



HAL
open science

A semiparametric family of bivariate copulas: dependence properties and estimation procedures

Cécile Amblard, Stéphane Girard

► **To cite this version:**

Cécile Amblard, Stéphane Girard. A semiparametric family of bivariate copulas: dependence properties and estimation procedures. IMS Annual Meeting and X Brazilian School of Probability, Jul 2006, Rio de Janeiro, Brazil. hal-00985320

HAL Id: hal-00985320

<https://inria.hal.science/hal-00985320>

Submitted on 29 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A SEMIPARAMETRIC FAMILY OF BIVARIATE COPULAS: DEPENDENCE PROPERTIES AND ESTIMATION PROCEDURES

Stéphane Girard, Université Grenoble 1.

Joint work with Cécile Amblard.

Outline

1. Definition and basic properties.
2. First sub-family, the case $\theta(1) = 0$.
3. Second sub-family, the case $\phi(1) = 0$.
4. Inference procedures.
5. Simulation results.
6. Real data.

1. Definition and basic properties.

Definition. Let I be the unit interval. The family is defined for all $(u, v) \in I^2$ by,

$$C_{\theta, \phi}(u, v) = uv + \theta[\max(u, v)]\phi(u)\phi(v).$$

where ϕ and θ are differentiable $I \rightarrow \mathbb{R}$ functions (vanishing at most on isolated points).

Theorem. $C_{\theta, \phi}$ is a copula if and only if ϕ and θ satisfy the following conditions:

- boundary conditions: $\phi(0) = 0$ and $(\phi\theta)(1) = 0$,
- θ is non increasing on I ,
- $\phi'(u)(\theta\phi)'(v) \geq -1$ for all $0 \leq u \leq v \leq 1$.

Remark. The family can be split in two sub-families according to $\theta(1) = 0$ or $\phi(1) = 0$.

Measure of association.

Let (X, Y) a random pair with joint distribution $H(x, y) = C(F(x), G(y))$. Spearman's Rho: probability of concordance minus the probability of discordance of two random pairs with respective joint cumulative law $C(F, G)$ and FG .

$$\rho = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3.$$

In the case of $C = C_{\theta, \phi}$, we have

$$\rho_{\theta, \phi} = 12 \left[\Phi^2(1)\theta(1) - \int_0^1 \Phi^2(t)\theta'(t) dt \right],$$

where $\Phi(t) = \int_0^t \phi(u) du$.

Remark.

- If $\theta(1) = 0$, then $\rho_{\theta, \phi} \geq 0$.
- If θ is a constant function, then $\rho_{\theta, \phi} = 12\theta\Phi^2(1)$.

Upper tail dependence.

The upper tail dependence coefficient is defined as

$$\lambda = \lim_{t \rightarrow 1} \mathbb{P}(F(X) > t | G(Y) > t) = \lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1 - u},$$

where \bar{C} is the survival copula, *i.e.* $\bar{C}(u, v) = 1 - u - v + C(u, v)$.

In the case where $C = C_{\theta, \phi}$, we have

$$\lambda_{\theta, \phi} = -\phi^2(1)\theta'(1).$$

Remark.

- If $\phi(1) = 0$, then $\lambda_{\theta, \phi} = 0$.
- If θ is a constant function, then $\lambda_{\theta, \phi} = 0$.

2. First sub-family, the case $\theta(1) = 0$.

Examples.

- Fréchet upper bound. Choosing $\phi(x) = x$ and $\theta(x) = (1 - x)/x$ yields $C_{\theta,\phi}(u, v) = M(u, v) = \min(u, v)$.
- Independent copula. $\theta(x) = 0$ yields $C_{\theta,\phi}(u, v) = \Pi(u, v) = uv$.
- Cuadras-Augé family: $\phi(x) = x$ and $\theta(x) = x^{-\alpha} - 1$, $0 \leq \alpha \leq 1$ yields

$$C_{\theta,\phi}(u, v) = \min(u, v)^\alpha (uv)^{1-\alpha} = M^\alpha(u, v) \Pi^{1-\alpha}(u, v),$$

which is the weighted geometric mean of M and Π .

Remark.

- $\theta(1) = 0$ and $\theta'(u) \leq 0$ imply $\theta(u) \geq 0$ for all $u \in I$.
- $0 \leq \rho_{\theta,\phi} \leq 1$ \longrightarrow Modelling of positive dependences.
- Lower (0) and upper bounds (1) of $\rho_{\theta,\phi}$ and $\lambda_{\theta,\phi}$ are reached respectively by the Π and M copulas.

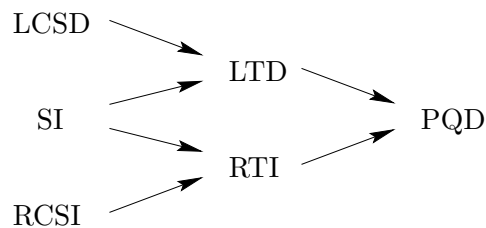
Dependence properties: definitions.

