# Dependence Models, Dependence Measures and Copulas

**Franco Pellerey**
**Dipartimento di Scienze Matematiche**
`franco.pellerey@polito.it`

Mainly based on lectures prepared by **Fabrizio Durante**
School of Economics and Management
Free University of Bozen-Bolzano (Italy)
`fabrizio.durante@unibz.it`
*http://sites.google.com/site/fbdurante*

(but all mistakes are mine!!)

# Outline

- Multivariate stochastic models
- Copulas: definition and first properties
- Examples of copulas
- Copulas and dependence
- Measures of association
- Archimedean copulas
- Some remarkable models
- Fitting copulas to data
- Bibliography

# Multivariate stochastic models

A **multivariate stochastic model** for a system formed by several components can be described by means of a random vector $\mathbf{X}$ defined on the probability space $(\Omega, \mathcal{F}, P)$

$$\mathbf{X} = (X_1, X_2, \ldots, X_d) : \Omega \longrightarrow \mathbb{R}^d$$

whose behavior is described by the $d$-dimensional distribution function $F_{\mathbf{X}} : \mathbb{R}^d \longrightarrow [0,1]$ such that

$$F_{\mathbf{X}}(\mathbf{x}) = P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d]$$

for all $\mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$.

Depending on the application, the $X_i$ may have a quite different interpretations.

- In Portfolio Management, the $X_i$ can represent (e.g., daily) returns of the assets constituting a portfolio of investments.
- In Credit Risk (or Reliability Theory), the $X_i$ take values on $\mathbb{R}^+$ and have the interpretation of time-to-default (lifetimes) of firms or components.
- In Hydrology, the $X_i$ may be random quantities related to the same meteorological event. For instance, to each storm event one can associate two r.v.s: intensity and duration of the raining event.

A stochastic model is a tool to enable both decisions to be made (prediction) and inferences to be drawn from previous knowledge.

The information about a multivariate stochastic model $\mathbf{X}$ contained in its d.f. is usually selected on the basis of:

- historical data (previously collected from similar models),
- theoretical considerations about the nature of the model,
- expert opinion.

The study about stochastic models usually aims at estimating certain system quantities, which describe the r.v. $\mathbf{X}$ as a whole.

- In Portfolio Management, it is of interest to evaluate the total value $Y$ of a portfolio, $Y = \sum_{i=1}^{d} w_i X_i$ for some weights $w_i \in [0;1]$ such that $\sum_{i=1}^{d} w_i = 1$, and hence to calculate the risk (e.g., $VaR_\alpha$) of this portfolio.
- In Credit Risk, one can be interested in the number of defaults in the given set of firms.
- In Reliability, one can be interested in the random lifetime of a $d$-components system, usually expressed as $Y = \phi(X_1, X_2, \ldots, X_d)$.

All these system quantities strongly depend on the distribution function $F_{\mathbf{X}}$ and serve to quantify the risk of a given system.

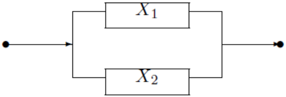
When a stochastic model $X = (X_1, X_2, \ldots, X_d)$ is given, two problems are of interest:

- to study the probabilistic behavior of each component $X_i$;
- to investigate their dependence or correlation structure.

Historically, while the univariate (marginal) behavior has been modelled by a number of different distributions (Gaussian, Weibull, Pareto, Exponential, etc.), the interdependencies among the components of the vector have been captured by simplified assumptions (e.g., independence) and/or numerical quantities (e.g., correlation coefficients).

## Why to investigate the dependence structure?

This is a very simple example. Consider a parallel system, composed by two components having random lifetimes $X_1$ and $X_2$, with $X_i \sim U[0, 1]$




Let $Y = X_1 \vee X_2$ be the lifetime of the system ($\vee = \mathrm{maximum}$). Consider three different cases:

1. The lifetimes $X_1$ and $X_2$ are independent;
2. The lifetimes $X_1$ and $X_2$ are such that $X_2 = X_1$ with probability 1 (maximal possible positive dependence);
3. The lifetimes $X_1$ and $X_2$ are such that $X_2 = 1 - X_1$ with probability 1 (maximal possible negative dependence).

What is the condition that maximizes (in some stochastic sense) the lifetime $Y$ of the system?

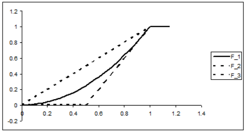
### Definition

Given the random lifetimes $Y_1$ and $Y_2$ we say that $Y_1$ is *stochastically smaller* than $Y_2$ (shortly, $Y_1 \leq_{st} Y_2$) if

$$F_{Y_1}(t) = P[Y_1 \leq t] \geq P[Y_2 \leq t] = F_{Y_2}(t) \quad \forall t \geq 0.$$

We have:

1. $P[Y_1 \leq t] = P[X_1 \leq t] \cdot P[X_2 \leq t] = t^2, \quad t \in [0,1];$
2. $P[Y_2 \leq t] = P[X_1 \leq t] = t \quad t \in [0,1];$
3. $P[Y_3 \leq t] = P[1 - t \leq X_1 \leq t] = 2t - 1, \quad t \in [1/2, 1].$




We have $Y_2 \leq_{st} Y_1 \leq_{st} Y_3$. Thus, negative dependence improves the reliability $Y$ of the system, while positive dependence reduces it.

## INDEPENDENCE OF EVENTS

Events are independent when the happening of any one of them does neither increase nor abate the probability of the rest. [T. Bayes, 1763]

The events $A_1, \dots, A_n$ are said to be *independent* if, for any $m$, $2 \le m \le n$, and for any pairwise distinct natural numbers $k_1, \dots, k_m$, the probability of the joint occurrence of the events is equal to the product of their probabilities:

$$P[A_{k_1} \cap \ldots \cap A_{k_m}] = P[A_{k_1}] \cdot \cdots \cdot P[A_{k_m}].$$

Hence, the conditional probability of each event given the occurrence of any combination of the others is equal to its unconditional probability.

## INDEPENDENCE OF RANDOM VARIABLES

The r.v.s $X_1, \ldots, X_d$ are said to be *independent* if, for any $x \in \mathbb{R}^d$, it holds:

$$F(x_1, \ldots, x_d) = F_1(x_1) \cdot \ldots \cdot F_d(x_d),$$

where, for each $i \in \{i, \ldots, d\}, X_i \sim F_i$.

In particular, if $(X_1, ..., X_d)$ has a density $f$, then for all $(x_1, ...., x_d) \in \mathbb{R}^d$,

$$f(x_1, ....., x_d) = f_1(x_1) \cdot ... \cdot f_d(x_d),$$

where, for $i \in \{1, \ldots, d\}$, $F_i(x) = \int_{-\infty}^{x} f_i(u)du$, i.e., $f_i$ is the density of $X_i$.

Warning: Pairwise independence does not necessarily entail independence.

## CORRELATION

The *correlation coefficient* is a numerical characteristic of the joint distribution of two random variables, expressing a relationship between them.

The correlation coefficient $\rho = \rho(X_1, X_2)$ for r.v.s $X_1, X_2$ with mathematical expectations $\mu_1, \mu_2$ and non-zero standard deviations $\sigma_1, \sigma_2$, respectively, is defined by

$$\rho = \frac{E[(X_1 - \mu_1)(X_2 - \mu_2)]}{\sigma_1 \sigma_2}$$

This correlation coefficient was developed by Karl Pearson from a related idea introduced by Francis Galton in the 19-th Century.

## PROPERTIES OF CORRELATION

- $\rho(X_1, X_2)$ is symmetric with respect to $X_1, X_2$:

$$\rho(X_1, X_2) = \rho(X_2, X_1).$$

- $\rho(X_1, X_2)$ is invariant under affine transformations, i.e., for all $a_1, a_2 > 0$ and $b_1, b_2 \in \mathbb{R}$,

$$\rho(a_1 X_1 + b_1, a_2 X_2 + b_2) = \rho(X_1, X_2).$$

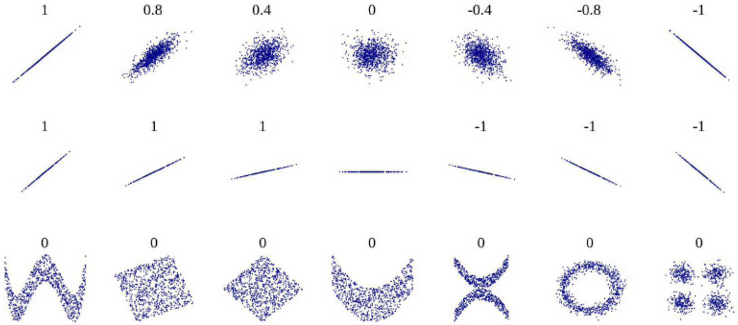
- If $X_1, X_2$ are independent, then $\rho(X_1, X_2) = 0$.
- $\rho(X_1, X_2) = \pm 1$ if and only if the relationship between the random variables is linear.

## PITFALLS OF CORRELATION

Possible disadvantages of Pearson's correlation are:

- $\rho(X_1, X_2)$ is only defined when the variances of $X_1$ and $X_2$ exist. (example: Pareto distributions may not have finite moments).
- $\rho(X_1, X_2) = 0$ does not imply that $X_1$ and $X_2$ are independent. Example: $X_1 \sim N(0, 1), X_2 = X_1^2$.
- The correlation coefficient together with the marginal behavior of $X_1$ and $X_2$ does not determine uniquely the joint distribution.
- For given univariate d.f.s $F_1$ and $F_2$ it is not always possible to construct a joint distribution $F$ with margins $F_1$ and $F_2$ and correlation coefficient equal to any $\rho \in [-1, 1]$.

