Dependence of non-continuous random variables

Johanna Nešlehová

Department of Mathematics
ETH Zurich
Switzerland
www.math.ethz.ch/~johanna

MOTIVATION

Risk processes for windstorm and floods claims

$$R_1(t) = \sum_{i=1}^{N_1(t)} X_i, \quad t \geq 0 \qquad\qquad R_2(t) = \sum_{i=1}^{N_2(t)} Y_i, \quad t \geq 0$$

MOTIVATION

Risk processes for windstorm and floods claims

$$R_1(t) = \sum_{i=1}^{N_1(t)} X_i, \quad t \geq 0 \qquad\qquad R_2(t) = \sum_{i=1}^{N_2(t)} Y_i, \quad t \geq 0$$

- R_1 and R_2 are dependent.
  - Dependence between the claim sizes X_i and Y_i,
    1. Modeling of X_i and Y_i separately $\rightsquigarrow F, G$
    2. Choosing a suitable copula $\mathcal{C} \rightsquigarrow \mathcal{C}(F, G)$
  - Dependence between the number of claims $N_1(t)$ and $N_2(t)$.
    1. Modeling of $N_1(t)$ and $N_2(t)$ separately $\rightsquigarrow F_{\boldsymbol{N}}, G_{\boldsymbol{N}}$
    2. Choosing a suitable copula $\mathcal{C}_{\boldsymbol{N}} \rightsquigarrow$ **?**

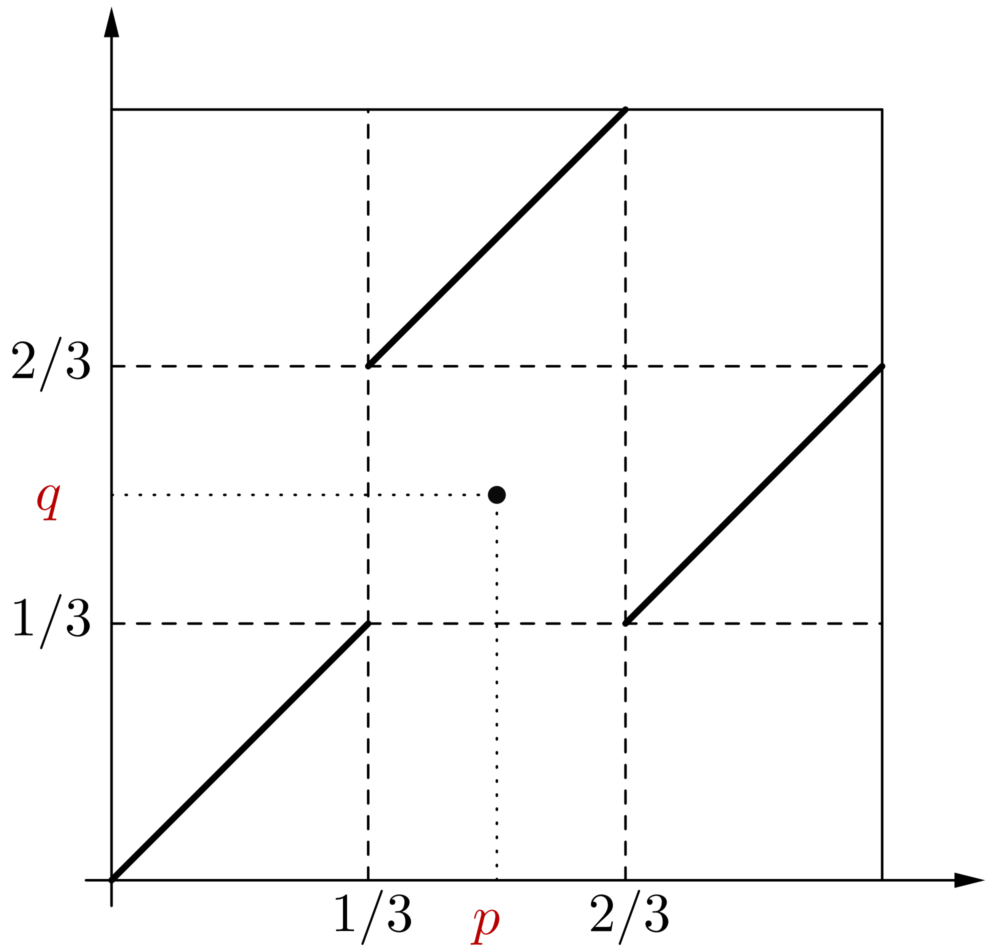
OUTLINE

- The "continuous" vs. "non-continuous" case: typical fallacies
- Dependence of non-continuous random variables
- Dependent risk processes
- References

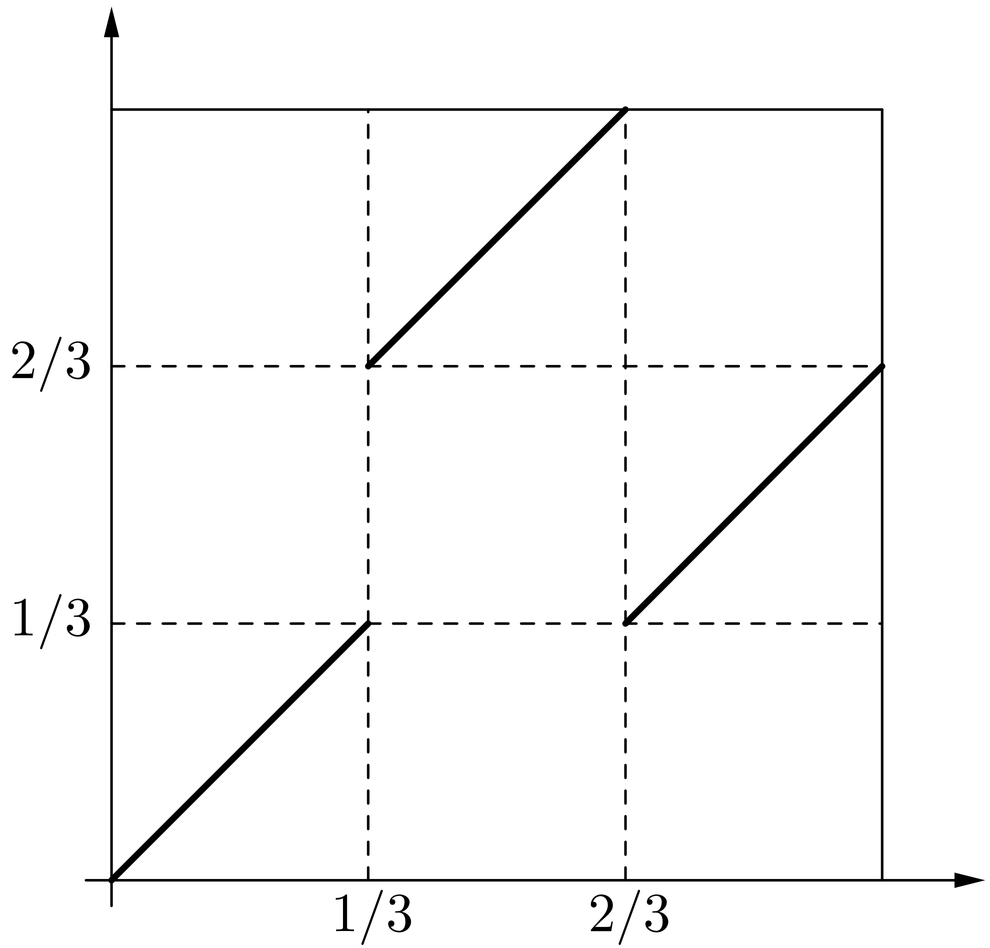
OUTLINE

- The "continuous" vs. "non-continuous" case: typical fallacies
- Dependence of non-continuous random variables
- Dependent risk processes
- References

PATHOLOGICAL EXAMPLE



PATHOLOGICAL EXAMPLE



- $(p, q) \in [0, 1/3] \times [0, 1/3]$:
  perfect positive dependence
- $(p, q) = (1/\sqrt{3}, 1/\sqrt{3})$:
  independence
- $(p, q) \in [2/3, 1] \times [2/3, 1]$:
  perfect negative dependence

([Marshall, 1996])

NON-CONTINUOUS CASE: PITFALLS

- There exist several copulas corresponding to one and the same cdf. They coincide only on the (closures) of the ranges of the marginal df.

