Les copules: introduction et estimation

M2MO Paris 7, le 18/12/2012

Xavier Milhaud

Co-responsable de l'Actuariat à l'ENSAE
xavier.milhaud@ensae.fr

Note: la reproduction de ce cours nécessite l'accord préalable de son auteur.

Plan du cours

1. Introduction
2. Estimation
3. Outils de simulation

Pourquoi les copules pour mesurer la dépendance?

Le + répandu pour mesurer la corrélation: coefficient Pearson

$$\rho_{X,Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$

où $\mu_X = E[X]$ et σ_X est l'écart-type de X.
Ce coefficient mesure la corrélation LINEAIRE. Oui car...

Exemple : considérons la v.a. X telle que $X \sim \mathcal{N}(0,1)$. Ainsi $\mu_X = 0$, et $\mu_{X^3} = 0$. Notons $Y = X^2$, on a

$$\begin{aligned}\rho_{X,Y} &= \frac{E[(X-\mu_X)(X^2-\mu_{X^2})]}{\sigma_X \sigma_{X^2}} \\ &= \frac{\mu_{X^3} - \mu_X \mu_{X^2}}{\sigma_X \sigma_{X^2}} = 0.\end{aligned}$$

Corrélation nulle alors que X et X^2 parfaitement corrélées!

Un peu d'Histoire...

$\rightarrow$ Outil récent né dans les années 1950, avec un fort développement depuis les années 1980;

$\rightarrow$ Modélisation de la dépendance entre les risques : définit une structure de dépendance entre des marginales (FdR de lois univariées);

$\rightarrow$ Principaux chercheurs sur le sujet : C. Genest (Université Laval, Québec, Canada) & al., Sklaar;

$\rightarrow$ Auteurs pour les applications...

- en Assurance : Frees et Valdez (1998),
- en Finance : Bouyé et al. (2000), Cherubini et al. (2004) et McNeil (2005).

Domaines d'application

Les copules sont généralement majoritairement utilisées dans les deux domaines que sont...

- la Finance:
  - risque de crédit pour la modélisation de la dépendance du risque de défaut,
  - ou pour les rendements de titres financiers;
- l'Assurance:
  - Vie : contrat décès sur 2 têtes,
  - Non-vie (IARD) :

    → assurance catastrophe naturelle;

    → réassurance.

Idée de base

On peut définir la fonction de répartition conjointe de X et Y de la manière suivante:

$$H_{(X,Y)}(x,y) = C_H(F_X(x), G_Y(y)),$$

où C est défini comme étant une copule.

→ Ceci permet d'introduire une dépendance entre X et Y, pas forcément au sens linéaire du terme...

→ Il y a distinction entre la définition de la relation de dépendance et celle des marginales.

Plaçons nous en dimension 2 pour simplifier

→ Soit 2 variables aléatoires X et Y de FdR F et G.
Une copule est une FdR de la loi jointe bivariée qui lie les marginales de X et Y (cf théorème de Sklaar).

→ Théorème fondamental de représentation de Sklaar:
Si H la FdR du couple (X,Y) et F et G les marginales des lois univariées X et Y (respectivement), alors il existe une copule C_H (unique si X, Y continues) telle que :

$$\forall (x,y) \in \mathbb{R}^2, H_{(X,Y)}(x,y) = C_H(F(x), G(y)).$$

→ En pratique, une copule C_H peut être vue comme une application de $[0,1]^2$ dans $[0,1]$ à partir de lois uniformes:

$$C_H(u,v) = P(F(x) \leq u, G(y) \leq v)$$

Démonstration

En effet, en posant $u = F(x)$ et $v = G(y)$ on a bien

$$(u, v) \in [0, 1] \times [0, 1];$$

Et $\forall (x, y) \in \mathbb{R}^2$;

$$\begin{aligned} C(F(x), G(y)) &= H_{(X,Y)}(x, y) \\ C(u, v) &= P(X \leq x, Y \leq y) \\ &= P(X \leq F^{-1}(u), Y \leq G^{-1}(v)) \\ &= P(F(X) \leq u, G(Y) \leq v) \end{aligned}$$

