Probit transformation for nonparametric kernel estimation of the copula

A. Charpentier (Université de Rennes 1 & UQAM),

joint work with G. Geenens (UNSW) & D. Paindaveine (ULB)

CIRM Workshop "Extremes - Copulas - Actuarial science", February 2016 http://arxiv.org/abs/1404.4414, to appear in *Bernoulli*

url: http://freakonometrics.hypotheses.org
email : charpentier.arthur@uqam.ca
y: @freakonometrics

Motivation

Consider some *n*-i.i.d. sample $\{(X_i, Y_i)\}$ with cumulative distribution function F_{XY} and joint density f_{XY} . Let F_X and F_Y denote the marginal distributions, and C the copula,

 $F_{XY}(x,y) = C(F_X(x), F_Y(y))$

so that

 $f_{XY}(x,y) = f_X(x)f_Y(y)\boldsymbol{c}(F_X(x),F_Y(y))$

We want a nonparametric estimate of c on $[0, 1]^2$.



Notations

Define uniformized *n*-i.i.d. sample $\{(U_i, V_i)\}$

 $U_i = F_X(X_i)$ and $V_i = F_Y(Y_i)$

or uniformized *n*-i.i.d. pseudo-sample $\{(\hat{U}_i, \hat{V}_i)\}$

$$\hat{U}_i = \frac{n}{n+1}\hat{F}_{Xn}(X_i) \text{ and } \hat{V}_i = \frac{n}{n+1}\hat{F}_{Yn}(Y_i)$$

where \hat{F}_{Xn} and \hat{F}_{Yn} denote empirical c.d.f.



Standard Kernel Estimate

The standard kernel estimator for c, say \hat{c}^* , at $(u, v) \in \mathcal{I}$ would be (see Wand & Jones (1995))

$$\hat{c}^*(u,v) = \frac{1}{n|\mathbf{H}_{UV}|^{1/2}} \sum_{i=1}^n \mathbf{K} \left(\mathbf{H}_{UV}^{-1/2} \begin{pmatrix} u - U_i \\ v - V_i \end{pmatrix} \right), \tag{1}$$

where $\mathbf{K}: \mathbb{R}^2 \to \mathbb{R}$ is a kernel function and \mathbf{H}_{UV} is a bandwidth matrix.

Standard Kernel Estimate

However, this estimator is not consistent along boundaries of $[0,1]^2$

$$\mathbb{E}(\hat{c}^*(u,v)) = \frac{1}{4}c(u,v) + O(h) \text{ at corners}$$

$$\mathbb{E}(\hat{c}^*(u,v)) = \frac{1}{2}c(u,v) + O(h) \text{ on the borders}$$

if **K** is symmetric and \mathbf{H}_{UV} symmetric. Corrections have been proposed, e.g. mirror reflection Gijbels (1990) or the usage of boundary kernels Chen (2007), but with mixed results.

Remark: the graph on the bottom is \hat{c}^* on the (first) diagonal.





Mirror Kernel Estimate

Use an enlarged sample: instead of only $\{(\hat{U}_i, \hat{V}_i)\},\$ add $\{(-\hat{U}_i, \hat{V}_i)\},\$ $\{(\hat{U}_i, -\hat{V}_i)\},\$ $\{(-\hat{U}_i, -\hat{V}_i)\},\$ $\{(\hat{U}_i, 2 - \hat{V}_i)\},\$ $\{(2 - \hat{U}_i, \hat{V}_i)\},$ $\{(-\hat{U}_i, 2 - \hat{V}_i)\},\$ $\{(2 - \hat{U}_i, -\hat{V}_i)\}\$ and $\{(2 - \hat{U}_i, 2 - \hat{V}_i)\}.$

See Gijbels & Mielniczuk (1990).

That estimator will be used as a benchmark in the simulation study.