- The dependence structures of the cdf. are generally not the same as the dependence structures of the extensions. For example:
  $H_1(x,y) \leq H_2(x,y)\ \forall x,y \in \mathbb{R} \not\Rightarrow \mathcal{C}_1(u,v) \leq \mathcal{C}_2(u,v)\ \forall u,v \in [0,1]$.

- Every *possible* copula remains invariant under strictly increasing and continuous transformations, but does not necessarily change in the same way as the copula in the "continuous" case if at least one of the transformations is decreasing.

- There exists a weak convergent sequence $\{(X_n, Y_n)\}_n$ of random vectors with even continuous marginals such that the corresponding sequence of copulas does not converge at any point in $(0,1)^2$.

- Any measure of association which depends only on the copula of the joint distribution function is a constant.

MEASURES OF CONCORDANCE: SOME EXAMPLES

- **Kendall's tau**

$$\tau(X,Y) = 4 \int_{\boldsymbol{I}^2} \mathcal{C}(u,v) d\mathcal{C}(u,v) - 1$$

- **Spearman's rho**

$$\rho(X,Y) = 3\Big[4 \int_{\boldsymbol{I}^2} \mathcal{C}(u,v) du\, dv - 1\Big] = 12 \int_{\boldsymbol{I}^2} \Big[\mathcal{C}(u,v) - uv\Big] du\, dv$$

- **Linear correlation coefficient**

$$\varrho(X,Y) = \frac{\int_{\mathbb{R}^2} \Big[\mathcal{C}(F(x),G(y)) - F(x)G(y)\Big] dx\, dy}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

KENDALL'S TAU AND SPEARMAN'S RHO IN THE "NON-CONTINUOUS" CASE?

"Naive" Idea: *work with some of the non unique fitting copula*

But: fitting copulas can differ considerably.

- $\mathsf{P}[(X-X^*)(Y-Y^*)>0]-\mathsf{P}[(X-X^*)(Y-Y^*)<0] \neq \int \mathcal{C}(u,v)\,d\mathcal{C}^*(u,v)-1$ for some fitting copula $\mathcal{C}$ and $\mathcal{C}^*$ of (X,Y) and (X^*,Y^*), respectively.

- for independent Bernoulli random variables X and Y,
  - $\tau(X,Y) \in [-3/4, 3/4]$,
  - $\rho(X,Y) \in [-13/16, 13/16]$.

- for comonotonic Bernoulli random variables X and Y,
  - $\tau(X,Y) \in [0,1]$,
  - $\rho(X,Y) \in [1/2, 1]$.

OUTLINE

- The “continuous” vs. “non-continuous” case: typical fallacies
- Dependence of non-continuous random variables
- Dependent risk processes
- References

DEPENDENCE STRUCTURE IN THE NON-CONTINUOUS CASE

In the continuous case: $\mathcal{C}$ is the cdf. of the transformed vector $(F_X(X), F_Y(Y))$.

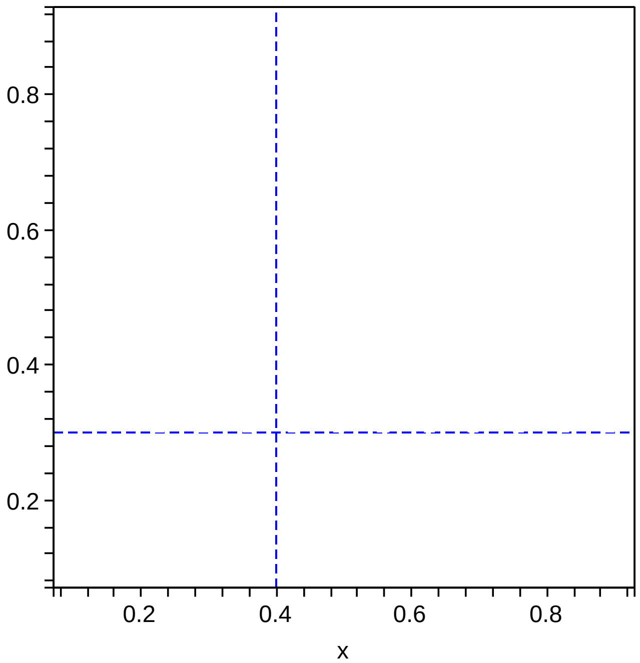
Idea: In the non-continuous case, use a different transformation of the marginals:

$$\psi(x,u) := \mathsf{P}[X < x] + u\mathsf{P}[X = x] = F(x-) + u\Delta F(x)$$

Result: for a random vector (U, V) with uniform marginals which is independent of (X, Y), $(\psi(X,U), \psi(Y,V))$ has uniform marginals and the corresponding unique copula is a possible copula of (X, Y).

Moreover: different dependence structure of (U, V) leads to different possible copulas of (X, Y).

EXAMPLES

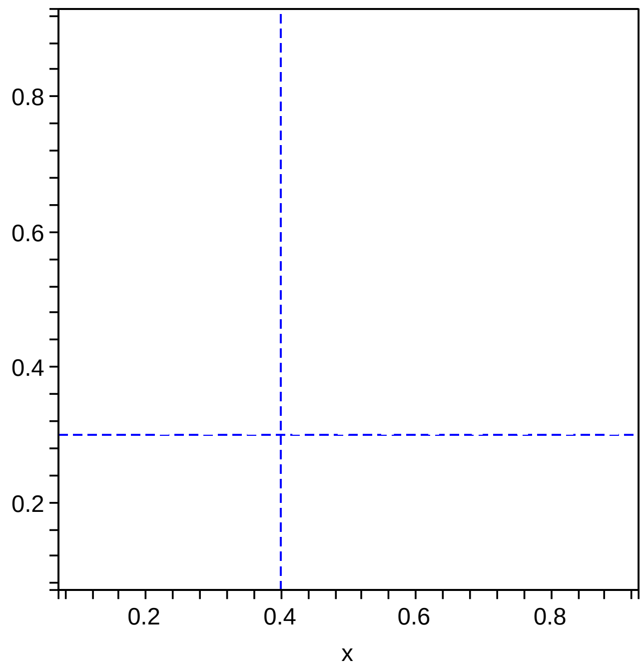


Situation for Bernoulli marginals with

$\mathsf{P}[X = 0] = 0.4$, $\mathsf{P}[Y = 0] = 0.3$ and

$$\mathsf{P}[X = 0, Y = 0] = 0.2.$$

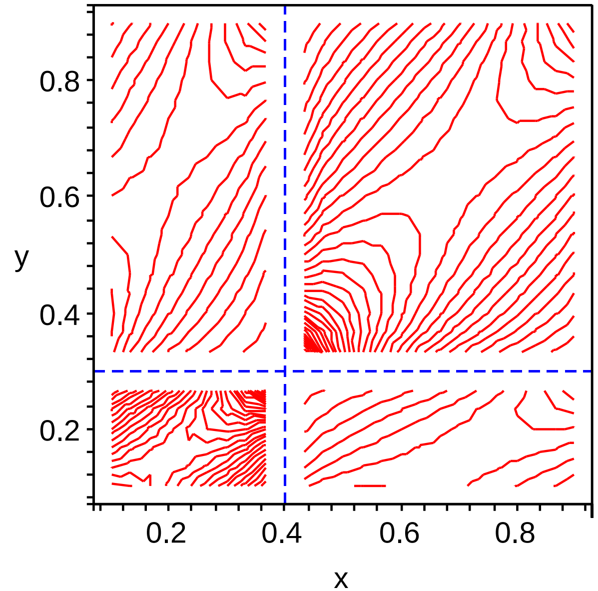
EXAMPLES



Situation for Bernoulli marginals with

P$[X = 0] = 0.4$, P$[Y = 0] = 0.3$ and

P$[X = 0, Y = 0] = 0.2$.



