

Inférence statistique des copules (deuxième partie)

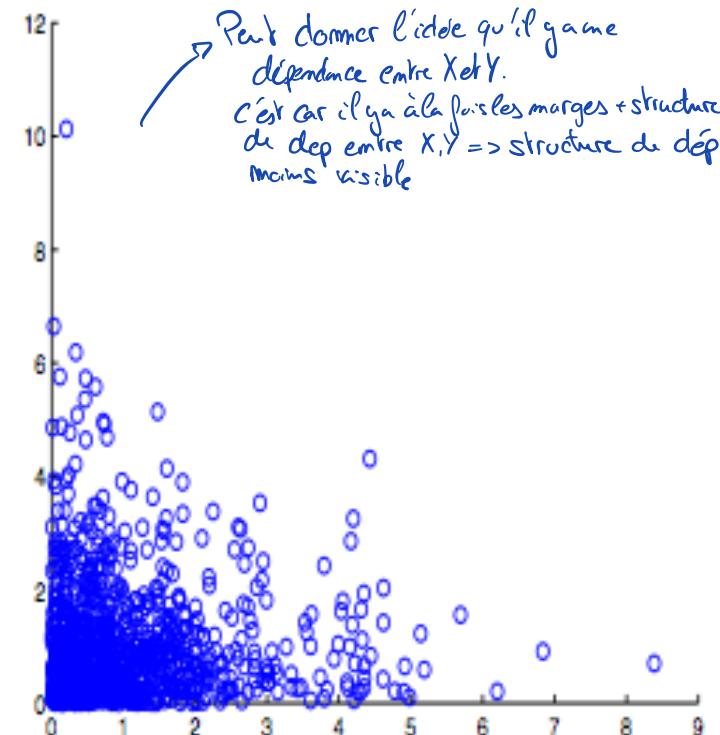
Esterina Masiello

M2 Actuariat 2019-2020

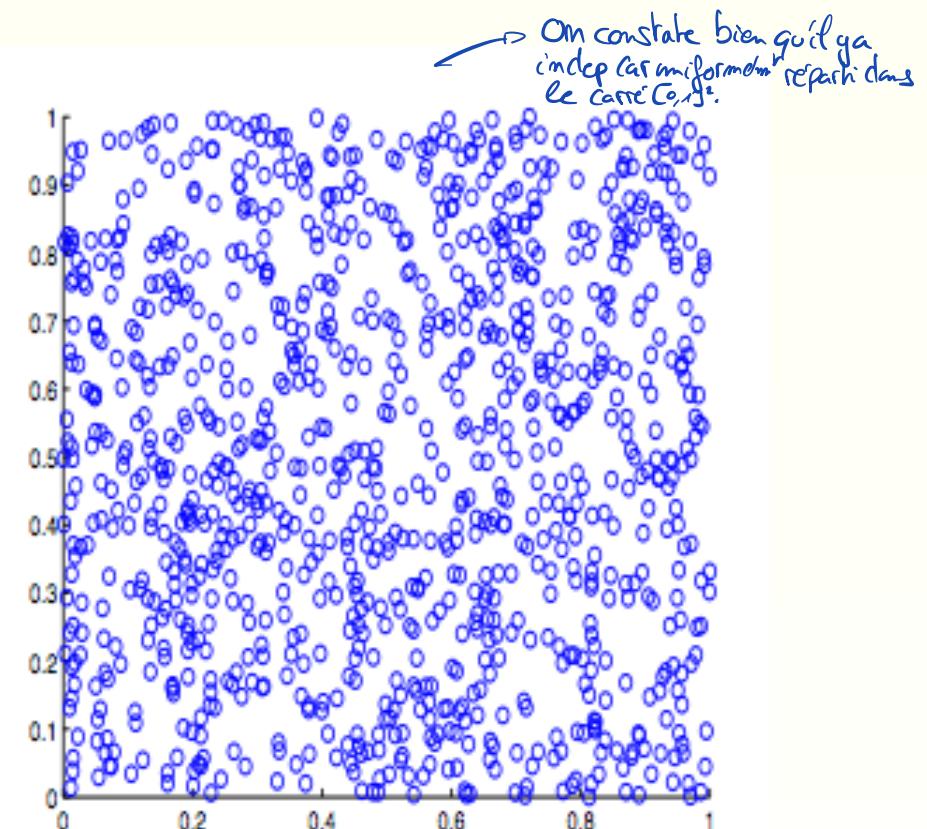
Outils de détection de la dépendance

- a) Diagramme de dispersion
 - b) Khi-plots
 - c) K-plots

a) Diagramme de dispersion



Peut donner l'idée qu'il y a une
dépendance entre X et Y.
C'est car il y a à la fois les marges + structure
de dép entre X,Y => structure du dép
moins visible



→ On constate bien qu'il ya
l'indépendance uniforme de la répartition dans
le Carré $C_{0,1}^2$.

CPL: si c'est structure de dép qui nous intéresse il faut faire rank-rank plot.

Diagramme de dispersion (gauche) et rank-rank plot (droite) pour 1000 paires de $Exp(1)$ indépendantes.

b) Khi-plots (Fisher et Switzer (1985, 2001))

Comment construire un khi-plot ?