## Examples of data sets with different correlations



### The multivariate Gaussian distribution

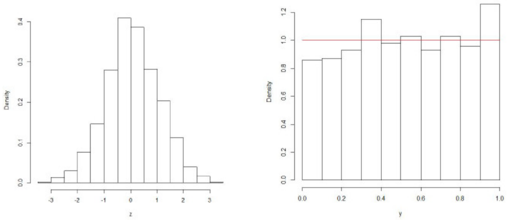
The *multivariate Gaussian* distributions have dominated the study of multivariate stochastic models for a long time. This is also due to the fact that they are uniquely represented by the parameters $\mu_i, \sigma_i$ and $\rho_{i,j}$.

However, several studies have recognized the necessity for examining alternatives to this setup (normal mixtures, elliptical, etc.). There are three main defects of these distributions.

- The tails of the univariate marginal distributions are too thin; they do not assign enough weight to extreme events.
- The joint tails of the distribution do not assign enough weight to joint extreme outcomes.
- The distribution has a strong form of symmetry, known as elliptical symmetry.

Let $X$ be a random variable on the probability space $(\Omega, \mathcal{F}, P)$, with distribution function $F$.

If $F$ is continuous, then $F(X)$ is uniformly distributed on $[0, 1]$.
Such a $F(X)$ is called *Probability Integral Transform* (shortly, PIT) of $X$.




Histogram of 1000 points from $Z \sim N(0, 1)$ (left) and histogram of the PIT of $Z$ (right).

**Quantile function**

Let $F$ be a univariate distribution function. We call *quantile inverse* of $F$ the function $F^{-1} : (0,1) \longrightarrow (-\infty, \infty)$ given by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$$

with the convention $\inf \emptyset = \infty$.

If $U$ is a random variable that is uniformly distributed on $[0,1]$, then $F^{-1}(U)$ has distribution function equal to $F$.

The previous result gives a procedure for simulating a random sample from a given d.f. $F$.

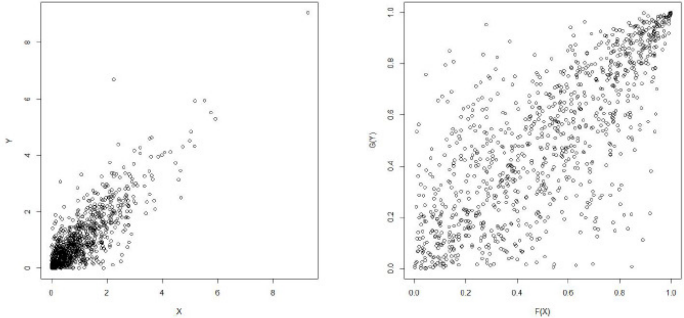
**Uniform representation of random vectors**

For r.v.s $X_1, X_2$ with continuous d.f.s $F_1, F_2$, both $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$ are uniformly distributed on $[0, 1]$. Thus,

$$\begin{aligned} F(x_1, x_2) &= P[X_1 \le x_1, X_2 \le x_2] \\ &= P[F_1(X_1) \le F_1(x_1), F_2(X_2) \le F_2(x_2)] \\ &= P[U_1 \le F_1(x_1), U_2 \le F_2(x_2)] \\ &= C(F_1(x_1), F_2(x_2)) \end{aligned}$$

where $C$ is the d.f. of $(U_1, U_2) = (F_1(X_1), F_2(X_2))$.

Since the d.f. C joins or couples a multivariate d.f. to its one-dimensional marginal d.f.s, we call it **copula**.
Plural: copulae or copulas.




Bivariate sample clouds of 2500 points from the d.f. $F = C(F_1, F_2)$ where $F_1, F_2 \sim Exp(1)$, the Spearman's $\rho$ is equal to $0.75$ (left), and the corresponding copula representation.

## Definition of copula

**Definition**

For every $d \geq 2$, a $d$-dimensional copula (shortly, a $d$-copula) is a $d$ dimensional distribution function whose univariate marginals are uniformly distributed on $[0, 1]$.

As a consequence of correspondence between d.f.s and r.v.s, to each copula $C$ there corresponds a random vector $\mathbf{U}$ defined on a suitable probability space such that the joint d.f. of $\mathbf{U}$ is given by $C$.

## Sklar's Theorem

**Theorem (Sklar, 1959)**

*Let $(X_1, ....X_d)$ be a r.v. with joint d.f. $F$ and univariate marginals $F_1, F_2, ...F_d$. Then there exists a d.f. $C : [0,1]^d \longrightarrow [0,1]$, called d-copula, such that, for all $\mathbf{x} \in \mathbb{R}^d$,*

$$F(x_1, x_2, ..., x_d) = C(F_1(x_1), F_2(x_2), ....., F_d(x_d)).$$

*$C$ is uniquely determined on $Range(F_1) \times .... \times Range(F_d)$ and, hence, it is unique, when $F_1, ...., F_d$ are continuous.*

Recall that $C$ is the d.f. of $(F_1(X_1), ....F_d(X_d))$, thus

$$C(u_1, ...u_d) = F(F_1^{-1}(u_1), ....., F_d^{-1}(u_d)).$$

## Gaussian copula and $t$-Student copula

The copula of the bivariate Normal distribution, also called *Gaussian copula*, is given by

$$C_\theta(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\Big(-\frac{s^2-2\theta st+t^2}{2(1-\theta^2)}\Big) ds dt,$$

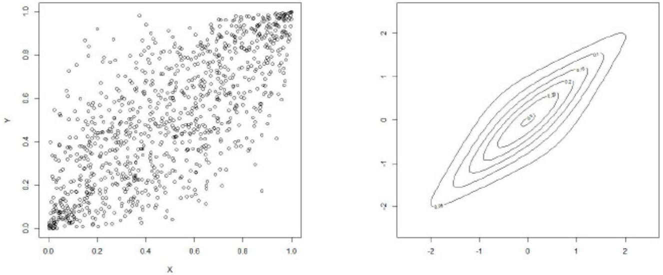
where $\theta \in [-1,1]$, and $\Phi^{-1}$ denotes the inverse of the univariate Normal distribution function.

The copula of the bivariate $t$-Student distribution with $\nu > 2$ degrees of freedom, also called *t-Student copula*, is given by

$$C_{\theta,\nu}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \Big(1+\frac{s^2-2\theta st+t^2}{\nu(1-\theta^2)}\Big)^{-(\nu+2)/2} ds dt,$$

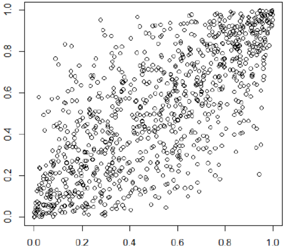
where $\theta \in [-1,1]$, and $t_\nu^{-1}$ denotes the inverse of the univariate $t$-distribution function.

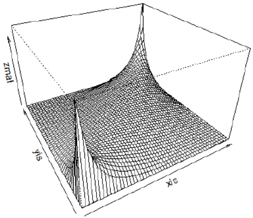
# Gaussian copula




Bivariate sample clouds of 1000 points from a Gaussian copula with $\rho = 0.75$ (left side) and contour plot of the density of the d.f. $F = C(F_1, F_2)$ where $F_1, F_2 \sim N(0,1)$, $C$ is a Gaussian copula (right figure).

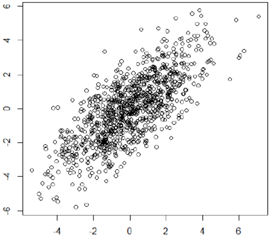
Pay attention at the difference between the Gaussian copula and the Gaussian distribution!!!

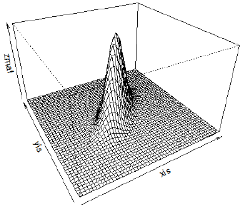



Gaussian copula

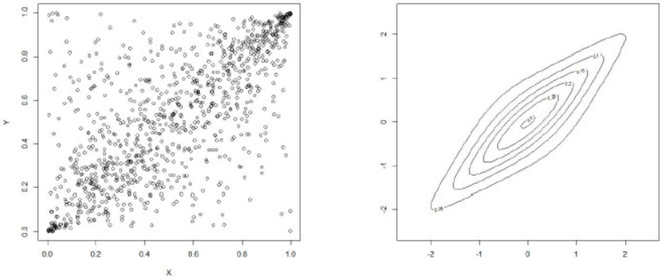
Pay attention at the difference between the Gaussian copula and the Gaussian distribution!!!