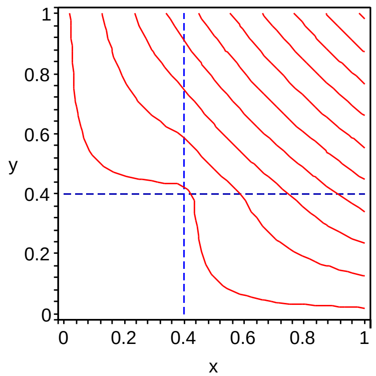
Copula of $(\psi(X, U), \psi(Y, V))$ where

(U, V) has a Gauss copula with parameter

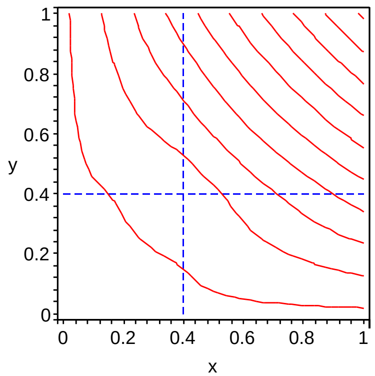
$\varrho = 0.6$.

THE STANDARD EXT. FOR BERNOULLI MARGINALS

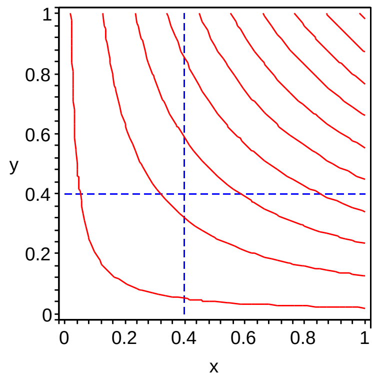
$\mathcal{C}(0.4, 0.4) = 0$



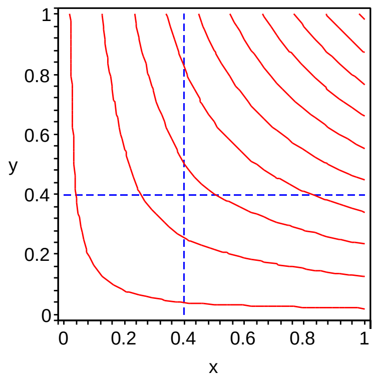
$\mathcal{C}(0.4, 0.4) = 0.5$



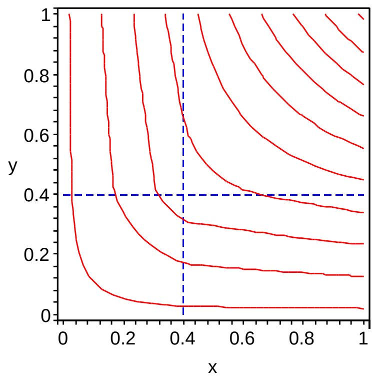
$\mathcal{C}(0.4, 0.4) = 0.16$



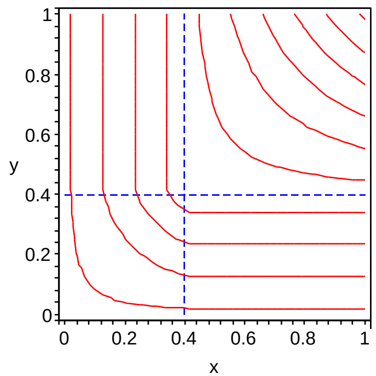
$\mathcal{C}(0.4, 0.4) = 0.2$



$\mathcal{C}(0.4, 0.4) = 0.3$



$\mathcal{C}(0.4, 0.4) = 0.4$



THE “STANDARD” TRANSFORMATION

U **and** V **independent**: the copula of $(\psi(X,U),\psi(Y,V))$ is the standard extension copula of Schweizer and Sklar:

$$\mathcal{C}_S(u,v) = (1-\lambda)(1-\mu)\mathcal{C}(a_1,b_1)+(1-\lambda)\mu\mathcal{C}(a_1,b_2)+\lambda(1-\mu)\mathcal{C}(a_2,b_1)+\lambda\mu\mathcal{C}(a_2,b_2)$$

with a_i being the least and largest element in $\overline{\operatorname{ran} F}$ satisfying $a_1 \le u \le a_2$, and b_i the least and the largest element in $\overline{\operatorname{ran} G}$ satisfying $b_1 \le v \le b_2$,

$$\lambda = \begin{cases} \frac{u-a_1}{a_2-a_1}, & \text{if } a_1 < a_2, \\ 1, & \text{if } a_1 = a_2; \end{cases} \quad \text{and} \quad \mu = \begin{cases} \frac{v-b_1}{b_2-b_1}, & \text{if } b_1 < b_2, \\ 1, & \text{if } b_1 = b_2; \end{cases}$$

- $\mathcal{C}_S$ corresponds to the linear interpolation of the unique subcopula as well as the unique copula of the linear interpolation of the joint distribution function.

THE “STANDARD” TRANSFORMATION CONT’D

U **and** V **independent**: the dependence structures between $\psi(X,U)$ and $\psi(Y,V)$ (i.e. $\mathcal{C}_S$) reflect certain aspects of the dependence structures between X and Y:

independence: X and Y are independent if and only if $\mathcal{C}_S$ is the independence copula.

stochastic ordering: (X^*, Y^*) stochastically smaller than (X, Y) if and only if $\mathcal{C}_S(u,v) \leq \mathcal{C}_S^*(u,v)$ for all $u, v \in [0,1]$.

monotone transformations of the marginals: $\mathcal{C}_S$ reacts on monotone transformations of the marginals in exactly the same way as does the unique copula in the “continuous” case.

difference between probabilities of concordance and discordance:

$$\mathsf{P}[(X-X^*)(Y-Y^*) > 0] - \mathsf{P}[(X-X^*)(Y-Y^*) < 0] = 4\int \mathcal{C}_S(u,v)\, d\mathcal{C}_S^*(u,v) - 1$$

THE “STANDARD” TRANSFORMATION CONT’D

However: $\mathcal{C}_S$ does not reflect everything:

co- and countermonotonicity: if X and Y are perfect monotonic dependent, $\mathcal{C}_S$ does not coincide with the Fréchet-Hoeffding bounds.

weak convergence: if $\{(X_n, Y_n)\}$ converges weakly to $\{(X, Y)\}$, then the sequence of the standard extension copulas does not need to converge pointwise. However, if $\{(X_n, Y_n)\}$ is a sequence of discrete random vectors such that the supports do not change with n, then the weak convergence of $\{(X_n, Y_n)\}$ to $\{(X, Y)\}$ is equivalent to

1. the weak convergence of the marginal distribution functions, and
2. the pointwise convergence of the standard extension copulas.