Attention à la confusion:

$$P(F(x) \leq u, G(y) \leq v) \neq P(X \leq u, Y \leq v)!$$

Définition et propriétés

Définition: une copule C est la fonction de répartition d'un vecteur de variables aléatoires $U = (U_1, U_2)$ dont toutes les composantes U_i obéissent à la loi uniforme $\mathcal{U}(0,1)$.

Remarque: en effet, si $X \sim F_X$ alors $F_X(X) = U \sim \mathcal{U}(0,1)$ (car $F_X(X) = F_X(F_X^{-1}(U)) = U$).

Propriétés: on a donc $C : [0,1]^2 \to [0,1]$, telle que:

1. $C(U_1, U_2)$ non decroissante sur $[0,1]^2$,
2. $C(U_1, U_2)$ continue a droite sur $[0,1]^2$,
3. $\lim\limits_{u_i \to 0} C(U_1, U_2) = 0$,
4. $\lim\limits_{u_i \to 1} C(U_1, U_2) = u_j$,
5. $\Delta_{a_1,b_1}\Delta_{a_2,b_2} C(U_1, U_2) = C(b_1, b_2) - C(b_1, a_2) - C(a_1, b_2) + C(a_1, a_2) \geq 0, \quad \forall a_1 \leq b_1, a_2 \leq b_2.$

Utilisation pratique des copules

- définir la structure de dépendance entre lois de proba.,
- mesurer la dépendance entre des v.a. données,
- construire des familles de lois bivariées.

Avantages:

- permet d'étendre les modèles à chocs communs,
- tous les résultats sont généralisables en dimension d.

Intérêts:

- décomposition lors de la définition des lois multivariées:
  - les marginales sont définies indépendamment les unes des autres, pas nécessairement de même type,
  - la structure de dépendance introduite via la copule ne dépend pas du choix de ces marges...
- invariance par transformation monotone.

Exemples de copules bivariées

- Copule d'indépendance:

$$C^I(u_1, u_2) = u_1 u_2,$$

- Copule de Farlie-Gumbel-Morgenstern ($\alpha \in [-1, 1]$):

$$C_\alpha(u_1, u_2) = u_1 u_2 + \alpha u_1 u_2 (1 - u_1)(1 - u_2),$$

- Copule de la borne supérieure de Fréchet:

$$C^+(u_1, u_2) = min(u_1, u_2),$$

- Copule de la borne inférieure de Fréchet:

$$C^-(u_1, u_2) = max(u_1 + u_2 - 1, 0),$$

- Copule de Fréchet (α et $\beta \in [0, 1]$):

$$C_{\alpha,\beta}(u_1, u_2) = \alpha C^-(u_1, u_2) + \beta C^+(u_1, u_2) + (1 - \alpha - \beta) C^I(u_1, u_2).$$

Exemple de familles de copules

Il existe aussi des familles de copules! Elles regroupent (comme on peut s'en douter) différentes copules...

- Copule de Franck ($\alpha \neq 0$):

$$C_\alpha^F(u_1, u_2) = -\frac{1}{\alpha}\ln(1 + \frac{(e^{-\alpha u_1 - 1})(e^{-\alpha u_2 - 1})}{e^{-\alpha - 1}}),$$