Étant donné une paire (X_i, Y_i) avec $i \in \{1, \dots, n\}$, on pose

$$H_i = \frac{1}{n-1} \# \{j \neq i : X_j \leq X_i, Y_j \leq Y_i\}$$

$$F_i = \frac{1}{n-1} \#\{j \neq i : X_j \leq X_i\}$$

$$G_i = \frac{1}{n-1} \#\{j \neq i, Y_j \leq Y_i\}$$

Représentation croisée

	$X < X_i$	$X > X_i$	Total
$Y > Y_i$.	.	$1 - G_i$
$Y < Y_i$	H_i	.	G_i
	F_i	$1 - F_i$	1

Sous l'indépendance, on s'attend à ce que

$$H_i = F_i G_i, \quad i = 1, \dots, n$$

C'est cette propriété que l'on cherche à exploiter pour construire un khi-plot !

$H(x,y) = F(x) G(y)$ em cas d'indépendance entre X et Y .

b) Khi-plot (cont.)

On trace les paires de points (λ_i, χ_i) où

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}}$$

et

$$\lambda_i = 4 \operatorname{signe}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i^2, \tilde{G}_i^2)$$

avec

$$\tilde{F}_i = F_i - 1/2$$

$$\tilde{G}_i = G_i - 1/2$$

pour $1 \leq i \leq n$.

Interpretation de λ_i

On vérifie facilement que

$$\lambda_i = 4 \operatorname{signe}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i^2, \tilde{G}_i^2) \in [-1, 1]$$

est une mesure de la distance entre le point (X_i, Y_i) et le centre du nuage de points.

Pour éviter les valeurs aberrantes on ne trace que les paires pour lesquelles

$$|\lambda_i| < 4 \left(\frac{1}{n-1} - \frac{1}{2} \right)^2$$

Interpretation de χ_i

Le coefficient χ_i mesure la corrélation entre les $n - 1$ paires (X_{ij}, Y_{ij}) de variables binaires déduites de l'échantillon original en fixant à la fois X_i et Y_i , puis en posant

$$X_{ij} = \begin{cases} 1 & \text{si } X_j \leq X_i \\ 0 & \text{sinon,} \end{cases} \quad \text{et} \quad Y_{ij} = \begin{cases} 1 & \text{si } Y_j \leq Y_i \\ 0 & \text{sinon,} \end{cases}$$

pour tous $j \neq i$.

Ainsi, $\chi_i \in [-1, 1]$ pour $i \in \{1, \dots, n\}$.

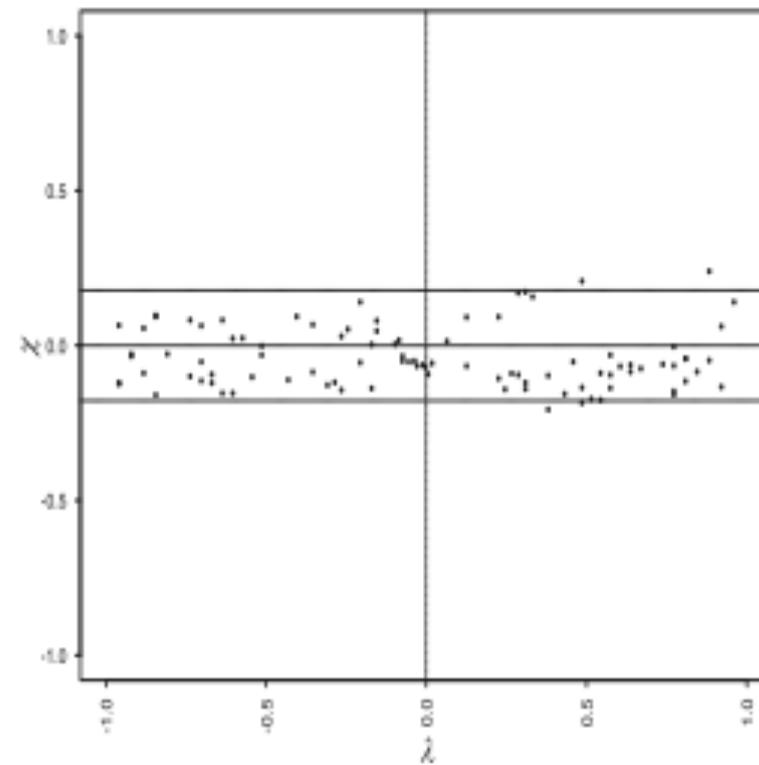
Propriété de χ_i

Remarquez aussi que

$$\sqrt{n}\chi_i = \sqrt{n} \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}}$$

est la racine carré de la statistique du χ^2 qui sert à tester l'indépendance dans des tables de fréquences.

b) Khi-plot (cont.)



Khi-plot pour 100 paires de variables indépendantes

c) K-plot ou Diagramme de Kendall (Genest et Boies, 2003)

- fondé sur les rangs
- inspiré par la notion de droite de Henry
- lié à la transformation intégrale de probabilité (TIP)

$H_0: X \sim N$ → Comme outils graphiques
 (histogramme, densité par noyaux, Q-Q plots)
 Dans Q-Q plots on représente les quantiles empiriques vs
 quantiles théoriques → Droite de Henry.
 Si points sur la droite → Indép.
 S'ils s'éloignent de la droite → Dép. positive/nég.

$$W = C(U, V) = H(X, Y)$$



Transformation intégrale de probabilité

En dimension $d = 1$, la TIP est

$$X \sim F(x) \longmapsto U = F(X) \sim U(0, 1)$$

Quand cette transformation est appliquée coordonnées par coordonnées, elle conduit au concept de copule :

$$(X, Y) \sim H(x, y) \longmapsto (U, V) = (F(X), G(Y))$$

et $(U, V) \sim C$ résume la structure de dépendance.