Assume X and Y are exchangeable. X and Y are

- Positively Quadrant Dependent (PQD) if $\mathbb{P}(X \leq x, Y \leq y) \geq \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ for all (x, y) .
- Left Tail Decreasing (LTD) if $\mathbb{P}(Y \leq y|X \leq x)$ is non-increasing in x for all y .
- Right Tail Increasing (RTI) if $\mathbb{P}(Y > y|X > x)$ is nondecreasing in x for all y .
- Stochastically Increasing (SI) if $\mathbb{P}(Y > y|X = x)$ is nondecreasing in x for all y .
- Left Corner Set Decreasing (LCSD) if $\mathbb{P}(X \leq x, Y \leq y|X \leq x', Y \leq y')$ is non-increasing in x' and y' for all (x, y) .
- Right Corner Set Increasing (RCSI) if $\mathbb{P}(X > x, Y > y|X > x', Y > y')$ is nondecreasing in x' and y' for all (x, y) .

Theorem. X and Y are:

- PQD iff $\phi(u)$ has a constant sign on I .
- LTD or LCSD iff either $\{\phi(u)/u$ is non increasing and $\forall u \in I, \phi(u) \geq 0\}$ or $\{\phi(u)/u$ is non decreasing and $\forall u \in I, \phi(u) \leq 0\}$.
- RTI or RCSI iff $\phi(u)/(1-u)$ and $\theta(u)\phi(u)/(1-u)$ are monotone.
- SI iff either $\{\phi$ and $\theta\phi$ are concave and $\forall u \in I, \phi(u) \geq 0\}$ or $\{\phi$ and $\theta\phi$ are convex and $\forall u \in I, \phi(u) \leq 0\}$.



Implications in the general case



Implications in the sub-family

3. Second sub-family, the case $\phi(1) = 0$.

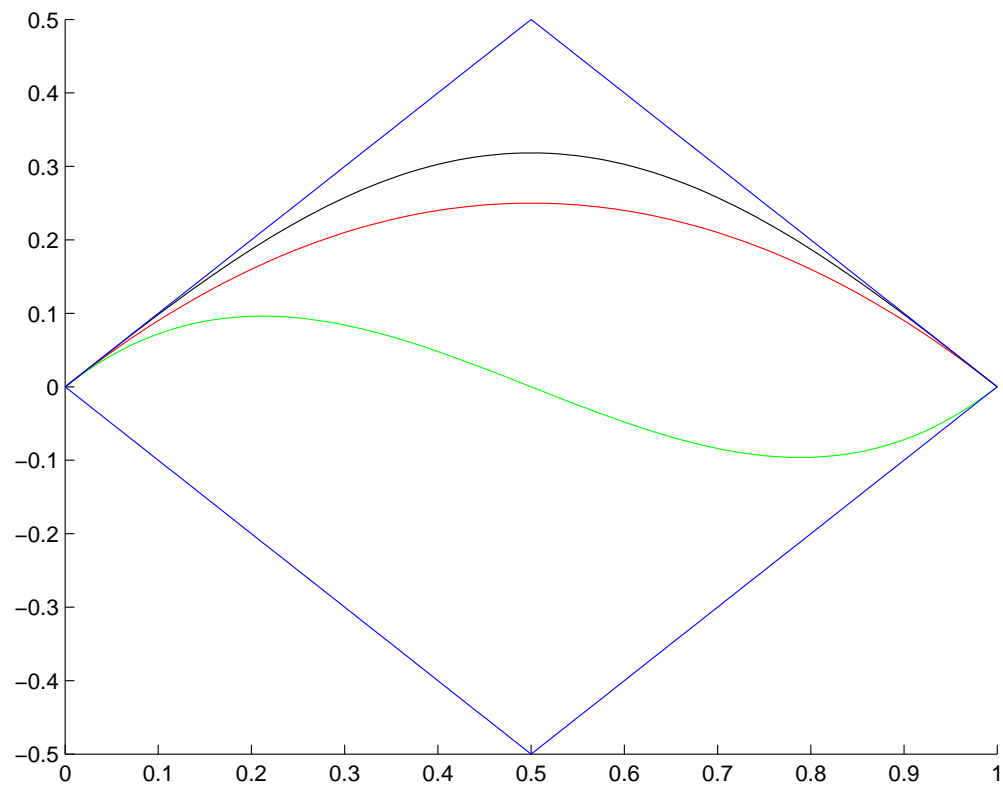
In this case, we restrict ourselves to a constant function θ , *i.e.* $\theta(x) = \theta \in [-1, 1]$.

Theorem. $C_{\theta, \phi}$ is a copula if and only if ϕ and θ satisfy the following conditions:

- boundary conditions: $\phi(0) = 0$ and $\phi(1) = 0$,
- $|\phi'(x)| \leq 1$ for all $x \in I$,
- $|\phi(x)| \leq \min(x, 1 - x)$, for all $x \in I$.

Examples.