Using Beta Kernels

Use a Kernel which is a product of beta kernels

$$\mathbf{K}_{\boldsymbol{x}_{i}}(\boldsymbol{u})) \propto \left(u_{1}^{\frac{x_{1,i}}{b}} [1-u_{1}]^{\frac{x_{1,i}}{b}}\right) \cdot \left(u_{2}^{\frac{x_{2,i}}{b}} [1-u_{2}]^{\frac{x_{2,i}}{b}}\right)$$

See Chen (1999).





Probit Transformation

See Devroye & Gyöfi (1985) and Marron & Ruppert (1994).

Define normalized *n*-i.i.d. sample $\{(S_i, T_i)\}$

$$S_i = \Phi^{-1}(U_i)$$
 and $T_i = \Phi^{-1}(V_i)$

or normalized *n*-i.i.d. pseudo-sample $\{(\hat{S}_i, \hat{T}_i)\}$

$$\hat{U}_i = \Phi^{-1}(\hat{U}_i) \text{ and } \hat{V}_i = \Phi^{-1}(\hat{V}_i)$$

where Φ^{-1} is the quantile function of $\mathcal{N}(0,1)$ (probit transformation).



Probit Transformation

 $F_{ST}(x,y) = C(\Phi(x), \Phi(y))$

so that

$$f_{ST}(x,y) = \phi(x)\phi(y)\boldsymbol{c}(\Phi(x),\Phi(y))$$

Thus

$$\boldsymbol{c}(u,v) = \frac{f_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}.$$

So use

$$\hat{c}^{(\tau)}(u,v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$



9

The naive estimator

Since we cannot use

$$\hat{f}_{ST}^*(s,t) = \frac{1}{n|\mathbf{H}_{ST}|^{1/2}} \sum_{i=1}^n \mathbf{K} \left(\mathbf{H}_{ST}^{-1/2} \binom{s-S_i}{t-T_i} \right),$$



where **K** is a kernel function and \mathbf{H}_{ST} is a bandwidth matrix, use

$$\hat{f}_{ST}(s,t) = \frac{1}{n|\mathbf{H}_{ST}|^{1/2}} \sum_{i=1}^{n} \mathbf{K} \left(\mathbf{H}_{ST}^{-1/2} \begin{pmatrix} s - \hat{S}_i \\ t - \hat{T}_i \end{pmatrix} \right).$$

and the copula density is

$$\hat{c}^{(\tau)}(u,v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$



The naive estimator

$$\hat{c}^{(\tau)}(u,v) = \frac{1}{n|\mathbf{H}_{ST}|^{1/2}\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \sum_{i=1}^{n} \mathbf{K} \left(\mathbf{H}_{ST}^{-1/2} \begin{pmatrix} \Phi^{-1}(u) - \Phi^{-1}(\hat{U}_{i}) \\ \Phi^{-1}(v) - \Phi^{-1}(\hat{V}_{i}) \end{pmatrix} \right)$$

as suggested in C., Fermanian & Scaillet (2007) and Lopez-Paz . *et al.* (2013). Note that Omelka . *et al.* (2009) obtained theoretical properties on the convergence of $\hat{C}^{(\tau)}(u, v)$ (not c).

Improved probit-transformation copula density estimators

When estimating a density from **pseudo-sample**, Loader (1996) and Hjort & Jones (1996) define a local likelihood estimator

Around $(s,t) \in \mathbb{R}^2$, use a polynomial approximation of order p for log f_{ST}

$$\log f_{ST}(\check{s},\check{t}) \simeq a_{1,0}(s,t) + a_{1,1}(s,t)(\check{s}-s) + a_{1,2}(s,t)(\check{t}-t) \doteq P_{\mathbf{a}_1}(\check{s}-s,\check{t}-t)$$

$$\log f_{ST}(\check{s},\check{t}) \simeq a_{2,0}(s,t) + a_{2,1}(s,t)(\check{s}-s) + a_{2,2}(s,t)(\check{t}-t) + a_{2,3}(s,t)(\check{s}-s)^2 + a_{2,4}(s,t)(\check{t}-t)^2 + a_{2,5}(s,t)(\check{s}-s)(\check{t}-t) \doteq P_{\mathbf{a}_2}(\check{s}-s,\check{t}-t).$$