$X \sim F$ Fonction quantile: $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$

Soit une v.a. $U \sim U(0,1)$. Alors $F(X) \sim U(0,1)$

On peut écrire que $\forall u \in [0,1]$, $P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$

Généralisation de la TIP

Une autre généralisation possible de la TIP consisterait à prendre :

$$(X, Y) \sim H(x, y) \longmapsto W = H(X, Y)$$

Cette possibilité conduit à un résumé fonctionnel de la structure de dépendance

$$K(w) = P(W \leq w)$$

$$w \in (0, 1).$$

La loi de Kendall

La fonction de répartition

$$\xrightarrow{\text{Loi de Kendall}} K(w) = P\{\underbrace{H(X, Y)}_W \leq w\}, \quad w \in (0, 1)$$

est appelée la loi de Kendall, parce que

$$E(W) = E\{H(X, Y)\} = \frac{\tau + 1}{4}$$

où τ est le tau de Kendall.

K ne dépend pas des marges

C'est évident, puisque pour tout $w \in (0, 1)$

$$\begin{aligned} K(w) &= P\{H(X, Y) \leq w\} \\ \text{Par Sklar} \quad \rightarrow &= P\{C(F(X), G(Y)) \leq w\} \\ &= P\{C(U, V) \leq w\} \end{aligned}$$

En d'autres termes, K est aussi la loi de $W = C(U, V)$.

Espérance de W

Puisque $W = C(U, V)$, on trouve tout de suite

$$\begin{aligned} E(W) &= E\{C(U, V)\} \\ &= \int_0^1 \int_0^1 C(u, v) dC(u, v) \end{aligned}$$

et

$$\tau(U, V) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1$$

$$\rightarrow E[W] = \frac{\tau+1}{4}$$

Calcul de K

Il est particulièrement simple lorsque la copule $C = C_\phi$ est archimédienne.

ϕ est le générateur de la copule archimédienne.

Dans ce cas, on obtient (Genest et Mackay (1986, CJS))

$$K(w) = w - \frac{\phi(w)}{\phi'(w)}$$

$w \in (0, 1)$.

Une COPULE ARCHIMEDIENNE s'écrit: $C_\alpha(u, v) = \phi_\alpha^{-1}(\phi_\alpha(u) + \phi_\alpha(v))$ où α est le paramètre de dépendance de la copule et $\phi: [0, 1] \rightarrow [0, +\infty]$ est une fonction décroissante convexe tel que $\phi(1) = 0$, appelée GENERATEUR de la copule archimédienne.

K dans le cas de l'indépendance

Lorsque $C(u, v) = uv$, on trouve $W = UV$ et donc

$$P(W \leq w) = \int_0^1 \int_0^1 1_{\{UV \leq w\}} dv du$$

On obtient finalement

$$P(W \leq w) = w - w \log w$$

K en cas de dépendance parfaite

Lorsque $U = V$ presque partout, $C(u, v) = \min(u, v)$ et donc

$$W = \min(U, V) = \min(U, U) = U \sim U(0, 1)$$

Lorsque $U = 1 - V$, alors $C(u, v) = \max(0, u + v - 1)$ et donc

$$W = \max(0, U + V - 1) = \max(0, 0) = 0$$

d'où

$$K_M(w) = w \text{ et } K_W(t) = 1$$

$C(u, v) = \underline{\min(u, v)}$ en cas de dépendance positive parfaite (variable COMONOTONES)

$\hookrightarrow M(u, v)$: BORNE SUPÉRIEURE de FRECHET-Hoeffding

Lorsque X et Y sont ANTIMONOTONES (parfaitemenr corréles négativement) alors:

$C(u, v) = \underline{\max(0, u+v-1)}$

$W(u, v)$: BORNE INFÉRIEURE de FRECHET-Hoeffding ⚠ IP ne s'agit pas du W précédent

Méthode de construction d'un K-plot

- Calculer

$$H_i = \frac{1}{n-1} \# \{j \neq i : X_j \leq X_i, Y_j \leq Y_i\}$$

- les ordonner : $H_{(1)}, \dots, H_{(n)}$
- Tracer les paires $(W_{i:n}, H_{(i)})$, où pour $i \in \{1, \dots, n\}$, $W_{i:n}$ est l'espérance de la i -ème statistique d'ordre dans un échantillon aléatoire de taille n de K_0 .

Qu'est-ce que K_0 ?

K_0 est la f.d.r. de la variable

$$H = H(X, Y) = C\{F(X), G(Y)\} = C(U, V)$$

sous l'hypothèse d'indépendance, i.e.