KENDALL'S TAU AND SPEARMAN'S RHO

1. (X, Y) and (X^*, Y^*) iid, governed by $\mathcal{C}_S$ (**Kendall's tau**):

$$\widetilde{\tau}(\mathcal{C}_S) = 4 \int_0^1 \int_0^1 \mathcal{C}_S(u, v)\, d\mathcal{C}_S(u, v) - 1$$

2. (X, Y) governed by $\mathcal{C}_S$ and X^*, Y^* independent (**Spearman's rho**):

$$\widetilde{\rho}(\mathcal{C}_S) = 12 \int_0^1 \int_0^1 [\mathcal{C}_S(u, v) - uv]\, d\mathcal{C}_S(u, v)$$

KENDALL'S TAU AND SPEARMAN'S RHO

1. (X, Y) and (X^*, Y^*) iid, governed by $\mathcal{C}_S$ (**Kendall's tau**):

$$\widetilde{\tau}(\mathcal{C}_S) = 4 \int_0^1 \int_0^1 \mathcal{C}_S(u, v)\, d\mathcal{C}_S(u, v) - 1$$

2. (X, Y) governed by $\mathcal{C}_S$ and X^*, Y^* independent (**Spearman's rho**):

$$\widetilde{\rho}(\mathcal{C}_S) = 12 \int_0^1 \int_0^1 [\mathcal{C}_S(u, v) - uv]\, d\mathcal{C}_S(u, v)$$

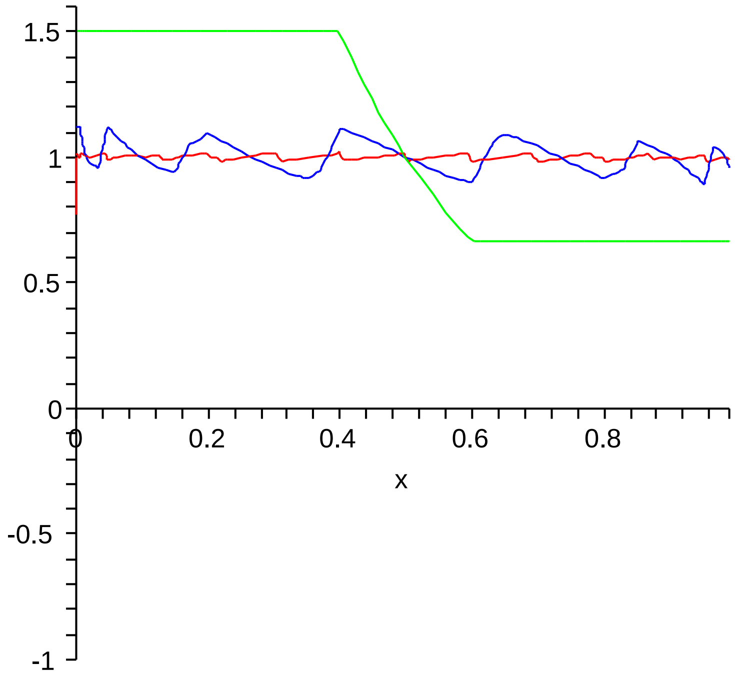
Fallacy 1 : $\widetilde{\tau}$ and $\widetilde{\rho}$ do not reach the bounds 1 and -1. Instead,

$$\widetilde{\tau}(\mathcal{W}_S) \le \widetilde{\tau}(\mathcal{C}_S) \le \widetilde{\tau}(\mathcal{M}_S) \quad \text{and} \quad \widetilde{\rho}(\mathcal{W}_S) \le \widetilde{\rho}(\mathcal{C}_S) \le \widetilde{\rho}(\mathcal{M}_S),$$

where $\mathcal{W}_S$ and $\mathcal{M}_S$ are standard extension copulas corresponding to the countermonotonic and comonotonic case, respectively.

Fallacy 2 :$-\widetilde{\tau}(\mathcal{W}_S) \neq \widetilde{\tau}(\mathcal{M}_S)$ and $-\widetilde{\rho}(\mathcal{W}_S) \neq \widetilde{\rho}(\mathcal{M}_S)$.

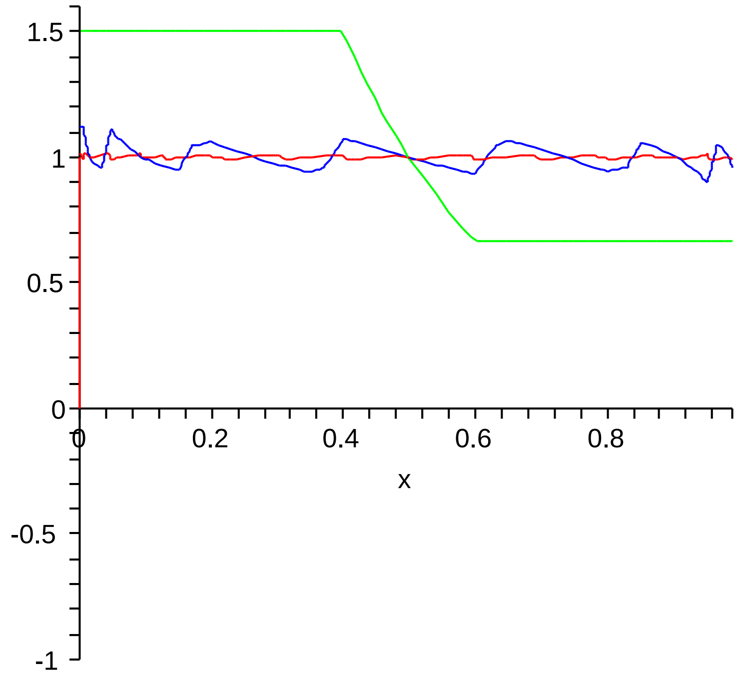
BOUNDS FOR BINOMIAL MARGINALS



$|\widetilde{\tau}(\mathcal{M}_S)/\widetilde{\tau}(\mathcal{W}_S)|$ for Binomial distributions

$F = \mathcal{B}(n, 0.4)$ and $G = \mathcal{B}(n, x)$ with

$n =$ 1, 4 and 10.



$|\widetilde{\rho}(\mathcal{M}_S)/\widetilde{\rho}(\mathcal{W}_S)|$ for Binomial distributions

$F = \mathcal{B}(n, 0.4)$ and $G = \mathcal{B}(n, x)$ with

$n =$ 1, 4 and 10.

LESS SHARP BOUNDS

Theorem. *Let X and Y have arbitrary distributions. Then*

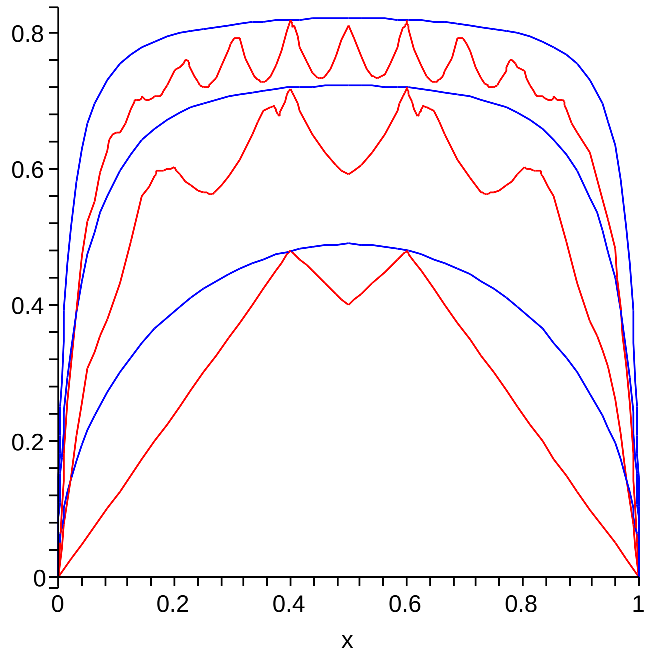
1. $|\widetilde{\tau}(X,Y)| \leq \sqrt{1 - \mathrm{E}\,\Delta F(X)}\sqrt{1 - \mathrm{E}\,\Delta G(Y)}$. *The bounds are attained if* $Y = T(X)$ *a.s. where* T *is a strictly monotone and continuous transformation on* $\operatorname{ran} X$.

2. $|\widetilde{\rho}(X,Y)| \leq \sqrt{1 - \mathrm{E}\,\Delta F(X)^2}\sqrt{1 - \mathrm{E}\,\Delta G(Y)^2}$. *The bounds are attained if* $Y = T(X)$ *a.s. where* T *is a strictly monotone and continuous transformation on* $\operatorname{ran} X$.