- $\phi(x) = \min(x, 1 - x)$: upper bound of the above theorem,
- $\phi(x) = x(1 - x)$: Farlie-Gumbel-Morgenstern family of copulas (Morgenstern, 1956), which contains all copulas with both horizontal and vertical quadratic sections (Quesada-Molina, Rodríguez-Lallena, 1995)
- $\phi(x) = x(1 - x)(1 - 2x)$: symmetric copulas with cubic sections (Nelsen *et al*, 1997),
- $\phi(x) = \pi^{-1} \sin(\pi x)$.



Upper bound, Farlie-Gumbel-Morgenstern, cubic sections, sinus.

Measure of association. The Spearman's Rho can be rewritten as:

$$\rho_{\theta,\phi} = 12\theta \left(\int_I \phi(u) du \right)^2,$$

and it follows that $-3/4 \leq \rho_{\theta,\phi} \leq 3/4$ for all $\theta \in [-1, 1]$. Similar bounds hold for the Kendall's Tau: $-1/2 \leq \tau_{\theta,\phi} \leq 1/2$.

Upper tail dependence. $\rho_{\theta,\phi} = 0$.

Dependence properties. Similar to the previous family in the case $\theta > 0$.

Symmetry properties: definitions.

- X is **symmetric** about a if $(X - a)$ and $(a - X)$ are identically distributed (id).
- X and Y are **exchangeable** if (X, Y) and (Y, X) are id.
- (X, Y) is **marginally symmetric** about (a, b) if X and Y are symmetric about a and b respectively.
- (X, Y) is **radially symmetric** about (a, b) if $(X - a, Y - b)$ and $(a - X, b - Y)$ are id.
- (X, Y) is **jointly symmetric** about (a, b) if the pairs $(X - a, Y - b)$, $(a - X, b - Y)$, $(X - a, b - Y)$ and $(a - X, Y - b)$ are id.

Theorem. In the $C_{\theta, \phi}$ family:

- If X and Y are id then X and Y are exchangeable.

Besides, if (X, Y) is marginally symmetric about (a, b) then:

- (X, Y) is radially symmetric about (a, b) if and only if
 either $\forall u \in I, \phi(u) = \phi(1 - u)$ or $\forall u \in I, \phi(u) = -\phi(1 - u)$.
- (X, Y) is jointly symmetric about (a, b) if and only if $\forall u \in I, \phi(u) = -\phi(1 - u)$.

4. Inference procedures.

Assumptions.

- We restrict ourselves to the second sub-family, with constant function θ :

$$C(u, v) = uv + \theta\phi(u)\phi(v).$$

→ Estimation of θ (scalar) and ϕ (univariate function).

→ Identifiability problem: (θ, ϕ) and $(\alpha\theta, \phi/\sqrt{\alpha})$ yield the same copula for all $\alpha > 0$.

- We focus on the PQD case: $\theta > 0$ and ϕ has a constant sign.

Under these assumptions, the family can be rewritten

$$C(u, v) = uv + \psi(u)\psi(v),$$

where $\psi(x) = \sqrt{\theta}|\phi(x)|$.

→ The estimation of C reduces to the estimation of ψ (positive univariate function).

Estimation of ψ

1) Preprocessing:

- $\{(x_i, y_i), i = 1, \dots, n\}$ a sample of (X, Y) from the cdf $H(x, y) = C(F(x), G(y))$.
- Rank transformations: $u_i = \text{rank}(x_i)/n$ and $v_i = \text{rank}(y_i)/n$.
 $\{(u_i, v_i), i = 1, \dots, n\}$ an approximate sample from the copula $C(u, v)$.
- Pseudo-observations $\{w_i = \max(u_i, v_i), i = 1, \dots, n\}$ from $C(w, w) = w^2 + \psi(w)$.

2) Projection estimate: linear combination of basis functions: $\{e_k, k \geq 1\}$

$$\hat{\psi}(w) = \sum_{k \geq 1} a_k e_k(w), \quad w \in I.$$

Choice of the set of functions:

- no orthogonality condition,
- boundary conditions $e_k(0) = e_k(1) = 0$ for all $k \geq 1$ so that $\hat{\psi}(0) = \hat{\psi}(1) = 0$.

3) Optimization problem: Define

- $w_{1,n} \leq \dots \leq w_{n,n}$, the ordered pseudo-observations,
- M and M' two matrices $M_{i,k} = e_k(w_{i,n})$, $M'_{i,k} = e'_k(w_{i,n})$, $k \geq 1$, $i \in \{1, \dots, n\}$,
- a and b two vectors $b_i = (i/(n+1) - w_{i,n}^2)^{1/2}$, a_i unknown, $i \in \{1, \dots, n\}$.

Definition of the estimator.

- $\hat{\psi}(w_{i,n}) = C(w_{i,n}, w_{i,n}) - w_{i,n}^2 \simeq i/(n+1) - w_{i,n}^2$ for $i = 1, \dots, n$ can be rewritten

$$\min_a \|Ma - b\|^2,$$

- $\hat{\psi}(w_{i,n}) \geq 0$ can be rewritten $Ma \geq 0$,
- $|\hat{\psi}(w_{i,n})| \leq 1$ can be rewritten $-1 \leq M'a \leq 1$.