Improved probit-transformation copula density estimators

Remark Vectors $\mathbf{a}_1(s,t) = (a_{1,0}(s,t), a_{1,1}(s,t), a_{1,2}(s,t))$ and $\mathbf{a}_2(s,t) \doteq (a_{2,0}(s,t), \ldots, a_{2,5}(s,t))$ are then estimated by solving a weighted maximum likelihood problem.

$$\begin{split} \tilde{\mathbf{a}}_{p}(s,t) &= \arg \max_{\mathbf{a}_{p}} \left\{ \sum_{i=1}^{n} \mathbf{K} \left(\mathbf{H}_{ST}^{-1/2} \begin{pmatrix} s - \hat{S}_{i} \\ t - \hat{T}_{i} \end{pmatrix} \right) P_{\mathbf{a}_{p}}(\hat{S}_{i} - s, \hat{T}_{i} - t) \\ &- n \iint_{\mathbb{R}^{2}} \mathbf{K} \left(\mathbf{H}_{ST}^{-1/2} \begin{pmatrix} s - \check{s} \\ t - \check{t} \end{pmatrix} \right) \exp \left(P_{\mathbf{a}_{p}}(\check{s} - s, \check{t} - t) \right) \, d\check{s} \, d\check{t} \right\}, \end{split}$$

The estimate of f_{ST} at (s,t) is then $\tilde{f}_{ST}^{(p)}(s,t) = \exp(\tilde{a}_{p,0}(s,t))$, for p = 1, 2.

The Improved probit-transformation kernel copula density estimators are

$$\tilde{c}^{(\tau,p)}(u,v) = \frac{\tilde{f}_{ST}^{(p)}(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$

Improved probit-transformation copula density estimators

For the local log-linear (p = 1) approximation

$$\tilde{c}^{(\tau,1)}(u,v) = \frac{\exp(\tilde{a}_{1,0}(\Phi^{-1}(u),\Phi^{-1}(v)))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$





Improved probit-transformation copula density estimators

For the local log-quadratic (p = 2) approximation

$$\tilde{c}^{(\tau,2)}(u,v) = \frac{\exp(\tilde{a}_{2,0}(\Phi^{-1}(u), \Phi^{-1}(v)))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$





Asymptotic properties

A1. The sample $\{(X_i, Y_i)\}$ is a *n*-i.i.d. sample from the joint distribution F_{XY} , an absolutely continuous distribution with marginals F_X and F_Y strictly increasing on their support;

(uniqueness of the copula)

Asymptotic properties

A2. The copula C of F_{XY} is such that $(\partial C/\partial u)(u, v)$ and $(\partial^2 C/\partial u^2)(u, v)$ exist and are continuous on $\{(u, v) : u \in (0, 1), v \in [0, 1]\}$, and $(\partial C/\partial v)(u, v)$ and $(\partial^2 C/\partial v^2)(u, v)$ exist and are continuous on $\{(u, v) : u \in [0, 1], v \in (0, 1)\}$. In addition, there are constants K_1 and K_2 such that

$$\left(\begin{array}{c|c} \left| \frac{\partial^2 C}{\partial u^2}(u,v) \right| &\leq \frac{K_1}{u(1-u)} & \text{for } (u,v) \in (0,1) \times [0,1]; \\ \left| \frac{\partial^2 C}{\partial v^2}(u,v) \right| &\leq \frac{K_2}{v(1-v)} & \text{for } (u,v) \in [0,1] \times (0,1); \end{array} \right)$$

A3. The density c of C exists, is positive and admits continuous second-order partial derivatives on the interior of the unit square \mathcal{I} . In addition, there is a constant K_{00} such that

$$c(u,v) \le K_{00} \min\left(\frac{1}{u(1-u)}, \frac{1}{v(1-v)}\right) \qquad \forall (u,v) \in (0,1)^2.$$

see Segers (2012).