For discrete marginals with finite support:

$$|\widetilde{\tau}(X,Y)| \leq \sqrt{1 - \sum_{i=1}^{m} p_i^2}\sqrt{1 - \sum_{j=1}^{n} q_j^2}\,,\ |\widetilde{\rho}(X,Y)| \leq \sqrt{1 - \sum_{i=1}^{m} p_i^3}\sqrt{1 - \sum_{j=1}^{n} q_j^3}$$

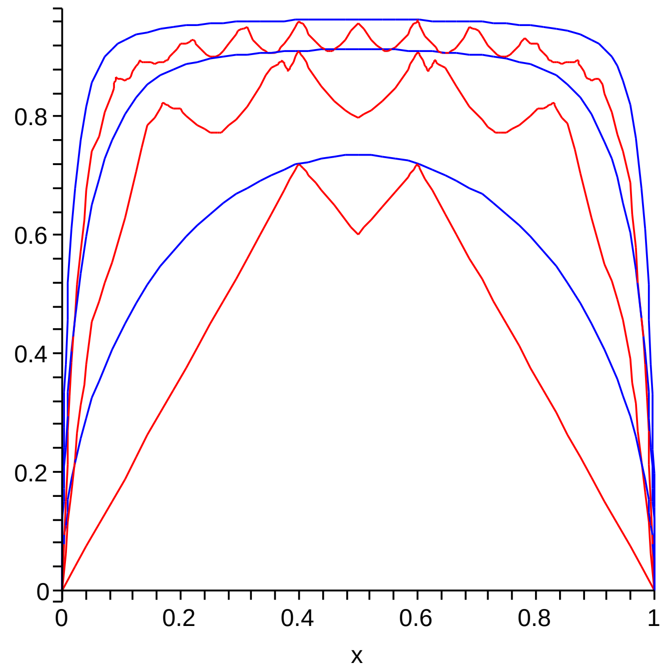
BOUNDS FOR BINOMIAL DISTRIBUTIONS



Exact and less sharp bounds for Kendall's tau

for Binomial distributions $F = \mathcal{B}(n, 0.4)$ and

$G = \mathcal{B}(n, x)$ with $n = 1,\ 4,\ 10.$



Exact and less sharp bounds for Spearman's rho

for Binomial distributions $F = \mathcal{B}(n, 0.4)$ and

$G = \mathcal{B}(n, x)$ with $n = 1,\ 4,\ 10.$

KENDALL'S TAU AND SPEARMAN'S RHO

$$\tau(X,Y) = \frac{4\int_0^1\int_0^1 \mathcal{C}_S(u,v)\, d\mathcal{C}_S(u,v) - 1}{\sqrt{[1-\mathrm{E}\,\Delta F(X)][1-\mathrm{E}\,\Delta G(Y)]}}$$

$$\rho(X,Y) = \frac{12\left(\int_0^1\int_0^1 \big(\mathcal{C}_S(u,v) - uv\big)\, d\mathcal{C}_S(u,v)\right)}{\sqrt{[1-\mathrm{E}\,\Delta F(X)^2][1-\mathrm{E}\,\Delta G(Y)^2]}}$$

satisfy (δ is either τ or ρ):

A1. symmetry: $\delta(X,Y) = \delta(Y,X)$,

A2. normalization: $-1 \le \delta(X,Y) \le 1$,

A3. independence: $\delta(X,Y) = 0$ if X and Y are independent,

KENDALL'S TAU AND SPEARMAN'S RHO contd

A4. range: $\delta(X,Y) = 1$ if $Y = T(X)$ a.s. for T strictly increasing and continuous and $\delta(X,Y) = -1$ if $Y = T(X)$ a.s. for T strictly decreasing and continuous,

A5. change of sign: If T is strictly monotone and continuous on $\operatorname{ran} X$, then

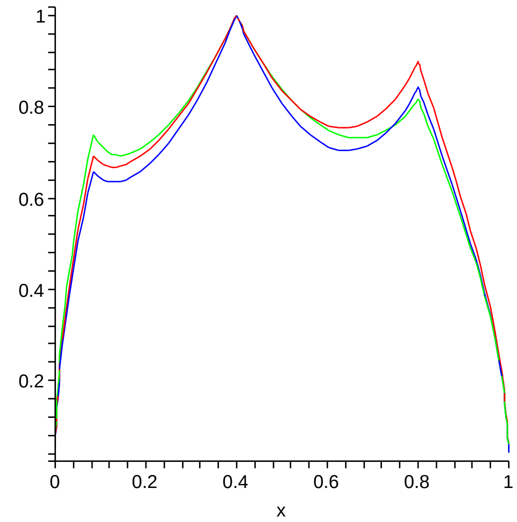
$$\delta(T(X),Y) = \begin{cases} \delta(X,Y), & \text{if } T \text{ increasing}, \\ -\delta(X,Y), & \text{if } T \text{ decreasing}, \end{cases}$$

A6. convergence: only under restrictions,

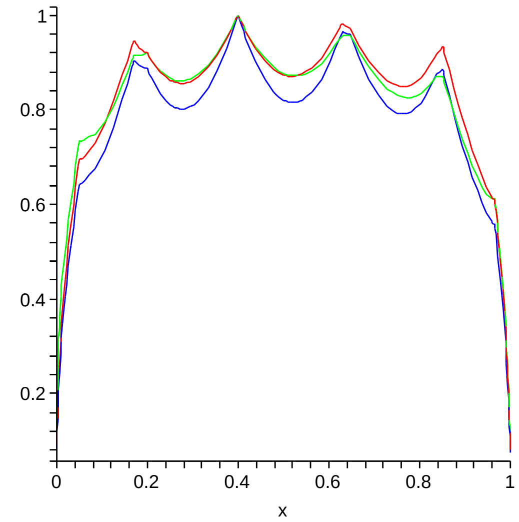
A7. coherence: If (X,Y) is stoch. smaller than (X^*,Y^*) then

$$\delta(X,Y) \geq \delta(X^*,Y^*).$$

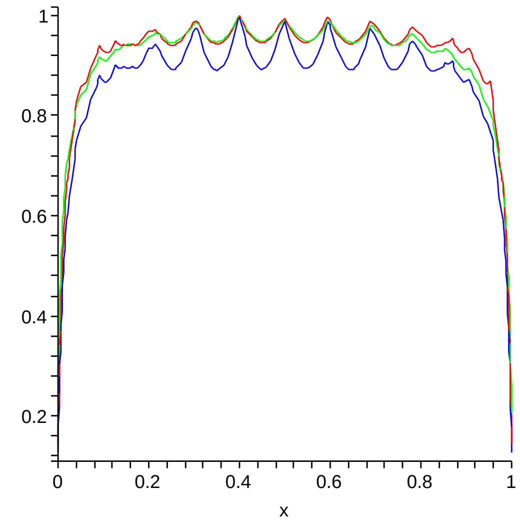
τ, ρ AND ϱ FOR BINOMIAL DISTRIBUTIONS



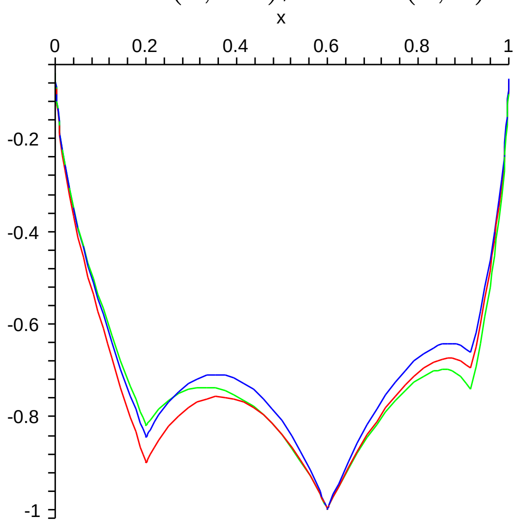
$F = \mathcal{B}(2, 0.4)$, $G = \mathcal{B}(2, x)$.



