# Probit transformation for nonparametric kernel estimation of the copula density

Arthur Charpentier [1], Gery Geenens [2] & Davy Paindaveine [3]

[1] *Département de mathématiques, Université du Québec à Montréal, Canada, charpentier.arthur@uqam.ca*
[2] *School of Mathematics and Statistics, University of New South Wales, Sydney, Australia, ggeenens@unsw.edu.au*
[3] *ECARES, Université Libre de Bruxelles, Belgique, dpaindav@ulb.ac.be*

**Résumé.** La modélisation de problèmes bivariés à l'aide de copules est devenu classique dans les statistiques modernes. Ici, nous nous intéresserons au problme de l'estimation non paramtrique de la densité d'une copule. Sans doute le plus populaire, l'estimateur à noyau n'est toutefois pas adapt pour les densits de copules dont le support est le carré unité, principalement parce qu'il est fortement touche par des problèmes de biais aux bords. En outre, les copules les plus courantes admettent des densités non bornes, et les mthodes à noyau ne sont pas pertinentes dans ce cas. Dans cet article, un estimateur de densit de la copule de type noyau est propos. Il est bas sur l'ide de transformer les marges (uniformes) de la densit de la copule en des distributions normales par une transformation de type probit. L'estimation de la densit dans le domaine transform, qui peut tre accompli sans problme aux limites, et on obtient un estimateur de la densité de la copule par transformation inverse. Plusieurs estimateurs construits autour de cette idée seront proposés, et les proprits asymptotiques de ces estimateurs sont donnés, ainsi qu'un moyen pratique de slection du paramètre de lissage.

**Mots-clés.** estimation de densité par noyau transformé; biais de bord; estimation de densité par vraisemblance locale; estimation de densité par log-polynômes locaux

**Abstract.** Copula modelling has become ubiquitous in modern statistics. Here, the problem of nonparametrically estimating a copula density is addressed. Arguably the most popular nonparametric density estimator, the kernel estimator is not suitable for the unit-square-supported copula densities, mainly because it is heavily affected by boundary bias issues. In addition, most common copulas admit unbounded densities, and kernel methods are not consistent in that case. In this paper, a kernel-type copula density estimator is proposed. It is based on the idea of transforming the uniform marginals of the copula density into normal distributions via the probit function, estimating the density in the transformed domain, which can be accomplished without boundary problems, and obtaining an estimate of the copula density through back-transformation. Although natural, a raw application of this procedure was, however, seen not to perform very well in the earlier literature. Here, it is shown that, if combined with local likelihood density estimation methods, the idea yields very good and easy to implement estimators, fixing boundary issues in a natural way and able to cope with unbounded copula densities.

**Keywords.** transformation kernel density estimator; boundary bias; local likelihood density estimation; local log-polynomial density estimation.

# 1 Notations et Définitions

On dispose d'un échantillon $\{(X_i, Y_i); i = 1, \ldots, n\}$ tiré suivant $F_{XY}$, avec

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)), \tag{1}$$

On suppose ici que les marges sont continues, de telle sorte que la copule $C$ est unique. On suppose de plus que la copule est à densité, et on pose

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v)$$

En posant $\{(U_i = F_X(X_i), V_i = F_Y(Y_i)); i = 1, \ldots, n\}$, l'estimateur à noyau de $c$, noté $\hat{c}^*$, est

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

où $K : R^2 \to R$ est un noyau bivarié et $H_{UV}$ une matrice symmétrique définie positive de fenêtre.

En pratique toutefois, comme les marges ne sont pas connues, il convient de travailler avec des 'pseudo-observations'

$$\hat{U}_i = \frac{n}{n+1} \hat{F}_{Xn}(X_i) \qquad et \qquad \hat{V}_i = \frac{n}{n+1} \hat{F}_{Yn}(Y_i) \tag{3}$$

où $\hat{F}_{Xn}(x) = \frac{1}{n} \sum_{j=1}^{n} 1\!\!1_{\{X_j \leq x\}}$ est la fonction de répartition empirique de $X$, comme discuté dans Charpentier *et al* (2007).

# 2 La Transformation Probit

Posons ici

$$S = \Phi^{-1}(U) \qquad and \qquad T = \Phi^{-1}(V),$$

où $\Phi$ est la fonction de répartition de la loi normale centrée réduite, et $\Phi^{-1}$ sa fonction quantile (i.e. la transformation probit). L'idée est alors de considérer des estimateurs de la forme

$$\hat{c}^{(\tau)}(u, v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \tag{4}$$

où l'exposant $(\tau)$ rappelle l'idée de la transformation.