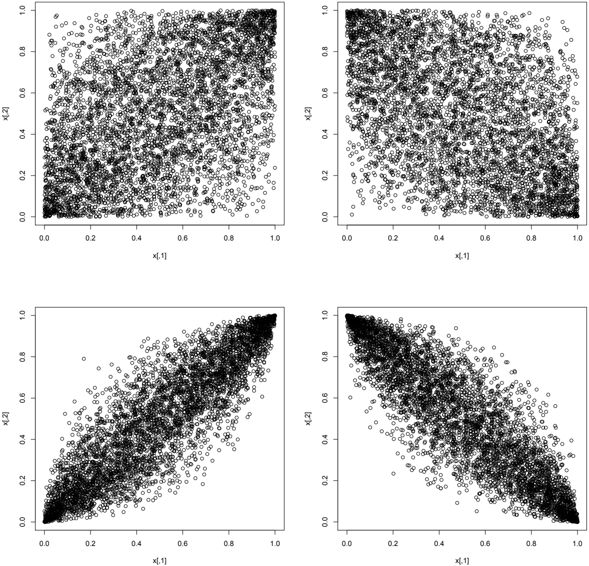
- Copule de Gumbel ($\alpha \geq 1$):

$$C_\alpha^G(u_1, u_2) = \exp(-[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{(1/\alpha)}),$$

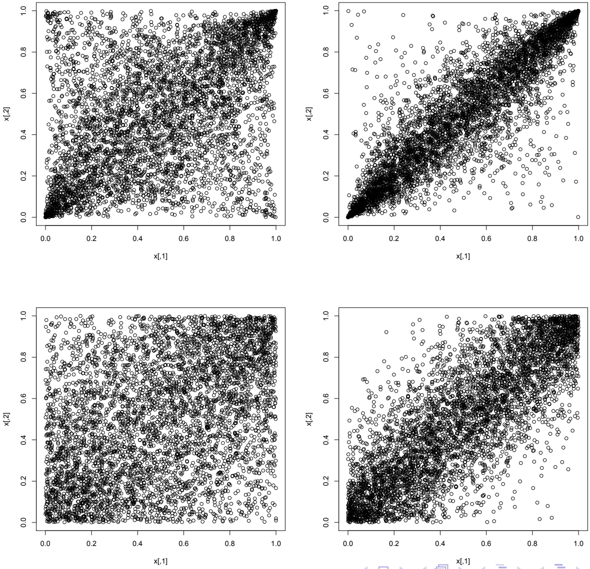
- Copule de Clayton ($\alpha > 0$):

$$C_\alpha^C(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}.$$

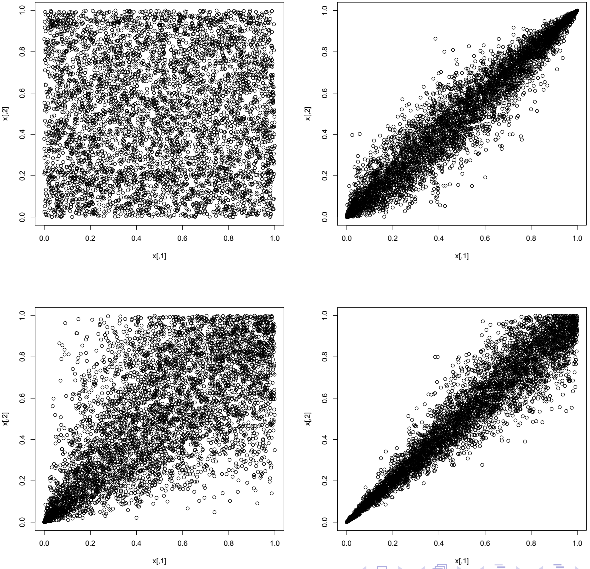
Copule normale ($\alpha = 0.5, -0.5, 0.9, -0.9$)



Copules de Student (2 prem.) et Franck ($\alpha = 2.5, 8$)



Copules de Gumbel ($\alpha = 1,6$) et Clayton ($\alpha = 2,10$)



Construction de copule

Pour construire une copule, il suffit de se servir du corollaire:
“Soit H une FdR bivariée avec des marginales continues F, G, et la copule C. Alors, ”

$$\forall (u,v) \in [0,1]^2, \quad C_H(u,v) = H_{(X,Y)}(F^{-1}(u), G^{-1}(v)).$$

Si on connait la distribution conjointe de (X,Y) ainsi que leur marginale, le tour est joué!