$$K(w) = P(UV \leq w) = w - w \log w, \quad w \in (0, 1)$$

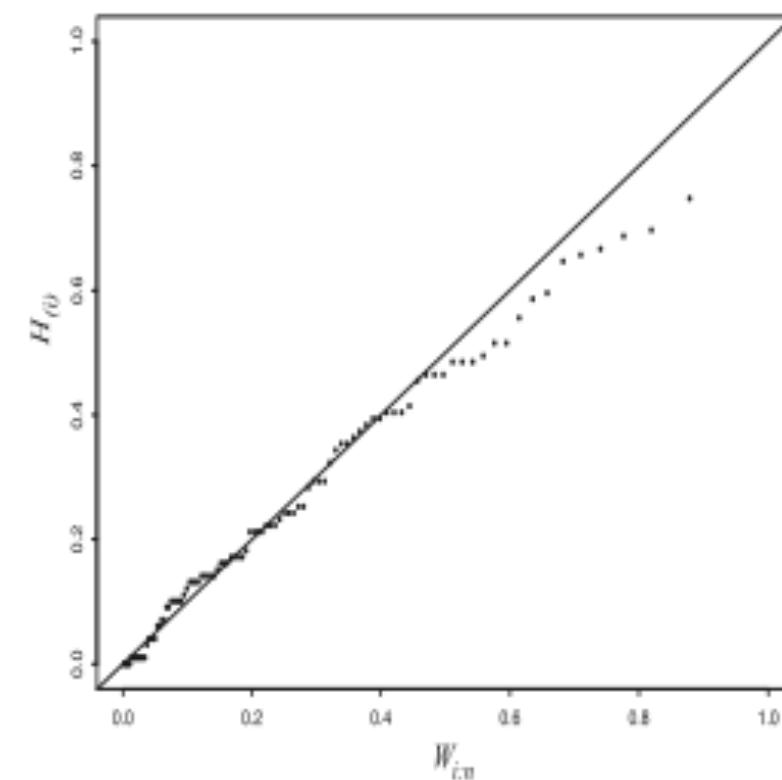
Calcul de $W_{i:n}$

En utilisant des résultats classiques sur les statistiques d'ordre, on obtient que $W_{i:n}$ est égal à

$$n \binom{n-1}{j-1} \int_0^1 w \{K_0(w)\}^{i-1} \{1 - K_0(w)\}^{n-i} dK_0(w)$$

pour tout $i \in \{1, \dots, n\}$

K-plot



K-plot de 100 paires indépendantes



Motivation pour les k-plots

Cette méthode fonctionne parce que K_n , la f.d.r. empirique de H_1, \dots, H_n

$$K_n(w) = \frac{1}{n} \sum_{i=1}^n 1_{\{H_i \leq w\}}, \quad w \in (0, 1)$$

est un estimateur asymptotiquement normal et convergent de

$$K(w) = P\{H(X, Y) \leq w\}$$

Propriétés des k-plots

Soit $\lfloor np \rfloor$ le plus petit entier supérieur ou égal à np . Alors

$$H_{(\lfloor np \rfloor)} = K_n^{-1}(p) \rightarrow K^{-1}(p)$$

et donc aussi

$$\lim_{n \rightarrow \infty} E(H_{(\lfloor np \rfloor)}) = \lim_{n \rightarrow \infty} W_{(\lfloor np \rfloor:n)} = K_0^{-1}(p)$$

sous l'hypothèse nulle d'indépendance.

Propriétés des k-plots

Quand n est assez grand, les paires $(W_{i:n}, H_{(i)})$ se concentrent le long de la courbe

$$p \longmapsto (K_0^{-1}(p), K^{-1}(p))$$

en d'autres mots, les points du k-plot se calent le long de la courbe

$$w \longmapsto K^{-1}\{K_0(w)\}$$

Le graphe tend à être linéaire quand $K = K_0$, comme c'est le cas sous l'hypothèse nulle.

Propriétés des k-plots

Tous les points du graphe se retrouvent sur l'abscisse ($p = 0$) quand les variables X et Y sont comonotones avec $\tau(X, Y) = -1$ parce qu'alors

$$K^{-1}(p) = 0$$

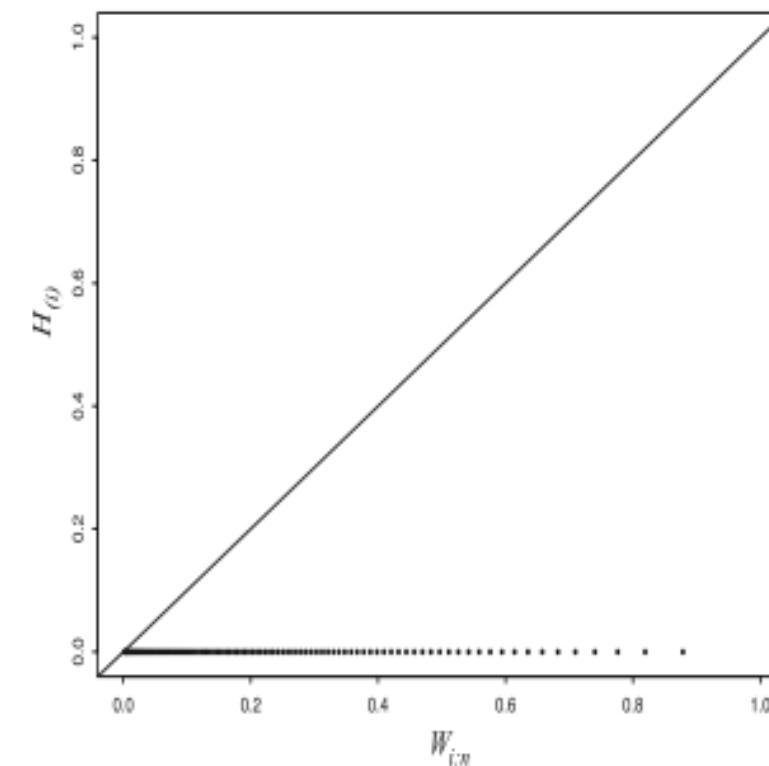
pour toute valeur possible de $p \in [0, 1]$.