L'estimateur naturel (dit naif) est construit via

$$\hat{f}^*_{ST}(s, t) = \frac{1}{n|H_{ST}|^{1/2}} \sum_{i=1}^{n} K\left(H_{ST}^{-1/2} s - S_i t - T_i\right) \tag{5}$$

en posant

$$\hat{c}^{(\tau)}(u,v) = \frac{1}{n|H_{ST}|^{1/2}\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}\sum_{i=1}^{n} K\left(H_{ST}^{-1/2}\Phi^{-1}(u) - \Phi^{-1}(\hat{U}_i)\Phi^{-1}(v) - \Phi^{-1}(\hat{V}_i)\right). \tag{6}$$

Moyennant quelques hypothèses techniques, on peut montrer que

$$\sqrt{nh^2}\left(\hat{c}^{(\tau)}(u,v) - c(u,v) - h^2 b(u,v)\right) \mathcal{L}\!\longrightarrow \mathcal{N}\left(0, \sigma^2(u,v)\right), \tag{7}$$

où

$$b(u,v) = \frac{1}{2}\left\{\frac{\partial^2 c}{\partial u^2}(u,v)\phi^2(\Phi^{-1}(u)) + \frac{\partial^2 c}{\partial v^2}(u,v)\phi^2(\Phi^{-1}(v))\right. \tag{8}$$

$$-3\left(\frac{\partial c}{\partial u}(u,v)\Phi^{-1}(u)\phi(\Phi^{-1}(u)) + \frac{\partial c}{\partial v}(u,v)\Phi^{-1}(v)\phi(\Phi^{-1}(v))\right) \tag{9}$$

$$\left. + c(u,v)\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2\right)\right\} \tag{10}$$

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

Cet estimateur peut être amélioré en supposant que $f_{ST}$ est log-linéaire

$$\log f_{ST}(\check{s},\check{t}) \simeq a_{1,0}(s,t) + a_{1,1}(s,t)(\check{s}-s) + a_{1,2}(s,t)(\check{t}-t) \doteq P_{\mathbf{a}_1}(\check{s}-s,\check{t}-t) \tag{11}$$

ou log-quadratique

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

où les vecteurs $\mathbf{a}_1(s,t) = (a_{1,0}(s,t), a_{1,1}(s,t), a_{1,2}(s,t))$ and $\mathbf{a}_2(s,t) \doteq (a_{2,0}(s,t),\ldots,a_{2,5}(s,t))$ sont obtenus en résolvant, pour $p = 1, 2$,

$$\tilde{\mathbf{a}}_p(s,t) = \arg\max_{\mathbf{a}_p}\left\{\sum_{i=1}^{n} K\left(H_{ST}^{-1/2}s - \hat{S}_i t - \hat{T}_i\right) P_{\mathbf{a}_p}(\hat{S}_i - s, \hat{T}_i - t)\right. \tag{12}$$

$$\left. - n_{R^2} K\left(H_{ST}^{-1/2}s - \check{s}t - \check{t}\right)\exp\left(P_{\mathbf{a}_p}(\check{s}-s,\check{t}-t)\right)\, d\check{s}\, d\check{t}\right\}, \tag{13}$$

En posant $\tilde{f}_{ST}^{(1)}(s,t) = \exp(\tilde{a}_{1,0}(s,t))$ dans le cas log-linéaire et $\tilde{f}_{ST}^{(2)}(s,t) = \exp(\tilde{a}_{2,0}(s,t))$ dans le cas log-quadratique, on pose

$$\tilde{c}^{(\tau,p)}(u,v) = \frac{\tilde{f}_{ST}^{(p)}(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \tag{14}$$

pour $p = 1$ et $p = 2$. La normalité asymptotique de cet estimateur peut être obtenue.

Enfin, une dernière amélioration est suggérée, inspirée de Geenens (2014), en posant

$$\hat{c}^{(\tau am)}(u,v) = \frac{\hat{f}_{ST}(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \times \frac{1}{1 + \frac{1}{2}h^2\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2 - 2\right)}. \tag{15}$$

# Bibliographie

[1] Charpentier, A., Fermanian, J.-D. and Scaillet, O. (2007), The estimation of copulas: theory and practice, 38 in: J. Rank (Ed.), Copulas: From Theory to Application in Finance, Risk Publications, London, pp. 35-60.
[2] Devroye, L. and Györfi, L., Nonparametric Density Estimation: the L1 View, Wiley, 1985.
[3] Geenens, G. (2014), Probit transformation for kernel density estimation on the unit interval, Journal of the American Statistical Association, in press.