Exemple: copule de Galambos: $\forall \theta \geq 1$, on a

$$H(x,y) = \exp(-[(x+y) - (x^{-\theta} + y^{-\theta})^{-1/\theta}])$$

admet pour marges $F(x) = G(x) = \exp(-x), \forall x \in [0, +\infty[$.

Soit donc $\forall u \in [0,1], F^{-1}(u) = G^{-1}(u) = -\ln(u)$, d'où:

$$C_\theta(u,v) = uv \exp(-[(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{-1/\theta})$$

1 Introduction

2 Estimation
- Inférence
- Non-paramétrique
- Semi-paramétrique
- Paramétrique
- Corrélation de rang
- Méthode des moments

3 Outils de simulation

Des données aux copules, l'inférence statistique

Estimation des marges:

Concrètement on a les observations

$$(x_1, y_1), (x_2, y_2), , (x_n, y_n)$$

équivalent à $H_{(X,Y)}(x, y) = C_H(F(x), G(y))$.

Question : comment choisir $C_H(u, v)$ (distribution de (U,V)) sachant que (U,V) = (F(x),G(y)) ?

(F et G ne sont pas évidemment pas connues dans la réalité)

Solution

Pour obtenir les observations

$$(u_1, v_1), (u_2, v_2), ..., (u_n, v_n),$$

on estime les marges par la fonction de répartition empirique des lois X et Y, i.e. :

$$\begin{aligned} F_n(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq x} \\ G_n(y) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{Y_i \leq y} \end{aligned}$$

Puis on pose

$$(U_i = F_n(X_i), V_i = G_n(Y_i))$$

.

Point technique important

$\rightarrow$ En fait cela revient à poser $U_i = \frac{R_i}{n}$ et $V_i = \frac{S_i}{n}$,

avec R_i et S_i qui sont les rangs des observations de X et Y.

$\rightarrow$ Propriétés:

i) Si X est v.a.c. de FdR F alors la variable $Y = F(X)$ est de loi uniforme $\mathcal{U}(0,1)$,

ii) Si F est une FdR continue d'inverse généralisée F^{-1} et X une v.a.c. telle que $X \sim \mathcal{U}(0,1)$, alors $F^{-1}(X) \sim F$!

$\rightarrow$ Csq: chacune des paires est bien de loi C par construction! Si $(X,Y) \sim H$ de marge F, G alors $H(F(X), G(Y)) \sim C$ (i);

$\rightarrow$ Réciproque:

Si $(U,V) \sim C$ alors $H(F^{-1}(U), G^{-1}(V)) \sim H$ $\quad$ (ii).

1 Introduction

2 Estimation

- Inférence
- Non-paramétrique
  - Copule empirique
- Semi-paramétrique
- Paramétrique
- Corrélation de rang
- Méthode des moments

3 Outils de simulation

Définition

$\rightarrow$ La copule empirique (Deheuvels, 1979) est définie de la même manière qu'une fonction de répartition empirique par:

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(F(X_i),G(Y_i))\leq(u,v)}.$$

$\rightarrow$ Résultats asymptotiques: $\sqrt{n}(C_n - C) \rightarrow \mathcal{N}(0,\sigma^2)$.

$\rightarrow$ Support de la copule: graphe des paires $\left(\frac{R_i}{n}, \frac{S_i}{n}\right)$;

$\rightarrow$ n ne doit être...

- *ni trop petit*: perte de la forme précise de la FdR,
- *ni trop grand*: temps de calcul trop important.

$\rightarrow$ Rq: bon outil graphique, insuffisant pour choisir la famille de copules adaptée à la structure de dépendance des données.