Tous les points tombent sur la courbe $K_0(p)$ quand X et Y sont comonotones avec $\tau(X, Y) = 1$, car alors

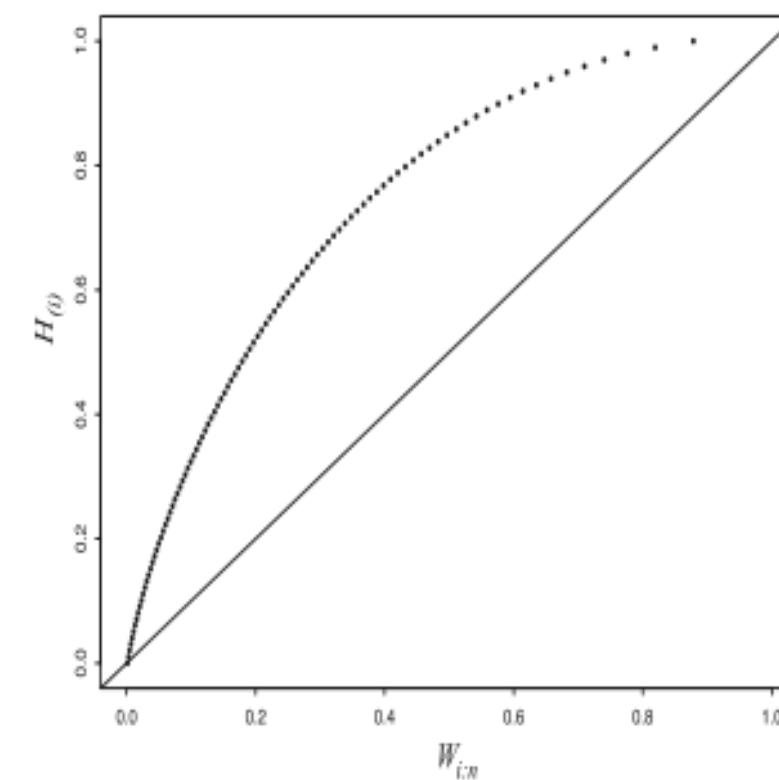
$$K^{-1}(p) = p$$

pour tout $p \in [0, 1]$.

