

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D
SEMANA						

Análise de Dependência por Meio de Cúpulas

Prof: Nikolai Valtchev Kolev

Se interessou por cúpulas faz 15 anos, "por causa da simulação sexual exótica".

19 a 20/03 Workshop em Cúpulas em Campinas

email: Kolev.ime@gmail.com

sala: A216

www.ime.usp.br/bcsnif *6th Brazilian 24-28/03
Satellite Copula Workshop

Nessa conferência*, ocorreram dois minicursos ministrados resp. por:

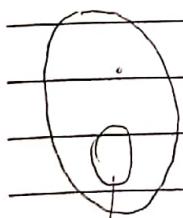
Bruno Drippe

Alane de Góes

19-20/03 | IMECC-UNICAMP

24-28/03 - Semana Santa

Olhando para trás:



Amostra

- ① população "normal" - não existe uma definição de normal
- ② amostra independente → Na verdade, é dependente → 2011*
- ③ $x_1, \dots, x_n \rightarrow$ dados não exalam o modelo

Modelagem Estatística → postulamos um modelo de,

que escreve a dependência, "existente"

- Um turco escreveu um artigo mostrando que a amostra sempre é dependente. Visita-lo em Izmir. Seu nome é Bayramoglu

Lei da poluição é linear. Ao modelar, podemos obter diferentes modelos

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

o que é óbvio absurdo.

~~dados escolhem o módulo~~

Não há nada na estatística nesse sentido, mas há um método computacional:

Minimum description length principle (MDL), 77. J. Rissanen

Sendo X, Y variáveis aleatórias em (Ω, \mathcal{F}, P) . Definimos

$$H(x, y) = P(X \leq x, Y \leq y)$$

Considerando

$$F(x) = P(X \leq x) = \lim_{y \rightarrow \infty} H(x, y)$$

$$G(y) = P(Y \leq y) = \lim_{x \rightarrow \infty} H(x, y)$$

Exemplo,

$$\text{Se } H(x, y) = (1 + e^{-x-y})^{-1}, \text{ então } F(x) = (1 + e^{-x})^{-1} \text{ e } G(y) = (1 + e^{-y})^{-1}$$

Se X é v.a. contínua. $F(X)$ é uma v.a.

↓ troca o argumento por uma v.a.

Daí,

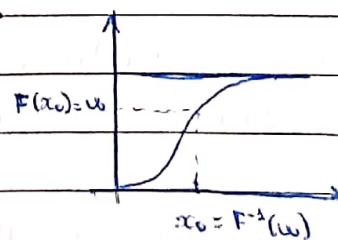
$$P(F(X) \leq t) = P(X \leq F^{-1}(t)) = F(F^{-1}(t)) = t$$