1 Introduction

2 Estimation
- Inférence
- Non-paramétrique
- Semi-paramétrique
  - CML
- Paramétrique
- Corrélation de rang
- Méthode des moments

3 Outils de simulation

Canonical Maximum Likelihood (CML)

$\rightarrow$ Cette méthode semi-paramétrique a été proposée par Oakes (1994), approfondie par Genest et al. (1995);

$\rightarrow$ Extension au traitement de données censurées par Shih et Louis (1995);

$\rightarrow$ Idée sous-jacente de la méthode: estimation non paramétrique des marges par la FdR empirique, puis estimation paramétrique du paramètre de copule.

$\rightarrow$ La CML est semblable à l'IFM (cf supra) mais pas besoin de spécifier les lois des marginales;

$\rightarrow$ Inconvénient majeur: perte d'information due à l'estimation non paramétrique des marges!

Mise en oeuvre

$\rightarrow$ Marginales remplacées par FdR empirique, puis le param. de la copule estimé par MV. En pratique,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x},$$

et θ est donc estimé en maximisant la log-vraisemblance:

$$\hat{\theta} = \underset{\theta \in \Theta}{argmax} \sum_{i=1}^{n} ln(c_\theta(F_n(x_i), G_n(y_i))).$$

où $c_\theta(u, v)$ est la densité de C.

$\rightarrow$ Facilité d'implémentation, invariance par transformation croissante des marges:

$$\hat{U}_i = F_n(X_i) \text{ et } \hat{V}_i = G_n(Y_i)$$

1 Introduction

2 Estimation

- Inférence
- Non-paramétrique
- Semi-paramétrique
- Paramétrique
  - ML
  - IFM
- Corrélation de rang
- Méthode des moments

3 Outils de simulation

Données de l'approche

$\rightarrow$ On dispose de l'échantillon aléatoire $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ de loi

$$H_{(X,Y)}(x,y) = C(F(x), G(y)), \forall (x,y) \in \mathbb{R}^2.$$

$\rightarrow$ On suppose des modèles paramétriques pour F, G et C_θ, où C_θ est une famille paramétrique de copules.

$\rightarrow$ Q.: comment estimer le paramètre de dépendance θ?

Exemple: considérons $F \sim N(\mu, \sigma^2)$, $G \sim Gamma(\nu, \beta)$. Et prenons C appartient à la famille FGM ($\alpha \in [-1,1]$):

$$C_\alpha(u_1, u_2) = u_1 u_2 + \alpha u_1 u_2 (1 - u_1)(1 - u_2),$$

a) Maximum Likelihood

L'estimation par maximum de vraisemblance nécessite d'introduire les quantités $f_\lambda, g_\nu, c_\theta$ où

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

qui sont les densités associées à F, G et C.

$\rightarrow$ D'autre part, $H(x,y) = C(F(x), G(y))$ donc

$$h(x,y) = c_\theta(F_\lambda(x), G_\nu(y)) f_\lambda(x) g_\nu(y),$$

d'où la log-vraisemblance à maximiser

$$l(\lambda,\nu,\theta) = \sum_{i=1}^{n} ln(f_\lambda(x_i)) + \sum_{i=1}^{n} ln(g_\nu(y_i)) + \sum_{i=1}^{n} ln(c_\theta(F_\lambda(x_i), G_\nu(y_i)))$$

Propriétés du MLE

$\rightarrow$ Les propriétés asymptotiques de l'estimateur du maximum de vraisemblance sont bien connues:

- il est sans biais
- convergent
- asymptotiquement normal

$\rightarrow$ Le temps de calcul est important si le nombre de paramètre est grand...

$\rightarrow$ Grande sensibilité au choix des marginales F et G (attention à ne pas se tromper!)