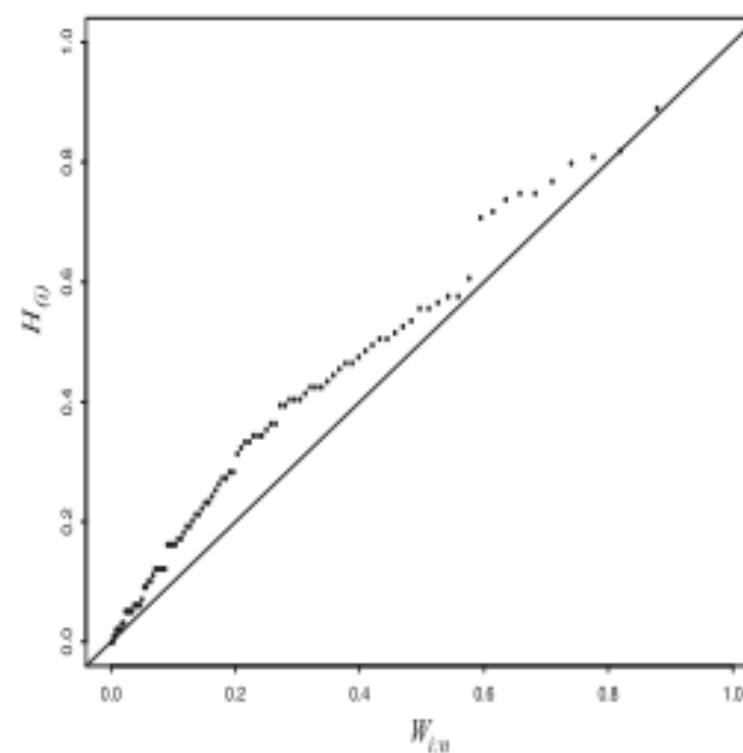
Borne inférieure de Fréchet-Hoeffding



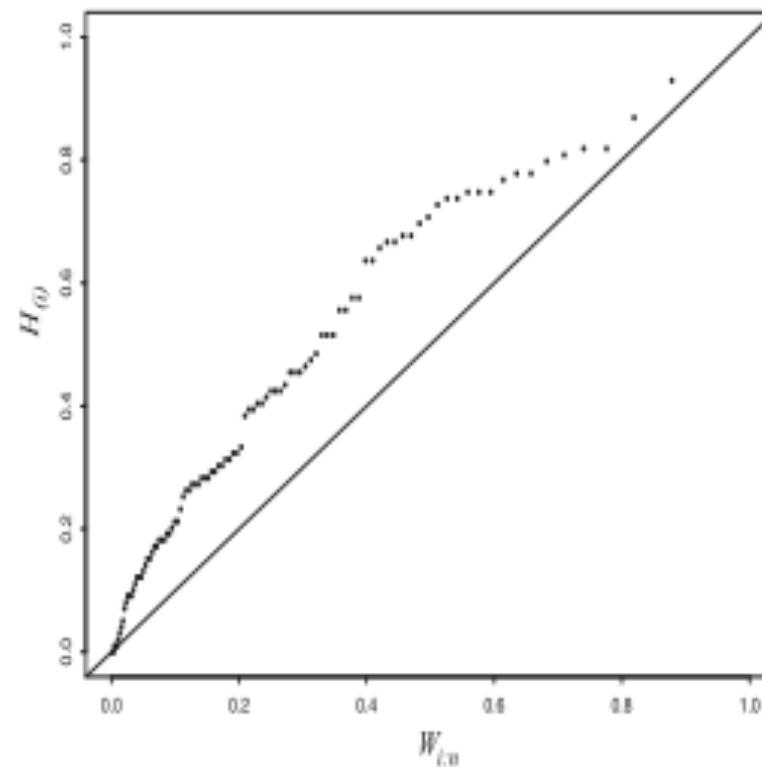
Borne supérieure de Fréchet-Hoeffding



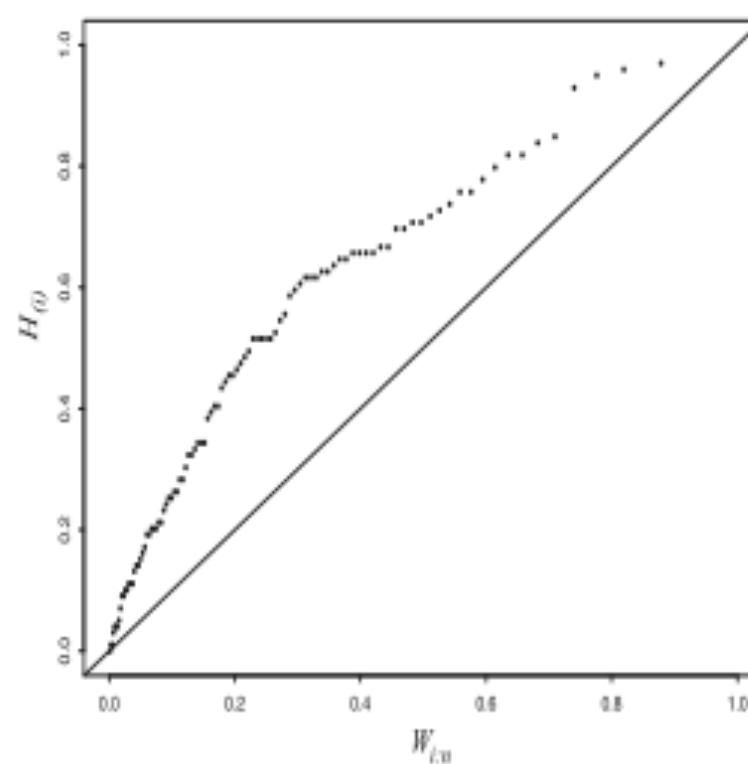
Copule de Clayton avec $\tau = 1/4$



Copule de Clayton avec $\tau = 1/2$



Copule de Clayton avec $\tau = 3/4$





Tests d'indépendance

Nous voulons tester l'hypothèse :

$$H_0 : H(x, y) = F(x)G(y)$$

pour tous $x, y \in \mathbb{R}$. Nous considérerons :

- 1) Deux tests classiques basés sur le coefficient de corrélation empirique
- 2) Tests non paramétriques
 - a) Test de Spearman
 - b) Test de Kendall
 - c) Test de Van der Waerden
- 3) Une petite parenthèse sur les statistiques linéaires des rangs

Le coefficient de corrélation linéaire empirique

Soit un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$.

Les tests classiques s'appuient sur la version empirique du coefficient de corrélation linéaire r

$$r_n = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{S_X^2 S_Y^2}}$$

1) Deux tests classiques d'indépendance

Sous $H_0 : r = 0$ et sous l'hypothèse $(X, Y) \sim N_2(\mu, \Sigma)$

*

$$\frac{r_n \sqrt{n-2}}{\sqrt{1-r_n^2}} \sim t_{n-2}$$

*

$$\sqrt{n-3} \left\{ \frac{1}{2} \log \left(\frac{1+r_n}{1-r_n} \right) - \frac{1}{2} \log \left(\frac{1+r}{1-r} \right) \right\} \sim N(0, 1)$$

Attention ! En dehors du cadre gaussien ces procédures ne sont plus optimales ! De plus, r n'existe pas toujours (e.g. Cauchy), peut être proche de 0 même en cas de forte dépendance, ne mesure que l'association linéaire, ...

2) Tests non paramétriques

- a) Test de Spearman
- b) Test de Kendall
- c) Test de Van der Waerden



a) Test de Spearman

La statistique du test est basée sur

$$\rho_n = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{j=1}^n (S_j - \bar{S})^2}}$$

Sous H_0 , $\sqrt{n-1}(\rho_n - \rho) \rightarrow_{\mathcal{L}} N(0, 1)$. Dans une autre version, on considère la statistique

$$B_n = \sum_{i=1}^n (R_i - S_i)^2$$

Sous H_0 , $B_n \rightarrow_{\mathcal{L}} N\left(\frac{n^3-n}{6}, \frac{n^2(n+1)^2(n-1)}{36}\right)$. On peut facilement montrer que $\rho_n = 1 - \frac{6B_n}{n^3-n}$.



b) Test basé sur le τ de Kendall

Le test s'appuie sur la statistique suivante :

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

Sous H_0

$$E(\tau_n) = 0$$

$$Var(\tau_n) = \frac{2(2n+5)}{9n(n-1)}$$

et, de plus, nous avons la normalité asymptotique.



c) Test de Van der Waerden

Le test s'appuie sur la valeur de la statistique

$$V_n = \frac{1}{n} \sum_{i=1}^n \phi^{-1} \left(\frac{R_i}{n+1} \right) \phi^{-1} \left(\frac{S_i}{n+1} \right)$$

Sous H_0

$$E(V_n) = 0$$

$$Var(nV_n) = \frac{1}{n-1} \left\{ \sum_{i=1}^n \phi^{-1} \left(\frac{i}{n+1} \right)^2 \right\}^2$$



3) Statistiques linéaires des rangs

Le ρ de Spearman et la statistique de Van der Waerden appartiennent à la famille des statistiques linéaires de rangs, dont la forme générale est

$$T_n^j = \frac{1}{n} \sum_{i=1}^n J\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)$$

où J est une fonction score $J : [0, 1]^2 \rightarrow R$ satisfaisant certaines propriétés.