↓
eventos equivalentes

F é não-decrescente

$F^{-1} \rightarrow$ não-decrescente

$$F(x) = u \rightarrow x = F^{-1}(u)$$

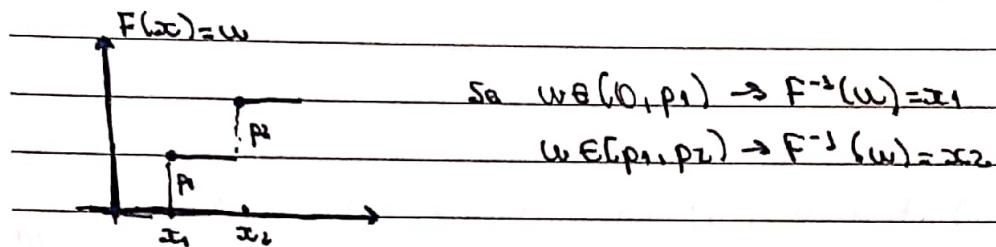


S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

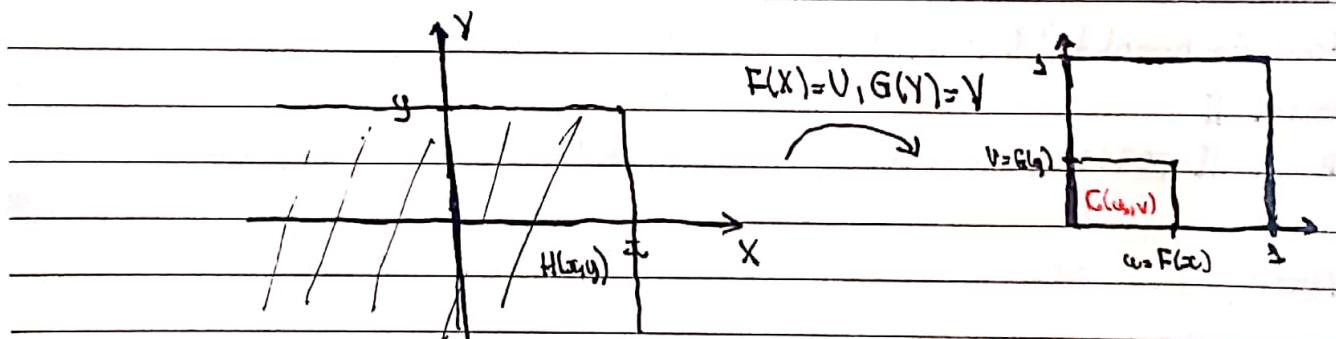
Como $V(t) = t \rightarrow F(X) \sim V(0,1)$

No caso discreto,



Abertura do evento: Joe Roberts and friends

Geometricamente,



Daí,

$$H(x,y) = C(F(x), G(y)) \quad (1)$$

↓
côpula.

Nessas condições,

$$\begin{aligned} x &= F^{-1}(u) \\ y &= G^{-1}(v) \end{aligned} \Rightarrow C(u,v) = H(F^{-1}(u), G^{-1}(v)) \quad (2)$$

$u, v \in [0,1]$

Exemplo:

$$\text{Sendo } F(y) := v = \frac{1}{1+e^{-y}} \Rightarrow 1+e^{-y} = v^{-1}$$

$$e^{-y} = v^{-1} - 1$$

$$-y = \ln(v^{-1} - 1)$$

$$y = -\ln\left(\frac{1-v}{v}\right) \Rightarrow y = -\ln\left(\frac{1-v}{v}\right)$$

SEMANA	S	T	Q	Q	S	S	D
	L	M	M	J	V	S	D

Se $H(x,y)$ é conhecido → obter copula C via (1)

$$C(u,v) = P(U \leq u, V \leq v) \text{ com } U, V \sim U(0,1).$$

Paul Embrechts

Se se conhece $C(u,v)$, $F(x)$ e $G(y)$ então podemos obter $H(x,y)$.

A relação (1) conhecida como Teorema de Sklar.

Paul Embrechts (ETH) 99.

Basel II

Basel III, 07/10. Nesse caso, a cópula entrou como método obrigatório.

Agora,

$$P(H(x,y) \leq t)$$

característica univariada de distribuição bivariada (dependência)
Variável Aleatória de Kendall

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Artigos que ele me passou.

1. C. Genest and A. -C. Favre (2007). Everything you always wanted to know about copulas modeling but were afraid to ask.

Journal of Hydrologic Engineering, 12, 347-368

2. G. Genest and B. Remillard (2006). Goodness-of-fit procedures for copula models based on probability integral transforms.

Scandinavian Journal of Statistics, 33, 337-368

3. G. Genest, B. Remillard, and D. Beaudoin (2009).

Goodness-of-fit tests for copulas: A review and power study.