$\rightarrow$ Rq: on estime tous les paramètres en même temps !

b) Inference Functions for Margins

→ Cette méthode utilise la propriété fondamentale des copules : séparation marginales / structure de dépendance;

→ Il s'agit d'une estimation par maximum de vraisemblance, mais en deux étapes:

1. estimation des paramètres des marginales;
2. estimation du paramètre de la copule en injectant les paramètres estimés des marginales dans l'expression de la vraisemblance de la copule;

→ *Avantage*: estimateur en deux temps asymptotiquement normal (Joe, 2005) sous certaines conditions;

→ *Inconvénient*: mauvais choix de marginales se propage dans l'estimation du paramètre de la copule...

Algorithme

1. Estimer les marges par (ML)

$$\hat{\lambda} = \arg\max_{\lambda \in \Lambda} \sum_{i=1}^{n} \ln(f_\lambda(x_i)) \qquad \hat{\nu} = \arg\max_{\nu} \sum_{i=1}^{n} ln(g_\nu(y_i))$$

2. Pour $i = 1, ..., n$, posons

$$\hat{U}_i = F_{\hat{\lambda}}(X_i) \text{ et } \hat{V}_i = G_{\hat{\nu}}(Y_i), \text{ où } \hat{\lambda}, \hat{\nu} \text{ sont les MLE.}$$

3. Maximiser $\sum_{i=1}^{n} \ln(c_\theta(\hat{U}_i, \hat{V}_i))$.

Même inconvénient:

- Propagation d'un mauvais choix de marginales dans l'estimation du paramètre de dépendance;

1 Introduction

2 Estimation
- Inférence
- Non-paramétrique
- Semi-paramétrique
- Paramétrique
- Corrélation de rang
- Méthode des moments

3 Outils de simulation

Mesures de dépendance non-linéaire

$\rightarrow$ Définitions: soient (X_1', X_2') indépendant de (X_1, X_2) et ayant les mêmes marginales que (X_1, X_2), on définit

- le rho de Spearman ($\rho \in [-1, 1]$):

$$\rho(X_1, X_2) = 3P((X_1 - X_1')(X_2 - X_2') > 0) - 3P((X_1 - X_1')(X_2 - X_2') < 0),$$

- le tau de Kendall ($\tau \in [-1, 1]$):

$$\tau(X_1, X_2) = P((X_1 - X_1')(X_2 - X_2') > 0) - P((X_1 - X_1')(X_2 - X_2') < 0).$$

$\rightarrow$ Intuition: les variables aléatoires évoluent-elles dans le même sens à chaque rang?

$\rightarrow$ Ou quelle est la proportion de paires (x_i, y_i) et (x_k, y_k) concordantes (i.e. $(x_i - x_k)(y_i - y_k) > 0$) et discordantes ? Sachant qu'il y a en tout $C_2^n = \frac{n(n-1)}{2}$ paires distinctes...

Relations entre Tau de Kendall et copule

La formule théorique liant ces deux grandeurs est :

$$\begin{aligned}\tau(X_1, X_2) &= 4\int_0^1\int_0^1 C(u_1, u_2)dC(u_1, u_2) - 1 \\ &= 4\mathbb{E}[C(U_1, U_2)] - 1.\end{aligned}$$

Ce qui donne:

- Copule de Clayton: $\tau(X_1, X_2) = \frac{\alpha}{\alpha+2}$,
- Copule Normale: $\tau(X_1, X_2) = \frac{2}{\pi}\arcsin(\alpha)$,
- Copule de Gumbel: $\tau(X_1, X_2) = \frac{\alpha-1}{\alpha}$,
- Copule de Fréchet: $\tau(X_1, X_2) = \frac{(\alpha-\beta)(\alpha+\beta+2)}{3}$,
- Copule FGM: $\tau(X_1, X_2) = \frac{\alpha}{9}$,
- Copule Marshall-Olkin: $\tau(X_1, X_2) = \frac{\alpha\beta}{\alpha-\alpha\beta+\beta}$.