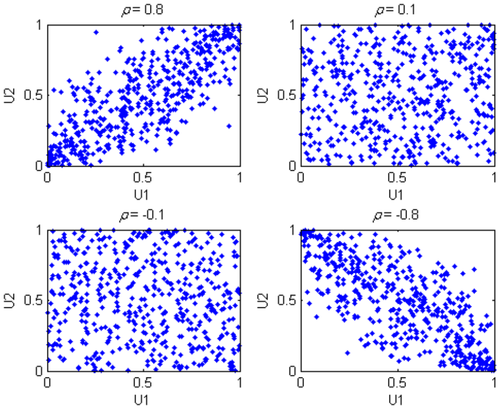

Gaussian distribution

# $t$-Student copula

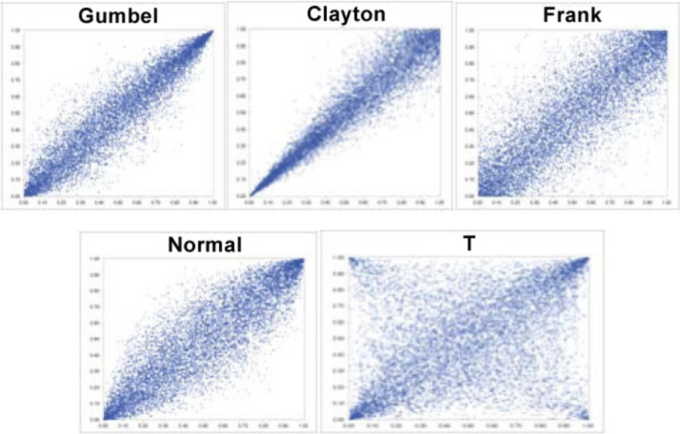



Bivariate sample clouds of 1000 points from $t$–Student copula (left side) and contour plot of the density of the d.f. $F = C(F_1, F_2)$ where $F_1, F_2 \sim N(0,1)$, $C$ is a $t$–Student copula ($\theta = 0.5, \nu = 2$) (right figure).

## Possible shapes of copulas



## Possible shapes of copulas



**Copulas for discrete r.v.s**

Several statistical models aim at describing the joint behavior of discrete r.v.s $X$ and $Y$ taking values on $\mathbb{N}$.

In these cases, Sklar's Theorem is still valid, but the copula of the model cannot be uniquely identified.

When dealing with count data, modeling and interpreting dependence through copulas should be done by using some caution.

For more details, see: C. Genest and J. Nelehová, A primer on copulas for count data, The Astin Bulletin 37 (2007) 475515.

## Viceversa!

**Theorem (Sklar, 1959)**

*If $F_1, F_2, \ldots F_d$ are univariate d.f.s, and if $C$ is any $d$copula, then $F : \mathbb{R}^d \longrightarrow [0,1]$ defined by*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots\ldots, F_d(x_d)).$$

*is a d-dimensional d.f. with marginals $F_1, F_2, \ldots F_d$.*

Thus, copulas allow to construct multivariate d.f.s where the margins could take very different forms. Moreover, such a method to construct d.f.s is not restrictive, due to Sklar's Theorem.

We can construct (and fit) parametric statistical models via a two-stage procedure:

- first, choose the univariate marginals $F_1^{\alpha_1}, F_2^{\alpha_2}, \ldots F_d^{\alpha_d}$;
- then, choose our favorite copula $C_\theta$,
- mix the two ingredients and obtain the model

$$F_{\overline{\alpha},\theta}(\mathbf{x}) = C_\theta(F_1^{\alpha_1}(x_1), F_2^{\alpha_2}(x_2), \ldots F_d^{\alpha_d}(x_d)).$$

**How to visualize dependence**

Plot your data $(X_i; Y_i), i = 1, ..., n$ in a bivariate graph.

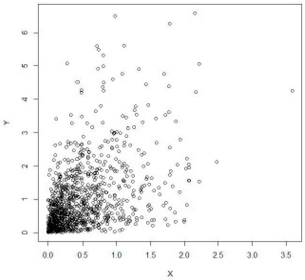
Transform each univariate set of observation by means of the respective empirical cumulative distribution function

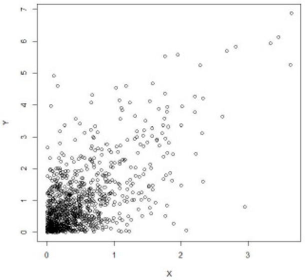
$$F_{X,n}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(X_i \leq x)} \qquad F_{Y,n}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(Y_i \leq x)}$$

Plot the pairs

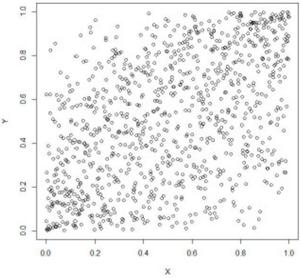
$$(U_i, V_i) = (F_{X,n}(X_i), F_{Y,n}(Y_i)).$$

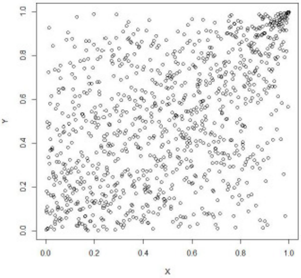
How to visualize dependence: scatterplots



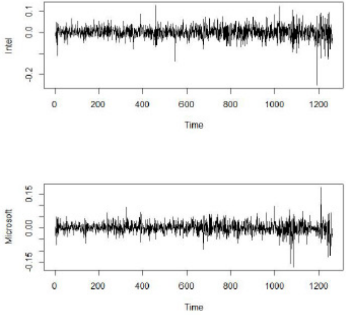



How to visualize dependence: normalized ranks



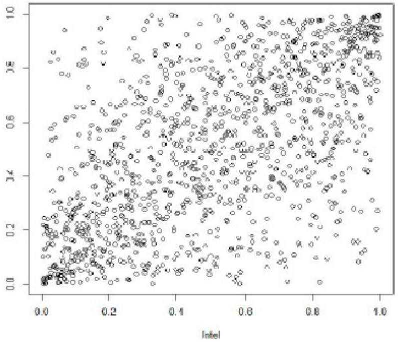



## How to visualize dependence: example




Univariate time series plots of five years of daily log-returns (from 1996 to 2000) of Intel (INTC) and Microsoft (MSFT) stocks.

How to visualize dependence: normalized ranks




Univariate normalized ranks of the time series of five years of daily log-returns (from 1996 to 2000) of Intel (INTC) and Microsoft (MSFT) stocks.

**The copula $M_d$**

Let $U$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Suppose that $U$ is uniformly distributed on $[0, 1]$.

Consider the r.v. $\mathbf{U} = (U, ...., U)$. Then, for all $\mathbf{u} \in [0, 1]^d$,

$$P[\mathbf{U} \leq \mathbf{u}] = P[U \leq \min\{u_1, u_2, ...., u_d\}] = \min\{u_1, u_2, ...., u_d\}.$$

Thus the d.f. given, for every $\mathbf{u} \in [0, 1]^d$, by

$$M_d(u_1, u_2, ...., u_d) := \min\{u_1, u_2, ...., u_d\}$$

is a copula, which will be called the *comonotone* copula.

**The copula $\Pi_d$**

Let $U_1, U_2, ..., U_d$ be independent r.v.s defined on the probability space $(\Omega, \mathcal{F}, P)$. Suppose that each $U_i$ is uniformly distributed on $[0,1]$.

Consider the r.v. $\mathbf{U} = (U_1, ...., U_d)$. Then, for all $\mathbf{u} \in [0,1]^d$,

$$P[\mathbf{U} \leq \mathbf{u}] = \prod_{i=1}^{d} P[U_i \leq u_i] = u_1 \cdot u_2 .... \cdot u_d.$$

Thus the d.f. given, for every $\mathbf{u} \in [0,1]^d$, by

$$\Pi_d(u_1, u_2, ...., u_d) := u_1 \cdot u_2 .... \cdot u_d$$

is a copula, which will be called the *independence* copula.

**The copula $W_2$**

Let $U$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Suppose that $U$ is uniformly distributed on $[0, 1]$.

Consider the random pair $\mathbf{U} = (U, 1 - U)$. Then, for all $(u_1, u_2) \in [0, 1]^2$,

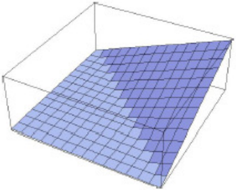
$$P[\mathbf{U} \leq \mathbf{u}] = P[U \leq u_1, 1 - U \leq u_2] = \max\{0, u_1 + u_2 - 1\}.$$

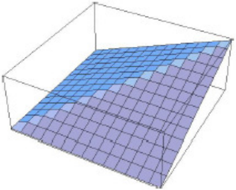
Thus the d.f. given, for every $(u_1, u_2) \in [0, 1]^2$, by

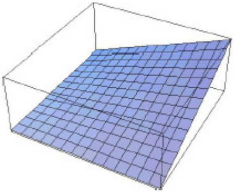
$$W_2(u_1, u_2) := \max\{0, u_1 + u_2 - 1\}$$

is a copula, which will be called the *countermonotone* copula.

