} = C_{su} * C_{ut} \tag{18}$$

Copulas and Markov processes IV

Theorem

A real valued stochastic process $X_t, t \in T$ is a Markov process if and only if for all $n \in \mathbb{N}^*$ and for all $t_1, \ldots, t_n \in T$ such that $t_1 < \cdots < t_n$ we have:

$$C_{t_1 \ldots t_n} = C_{t_1 t_2} \star C_{t_2 t_3} \star \cdots \star C_{t_{n-1} t_n} \tag{19}$$

where $C_{t_1 \ldots t_n}$ is the copula of $(X_{t_1}, \ldots, X_{t_n})$ and $C_{t_k t_{k+1}}$ the copula of $(X_{t_k}, X_{t_{k+1}})$.

This result has been generalized [Ibragimov] to a Markov process of order k, i.e. such that:

$$\mathbb{P}(X_t < x | X_{t_1}, \ldots, X_{t_{n-k}}, X_{t_{n-k+1}}, \ldots, X_{t_n}) = \mathbb{P}(X_t < x | X_{t_1}, \ldots, X_{t_{n-k}}) \tag{20}$$

for all $t, t_i \in T$ such that $t_1 < \cdots < t_{n-k} < t_{n-k+1} < \cdots < t_n < t$ and $x \in \mathbb{R}$.

References I

D. Berg.
Copula goodness-of-fit tests: an overview and power comparison.
Technical Report 5, University of Oslo, October 2007.

D. Fermanian.
Goodness-of-fit tests for copulas.
Journal of Multivariate Analysis, 95:119–152, 2005.

C. Genest, B. Rémillard, and D. Beaudoin.
Goodness-of-fit tests for copulas: A review and a power study.
Insurance Mathematics and Economics, 44:199–213, 2009.

R. Ibragimov.
Copula-based characterizations for higher-order markov processes.
3:819–846, 2009.

H. Joe.
Multivariate models and dependence concepts.
Chapman & Hall, 1997.

References II

D. Kurowicka and R. Cooke.
Uncertainty Analysis with High Dimensional Dependence Modelling.
Wiley series in probability and statistics. John Wiley & Sons, 2006.

T. Mikosch.
Copulas: tales and facts.
Extremes, 9:3–20, 2006.

R. B. Nelsen.
An Introduction to Copulas.
Springer Series in Statistics. Springer-Verlag New York Inc., 2nd edition, 2006.

M. Sklar.
Fonctions de répartition à n dimensions et leurs marges.
Publication de l'Institut Statistique Universitaire Paris, 8:229–231, 1959.