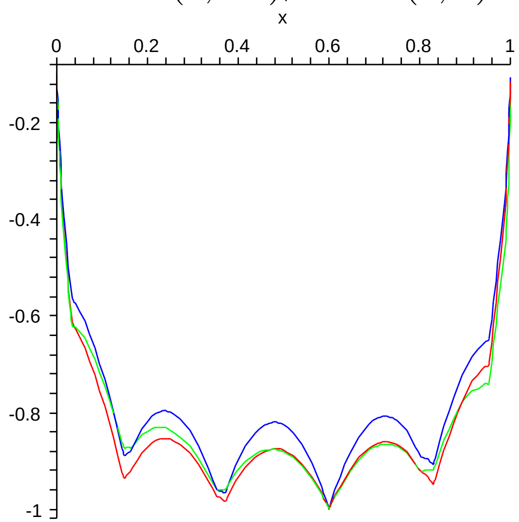
$F = \mathcal{B}(4, 0.4)$, $G = \mathcal{B}(4, x)$.



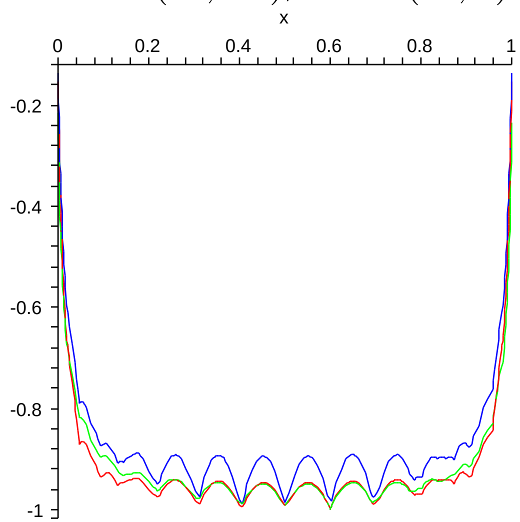
$F = \mathcal{B}(10, 0.4)$, $G = \mathcal{B}(10, x)$.



$F = \mathcal{B}(2, 0.4)$, $G = \mathcal{B}(2, x)$.



$F = \mathcal{B}(4, 0.4)$, $G = \mathcal{B}(4, x)$.



$F = \mathcal{B}(10, 0.4)$, $G = \mathcal{B}(10, x)$.

EMPIRICAL DISTRIBUTIONS

Theorem. *Let* $\{x_k, y_k\}_{k=1}^n$ *denote a sample of size* n *from a bivariate distribution* H. *Then*

1. τ *corresponding to the empirical distribution function* $\widehat{H}_n$ *equals the sample version of Kendall's tau,*

$$\widehat{\tau} = \frac{\#[\text{concordant pairs}] - \#[\text{discordant pairs}]}{\sqrt{\binom{n}{2} - u}\sqrt{\binom{n}{2} - v}},$$

where $u = \sum_{k=1}^{r} \binom{u_k}{2}$ *and* $v = \sum_{l=1}^{s} \binom{v_l}{2}$.

2. ρ *corresponding to the empirical distribution function* $\widehat{H}_n$ *equals the sample version of Spearman's rho,*

$$\widehat{\rho} = \frac{\sum_{k=1}^{n}(\overline{R}(x_k) - \overline{R}_x)(\overline{R}(y_k) - \overline{R}_y)}{\sqrt{\sum_{k=1}^{n}(\overline{R}(x_k) - \overline{R}_x)^2 \sum_{k=1}^{n}(\overline{R}(y_k) - \overline{R}_y)^2}},$$

where $\overline{R}_x$ *and* $\overline{R}_y$ *are given by* $\overline{R}_x = \frac{1}{n}\sum_{i=1}^{n}\overline{R}(x_i)$ *and* $\overline{R}_y = \frac{1}{n}\sum_{j=1}^{n}\overline{R}(y_j)$.

OUTLINE

- The “continuous” vs. “non-continuous” case: typical fallacies
- Dependence of non-continuous random variables
- Dependent risk processes
- References

APPLICATIONS TO POINT PROCESSES

- Let N be a non-negative integer valued random variable ("number of occurrences") independent of $\{\boldsymbol{X_i}\}$ iid, d-dimensional random vectors ("occurrence times", "losses", "claim sizes") with values in (a subset of) $\mathbb{R}^d$. A (finite) point process is given by

$$\xi := \sum_{i=1}^{N} \varepsilon_{\boldsymbol{X_i}}.$$

- The intensity measure of ξ is a measure given by $\mathrm{E}[\xi(A)]$ for a Borel set A.

- If N is Poisson, then ξ is a (finite) Poisson point process. It is homogeneous, if $\mathrm{E}\,\xi$ is a multiple of the Lebesgue measure.

- If ξ_1, ξ_2 are independent Poisson point processes, then the superposition $\xi_1 + \xi_2$ is a Poisson point process with intensity $\mathrm{E}\,\xi_1 + \mathrm{E}\,\xi_2$.

CONSTRUCTION METHOD I. N FIXED

Let $\xi = \sum_{i=1}^{N} \varepsilon_{\boldsymbol{X}_i}$ be a Poisson process with iid d-dimensional event points $\boldsymbol{X}_i = (X_i(1), \ldots, X_i(d))$ whose joint distribution for each i is given through a copula $\mathcal{C}_{\boldsymbol{X}}$. The marginal processes $\xi(k) = \sum_{i=1}^{N} \varepsilon_{X_i(k)}$, $k = 1, \ldots, d$ are Poisson, but dependent.

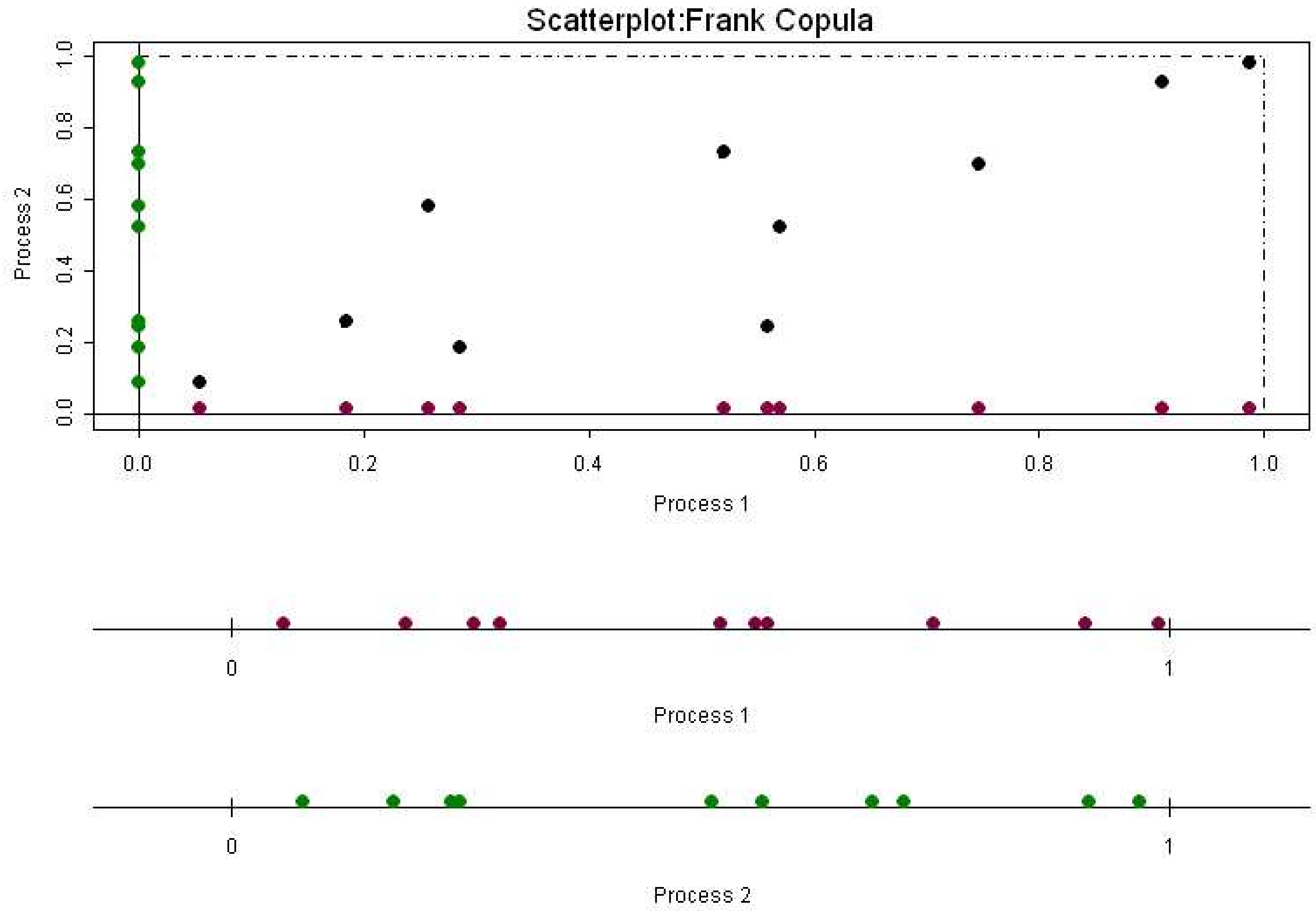
For all Borel sets A and B, the correlation of each two $\xi(k)$, $\xi(l)$ is given by