Asymptotic properties

Assume that $\mathbf{K}(z_1, z_2) = \phi(z_1)\phi(z_2)$ and $\mathbf{H}_{ST} = h^2 \mathbf{I}$ with $h \sim n^{-a}$ for some $a \in (0, 1/4)$. Under Assumptions A1-A3, the 'naive' probit transformation kernel copula density estimator at any $(u, v) \in (0, 1)^2$ is such that

$$\sqrt{nh^2}\left(\hat{c}^{(\tau)}(u,v) - c(u,v) - h^2 \frac{b(u,v)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}\right) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(0,\sigma^2(u,v)\right),$$

where
$$b(u,v) = \frac{1}{2} \left\{ \frac{\partial^2 c}{\partial u^2}(u,v)\phi^2(\Phi^{-1}(u)) + \frac{\partial^2 c}{\partial v^2}(u,v)\phi^2(\Phi^{-1}(v)) - 3\left(\frac{\partial c}{\partial u}(u,v)\Phi^{-1}(u)\phi(\Phi^{-1}(u)) + \frac{\partial c}{\partial v}(u,v)\Phi^{-1}(v)\phi(\Phi^{-1}(v))\right) + c(u,v)\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2\right) \right\}$$
(2)

and
$$\sigma^{2}(u, v) = \frac{c(u, v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$

The Amended version

The last unbounded term in b be easily adjusted.

$$\hat{c}^{(\tau \text{am})}(u,v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \times \frac{1}{1 + \frac{1}{2}h^2\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2\right)}.$$

The asymptotic bias becomes proportional to

$$b^{(\mathrm{am})}(u,v) = \frac{1}{2} \left\{ \frac{\partial^2 c}{\partial u^2}(u,v)\phi^2(\Phi^{-1}(u)) + \frac{\partial^2 c}{\partial v^2}(u,v)\phi^2(\Phi^{-1}(v)) - 3\left(\frac{\partial c}{\partial u}(u,v)\Phi^{-1}(u)\phi(\Phi^{-1}(u)) + \frac{\partial c}{\partial v}(u,v)\Phi^{-1}(v)\phi(\Phi^{-1}(v))\right) \right\}$$

A local log-linear probit-transformation kernel estimator

$$\tilde{c}^{*(\tau,1)}(u,v) = \tilde{f}_{ST}^{*(1)}(\Phi^{-1}(u), \Phi^{-1}(v)) / \left(\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\right)$$

Then

$$\sqrt{nh^2}\left(\tilde{c}^{*(\tau,1)}(u,v) - c(u,v) - h^2 \frac{b^{(1)}(u,v)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}\right) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(0, \, \sigma^{(1)^2}(u,v)\right),$$

where
$$b^{(1)}(u,v) = \frac{1}{2} \left\{ \frac{\partial^2 c}{\partial u^2}(u,v)\phi^2(\Phi^{-1}(u)) + \frac{\partial^2 c}{\partial v^2}(u,v)\phi^2(\Phi^{-1}(v)) - \frac{1}{c(u,v)} \left(\left\{ \frac{\partial c}{\partial u}(u,v) \right\}^2 \phi^2(\Phi^{-1}(u)) + \left\{ \frac{\partial c}{\partial v}(u,v) \right\}^2 \phi^2(\Phi^{-1}(v)) \right) - \left(\frac{\partial c}{\partial u}(u,v)\Phi^{-1}(u)\phi(\Phi^{-1}(u)) + \frac{\partial c}{\partial v}(u,v)\Phi^{-1}(v)\phi(\Phi^{-1}(v)) \right) - 2c(u,v) \right\}$$

Using a higher order polynomial approximation

Locally fitting a polynomial of a higher degree is known to reduce the asymptotic bias of the estimator, here from order $O(h^2)$ to order $O(h^4)$, see Loader (1996) or Hjort (1996), under sufficient smoothness conditions.