Insurance: Mathematics & Economics, 44, 199-213

06/03/2012

"Insistência, dedicações ajudam a encontrar o caminho certo para a resolução de um problema"

Resumo aula anterior

(X, Y) continua

$$H(x, y) = P(X \leq x, Y \leq y)$$

$$F(x) = P(X \leq x) = \lim_{y \rightarrow \infty} H(x, y)$$

$$G(y) = P(Y \leq y) = \lim_{x \rightarrow \infty} H(x, y)$$

$$H(x, y) = C(F(x), G(y)) \quad (1)$$

$$F(x) = u, \in [0, 1] \text{ e } G(y) = v, \in [0, 1]$$

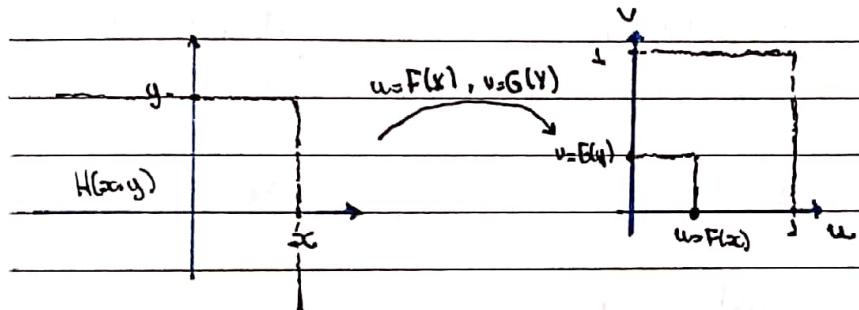
S	T	Q	Q	S	S	D

SEMANA
L M M J V S D

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \quad (1)$$

(1) e (2) são as bases do Teorema de Sklar.

Se X e Y são contínuas então a cópula $C(\cdot, \cdot)$ será única.
Caso contrário, $C(\cdot, \cdot)$ não será única.



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

$$G(Y) = V \sim U(0,1)$$

$$C(u, v) = P(U \leq u, V \leq v) \text{ com } u, v \in [0,1]$$

↳ f.d. conjunta.

$C(u, v)$ é uma f. não decrescente em cada argumento.

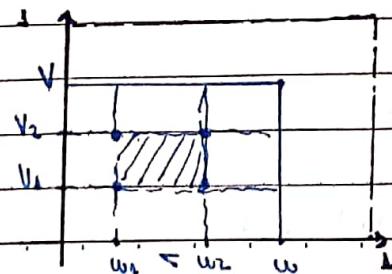
Então $C(u, v)$ é f.d.c., i.e.

$$C(0,0) = 0 \Rightarrow \lim_{u \rightarrow 0} C(u, v) = 0$$

$$\lim_{u \rightarrow 1} C(u, v) = C(1, v) = P(V \leq v) = v$$

$$\lim_{v \rightarrow 1} C(u, v) = C(u, 1) = P(U \leq u) = u$$

$$\lim_{u, v \rightarrow 1} C(u, v) = 1$$



S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Se $u_1 \leq u_2 \Rightarrow C(u_2, v_2) - C(u_2, v_1) = C(u_1, v_1) + C(u_1, v_2) \geq 0$
 $v_1 \leq v_2$

Def. 1

Def 2: Cúpula é f.d. conjunta de 2 v.a. $U(0,1)$

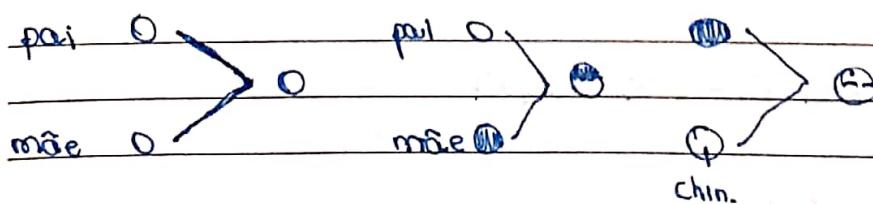
$C(f(x), G(y)) \Rightarrow$ cúpula da dependência entre X e Y [independente] de marginais $F(x)$ e $G(y)$.