$$\varrho(\xi(k)(A), \xi(l)(B)) = \frac{Q_{kl}(A \times B)}{\sqrt{Q_k(A) Q_l(B)}}, \quad k, l = 1, \ldots, d.$$

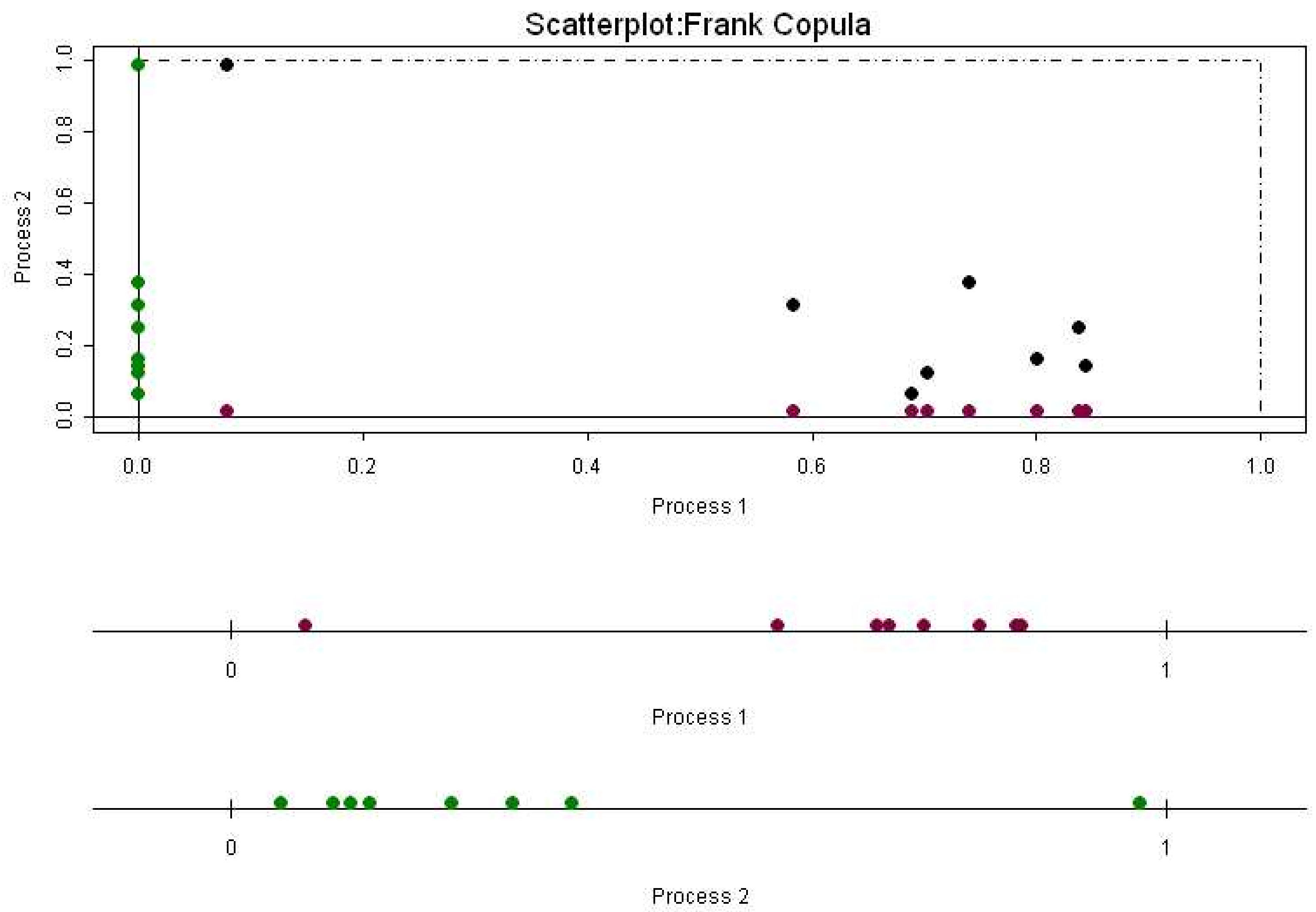
where Q_{kl} is the joint distribution of $(X_i(k), X_i(l))$ and Q_k, Q_l the marginals of $X_i(k)$ and $X_i(l)$, respectively.

- The correlation is always non-negative and does not depend on N,
- ϱ is not zero even if $X_i(k)$ and $X_i(l)$ are independent.

CONSTRUCTION METHOD I.: ILLUSTRATION



CONSTRUCTION METHOD I.: ILLUSTRATION



CONSTRUCTION METHOD II. VARIABLE N's

Here we construct dependent Poisson random variables $N_1, \ldots N_d$ using an appropriate copula $\mathcal{C}_{\boldsymbol{N}}$. Secondly, the occurrence time points $X_i(k)$ are again generated as (possibly dependent) margins of a d-dimensional time-event point $\boldsymbol{X}_i = (X_i(1), \ldots, X_i(d))$. The distribution function of $\boldsymbol{X}_i$ might again be described by a suitable copula $\mathcal{C}_{\boldsymbol{X}}$. In this way, we obtain d dependent processes $\xi(k) = \sum_{i=1}^{N_k} \varepsilon_{X_i(k)}$, $k = 1, \ldots, d$.