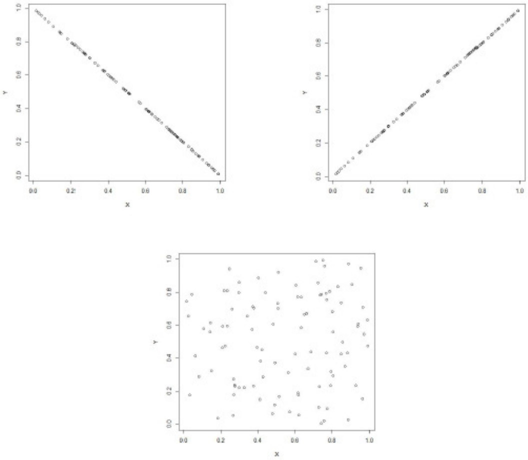
# The copulas $M_2, \Pi_2, W_2$







## The copulas $M_2, \Pi_2, W_2$



**Fréchet family**

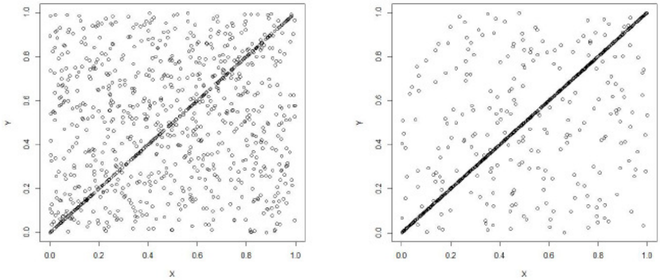
The standard expression for members of this family is

$$C_\theta(\mathbf{u}) = \theta\Pi_d(\mathbf{u}) + (1-\theta)M_d(\mathbf{u})$$

where $\mathbf{u} \in [0,1]^d$, and $\theta \in [0,1]$ is a dependence parameter.

If $U_1, U_2, ..., U_d$ are r.v.s with copula $C_\theta$, then they are independent for $\theta = 1$. Moreover, $C_0 = M_d$.

# Fréchet family




Bivariate sample clouds of 1000 points from a Frechét copula with $\theta = 0.75$ (left) and $\theta = 0.25$ (right).

**Absolutely continuous copulas**

**Definition**

A copula $C$ is absolutely continuous if it can be expressed in the form

$$C(\mathbf{u}) = \int_{[\mathbf{0},\mathbf{u}]} c(\mathbf{t})d\mathbf{t}$$

for a suitable integrable function $c : [0,1]^d \longrightarrow \mathbb{R}^+$.

The function $c$ is called the *density* of $C$. For continuously derivable functions up to order $d$, we have

$$c(\mathbf{u}) = \frac{\partial^d C}{\partial u_1 \partial u_2 ... \partial u_d}(\mathbf{u})$$

Example: $\Pi_d$ is absolutely continuous with density $c(\mathbf{u}) = 1$.

## Example: FGM family

The standard expression for members of *Farlie-Gumbel-Morgenstern* (FGM) family of copulas is

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v),$$

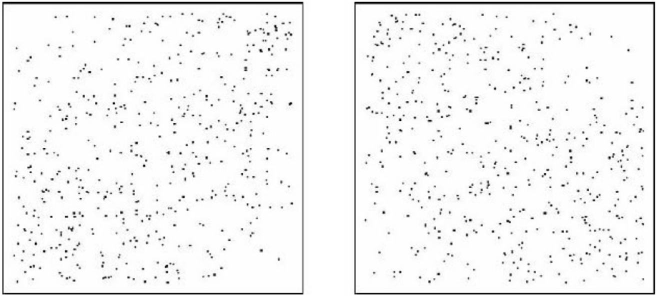
where $(u,v) \in [0,1]^2$, and $\theta \in [-1,1]$ is a dependence parameter.

- If $U$ and $V$ are r.v.s with copula $C_\theta$, then they are independent for $\theta = 0$, i.e. $C_0 = \Pi_2$.
- If $U$ and $V$ are r.v.s with copula $C_\theta$, then they are positively dependent for $\theta > 0$.
- If $U$ and $V$ are r.v.s with copula $C_\theta$, then they are negatively dependent for $\theta < 0$.

Every member of the FGM family is absolutely continuous, and its density has a simple expression given by

$$c_\theta(u,v) = 1 + \theta(1-2u)(1-2v).$$

## Example: FGM family



Scatterplots for FGM copulas with $\theta = 1$ (left) and $\theta = -1$ (right).

**Example: Marshall-Olkin family**

Let $Y_1, Y_2$ and $Y_3$ be three independent exponentially distributed random variables, having parameters $\lambda_1, \lambda_2$ and $\lambda_3$, respectively (with $\lambda_i > 0$). Consider the vector

$$(X_1, X_2) = (Y_1 \wedge Y_3, Y_2 \wedge Y_3). \qquad [\wedge = \mathrm{minimum}]$$

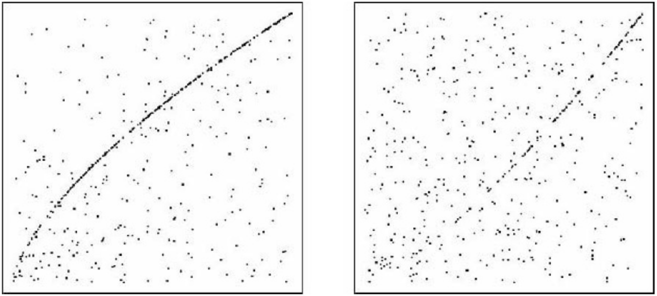
The variables $X_1$ and $X_2$ are clearly dependent, and the joint distribution is known as the Marshall Olkin bivariate exponential distribution. The corresponding (survival) copula is the *Marshall Olkin* copula, defined as

$$C_{\alpha,\beta}(u,v) = uv \min\{u^{-\alpha}, v^{-\beta}\} = \begin{cases} u^{1-\alpha}v & \text{if } u^\alpha \geq v^\beta, \\ uv^{1-\alpha} & \text{if } u^\alpha \leq v^\beta \end{cases},$$

where $\alpha = \lambda_3/(\lambda_1 + \lambda_3)$ and $\beta = \lambda_3/(\lambda_2 + \lambda_3)$.

This copula always describes positive dependence, and it is not absolutely continuous.

## Example: MO family



Scatterplots for MO copulas with $(\alpha, \beta) = (1/2, 3/4)$ (left) and with $(\alpha, \beta) = (1/3, 1/4)$ (right).

## Why to investigate the dependence structure?

Recipe for a disaster: Li's model for credit portfolio

FINANCIAL TIMES

April 24, 2009

**Of couples and copulas: the formula that felled Wall Street**

Consider a credit portfolio of d bonds issued by some companies $Z_i$. The value of the portfolio depends on the joint d.f. of $\mathbf{T} = (T_1, ...., T_d)$, where $T_i$ represents the default time of $Z_i$.

For describing the joint behavior of $\mathbf{T}$, Li (2000) [Li, David X., "On Default Correlation: A Copula Function Approach." Journal of Fixed Income 9, 43-54] considered the following approach:

$$P[T_1 \leq t_1, ...., T_d \leq t_d] = C^{Ga}(F_1(t_1), ..., F_d(t_d));$$

where $C^{Ga}$ is the Gaussian copula and $F_i \sim \mathrm{Exp}(\lambda_i)$.

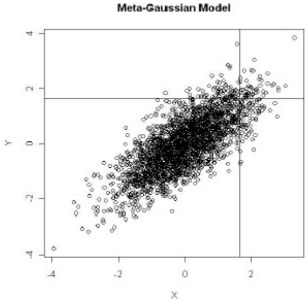
The pitfalls of the Li's model:

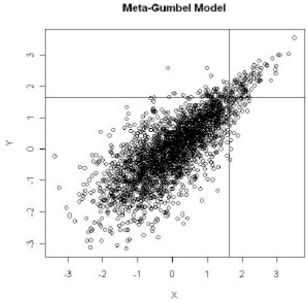
For a discussion, see
`http://sites.google.com/site/fbdurante/discussions`

> One of the main disadvantages of the model is that it does not adequately model the occurrence of defaults in the underlying portfolio of corporate bonds. In times of crisis, corporate defaults occur in clusters, so that if one company defaults then it is likely that other companies also default within a short time period. Under the Gaussian copula model, company defaults become independent as their size of default increases. [Donnelly and Embrechts, 2010]

> Very few people understand the essence of the model. The current copula framework gains its popularity owing to its simplicity. [Li, 2005]

Behavior of corporate defaults under different scenarios




Bivariate sample clouds of 2500 points from the d.f. $F = C(F_1, F_2)$ where $F_1, F_2 \sim N(0,1)$, the Spearmans $\rho$ is equal to 0,75, and $C$ is a Gaussian copula (left figure) or a Gumbel copula (right figure). The points in the right-upper side of the figure are, respectively, 51 and 92.

**Rank-invariant property**

**Theorem**

*Let $(X_1, ..., X_d)$ be a random vector with continuous d.f. $F$, univariate marginals $F_1, ...., F_d$, and copula $C$. Let $h_1, h_2, ..., h_d$ be strictly increasing functions from $\mathbb{R}$ to $\mathbb{R}$. Then $C$ is also the copula of the random vector*

$$(h_1(X_1), ...., h_d(X_d)).$$

As a consequence, the study of rank statistics, i.e., the study of properties invariant under such transformations, may be characterized as the study of copulas and copula-invariant properties.

**Copulas and independence**

Copulas describe several dependence properties of a random vector, independently from the marginal behavior.

**Theorem**

*Let $(X_1, ..., X_d)$ be a random vector with continuous d.f. $F$, univariate marginals $F_1, ...., F_d$, and copula $C$. Then $X_1, ..., X_d$ are independent iff the copula $C$ is the independence copula, i.e., iff $C = \Pi_d$.*

**Copulas and monotone dependence**

The continuous r.v.s $X_1, ..., X_d$ are called *comonotonic* if they admit as copula $M_d$.