Relations entre Rho de Spearman et copule

La formule théorique liant ces deux grandeurs est :

$$\begin{aligned}\rho(X_1, X_2) &= 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3 \\ &= 12 \int_0^1 \int_0^1 u_1\, u_2\, dC(u_1, u_2) - 3.\end{aligned}$$

Ce qui donne:

- Copule Normale: $\rho(X_1, X_2) = \frac{6}{\pi} \arcsin(\frac{1}{2}\alpha)$,
- Copule de Gumbel: $\rho(X_1, X_2) = \frac{\alpha - 1}{\alpha}$,
- Copule de Fréchet: $\rho(X_1, X_2) = \alpha - \beta$,
- Copule FGM: $\rho(X_1, X_2) = \frac{\alpha}{3}$,
- Copule Marshall-Olkin: $\rho(X_1, X_2) = \frac{3\alpha\beta}{2\alpha - \alpha\beta + 2\beta}$.

1 Introduction

2 Estimation
- Inférence
- Non-paramétrique
- Semi-paramétrique
- Paramétrique
- Corrélation de rang
- Méthode des moments

3 Outils de simulation

Notion de dépendance positive par quadrant

→ Interprétation de la PQD:
la probabilité que les variables prennent simultanément des valeurs élevées est plus grande si elles sont PQD que en cas d'indépendance.

→ Lien évident avec les mesures du tau de Kendall et du Rho de Spearman. Basée sur l'inversion de mesures de dépendance (tau de Kendall, rho de Spearman).

→ X_1 et X_2 positivement dépendant par quadrant (PQD) s'écrit:

$$(X_1, X_2) \text{ est PQD si } P(X_1 > x_1, X_2 > x_2) \geq \underbrace{P(X_1 > x_1)P(X_2 > x_2)}_{\text{cas d'independance}}.$$

Propriétés et mise en oeuvre de la méthode

→ Propriété: si la famille C_θ est ordonnée par PQD, i.e.

$$\theta_1 < \theta_2 \Rightarrow C_{\theta_1} < C_{\theta_2}$$

Alors $\tau = \phi(\theta)$ et $\rho = \gamma(\theta)$ sont des fonctions croissantes de θ (où τ et ρ sont les mesures de Kendall et Spearman).

→ Puisque ϕ et γ sont croissantes en le paramètre de la copule, et τ_n et ρ_n sont des mesures de dépendance non-paramétriques:

$$\theta_n = \phi^{-1}(\tau_n) \quad \text{ou} \quad \theta_n = \gamma^{-1}(\rho_n).$$

→ En pratique, on utilise très souvent le tau de Kendall car des expressions simples existent pour la plupart des copules.

→ Remarque: la plupart des familles de copules satisfont la propriété (PQD)!

Intervalle de confiance: cas du tau de Kendall

$\rightarrow$ Expression analytique de τ_n (différence entre proportion de couples concordants et proportion de couples discordants):

$$\tau_n = \frac{2}{n(n-1)} \sum_i \sum_{j, j \neq i} signe\left[(x_i - x_j)(y_i - y_j)\right],$$

$signe\left[(x_i - x_j)(y_i - y_j)\right] = 1$ si paires concordantes, -1 sinon.

Ou bien $\tau_n = \frac{4}{n(n-1)} \sum_{i<j} \mathbb{1}_{((X_i - X_j)(Y_i - Y_j) > 0)} - 1.$

$\rightarrow$ L'intervalle de confiance de l'estimation du paramètre de la copule est donné par le comportement asymptotique de τ_n;

$\rightarrow$ τ_n vérifie asymptotiquement $\sqrt{n}\,\frac{\tau_n - \tau}{4S} \sim \mathcal{N}(0, 1)$ où

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (W_i + \tilde{W}_i - 2\bar{W}_i)^2, \qquad \text{avec...}$$