→ Constrained least-square problem.

Estimation of the Spearman's rho

Recall that

$$\rho_{\theta,\phi} = 12\theta \left(\int_I \phi(u) du \right)^2 = 12 \left(\int_I \psi(u) du \right)^2.$$

Replacing ψ by $\hat{\psi}$ yields the following semi-parametric estimator:

$$\hat{\rho}_{\text{SP}} = 12 \left(\sum_{k \geq 1} a_k \beta_k \right)^2,$$

where we have introduced $\beta_k = \int_I e_k(u) du$.

Another solution: adapt the nonparametric estimator of the Kendall's Tau introduced in (Genest, Rivest, 1993) to obtain

$$\hat{\rho}_{\text{NP}} = \frac{6}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}\{u_j < u_i, v_j < v_i\} - \frac{3}{2},$$

Estimation of high probability regions

Definition. The α -quantile of the copula C is defined by

$$Q_\alpha = \inf\{\lambda(S) : \mathbb{P}(S) \geq \alpha, S \subset I^2\}, \quad 0 < \alpha \leq 1,$$

where λ is the Lebesgue measure on I^2 .

Partitions. $\{I_k, k = 1, \dots, N\}$ be the equidistant N -partition of I , $K_{k,\ell} = I_k \times I_\ell$ the associated $N \times N$ grid. Denote $\delta_{k,\ell} \in \{0, 1\}$, $(k, \ell) \in \{1, \dots, N\}^2$.

Estimator: $\hat{Q}_\alpha = \bigcup_{k,\ell} K_{k,\ell} \mathbf{1}\{\delta_{k,\ell} = 1\}$.

Optimization problem. The $\delta_{k,\ell}$ are defined by

$$\min \sum_{k=1}^N \sum_{\ell=1}^N \delta_{k,\ell},$$

under the constraints $\delta_{k,\ell} \in \{0, 1\}$ and $\sum_{k=1}^N \sum_{\ell=1}^N \delta_{k,\ell} \hat{P}(K_{k,\ell}) \geq \alpha$,

where $\hat{P}(K_{k,\ell})$ is an estimation of the probability $P(K_{k,\ell})$.

Algorithm.

- First step: sort the $\widehat{P}(K_{k,\ell})$ in decreasing order to obtain the sequence \tilde{P}_τ , $\tau = 1, \dots, N^2$.
- Second step: Computation of the number of subsets of the partition:

$$J = \min \left\{ j, \sum_{\tau=1}^j \tilde{P}_\tau \geq \alpha \right\}.$$

- Third step: selection of the J first subsets: $\delta_{k,\ell} = 1$ if $1 \leq \tau(k, \ell) \leq J$,

Estimation of $P(K_{k,\ell})$. Two solutions:

- Semi-parametric estimate based on $\widehat{\psi}$

$$\widehat{P}_{\text{SP}}(K_{k,\ell}) = \frac{1}{N^2} + \left(\widehat{\psi} \left(\frac{k}{N} \right) - \widehat{\psi} \left(\frac{k-1}{N} \right) \right) \left(\widehat{\psi} \left(\frac{\ell}{N} \right) - \widehat{\psi} \left(\frac{\ell-1}{N} \right) \right).$$

- Nonparametric estimate

$$\widehat{P}_{\text{NP}}(K_{k,\ell}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{(u_i, v_i) \in K_{k,\ell}\}.$$

5. Simulation results.

Numerical experiments on the family of copulas C_k generated by the set of functions

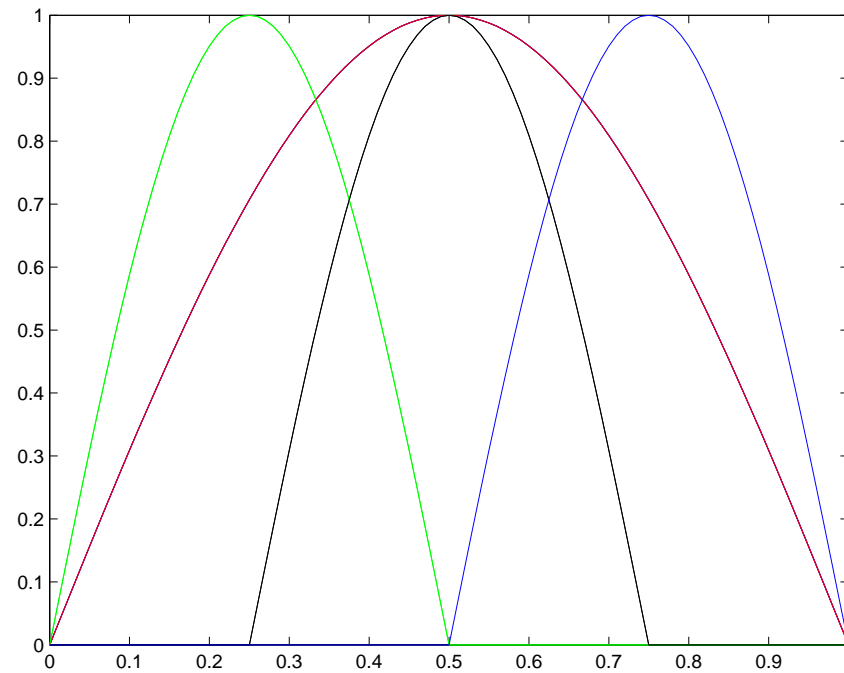
$$\forall k \geq 1, \psi_k(x) = 1 - (x^k + (1-x)^k)^{1/k}, x \in I.$$