Plusieurs exemples : Blest, Blomqvist, Gini, Savage, ...

Des conditions sur la fonction J assurent la normalité asymptotique de T_n sous H_0 .

Tests localement optimaux

Lequel de ces tests est le meilleure ?

- leur lois sont toutes libres sous H_0
 - leur puissance dépend du choix de l'hypothèse alternative
 - selon la forme de la copule sous l'hypothèse alternative certains se révèlent mieux que d'autres

Test localement le plus puissant (LPP)

Il est possible d'identifier le test LPP comme étant de la forme

$$T_n^* = \frac{1}{n} \sum_{i=1}^n T(R_i, S_i)$$

où

$$T(r, s) = E \left[\frac{\dot{c}_{\theta_0}}{c_{\theta_0}}(B_{(r)}, B_{(s)}^*) \right]$$

et

$$B_{(r)} \sim Beta(r, m - r + 1) \quad \perp \quad B_{(s)}^* \sim Beta(s, m - s + 1)$$

Tests d'adéquation pour les modèles de copule : principes généraux

Il s'agit de tester l'hypothèse nulle suivante :

$$H_0 : C \in (C_\theta)$$

à partir d'un échantillon aléatoire.

Nous considérons:

- a) tests basés sur C_n
- b) tests basés sur K_n



Stratégie générale

- On suppose que l'hypothèse $H_0 : C \in (C_\theta)$ est vraie.
- On estime θ par θ_n sous cette hypothèse.
- On compare C_{θ_n} à un estimateur de la copule \hat{C} .
- On mesure la “distance” observée $\|C_{\theta_n} - \hat{C}\|$.

On s'attend à avoir...

Idéalement, la procédure d'adéquation devrait être :

- indépendante des marges, c'est-à-dire fondée sur les rangs ;
- conceptuellement simple et facile à implémenter ;
- libre de choix arbitraires (e.g., paramètres de lissage) ;
- convergente et puissante.



a) Tests basés sur C_n

Sans doute, le test le plus naturel consiste à mesurer

$$D_n = n \int_{[0,1]^2} \{C_{\theta_n}(u, v) - C_n(u, v)\}^2 dC_n(u, v)$$

où

- C_n est la copule empirique de Deheuvels ;
- θ_n est un estimateur basé sur les rangs de θ , obtenu par exemple par inversion du tau de Kendall.

Formule simplifiée pour D_n

La statistique D_n de Cramér-von Mises se réduit à

$$\sum_{i=1}^n \left\{ C_{\theta_n} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - C_n \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right\}^2$$



D'autres possibilités

D'autres distances pourraient être envisagées, e.g., L_1 ou L_∞ .

En particulier, la statistique de Kolmogorov-Smirnov mesure

$$\max_{1 \leq i \leq n} \left| C_{\theta_n} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - C_n \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right|$$



Avantages de D_n

- Elle est conceptuellement simple et basée sur les rangs.
- Elle ne fait intervenir aucun choix arbitraire.
- Elle est convergente, à cause du comportement asymptotique du processus de copule.



Inconvénients de D_n

- Sa loi asymptotique est complexe.
- La puissance du test reste à évaluer.

Ces inconvénients ont conduit Fermanian (2005, JMVA) à proposer des procédures d'ajustement basées sur des estimateur à noyau de la densité de copule.

Voir aussi Genest et Rémillard (2008, AIHP) et Genest, Rémillard et Beaudoin (2008, IME)

Bootstrap paramétrique - Algorithme (partie I)

- a) On convertit l'échantillon en paire de rangs (R_i, S_i) ,
 $i \in \{1, \dots, n\}$
- b) On calcule $\hat{U}_i = \frac{R_i}{n+1}$, $\hat{V}_i = \frac{S_i}{n+1}$ pour $i = 1, \dots, n$ et on pose

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1_{\{\hat{U}_i \leq u, \hat{V}_i \leq v\}}, \quad u, v \in (0, 1)$$

- c) On estime θ par θ_n et on calcule D_n .

Bootstrap paramétrique - Algorithme (partie II)

On fixe N et on répète les étapes suivantes pour tout $k \in \{1, \dots, N\}$:

- i) générer un échantillon aléatoire $X_{1,k}^*, \dots, X_{n,k}^*$ de la copule C_{θ_n} et en déduire les paires de rangs correspondantes $(R_{1,k}^*, S_{1,k}^*), \dots, (R_{n,k}^*, S_{n,k}^*)$, $i \in \{1, \dots, n\}$
- ii) calculer $\hat{U}_{i,k}^* = \frac{R_{i,k}^*}{n+1}$, $\hat{V}_{i,k}^* = \frac{S_{i,k}^*}{n+1}$ pour $i = 1, \dots, n$ et poser

$$C_{n,k}^*(u, v) = \frac{1}{n} \sum_{i=1}^n 1_{\{\hat{U}_{i,k}^* \leq u, \hat{V}_{i,k}^* \leq v\}}, \quad u, v \in (0, 1)$$

Bootstrap paramétrique - Algorithme (partie III)