If f_{ST} admits continuous fourth-order partial derivatives and is positive at (s, t), then

$$\begin{split} &\sqrt{nh^2} \left(\tilde{f}_{ST}^{*(2)}(s,t) - f_{ST}(s,t) - h^4 b_{ST}^{(2)}(s,t) \right) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N} \left(0, \sigma_{ST}^{(2)} \right), \\ \text{where } \sigma_{ST}^{(2)} \left(s,t \right) &= \frac{5}{2} \, \frac{f_{ST}(s,t)}{4\pi} \text{ and} \\ & b_{ST}^{(2)}(s,t) = -\frac{1}{8} f_{ST}(s,t) \\ &\times \left\{ \left(\frac{\partial^4 g}{\partial s^4} + \frac{\partial^4 g}{\partial t^4} \right) + 4 \left(\frac{\partial^3 g}{\partial s^3} \frac{\partial g}{\partial s} + \frac{\partial^3 g}{\partial t^3} \frac{\partial g}{\partial t} + \frac{\partial^3 g}{\partial s^2 \partial t} \frac{\partial g}{\partial t} + \frac{\partial^3 g}{\partial s \partial t^2} \frac{\partial g}{\partial s} \right) + 2 \frac{\partial^4 g}{\partial s^2 \partial t^2} \right\} (s,t), \\ \text{with } g(s,t) &= \log f_{ST}(s,t). \end{split}$$

Using a higher order polynomial approximation

A4. The copula density $c(u, v) = (\partial^2 C / \partial u \partial v)(u, v)$ admits continuous fourth-order partial derivatives on the interior of the unit square $[0, 1]^2$. Then

$$\sqrt{nh^2} \left(\tilde{c}^{*(\tau,2)}(u,v) - c(u,v) - h^4 \frac{b^{(2)}(u,v)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \right) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N} \left(0, \, \sigma^{(2)^2}(u,v) \right)$$

where
$$\sigma^{(2)^2}(u,v) = \frac{5}{2} \frac{c(u,v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$$

Improving Bandwidth choice

Consider the principal components decomposition of the $(n \times 2)$ matrix $[\hat{S}, \hat{T}] = M$.

Let $W_1 = (W_{11}, W_{12})^{\mathsf{T}}$ and $W_2 = (W_{21}, W_{22})^{\mathsf{T}}$ be the eigenvectors of $\boldsymbol{M}^{\mathsf{T}} \boldsymbol{M}$. Set

$$\begin{pmatrix} Q \\ R \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} S \\ T \end{pmatrix} = \mathbf{W} \begin{pmatrix} S \\ T \end{pmatrix}$$

which is only a linear reparametrization of \mathbb{R}^2 , so an estimate of f_{ST} can be readily obtained from an estimate of the density of (Q, R)Since $\{\hat{Q}_i\}$ and $\{\hat{R}_i\}$ are empirically uncorrelated, consider a diagonal bandwidth matrix $\mathbf{H}_{QR} = \mathbb{R}$ $\operatorname{diag}(h_Q^2, h_R^2)$.



Improving Bandwidth choice

Use univariate procedures to select h_Q and h_R independently

Denote $\tilde{f}_Q^{(p)}$ and $\tilde{f}_R^{(p)}$ (p = 1, 2), the local log-polynomial estimators for the densities

 h_Q can be selected via cross-validation (see Section 5.3.3 in Loader (1999))

$$h_Q = \arg\min_{h>0} \left\{ \int_{-\infty}^{\infty} \left\{ \tilde{f}_Q^{(p)}(q) \right\}^2 dq - \frac{2}{n} \sum_{i=1}^n \tilde{f}_Q^{(p)}(\hat{Q}_i) \right\},\$$

where $\tilde{f}_{Q(-i)}^{(p)}$ is the 'leave-one-out' version of $\tilde{f}_Q^{(p)}$.