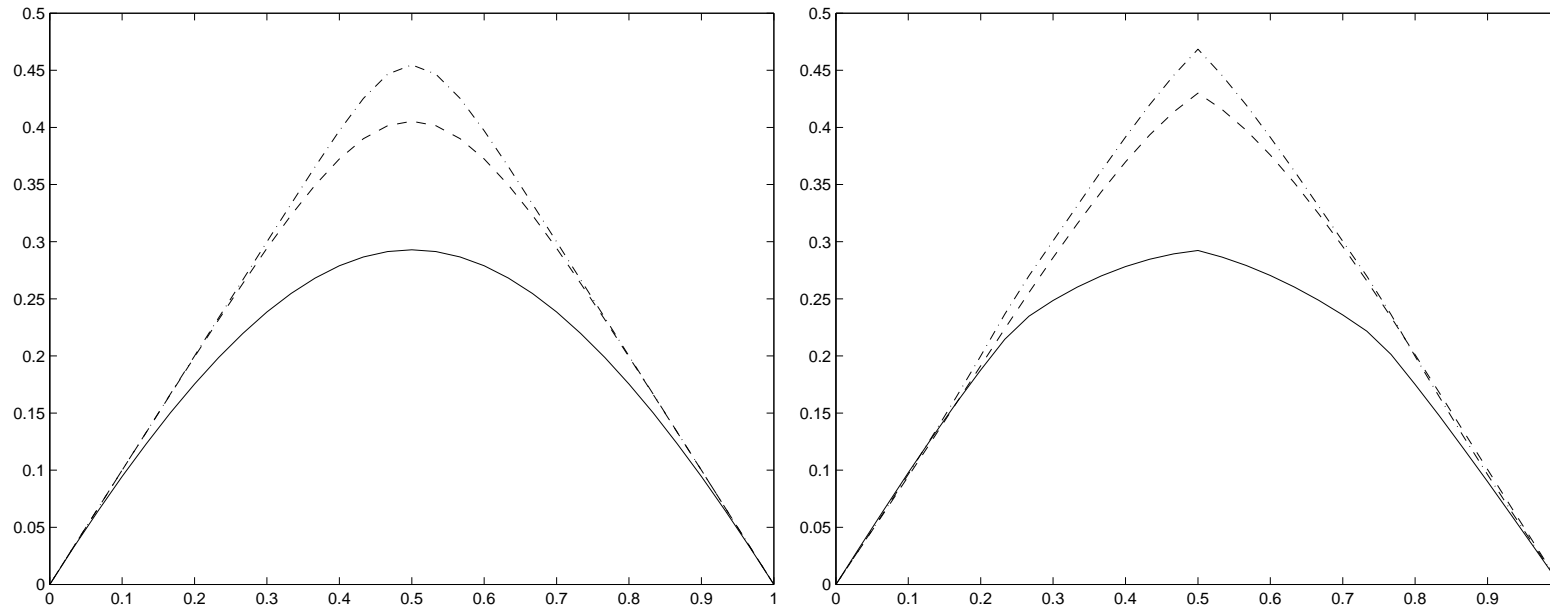
- When $k = 1$, C_1 : uniform distribution on I^2 . Spearman's Rho $\rho_1 = 0$.
- When $k \rightarrow \infty$, $\psi_k(x) \rightarrow \psi_\infty(x) = \min(x, 1-x)$ for all $x \in I$.
 C_∞ : mixture of two uniform distributions on the squares $[0, 1/2]^2$ and $[1/2, 1]^2$ with mixing parameter $1/2$. Spearman's Rho $\rho_\infty = 3/4$ (the maximum value in the sub-family).
- When $1 < k < \infty$, bivariate distribution “interpolating” between the two previous ones.

Chosen basis of functions:

$$e_{s,\ell}(x) = \sin\left(\frac{\pi}{2}(2^{s+1}x - \ell)\right) \mathbf{1}\{2^{s+1}x \in [\ell, \ell + 2]\},$$

s is a scale parameter, ℓ is a location parameter.