Se $F(x)$, $G(y)$ e $C(\cdot, \cdot)$ são conhecidos \Rightarrow podemos restituir a f.d. conjunta $H(x, y)$

(*) Problema em aberto

X	Y	As marginais têm sua influência em dependência <u>existe</u> , se verdadeira genuína.
$F(x)$	$G(y)$	

Exemplo da vida real



$$\Delta(x,y) = \frac{H(x,y)}{F(x)G(y)} \begin{cases} > 1 & \text{dep. positiva} \\ < 1 & \text{dep. negativa} \end{cases}$$

Sibuya, 1960

Uma das primeiras funções utilizadas para a análise de val. extremas

SEMANA	S	T	Q	Q	S	S	D
	L	M	M	J	V	S	D

Perceber que

$$\Delta(x,y) = H(x,y) = C(F(x), G(y)) = \underline{C(w, v)}$$

$$F(x)G(y) \quad F(x) G(y) \quad wv$$

Suponhamos (95%)

$$\begin{cases} H_0: \text{Valor de } A \text{ vai crescer} \\ H_A: \end{cases}$$

Econometria

5% = 1 dia por mês, o valor vai crescer Svetozar Rachev

(Bovespa uso o col. decora)

$$H(x,y) = P(X \leq x, Y \leq y) \leq P(X \leq x) + F(x)$$

$$\leq P(Y \leq y) = G(y)$$

Daí,

$$H(x,y) \leq \min(F(x), G(y))$$

Por outro lado,

$$0 \leq P(X > y, Y > z) = 1 - P(X \leq x) - P(Y \leq y) + P(X \leq x, Y \leq y)$$

$$\underbrace{H(x,y)}_{F(x) + G(y) - 1} \leq H(x,y)$$

$$F(x) + G(y) - 1 \leq H(x,y)$$

Com $H(x,y) \geq 0$, segue que

$$\max(0, F(x) + G(y) - 1) \leq H(x,y)$$

Assim,

$$\max(F(x) + G(y) - 1, 0) \leq H(x,y) \leq \min(F(x), G(y))$$

(Daí a noção de limites).

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Limites de Hölding (1948) e Frechet (1951)

Fazendo $w = F(x)$ e $v = G(y)$

$$\min_{W(w,v)} (w+v-1, 0) \leq C(w,v) \leq \min_{M(w,v)} (w, v)$$

fronteira

Inferior

fronteira

superior

No caso de independência,

$$H(x,y) = F(x)G(y) \quad H(x,y)$$



$$C(w,v) = w.v = \Pi(w,v)$$

$$H(x,y) \xrightarrow{\text{sóntica}} C(w,v)$$

$$H(x,y) \leftrightarrow C(w,v)$$

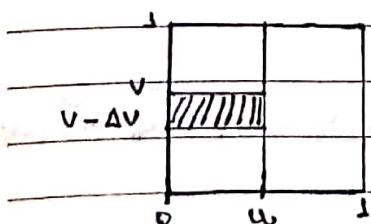
← se X e Y tem mesma distribuição.

Segundo $C(w,v) = \min(w,v)$ para todo $(w,v) \in [0,1]^2 \Rightarrow P(U=Y)=1$?

$$C_{U|V}(w|v) = P(U \leq w | V=v) = \lim_{\Delta v \rightarrow 0} P(U \leq w | v-\Delta v \leq V \leq v) =$$

$$= \lim_{\Delta v \rightarrow 0} \frac{P(U \leq w, v-\Delta v \leq V \leq v)}{P(v-\Delta v \leq V \leq v)} = \lim_{\Delta v \rightarrow 0} \frac{P(U \leq w, V \leq v) - P(U \leq w, V \leq v-\Delta v)}{\Delta v} =$$

$$\lim_{\Delta v \rightarrow 0} \frac{C(w,v) - C(w,v-\Delta v)}{\Delta v} = *$$



Desse modo, $\lim_{\Delta v \rightarrow 0} [C(w,v) - C(w,v-\Delta v)] = 0$.