**Theorem**

*Let $(X_1, ..., X_d)$ be a random vector with continuous d.f. $F$, univariate marginals $F_1, ...., F_d$. Then $X_1, ..., X_d$ are comonotonic iff*

$$(X_1, ..., X_d) =_{st} (h_1(X), ...., h_d(X)) \quad [=_{st} \textit{ means equality in distribution}]$$

*for some $X$ and increasing functions $h_1, h_2, ..., h_d$.*

### Copulas and countermonotone dependence

The continuous r.v.s $X_1, X_2$ are called *countermonotonic* if they admit as copula $W_2$.

**Theorem**

*Let $(X_1, X_2)$ be a random vector with continuous d.f. $F$, univariate marginals $F_1, F_2$. Then $X_1, X_2$ are countermonotonic iff*

$$X_2 =_{st} h(X_1)$$

*for some decreasing function $h$.*

## Bounds for copulas

**Theorem (Fréchet-Hoeffding bounds)**

*For any copula $C$ and for all $\mathbf{u} \in [0,1]^d$,*

$$W_d(\mathbf{u}) \leq C(\mathbf{u}) \leq M_d(\mathbf{u}),$$

*where*

$$W_d(\mathbf{u}) = \max\left\{\sum_{i=1}^{d} u_i - d + 1, 0\right\}$$

*These bounds are the best-possible, in the sense that*

$$\inf_{C \in \mathcal{C}_d} C(\mathbf{u}) = W_d(\mathbf{u}) \quad \text{and} \quad \sup_{C \in \mathcal{C}_d} C(\mathbf{u}) = M_d(\mathbf{u})$$

Note that $W_d$ is a copula only for $d = 2$.

## Concordance and Concordance Order

Informally, two random variables are said to be concordant if large values of one of them correspond to large values of the other, and if small values of one correspond to small values of the other.

Let $(x_i, y_i)$ and $(x_j, y_j)$ denote two observations from a vector $(X, Y)$ of continuous random variables.

We say that $(x_i, y_i)$ and $(x_j, y_j)$ are *concordant* if $x_i < x_j$ and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. Similarly, we say that $(x_i, y_i)$ and $(x_j, y_j)$ are *discordant* if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$.

Thus, $(x_i, y_i)$ and $(x_j, y_j)$ are concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$.

We can assert that a vector $(X, Y)$ is *concordant* if, considered two observation from it, there is an high probability of concordance, and that it is *discordant* if there is an high probability of discordance

We can formalize the concept of concordance by means of a partial order on the set of copulas, suggested also by the Fréchet-Hoeffding bounds described previously.

**Definition**

If $C_1$ and $C_2$ are copulas, we say that $C_1$ is *smaller than* $C_2$ in *concordance*, and we write $C_1 \prec_c C_2$, if

$$C_1(u, v) \leq C_2(u, v) \text{ for all } (u, v) \in [0, 1]^2.$$

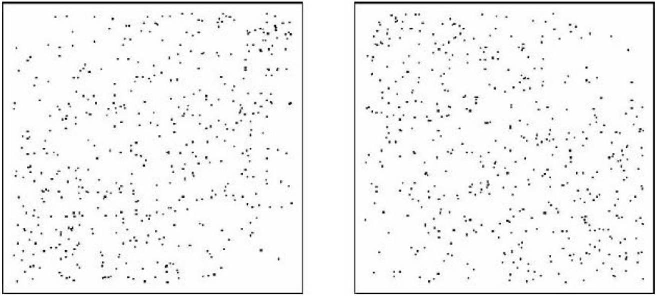
The inequality $C_1(u, v) \leq C_2(u, v)$ means that the probability mass concentrated in the lower orthant

$$L_{(u,v)} = \{(x, y) \in [0, 1]^2 : x \leq u, y \leq v\}$$

is smaller for $C_1$ than for $C_2$, whatever we choose $(u, v) \in [0, 1]^2$. Thus, for $C_1$ we have a lower probability, with respect to $C_2$, of extracting a point that is in the left-down corner of the support, i.e., a point in a region of concordance.

Note that in the case $d = 2$ this order is equivalent to the order defined considering upper orthants instead of lower orthants.

**Example:**



Scatterplots for FGM copulas with $\theta = 1$ (left) and $\theta = -1$ (right). Observe that the first one is clearly more concordant than the second one.

An interesting fact of the concordance order is that it can be used to compare the degree of positive dependence in vectors having different marginal distribution.

**Definition**

Let $\mathbf{X} = (X_1, X_2)$ and $\mathbf{Y} = (Y_1, Y_2)$ be two 2-dimensional random vectors having copulas $C_{\mathbf{X}}$ and $C_{\mathbf{Y}}$, respectively. Then $\mathbf{X}$ is said to be smaller than $\mathbf{Y}$ in the *concordance order* (denoted by $\mathbf{X} \prec_c \mathbf{Y}$) if

$$C_{\mathbf{X}}(u_1, u_2) \le C_{\mathbf{Y}}(u_1, u_2) \quad \forall (u_1, u_2) \in [0, 1]^2.$$

It is useful to observe that, in order to be comparable in concordance, the vectors $\mathbf{X}$ and $\mathbf{Y}$ do not need to have the same marginal distributions.

When $\mathbf{X}$ and $\mathbf{Y}$ have the same marginal distributions then this comparison is commonly denoted as *Positive Orthant Dependence* (shortly, $\prec_{POD}$ ). Note that $\mathbf{X} \prec_{POD} \mathbf{Y} \quad \Rightarrow \quad \mathbf{X} \prec_c \mathbf{Y}$, but not the viceversa.

## Positive and Negative Dependence

Now, it is possible to define positive, or negative, dependence properties of a vector $\mathbf{X} = (X_1, X_2)$ by means of a comparisons with respect to its independent counterpart.

### Definition

Let $\mathbf{X} = (X_1, X_2)$ be a random vector having copula $C_{\mathbf{X}}$. Then $\mathbf{X}$ is said to be *Positive Orthant Dependent*, shortly POD, (and, informally, it is said to have *positive dependence*) if $\mathbf{X}^{\perp} \prec_c \mathbf{X}$, where $\mathbf{X}^{\perp}$ denotes the independent version of $\mathbf{X}$. Equivalently, $\mathbf{X}$ is POD if

$$uv \leq C_{\mathbf{X}}(u, v) \quad \forall (u, v) \in [0, 1]^2,$$

i.e., if

$$F_1(t)F_2(s) \leq F_{\mathbf{X}}(t, s) \quad \forall (t, s) \in \mathbb{R}^2.$$

Similarly is the defined the *Negative Orthant Dependent* (NOD) property.

Among other useful properties, we remind the following:

- $(X_1, X_2) \prec_c (Y_1, Y_2)$ implies

$$E[\phi_1(X_1)\phi_2(X_2)] \leq E[\phi_1(Y_1)\phi_2(Y_2)]$$

  for all increasing functions $\phi_i$ such that the two expectations exist.
- If $(X_1, X_2)$ is POD [NOD] then

$$E[\phi_1(X_1)]E[\phi_2(X_2)] \leq \ [\geq]\ E[\phi_1(X_1)\phi_2(X_2)]$$

  for all increasing functions $\phi_i$ such that the expectations exist.

## Rank correlation

Pearsons correlation coefficient depends on the underlying distribution, and may even not exist (when there is no finite second moment).

In the literature, non-parametric and robust alternatives have been proposed, which are only based on the ranks of the observations. These are called *rank correlation coefficients*, or *concordance measures*, or *measures of association*.

Rank correlation coefficients share some of the properties of Pearsons correlation coefficient:

- they are symmetric,
- lie between -1 and 1,
- if $X_1$ and $X_2$ are independent, their rank correlation is equal to 0.

Moreover, since they are based on ranks, rank correlations are invariant with respect to increasing transformations.

## Kendall's concordance index $\tau$

Let $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ denote a random sample of $n$ observations from a vector $(X, Y)$ of continuous random variables.

There are $\binom{n}{2}$ distinct pairs $(x_i, y_i)$ and $(x_j, y_j)$ of observations in the sample, and each pair is either concordant or discordant (i.e., such that $(x_i - x_j)(y_i - y_j) > 0$ or $(x_i - x_j)(y_i - y_j) < 0$. Let $c$ denote the number of concordant pairs and $d$ the number of discordant pairs. Then *Kendall's* $\tau$ for the sample is defined as

$$\tau = \frac{c-d}{c+d} = \frac{(c-d)}{\binom{n}{2}}$$

Equivalently, $\tau$ is the probability of concordance minus the probability of discordance for a pair of observations $(x_i, y_i)$ and $(x_j, y_j)$ that is chosen randomly from the sample.

The population version of Kendall's $\tau$ for a vector $(X, Y)$ of continuous random variables with joint distribution function $F$ is defined similarly. Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be independent and identically distributed random vectors, each with joint distribution function $F$. Then the population version of Kendall's tau is defined as the probability of concordance minus the probability of discordance:

$$\tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

After some calculations it can be proved that

$$\tau_{X,Y} = 4 \int\int_{[0,1]^2} C(u, v) dC(u, v) - 1,$$

where $C$ is the copula of $(X, Y)$.

## Spearman's concordance index $\rho$

As with Kendall's $\tau$, the population version of the measure of association known as *Spearman's* $\rho$ is based on concordance and discordance.

$$\rho = \rho_{X,Y} = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}, \qquad d_i = \mathrm{rank}(x_i) - \mathrm{rank}(y_i).$$

To obtain the population version of this measure, let $(X_1, Y_1), (X_2, Y_2)$ and $(X_3, Y_3)$ be three independent random vectors with a common joint distribution function $F$ and copula $C$.