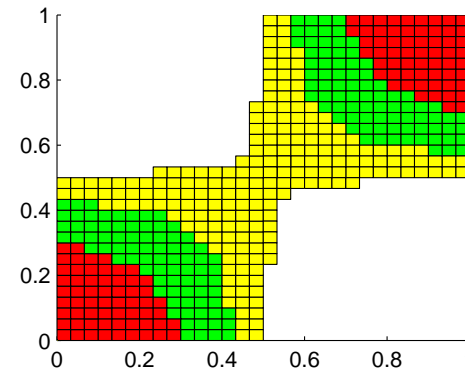
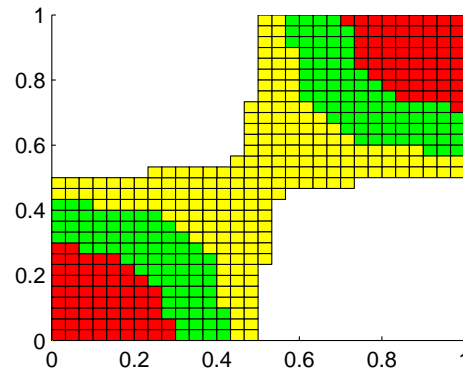
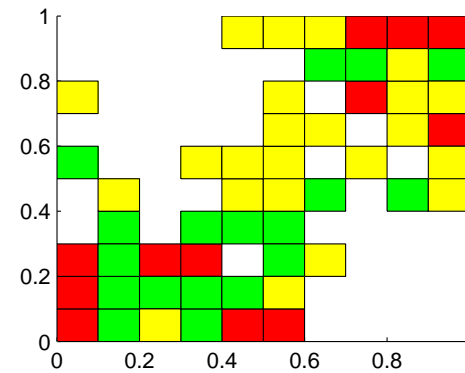
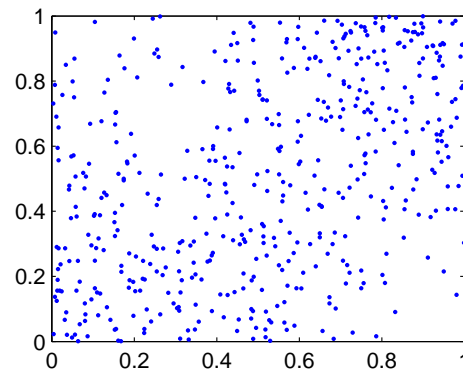




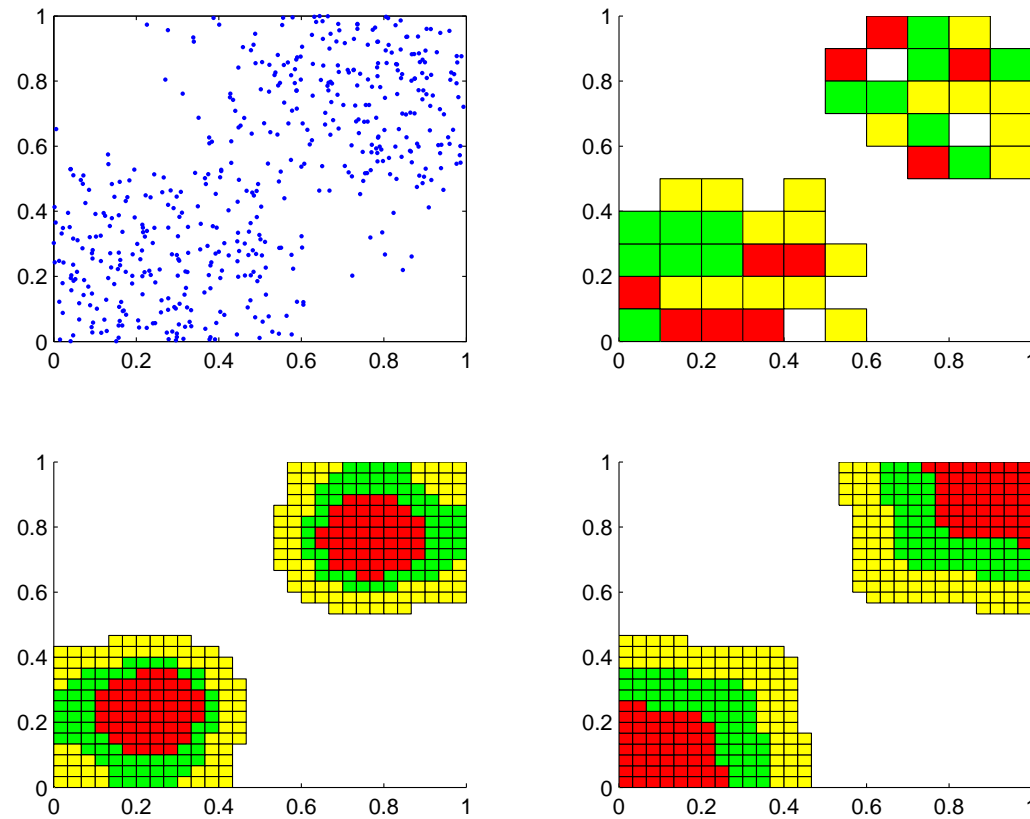
True functions $\psi_k(x)$, $k \in \{2, 4, 8\}$ – Estimated functions $\hat{\psi}_k(x)$, $k \in \{2, 4, 8\}$, $n = 100$.