- iii) estimer θ par $\hat{\theta}_{n,k}^*$ par la même méthode que précédemment, mais en utilisant $\hat{U}_{i,k}^* = \frac{R_{i,k}^*}{n+1}$ et $\hat{V}_{i,k}^* = \frac{S_{i,k}^*}{n+1}$ au lieu de $\hat{U}_i = \frac{R_i}{n+1}$ et $\hat{V}_i = \frac{S_i}{n+1}$
- iv) calculer la valeur de la statistique

$$D_{n,k}^* = \sum_{i=1}^n \{C_{n,k}^*(\hat{U}_{i,k}^*, \hat{V}_{i,k}^*) - C_{\hat{\theta}_{n,k}^*}(\hat{U}_{i,k}^*, \hat{V}_{i,k}^*)\}^2$$

Bootstrap paramétrique

Une estimation du seuil critique du test est alors donnée par

$$L = D_{n,\lfloor(1-\alpha)N\rfloor:N}^*$$

où $D_{n,1:N}^* < \dots < D_{n,N:N}^*$ sont les statistiques d'ordre et $\lfloor x \rfloor$ est la partie entière de x .

Règle de décision : on n'accepte pas H_0 si $D_n > L$.

Une approximation de la p-value pour la statistique de Cramer-Von Mises est donnée par

$$\frac{1}{N} \sum_{k=1}^N \mathbf{1}_{\{D_{n,k}^* > D_n\}}$$

Difficulté : le calcul de C_{θ_n} ou de $C_{\theta_{n,k}^*}$ peut nécessiter une intégration numérique.

b) Tests basés sur K_n (Wang et Wells, 2000, JASA)

Le test s'appuie sur la statistique

$$S_{n\xi} = \int_{\xi}^1 \{\mathcal{K}_n(w)\}^2 dw$$

avec $\mathcal{K}_n(w) = \sqrt{n}(K_n(w) - K_{\theta_n}(w))$ le processus de Kendall et $\xi \in (0, 1)$ une valeur arbitraire.

N.B. Ce test a été proposé pour le cas Archimédien bivarié et en présence de données censurées.

Problèmes

- La loi limite de $S_{n\xi}$ est complexe
- La procédure bootstrap proposée par Wang et Wells ne fonctionne pas; on ne peut donc calculer une p-value. Wang et Wells suggèrent de retenir le modèle qui correspond à la valeur la plus faible de $S_{n\xi}$.



Variantes de $S_{n\xi}$ (Genest et al., 2006, SJOS)

Les statistiques proposées sont :

$$S_n = \int_0^1 |\mathcal{K}_n(w)|^2 k_{\theta_n}(w) dw$$

et

$$T_n = \sup_{0 \leq w \leq 1} |\mathcal{K}_n(w)|$$

avec $\mathcal{K}_n(w) = \sqrt{n}(K_n(w) - K_{\theta_n}(w))$, $w \in (0, 1)$.



Avantages de S_n et T_n

- On dispose de formules simples pour S_n et T_n qui dépendent des rangs (ce n'est pas le cas pour $S_{n\xi}$)
- Pas de constante ξ (problème du choix et de l'influence de cette constante sur la distribution asymptotique de $S_{n\xi}$ pas traité par Wang et Wells (2000))
- On ne se restreint pas au cas de la copule bivariée archimédienne
- Il existe une procédure de Bootstrap paramétrique qui permet d'approximer la p-value.

Expressions simplifiées

$$T_n = \sqrt{n} \max_{i=0,1; 0 \leq j \leq n-1} \left| K_n \left(\frac{j}{n} \right) - K_{\theta_n} \left(\frac{j+1}{n} \right) \right|$$

$$\begin{aligned} S_n &= \frac{n}{3} + n \sum_{j=1}^{n-1} K_n^2 \left(\frac{j}{n} \right) \left[K_{\theta_n} \left(\frac{j+1}{n} \right) - K_{\theta_n} \left(\frac{j}{n} \right) \right] \\ &\quad - n \sum_{j=1}^{n-1} K_n \left(\frac{j}{n} \right) \left[K_{\theta_n}^2 \left(\frac{j+1}{n} \right) - K_{\theta_n}^2 \left(\frac{j}{n} \right) \right] \end{aligned}$$

Procédure bootstrap paramétrique

- Estimer θ par un estimateur convergent $\hat{\theta}_n$
- Générer N échantillons aléatoires de taille n de la loi $C_{\hat{\theta}_n}$ et, pour chacun d'entre eux, estimer θ par la même méthode que précédemment
- Déterminer ensuite la valeur de la statistique du test
- On note $S_{1:N}^* \leq \dots \leq S_{N:N}^*$ les valeurs ordonnées de la statistique du test calculée à l'étape précédente. Une estimation de la valeur critique du test au niveau α associée à la statistique S_n est alors donnée par $S_{[(1-\alpha)N]:N}^*$ et une estimation de la p-value associée à la valeur observée de la statistique S_n est donnée par

$$\frac{1}{N} \# \{j : S_j^* \geq S_n\}$$



D'autres possibilités que nous ne traiterons pas...

- Tests basés sur un estimateur à noyau de la densité de la copule (Fermanian (2005, JMVA)) mais choix de h et de K (Scaillet (2007, JMVA)). Pour une résolution des problèmes au bord voir aussi Omelka et al. (2009, Annals of Statistics).
- Tests basés sur un estimateur par ondelettes de la copule (Gayraud et Tribouley (2011, Test)).
- Tests pour des familles particulières de copules (famille archimédienne (Hering et Hofert (2010)), copules des valeurs extremes (Ghoudi et al. (1998, CJS)), etc.).