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Observação:

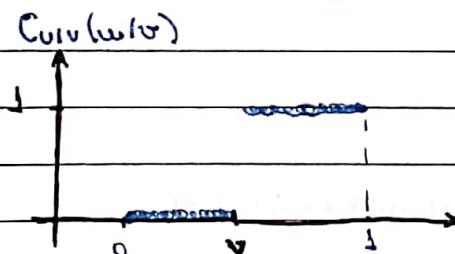
$$\lim_{\Delta v \rightarrow 0} \frac{C(w, v) - C(w, v - \Delta v)}{\Delta v} = \frac{\partial C(w, v)}{\partial v} = C_{w|v} \text{ quando absolutamente contínuas}$$

Se $C(w, v) = \min(w, v)$

$$\lim_{\Delta v \rightarrow 0} \frac{\min(w, v) - \min(w, v - \Delta v)}{\Delta v} = \begin{cases} (v - (v - \Delta v)) / \Delta v = 1 & \text{se } w > v \\ (w - w) / \Delta v = 0 & \text{se } w < v \end{cases} \xrightarrow{\Delta v \rightarrow 0} w < v \Rightarrow w < v - \Delta v$$

Se $C(w, v) = \min(w, v)$, então

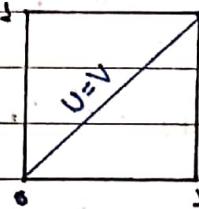
$$C_{w|v}(w|v) = \begin{cases} 1 & \text{se } w > v \\ 0 & \text{se } w \leq v \end{cases}$$



Logo,

$$C_{w|v}(v|v) = 1 \text{ quase certamente}$$

$$P(U=v) = 1 \Leftrightarrow U=v \text{ quase certamente}$$



Então a massa de probabilidade de $C(w, v) = \min(w, v)$ é concentrada na diagonal principal $w=v$ para todo $w, v \in [0, 1]$.

Tarefa.

Fazer para $C(w, v) = \max(w + v - 1, 0) \Rightarrow$ resultado é $\boxed{U = 1 - V \Leftrightarrow P(U = 1 - V) = 1}$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Causalidade de Granger (Pesquisar)

Causalidade de Granger (lucasmassaroppe@gmail.com)

Multiple Times Series, Lütkepohl.

12/03/2013

$$H(x, y) = P(X \leq x, Y \leq y)$$

$$F(x) = P(X \leq x)$$

$$G(y) = P(Y \leq y)$$

(x, y)

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y))$$

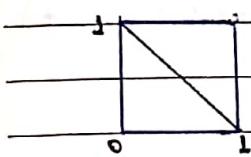
$$w = F(x)$$

$$v = G(y)$$

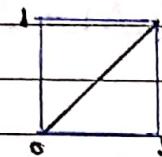
$$\max(w + v - 1, 0) \leq C(w, v) \leq \min(w, v)$$

$$P(U=V)=1$$

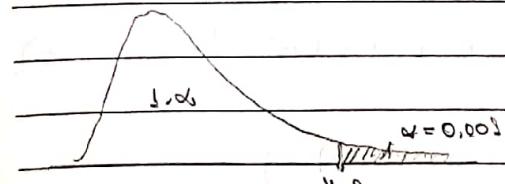
$$P(U \neq V)=0$$



contra-monotonicos



monotonicos



• Explicação do valor Var α para exemplificação do exemplo

$$\text{Período: } S_n = Y_1 + \dots + Y_n$$

No caso bivariado, simples. Mas se mais, a coisa complica. Adé no caso em que Y_1, \dots, Y_n são i.i.d. com distrib. $U(0, 1)$.

Utilizamos o T.C.L, mas não tem mais!

PS: Tudo acima vale para (X, Y) contínuas.

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

Relembrando o Teorema de Sklar:

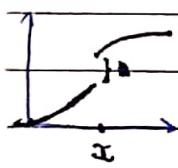
$$H(x, y) = C(F(x), G(y))$$

$$C(u, v) = H(F^{-1}(u), G^{-1}(v))$$

\exists , \exists

A cópula é única quando X e Y são contínuas.