Intervalle de confiance (2)

$$
\begin{aligned}
W_i &= \frac{1}{n} \#\{J \in \{1, ..., n\} : X_i \geq X_J, Y_i \geq Y_J\} \\
\tilde{W}_i &= \frac{1}{n} \#\{J \in \{1, ..., n\} : X_i \leq X_J, Y_i \leq Y_J\} \\
\bar{W}_i &= \frac{1}{n} \sum_{i=1}^{n} W_i
\end{aligned}
$$

→ Finalement le paramètre de copule peut être estimé grâce au théorème de Slutsky (plus couramment méthode *delta*);

→ Théorème de Slutsky: en notant ϕ' la dérivée de ϕ,

$$\tilde{\theta}_n \simeq N(\theta, \frac{1}{n}\{4S\phi^{-1'}(\tau_n)\}^2),$$

où $\tilde{\theta}$ est l'estimateur de θ par la méthode des moments.

Intervalle de confiance (3)

$\rightarrow$ Par conséquent, l'intervalle de confiance asymptotique de θ au seuil $100.(1-\alpha)\%$ est:

$$\left[\tilde{\theta}_n - z_{1-\alpha/2}\frac{1}{\sqrt{n}}4S|\phi^{'-1}(\tau_n)|\ ;\ \tilde{\theta}_n + z_{1-\alpha/2}\frac{1}{\sqrt{n}}4S|\phi^{'-1}(\tau_n)|\right]$$

avec z quantile de la gaussienne centrée réduite $\mathcal{N}(0,1)$.

$\rightarrow$ Remarque: la loi asymptotique du rho de Spearman reste gaussienne, mais la variance de la loi diffère de celle du tau de Kendall.

1 Introduction

2 Estimation

3 Outils de simulation
- Généralités
- Inversion FdR

De nombreux outils en ligne...

Il existe un large panel d'outil permettant de se familiariser avec l'utilisation de copules. Pour la simulation, il y a notamment:

→ les méninges: il vous suffit d'utiliser la technique d'inversion de la fonction de répartition empirique...

→ le logiciel libre R, et sa librairie copula:

- R website: http://www.r-project.org ;
- Package mirror: http://cran.r-project.org ;
- Aide ou liens utiles:
  - http://www.statmethods.net
  - http://forums.cirad.fr/logiciel-R/index.php

→ Mais aussi d'autres logiciels mathématiques et statistiques: Scilab, Matlab...

Méthode d'inversion de la FdR

Prenons un exemple: simulation de la copule Normale en dimension 2. La copule Normale s'écrit

$$C_\alpha^N(u_1, u_2) = H_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \quad \alpha \in [-1, 1],$$

pour $u_i \in [0, 1]$, i=1,2. Φ^{-1} est la FdR inverse d'une $\mathcal{N}(0, 1)$, et H_α est la FdR de la normale bivariée avec corrélation α.

Algorithme:

1. simulation d'un couple de variables aléatoires normale standard (Z_1, Z_2) avec un coefficient de corrélation α,
2. calcul de $U_i = \phi(Z_i), \quad i = 1, 2$ où ϕ est la FdR d'une loi normale standard,
3. déduction des X_i par $X_i = F_{X_i}^{-1}(U_i), \quad i = 1, 2.$

Justification de l'algo

$\rightarrow$ A partir du thérème de Sklaar, on a le corollaire suivant: Soit H une FdR bivariée avec des marginales continues F, G, et la copule C. Alors,

$$\forall (u,v) \in [0,1]^2, \quad C_H(u,v) = H_{(X,Y)}(F^{-1}(u), G^{-1}(v)).$$

$\rightarrow$ Ici $H = H_\alpha$, $F = G = \phi$, $Z_1 = F^{-1}(u)$ et $Z_2 = G^{-1}(v)$.

1. on simule la structure de dépendance H pour obtenir $H(Z_1, Z_2)$,
2. on calcule $U = F(Z_1)$ et $V = G(Z_2)$,
3. on revient à la déf. $X_1 = F^{-1}(U)$, $X_2 = G^{-1}(V)$.