k	$\rho_k \times 10^{-2}$	$\text{mean}(\hat{\rho}_{\text{SP}}) \times 10^{-2}$	$\text{mean}(\hat{\rho}_{\text{NP}}) \times 10^{-2}$
1	0	0.81	0.18
2	42.5	43.0	41.2
4	66.4	65.8	64.3
6	71.2	70.6	68.8
8	72.8	72.1	70.2

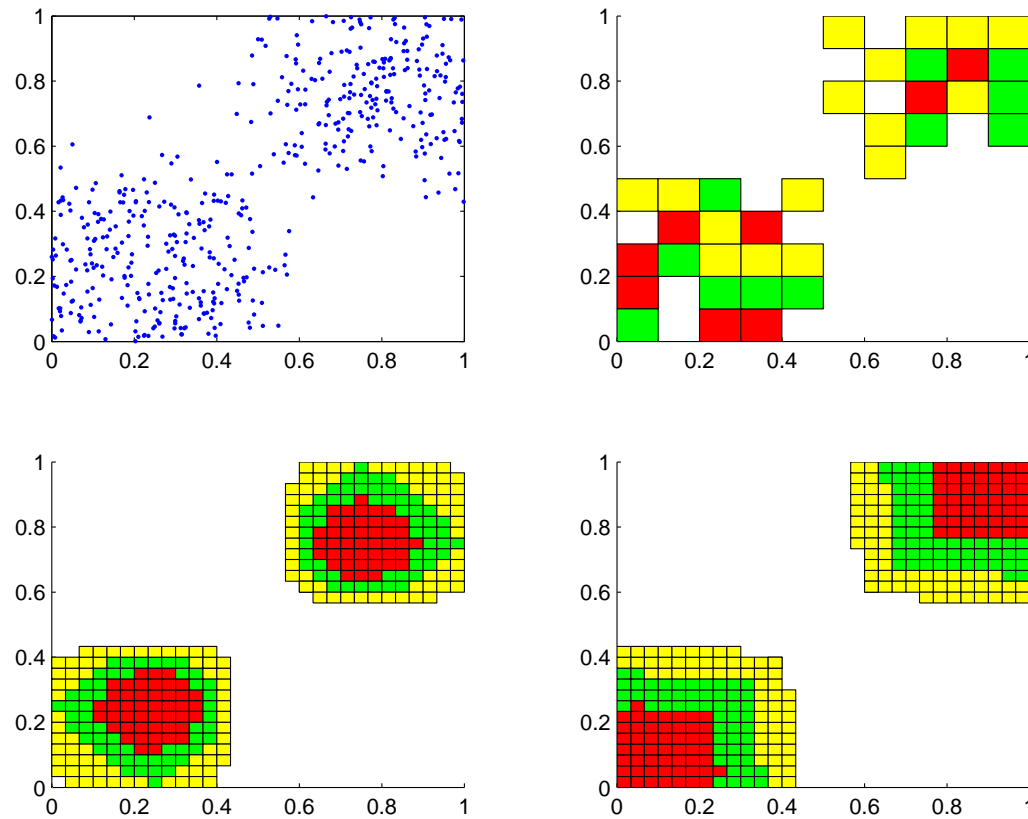
Estimation of the generating function and of the Spearman's Rho (ρ_k). The mean value of the estimates $\hat{\rho}_{\text{SP}}$ and $\hat{\rho}_{\text{NP}}$ are evaluated on 100 repetitions.



Estimation of high probability regions Q_α from C_2 . Red: $\alpha = 0.25$, green: $\alpha = 0.5$, yellow: $\alpha = 0.75$. Top left: simulated sample, top right: nonparametric estimate, bottom left: semiparametric estimate, bottom right: semiparametric estimate with the true function ψ , ($n = 500$).



Estimation of high probability regions Q_α from C_4 . Red: $\alpha = 0.25$, green: $\alpha = 0.5$, yellow: $\alpha = 0.75$. Top left: simulated sample, top right: nonparametric estimate, bottom left: semiparametric estimate, bottom right: semiparametric estimate with the true function ψ , ($n = 500$).

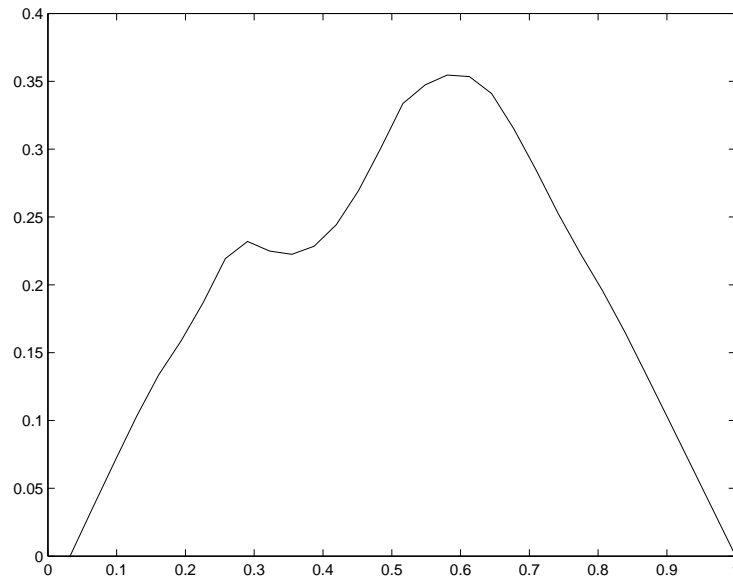


Estimation of high probability regions Q_α from C_8 . Red: $\alpha = 0.25$, green: $\alpha = 0.5$, yellow: $\alpha = 0.75$. Top left: simulated sample, top right: nonparametric estimate, bottom left: semiparametric estimate, bottom right: semiparametric estimate with the true function ψ , ($n = 500$).