Graphical Comparison (loss ALAE dataset)



Simulation Study

M = 1,000 independent random samples $\{(U_i, V_i)\}_{i=1}^n$ of sizes n = 200, n = 500and n = 1000 were generated from each of the following copulas:

- · the independence copula (i.e., U_i 's and V_i 's drawn independently);
- \cdot the Gaussian copula, with parameters $\rho=0.31,\,\rho=0.59$ and $\rho=0.81;$
- · the Student t-copula with 4 degrees of freedom, with parameters $\rho = 0.31$, $\rho = 0.59$ and $\rho = 0.81$;
- · the Frank copula, with parameter $\theta = 1.86$, $\theta = 4.16$ and $\theta = 7.93$;
- · the Gumbel copula, with parameter $\theta = 1.25$, $\theta = 1.67$ and $\theta = 2.5$;
- · the Clayton copula, with parameter $\theta = 0.5$, $\theta = 1.67$ and $\theta = 2.5$.

(approximated) MISE relative to the MISE of the mirror-reflection estimator (last column), n = 1000. Bold values show the minimum MISE for the corresponding copula (non-significantly different values are highlighted as well).

n = 1000	$\hat{c}^{(\tau)}$	$\hat{c}^{(au \mathrm{am})}$	$ ilde{c}^{(au,1)}$	$\tilde{c}^{(au,2)}$	$\hat{c}_1^{(eta)}$	$\hat{c}_2^{(eta)}$	$\hat{c}_1^{(B)}$	$\hat{c}_2^{(B)}$	$\hat{c}_1^{(p)}$	$\hat{c}_2^{(p)}$	$\dot{\epsilon}_3^{(p)}$
Indep	3.57	2.80	2.89	1.40	7.96	11.65	1.69	3.43	1.62	0.50	0 14
Gauss2	2.03	1.52	1.60	0.76	4.63	6.06	1.10	1.82	0.98	0.66	0.89
Gauss4	0.63	0.49	0.44	0.21	1.72	1.60	0.75	0.58	0.62	0.99	2.93
Gauss6	0.21	0.20	0.11	0.05	0.74	0.33	0.77	0.37	0.72	1.21	2.83
$\operatorname{Std}(4)2$	0.61	0.56	0.50	0.40	1.57	1.80	0.78	0.67	0.75	1.01	1.88
$\operatorname{Std}(4)4$	0.21	0.27	0.17	0.15	0.88	0.51	0.75	0.42	0.75	1.12	2.07
$\operatorname{Std}(4)6$	0.09	0.17	0.08	0.09	0.70	0.19	0.82	0.47	0.90	1.17	1.90
Frank2	3.31	2.42	2.57	1.35	7.16	9.63	1.70	2.95	1.31	0.45	0 49
Frank4	2.35	1.45	1.51	0.99	4.42	4.89	1.49	1.65	0.60	0.72	6.14
Frank6	0.96	0.52	0.45	0.44	1.51	1.19	1.35	0.76	0.65	1.58	7.25
Gumbel2	0.65	0.62	0.56	0.43	1.77	1.97	0.82	0.75	0.83	1.03	1.52
Gumbel4	0.18	0.28	0.16	0.19	0.89	0.41	0.78	0.47	0.81	1.10	1.78
Gumbel6	0.09	0.21	0.10	0.15	0.78	0.29	0.85	0.58	0.94	1.12	1.63
Clayton2	0.63	0.60	0.51	0.34	1.78	1.99	0.78	0.70	0.79	1.04	1.79
Clayton4	0.11	0.26	0.10	0.15	0.79	0.27	0.83	0.56	0.90	1.10	1.50
Clayton6	0.11	0.28	0.08	0.15	0.82	0.35	0.88	0.67	0.96	1.09	1.36