Se X cont. $\Rightarrow F(x) \sim U(0,1)$



(?)

$$F_1(x, \lambda) = D(X \leq x) + \lambda P(X = x), \quad \lambda \in [0,1]$$

Se $\lambda = 0$ ou 1 se X é contínua $\Rightarrow F_1(x, \lambda) = F(x)$.

$$F_1(x, \lambda) = F(x) = P(X \leq x) \text{ se } X \text{ é contínua}$$

$\Delta \sim U(0,1)$ e Δ independente de X

$$F_1(X, \Delta) = F(X-) + \Delta [F(X) - F(X-)], \quad F_1(X, \Delta) \sim U(0,1)? \quad (2009)$$

(sim, prova futura) (2004, prova) (livro português)

Daí,

$$\boxed{F_1(X, \Delta) = \Delta F(X) + (1-\Delta) F(X-)} \quad \text{Ferguson (1967)}$$

$F_1(X, \Delta) \sim ?$

Seja $\alpha(x)$ e $\beta(y)$ \Rightarrow funções contínuas crescentes (não-decrescentes).

$(x, y) \rightarrow$ cópula correspondente à (X, Y)

$$(x, y) \rightarrow (\alpha(x), \beta(y)) \quad \left\{ \begin{array}{l} \Rightarrow C_{X,Y}(x, y) = C_{\alpha(X), \beta(Y)}(x, y) \\ \end{array} \right.$$

$$\text{Se } (x, y) \xrightarrow{\text{marg.}} F_X, G_Y \rightarrow C_{X,Y}(x, y) = C(F_X^{-1}(x), G_Y^{-1}(y))$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

$$(\alpha(X), \beta(Y)) \xrightarrow{\text{marg.}} F_1, G_1 \rightarrow C_{\alpha(X), \beta(Y)}(w, v)$$

$$w = F_1(x) = P(\alpha(X) \leq x) = P(X \leq \alpha^{-1}(x)) = F_X(\alpha^{-1}(x)) \quad \forall x \in (-\infty, \infty)$$

$$v = G_1(y) = P(\beta(Y) \leq y) = P(Y \leq \beta^{-1}(y)) = F_Y(\beta^{-1}(y)) \quad \forall y \in (-\infty, \infty)$$

Agora,

$$C_{\alpha(X), \beta(Y)}(w, v) = P(\alpha(X) \leq F^{-1}(w), \beta(Y) \leq G^{-1}(v)) =$$

$$= P(\alpha(X) \leq x, \beta(Y) \leq y) =$$

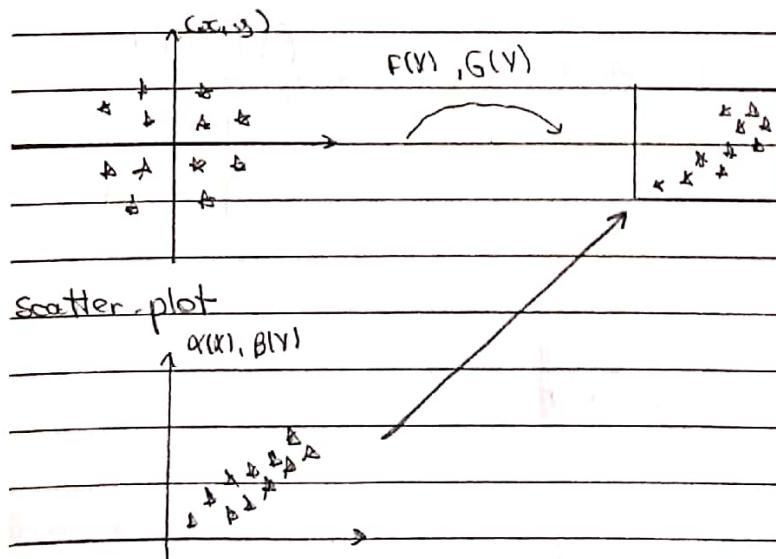
$$= P(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) =$$

$$= H(\alpha^{-1}(x), \beta^{-1}(y)) =$$

$$= C_{X,Y}(F_X(\alpha^{-1}(x)), G_Y(\beta^{-1}(y))) =$$

$$= C_{X,Y}(w, v)$$

Com o intermediário acima, provamos que $C_{X,Y}(w, v) = C_{\alpha(X), \beta(Y)}(w, v)$ para α e β funções não-decrescentes.