The population version $\rho_{X,Y}$ of Spearman's coefficient is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors $(X_1, Y_1)$ and $(X_2, Y_3)$, i.e., a pair of vectors with the same margins, but one vector has distribution function $F$, while the components of the other are independent:

$$\rho = \rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

After some calculations it can be proved that

$$\rho_{X,Y} = 12 \int\int_{[0,1]^2} C(u, v)\ du\ dv - 3.$$

Also, it can be proved that

$$\rho_{X,Y} = 12 \int\int_{[0,1]^2} [C(u, v) - uv]\ du\ dv.$$

The integral in the last expression represents the volume under the graph of the copula and over the unit square, and hence $\rho_{X,Y}$ is a scaled volume under the graph of the copula (scaled to lie in the interval $[-1, 1]$).

Thus $\rho_{X,Y}$ is a measure of average distance between the distribution of $(X, Y)$ (as represented by $C$) and independence (as represented by the copula $\Pi_2$).

**Example: FGM family**

Consider the FGM family of copulas

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v),$$

where $(u, v) \in [0, 1]^2$, and $\theta \in [-1, 1]$ is a dependence parameter.

Two useful relationships exist between $\theta$ and, respectively, Kendall's $\tau$ and Spearman's $\rho$:

$$\tau(\theta) = \frac{2\theta}{9} \quad \text{and} \quad \rho(\theta) = \frac{\theta}{3}.$$

Therefore, the values of $\tau(\theta)$ have a range of $[-2/9, 2/9]$, and the values of $\rho(\theta)$ have a range of $[-1/9, 1/9]$.

These limited intervals restrict the usefulness of this family for modelling.

## Multivariate measures of rank correlation

In the literature, several measures have been proposed to quantify the amount of dependence of a given vector $\mathbf{X}$ of dimension $d$.

Most of them are based on the copula $C$ of $\mathbf{X}$, like Kendall's $\tau$:

$$\tau_{\mathbf{X}} = \frac{1}{2^{d-1}} \left( 2^d \int_{[0,1]^d} C(\mathbf{u}) dC(\mathbf{u}) - 1 \right),$$

and Spearman's $\rho$:

$$\rho_{\mathbf{X}} = \frac{d+1}{2^d - (d+1)} \left( 2^d \int_{[0,1]^d} C(\mathbf{u}) d\mathbf{u} dv - 1 \right).$$

Under these definitions, it is easy to verify that $\mathbf{X} \prec_c \mathbf{Y}$ implies both

$$\tau_{\mathbf{X}} \leq \tau_{\mathbf{Y}} \quad \text{and} \quad \rho_{\mathbf{X}} \leq \rho_{\mathbf{Y}}$$

## Gini's "Indice di cograduazione semplice"

If $p_i$ and $q_i$ denote the ranks in a sample of size n of two continuous random variables $X$ and $Y$, respectively, then the Gini's index is defined as

$$\gamma_{X,Y} = \frac{\sum_{i=1}^{n} |p_i + q_i - n - 1| - \sum_{i=1} |p_i - q_i|}{\mathrm{int}(n^2/2)}.$$

The population version of $\gamma_{X,Y}$ can be computed letting the sample size $n$ go to $\infty$. After some calculations it can be proved that

$$\gamma_{X,Y} = 2 \int\int_{[0,1]^2} (|u + v - 1| - |u - v|) dC(u, v).$$

To better understand the Gini's index, let us define the operator $Q$ as follows.

Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be two r.v.s having the same marginal distributions and copulas $C_1$ and $C_2$, respectively. Let

$$Q = Q(X, Y) = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

It can be proved that $Q$ only depends on $C_1$ and $C_2$, and that

$$Q = Q(C_1, C_2) = Q(C_2, C_1) = 4 \int\int_{[0,1]^2} C_1(u, v) dC_2(u, v) - 1.$$

We have, e.g., $\tau_C = Q(C, C)$,

$$Q(M_2, M_2) = 1 \quad Q(\Pi_2, \Pi_2) = 0 \quad Q(W_2, W_2) = -1,$$

$$Q(C, M_2) \in [0, 1] \quad Q(C, \Pi_2) \in [-1/3, 1/3] \quad Q(C, W_2) \in [-1, 0],$$

**Theorem**

$$\gamma_{X,Y} = \gamma_C = Q(C, M_2) + Q(C, W_2).$$

*Thus $Q$ measures a concordance relationship, or distance, between $C$ and monotone dependence, as represented by the copulas $M_2$ and $W_2$.*

## Schweizer and Wolff's index

Recall that for continuous random variables $X$ and $Y$ with copula $C$, Spearman's $\rho$ can be written as

$$\rho_{X,Y} = 12 \int \int_{[0,1]^2} [C(u, v) - uv] dudv.$$

If in the integral above, we replace the difference $[C(u, v) - uv]$ with the absolute difference $|C(u, v) - uv|$, then we have a measure based upon the distance between the graphs of $C$ and $\Pi_2$. It is the Schweizer and Wolff's index $\sigma_C$

$$\sigma_C = 12 \int \int_{[0,1]^2} |C(u, v) - uv| dudv.$$

The interesting fact of this measure of concordance is that, among all those we have seen until now, it assumes value 0 if and **only if** the variables $X$ and $Y$ are independent.

**Why are families of copulas useful?**

If we have a collection of copulas, then, as a consequence of Sklar's theorem, we automatically have a collection of multivariate distributions with whatever marginal distributions we desire. Clearly this can be useful in modeling and simulation. [R.B. Nelsen, An Introduction to Copulas, 2006]

It is necessary to construct dependency models that reflect observed and expected dependencies without formalizing the structure of those dependencies with cause-effect models.
The theory of copulas provides a comprehensive modelling tool that can reflect dependencies in a very flexible way. [International Actuarial Association, 2004]

## Archimedean copulas

We call Archimedean generator any $\psi : [0, \infty] \longrightarrow [0, 1]$ such that

- it is decreasing and continuous;
- $\psi(0) = 1$;
- $\lim_{t\to\infty} \psi(t) = 0$;
- it is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$.

By convention, $\psi(+\infty) = 0$ and

$$\psi^{-1}(0) = \inf\{t \geq 0 : \psi(t) = 0\}$$

where $\psi^{-1}$ denotes the pseudoinverse of $\psi$.

For example, $\psi$ can be the survival function $\overline{F}(t) = P[X > t]$ for some univariate random variable $X$.

A $d$-dimensional copula $C$ is called *Archimedean* if it admits the representation

$$C_\psi(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \ldots.\psi^{-1}(u_d))$$

for all $\mathbf{u} \in [0,1]^d$ and for some Archimedean generator with $\psi(0) = 1$.

**Theorem**

*Let $\psi$ be an Archimedean generator. Let $C_\psi$ be the function defined above. If $\psi$ is $d$-monotone, i.e. if*

$$(-1)^k \psi^{(k)}(t) \geq 0 \quad \forall k \in \{0, 1, \ldots, d\}$$

*for all $t \geq 0$, then $C_\psi$ is a $d$-copula.*
*In particular, for $d = 2$, $C_\psi$ is a copula if and only if $\psi$ is decreasing and convex.*

**Gumbel-Barnett copula ($d = 2$)**

The *Gumbel-Barnett* copula is the Archimedean copula having generator $\psi_\theta(t) = \exp(\frac{1-\exp(t)}{\theta})$, with $\theta \in (0, 1]$.

$$C(u, v) = uv \exp(-\theta \ln u \ln v).$$

For $\theta \longrightarrow 0$ it converges to the independence copula $\Pi_2$, while for other values of $\theta$ it describes negative dependence (being $uv \exp(-\theta \ln u \ln v) \leq uv$ for $u, v \leq 0$ and $\theta > 0$

This is the copula of the *Gumbel multivariate exponential* distribution: the bivariate vector $(X, Y)$ has the Gumbel multivariate exponential distribution if

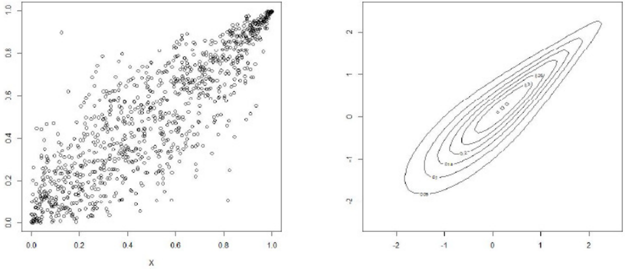
$$\overline{F}(t, s) = \exp(-\lambda t - \lambda s - \lambda^2 \theta ts),$$

with $\lambda > 0, \theta \in [0, 1]$ (and for $t, s \geq 0$).

**Gumbel-Hougaard copula** The *Gumbel-Hougaard* copula is the Archimedean copula having generator $\psi(t) = \exp(-u^{1/\theta})$, with $\theta \in [1, +\infty)$.

$$C(\mathbf{u}) = \exp\left(-\left(\sum_{i=1}^{d}(-\ln(u_i))^{\theta}\right)^{1/\theta}\right).$$

For $\theta = 1$ we obtain the independence copula as a special case, and the limit for $\theta \longrightarrow \infty$ is the comonotonicity copula. For other values of $\theta$ it describes positive dependence.