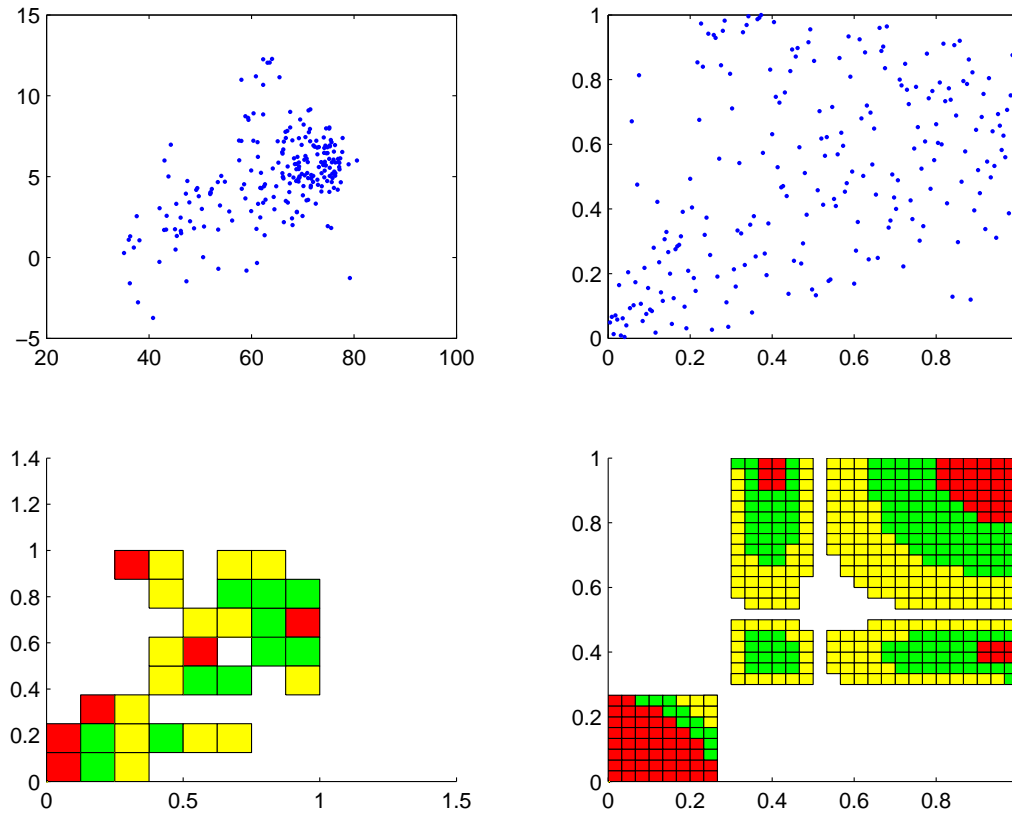
6. Real data.

$n = 225$ countries, two variables: X , the life expectancy at birth (years) in 2002 of the total population and Y , the difference between the life expectancy at birth of women and men. <http://www.odci.gov/cia/publications/factbook/>.

According to the PQD test proposed in (Scaillet, 2004), these data are PQD.



$$\hat{\rho}_{\text{NP}} = 52.4\%$$
$$\hat{\rho}_{\text{SP}} = 40.7\%$$



Estimation of high probability regions Q_α from real data. Red: $\alpha = 0.25$, green: $\alpha = 0.5$, yellow: $\alpha = 0.75$. Top left: real data, top right: real data after rank transformation, bottom left: nonparametric estimate, bottom right: semiparametric estimate.

Further work.

- Goodness of fit test.
- Study of the sub-family $\phi(1) = 0$ without the assumption that θ is a constant function.
(what is the lower bound of $\rho_{\theta, \phi}$?)
- Estimation of the function θ in the general case.

References.

- C. Amblard and S. Girard. A new bivariate extension of FGM copulas, *Metrika*, 70, 1–17, 2009.
- C. Amblard and S. Girard. Estimation procedures for a semiparametric family of bivariate copulas, *Journal of Computational and Graphical Statistics*, 14, 1–15, 2005.
- C. Amblard and S. Girard. Symmetry and dependence properties within a semiparametric family of bivariate copulas, *Nonparametric Statistics*, 14, 715–727, 2002.
- C. Amblard and S. Girard. A semiparametric family of symmetric bivariate copulas, *Comptes-Rendus de l'Académie des Sciences*, t. 333, Série I:129–132, 2001