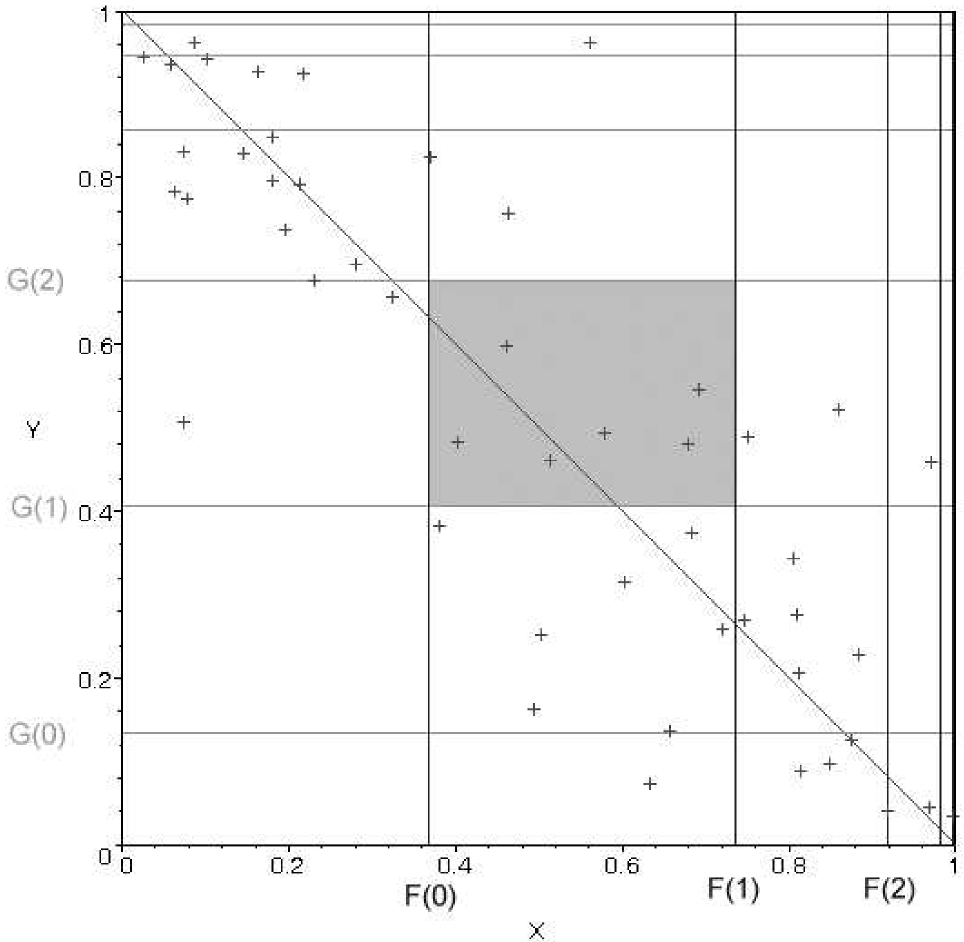
In case $X_i(k)$ are mutually independent, the correlation is given by

$$\varrho(\xi(k)(A), \xi(l)(B)) = \varrho(N_k, N_l)\sqrt{Q_k(A)Q_l(B)}, \quad k, l = 1, \ldots, d.$$

- The correlation is crucially influenced by the correlation of $N_1, \ldots N_d$,
- ϱ negative is possible.

CONSTRUCTION METHOD II.: ILLUSTRATION

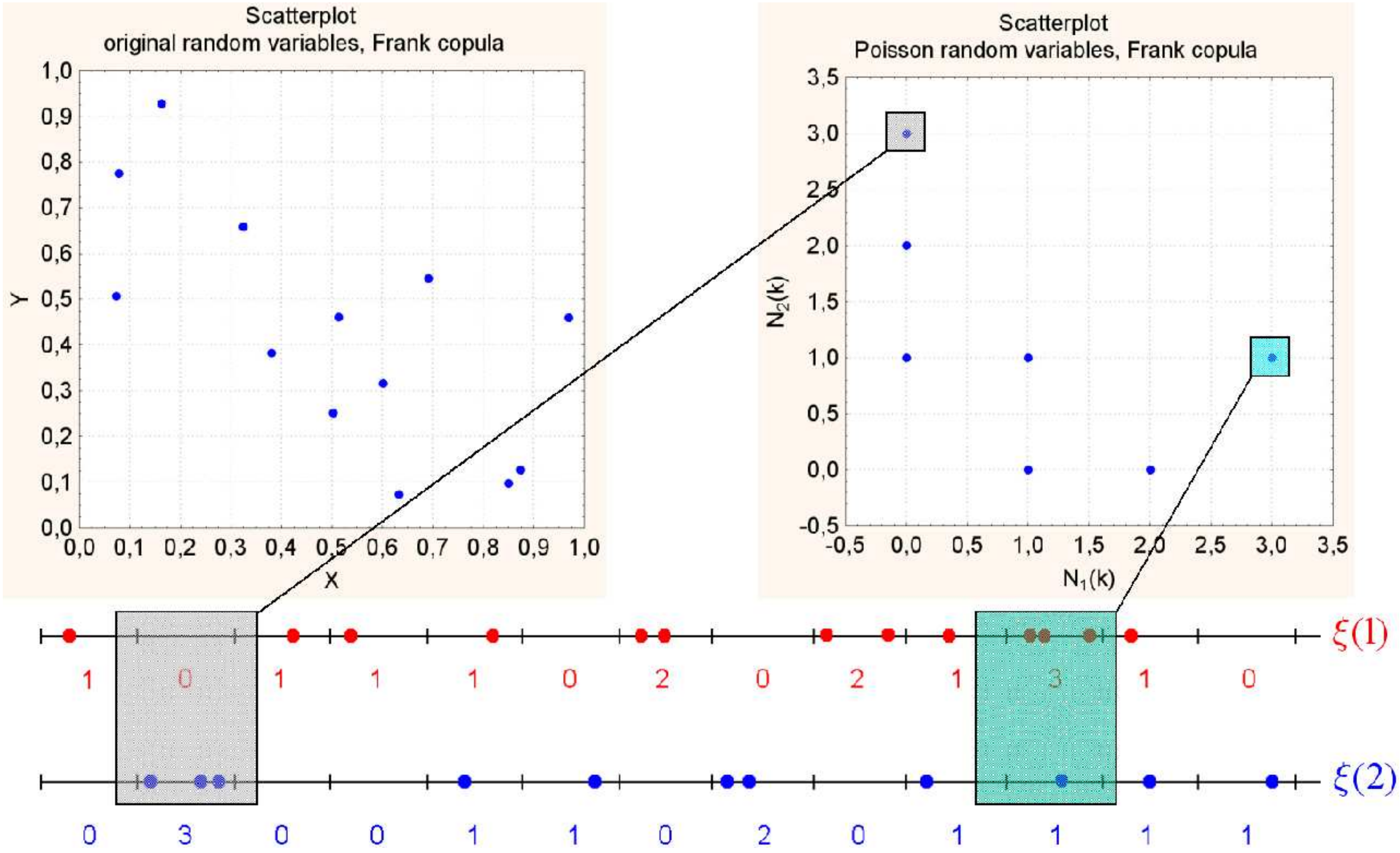
Simulation of negatively correlated Poisson random variables from a Frank Copula. All points in the **gray** rectangle correspond to the simulated Pair $(1,2)$.



- $N_1 \sim \mathcal{P}(1)$ with cdf F,

- $N_2 \sim \mathcal{P}(2)$ with cdf G,

CONSTRUCTION METHOD II.: ILLUSTRATION



OUTLINE

- The “continuous” vs. “non-continuous” case: typical fallacies
- Dependence of non-continuous random variables
- Dependent risk processes
- References

References

[Denuit and Lambert, 2005] Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93:40–57.

[Hoeffding, 1940] Hoeffding, W. (1940). Maßstabinvariante korrelationstheorie für diskontinuierliche verteilungen. *Archiv für mathematische Wirtschafts- und Sozialforschung*, VII.(2):4–70.

[Marshall, 1996] Marshall, A. W. (1996). Copulas, marginals and joint distributions. In Rüschendorf, L., Schweizer, B., and Taylor, M. D., editors, *Distributions with Fixed Marginals and Related Topics*, pages 213–222. Institute of Mathematical Statistics, Hayward, CA.

[Neslehova, 2004] Neslehova, J. (2004). *Dependence of Non-Continuous Random Variables*. PhD thesis, Carl von Ossietzky Universität Oldenburg.

[Neslehova and Pfeifer, 2004] Neslehova, J. and Pfeifer, D. (2004). Modeling and generating dependent risk processes for IRM and DFA. *Astin Bulletin*, 34(2):333–360.