Bivariate sample clouds of 1000 points from Gumbel-Hougaard copula with $\theta = 3$ (left) and contour plot of the density of the d.f. $F = C_\theta(F_1, F_2)$ where $F_1, F_2 \sim N(0, 1)$ (right).

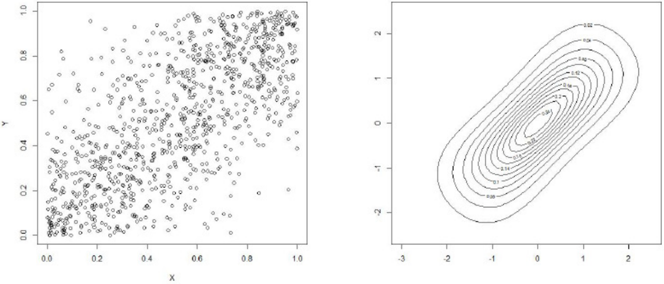
**Frank copula**

The standard expression for members of this family of $d$copulas is

$$C_\theta(\mathbf{u}) = -\frac{1}{\theta} \ln \left( 1 + \frac{\prod_{i=1}^{d} (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{d-1}} \right),$$

where $\theta > 0$. The limiting case $\theta = 0$ corresponds to $\Pi_d$. For the case $d = 2$, the parameter $\theta$ can be extended also to the case $\theta < 0$, allowing description of negative dependence.

The Archimedean generator is given by $\psi(t) = -\frac{1}{\theta} \ln(1 - (1 - e^{-\theta})e^{-t})$.




Bivariate sample clouds of 1000 points from Frank copula with $\theta = 5.7$ (left) and contour plot of the density of the d.f. $F = C_\theta(F_1, F_2)$ where $F_1, F_2 \sim N(0, 1)$ (right).

## Clayton copula

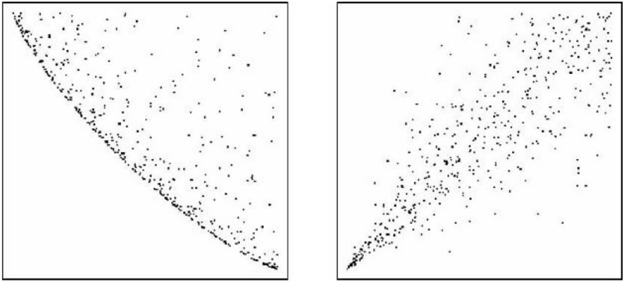
The standard expression for members of this family of $d$copulas is

$$C_\theta(\mathbf{u}) = \max\left\{\left(\sum_{i=1}^{d} u_i^{-\theta} - (d-1)\right)^{-\frac{1}{\theta}}, 0\right\},$$

where $\theta > \frac{-1}{d-1}, \theta \neq 0$. The limiting case $\theta = 0$ corresponds to $\Pi_d$. For $\theta > 0$, it describes positive dependence, and viceversa for $\theta < 0$. In general, for negative $\theta$ is should be used with some caution.

The Archimedean generator is given by $\psi(t) = (\max\{1 + \theta t, 0\})^{-\frac{1}{\theta}}$.

The Clayton copula satisfies a really important property, which will be described later



Scatterplots for Clayton copulas with $\theta = -0.8$ (left) and $\theta = 4$ (right).

**Kendall's $\tau$ for Archimedean copulas**

For different purposes it is useful to observe that, for Archimedean copulas, the following relationship holds.

$$\tau_C = 1 + 4 \int_0^1 \frac{\psi(t)}{\psi'(t)} dt$$

Thus, for example, for the Clayton family $C_\theta$ it holds

$$\tau_{C_\theta} = \frac{\theta}{\theta + 2}$$

.

# Some remarkable models

## Frailty models

Extensively considered in stochastic modeling of dependence to describe random lifetimes subjected to common random environments. According to these models, the *frailty* (an unobservable random variable that describes environmental factors) acts simultaneously on the distribution functions of the lifetimes.

### Definition

A vector of random lifetimes $(X, Y)$ is said to be described by a *bivariate frailty model* if its joint survival function is defined as

$$\overline{F}_{(X,Y)}(t,s) = P[X > t, Y > s] = E_\Theta\left[\overline{G}_X^\Theta(t)\overline{G}_Y^\Theta(s)\right], \quad \text{with } t, s \in \mathbb{R}^+,$$

where $\Theta$ is an environmental random frailty taking values in a set $\mathcal{A} \subseteq \mathbb{R}^+$ and the survival functions $\overline{G}_X$ and $\overline{G}_Y$ (that can be any pair of survival functions) are commonly called *baseline survival functions*.

- Note that $\overline{G}_X$ and $\overline{G}_Y$ are different from the marginal survival functions $\overline{F}_X$ and $\overline{F}_Y$ of $X$ and $Y$, respectively, unless $\Theta = 1$ a.s..

Frailty models are strictly related with Archimedean copulas, since joint survival functions defined as above can be rewritten as

$$\overline{F}_{(X,Y)}(t,s) = \psi(R_X(t) + R_Y(s)), \quad \text{with } t,s \in \mathbb{R}^+,$$

where

$$\psi(x) = E_\Theta[\exp(-x\Theta)],$$

while

$$R_X(t) = -\ln \overline{G}_X(t) \quad \text{and} \quad R_Y(s) = -\ln \overline{G}_Y(s).$$

From this expression of the survival function of $(X, Y)$ it is easy to verify that the vector has an Archimedean (survival) copula, whose generator is the completely monotone function $\psi^{-1}$ (i.e., $d$-monotone for every $d \in \mathbb{N}^+$). In fact, observing that

$$\overline{F}_1 = \psi(R_1(t)) \quad \text{and} \quad \overline{F}_1^{-1}(u) = R_1^{-1}(\psi^{-1}(u)),$$

one has

$$\begin{aligned}
\hat{C}_{\mathbf{X}}(u,v) &= \overline{F}(\overline{F}_1^{-1}(u), \overline{F}_2^{-1}(v)) \\
&= \psi[R_1(R_1^{-1}(\psi^{-1}(u))) + R_2(R_2^{-1}(\psi^{-1}(v)))] \\
&= \psi(\psi^{-1}(u) + \psi^{-1}(v)).
\end{aligned}$$

With frailty models it is easy to handle dependence, and evolution of dependence during time. In fact, let

$$\mathbf{X} = (X_1, X_2) \sim \overline{F}(x, y) = \psi(R_1(x) + R_2(y)),$$

and let

$$\mathbf{X_t} = [(X_1 - t, X_2 - t) | X_1 > t, X_2 > t].$$

It is easy to prove that $\mathbf{X_t}$ has joint survival function

$$\overline{F}_t(x, y) = \psi_t(R_{1,t}(x) + R_{2,t}(y))$$

where

$$R_{i,t}(x) = R_i(t + x) - R_i(t) \quad \text{and} \quad \psi_t(x) = \frac{\psi(x + R_1(t) + R_2(t))}{\psi(R_1(t) + R_2(t))}.$$

Thus, the vector of residual lifetimes $\mathbf{X}_t$ has again an Archimedean survival copula, whose generator is $\psi_t(x)$.

In the particular case that $\hat{C}_{\mathbf{X}}$ is a Clayton copula, i.e., if its generator is

$$\psi(x) = (\max\{1 + \alpha x, 0\})^{-\frac{1}{\alpha}},$$

then it can be proved that equality

$$\psi_t(x) = \psi(x)$$

holds true for any positive $\theta$ and $t$, and for any $R_1$ and $R_2$, so that

$$\hat{C}_{\mathbf{X}_t} = \hat{C}_{\mathbf{X}}.$$

Thus, the Clayton copula preserves the dependence structure along time!

The Clayton copula occurs when the random environmental parameter $\Theta$ has a Gamma distribution with shape parameter $\alpha$ and scale parameter $\lambda = 1$.

## Generalized Marshall-Olkin models

**Definition**

Let $Y_1$, $Y_2$ and $Y_3$ be three independent random lifetimes (i.e., non-negative random variables). Then the vector

$$\mathbf{X} = (X_1, X_2) = (Y_1 \wedge Y_3, Y_2 \wedge Y_3)$$

is said to be described by a *Generalized Marshall-Olkin* (GMO) model.

The vector $\mathbf{X} = (X_1, X_2)$ defined has above has a (survival) copula $\hat{C}_{\mathbf{X}}$ called GMO copula, which is a generalization of the Marshall-Olkin copula.

- Under appropriate assumptions, it can be shown that $\hat{C}_{\mathbf{X}} = \hat{C}_{\mathbf{X}_t}$ for all $t \geq 0$, i.e., that the vector preserves the dependence structure along time. This happens for example for the Marshall-Olkin bivariate exponential distribution.
- It can be proved that if $Y_3 \leq_{st} \widetilde{Y}_3$, then

$$(Y_1 \wedge \widetilde{Y}_3, Y_2 \wedge \widetilde{Y}_3) \prec_c (Y_1 \wedge Y_3, Y_2 \wedge Y_3)$$

## Linear (additive) models

**Definition**

Let $Y_1$, $Y_2$ and $Y_3$ be three independent random variables. Then the vector

$$\mathbf{X} = (X_1, X_2) = (Y_1 + Y_3, Y_2 + Y_3)$$

is said to be described by a *Linear* (or *Additive*) model.

Unfortunately, it is not possible (until now...) to provide a general expression of the copula $C_{\mathbf{X}}$ defined on the vector $\mathbf{X} = (X_1, X_2)$.

- It can be proved that if $Y_3 \leq_{disp} \widetilde{Y}_3$, then

$$\tau_{(Y_1+Y_3, Y_2+Y_3)} \leq \tau_{(Y_1+\widetilde{Y}_3, Y_2+\widetilde{Y}_3)}.$$

  Here $\leq_{disp}$ denotes the dispersive order, which is a comparison among random variables based on their variability.

**Dependence and ranks**

Suppose that a random sample $(X_1, Y_1), ..., (X_n, Y_n)$ is given from a pair of continuous rvs $(X, Y)$.

Consider the pairs of ranks $(R_1, S_1), ..., (R_n, S_n)$, where $R_i$ stands for the rank of $X_i$ among $\{X_1, ..., X_n\}$ and $S_i$ stands for the rank of $Y_i$ among $\{Y_1, ..., Y_n\}$.

Assume that ranks are unambiguously defined, because ties occur with probability zero under the assumption of continuity for $X$ and $Y$.

## Empirical copula

The *empirical copula* associated with the random sample $(X_1, Y_1), ..., (X_n, Y_n)$ is defined by

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1} \left( \frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right).$$

Notice that $C_n$ is a jump function, and therefore not a copula itself. If the copula of $(X, Y)$ admits continuous first-order partial derivatives on $(0,1)^2$ then

$$\mathbb{C}_n(u, v) = \sqrt{n}(C_n(u, v) - C(u, v)), \quad u, v \in [0, 1]$$

converges weakly as $n \to \infty$ to a centered Gaussian process. Thus:

Statistical procedures for goodness of fit tests can be applied, similarly as for the Kolmogorov-Smirnov test for univariate distribution functions.

**Inference for copula models**

- Model selection: descriptive statistics
- Model fitting: estimation
- Model validation: selection and goodness-of-fit tests

## Sample version of Kendall's $\tau$ and Spearman's $\rho$

We have seen it already. A different (but equivalent) expression is

$$\tau_n = \frac{4}{n(n-1)} P_n - 1,$$

where $P_n$ is the number of concordant pairs.

The sample version of the Spearman's $\rho$ is

$$\rho_n = \frac{12}{(n+1)(n-1)} \sum_{i=1}^{n} R_i S_i - 3\frac{n+1}{n-1}.$$

When the random sample comes from the copula $\Pi_2$, one has

$$\rho_n \sim N\left(0, \frac{1}{n-1}\right),$$

and this can be used to test the null hypothesis of independence.

## Estimation

Given random sample $(X_1, Y_1), ..., (X_n, Y_n)$ from $(X, Y)$, suppose that the joint d.f. can be described via

$$F(x, y) = C_\theta(F_\alpha(x), G_\beta(y))$$

where $F_\alpha$ and $G_\beta$ are some suitable marginals, and a parametric family $\{C_\theta, \theta \in T\}$ of copulas is being considered for the dependence between $X$ and $Y$.

*How should $\theta$ be estimated?*

## Illustration

Suppose that the dependence of $(X, Y)$ is appropriately modelled by the Farlie-Gumbel-Morgenstern family of copulas given by

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v),$$

where $(u, v) \in [0, 1]^2$, and $\theta \in [-1, 1]$ is a dependence parameter.

It is known that, for this family, the value of Kendall's $\tau$ is given by

$$\tau_c = \frac{2\theta}{9}.$$

Therefore, an appropriate estimation of $\theta$ is

$$\widehat{\theta} = \frac{9\tau_n}{2}.$$

where $\tau_n$ is the sample version of the Kendall's $\tau$.

**Estimation based on moments' method**

In general, suppose that the dependence of $(X, Y)$ is appropriately modelled by a family of copulas $\{C_\theta, \theta \in T\}$ such that

$$\theta = g(\tau)$$

for some function $g$.

Then, $\widehat{\theta}_n = g(\tau_n)$ may be referred to as the Kendall-based estimator of $\theta$, that can be constructed as an adaptation of the celebrated method of moments.

Notice that a similar procedure can be based on Spearman's $\rho$.

## ML Estimation

Otherwise, suppose that a parametric family $\{C_\theta, \theta \in T\}$ of absolutely continuous copulas is being considered as a model for the dependence between $X$ and $Y$.

Suppose, moreover, that $X$ and $Y$ are modelled by univariate d.f.s $F_\alpha$ and $G_\beta$ with density $f_\alpha$ and $g_\beta$.

In order to estimate the parameters, three methods will be considered:

- *maximum likelihood* (ML),
- *Inference from margins* (IFM),
- *maximum pseudolikelihood* (MPL).

## Maximum likelihood

The ML method consists in maximizing

$$\mathfrak{L}(\theta, \alpha, \beta) = \sum_{i=1}^{n} \ln(f_{\theta,\alpha,\beta}(X_i, Y_i))$$

where $f$ is the density of the model. This can be rewritten as

$$\mathfrak{L}(\theta, \alpha, \beta) = \sum_{i=1}^{n} \ln(c_\theta(F_\alpha(X_i), G_\beta(Y_i)) f_\alpha(X_i) g_\beta(Y_i))$$

This approach:

- is easy to implement;
- yields an asymptotically Gaussian, unbiased estimate.

However, it may produce wrong estimation of dependence structure if the margins are incorrectly specified.

**Inference from margins (IFM)**

The IFM method is divided into two parts:

- first, each marginal is estimated separately;
- then, the IFM estimator of $\theta$ involves maximizing

$$\mathfrak{L}(\theta) = \sum_{i=1}^{n} \ln(C_\theta(\widehat{F}(X_i), \widehat{G}(Y_i)))$$

with $\widehat{F} = F_{\alpha_n}$ and $\widehat{G} = G_{\beta_n}$ , where $\alpha_n$ and $\beta_n$ are standard maximum likelihood estimates of the parameters of the marginal df's.

## The MPL estimator

The method of maximum pseudo-likelihood simply involves maximizing a rank-based log-likelihood of the form

$$\mathfrak{L}(\theta) = \sum_{i=1}^{n} \ln \left( c_\theta \left( \frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right).$$

Here, notice that the marginal d.f.'s are estimated parametrically by their sample empirical distributions. Therefore, it can be applied even if the marginal d.f.'s are not known.

Such estimators are robust to misspecification in the marginal model.

**Goodness-of-fit tests**

In typical modelling, the user has a choice between several different dependence structures for the data at hand.

Suppose that different copulas $\{C_1, ...C_k\}$, belonging to different families, were fitted by some arbitrary method. It is then natural to ask, according to some specific criteria chosen by the practitioner,

*what is the copula that best fits the observations?*

## Graphical diagnostic

When dealing with bivariate data, the most intuitive way of checking the adequacy of the copula $C_i$ with respect to our model would be to compare:

- the scatter plot of the pairs

$$\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)$$

  with
- an artificial data set of the same size generated from the copula $C_i$.

**Goodness-of-fit tests for copulas**

Several goodness-of-fit tests have been recently introduced for copulas. The main interesting one is based on the Cramér-von Mises distance

$$S_j = \sum_{i=1}^{n} \left( C_n\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right) - C_j\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)\right)^2$$

Other popular ones are based on different distances, like the maximum absolute difference, or the integral of the absolute difference.

# Bibliography

## MONOGRAPHS


- H. Joe. *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman  Hall, London, 1997.
- R. B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition, 2006.
- J.-F. Mai and M. Scherer. *Simulating copulas*. World Scientific, Singapore, 2012.
- U. Cherubini, E. Luciano, and W. Vecchiato. *Copula methods in finance*. Wiley Finance Series. John Wiley  Sons Ltd., Chichester, 2004.
- P. Jaworski, F. Durante, W. Härdle, and T. Rychlik, editors. Copula Theory and its Applications, volume 198 of Lecture Notes in Statistics - Proceedings. Springer, Berlin Heidelberg, 2010. [with *Copula Theory: an Introduction*, by F. Durante and C. Sempi]
- P. K. Trivedi and D. M. Zimmer. Copula Modeling: *An Introduction for Practitioners*. Econometrics. Vol. 1, No 1 (2005) 1-111.

## REMARKABLE PAPERS


- P. Embrechts, F. Lindskog, and A. J. McNeil. Modelling dependence with copulas and applications to risk management. In S.T. Rachev, editor, *Handbook of heavy tailed distributions in finance*, pages 329-384, Amsterdam, 2003. Elsevier
- C. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, 12(4): 347-368, 2007.
- T. Mikosch. Copulas: tales and facts. *Extremes*, 9(1): 3-20, 2006.
- A.J. Patton. A review of copula models for economic time series. *J. Multivariate Anal.*, 110: 4-18, 2012.
- C. Genest and J. Nelehová. A primer on copulas for count data. *Astin Bull.*, 37(2): 475-515, 2007.
- C. Genest, B. Rémillard and D. Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Math. and Econ.*, 44(2): 199-213, 2009.
- Malevergne, Y. and D. Sornette. Testing the gaussian copula hypothesis for financial assets dependence. *Quantitative Finance* 3, 231-250, 2003
- T. Ledwina and G. Wylupek. Validation of positive quadrant dependence. *Insurance: Math. and Econ.*, 56: 38-47, 2014.