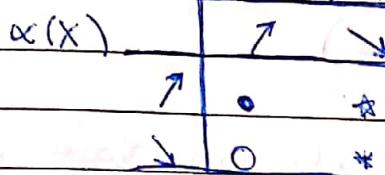
19/03 - INECC 02/04 - Buenos Aires

20/03 - INECC

26 e 27/03 - Semana Santa. Maresias

SEMANA	S	T	Q	Q	S	S	D
	L	M	M	J	V	S	D

$\beta(Y)$



$$\alpha(x,y), \beta(y) \quad (u,v) = v + C_{x,y} (1-u, v) \quad \text{tr* them exist}.$$

$F(X) \rightarrow U(0,1)$

$G(Y)$

$$P(H(X,y) \leq t) = K(t)$$

univariate \rightarrow V.D. de Kendall

characteristic

of bivariate

dependence

Introduzida por Genest (2001)

Nelsen

Os dois consideraram a mesma coisa
ao mesmo tempo. No entanto, maneira
de Nelsen é um pouco mais detalhada,
mais interessante.

$$p(x,y) = \iint u v r d(C(u,v)) - 3$$

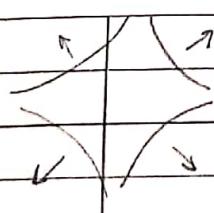
global não é bom, mas olhando para cada subpopulação.

I.C.

$$P = P(a \leq U \leq b)$$

Bivariate confidence Intervals

Caminho



Se olhar ao global, perdemos informação.
Então, começamos a olhar para a
dependência direcional.

S	T	Q	Q	S	S	D	
L	M	M	J	V	S	D	SEMANA

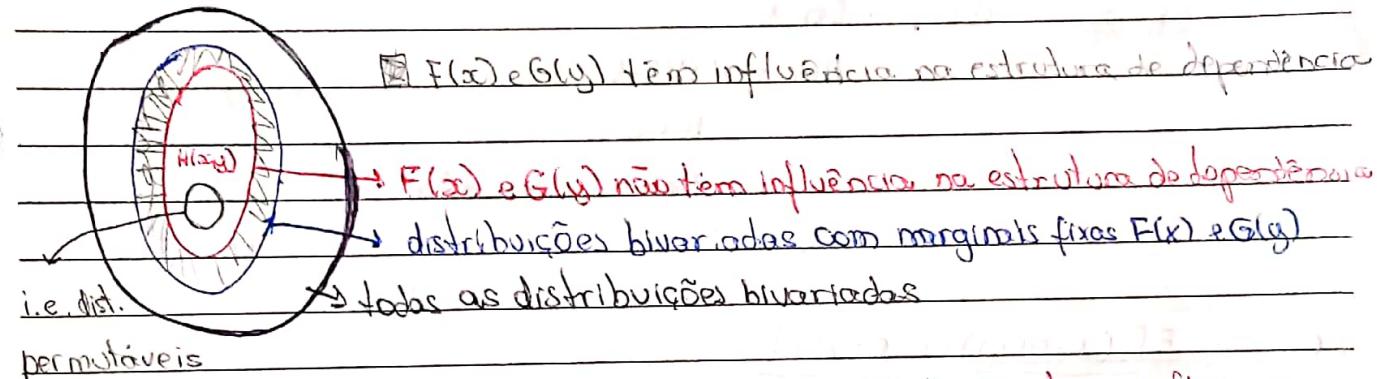
13/03/2013

08/04 → 59-64 Copula construction (geom. methods)

08/04 → 76-84 Copulas with given horizontal and vertical sections

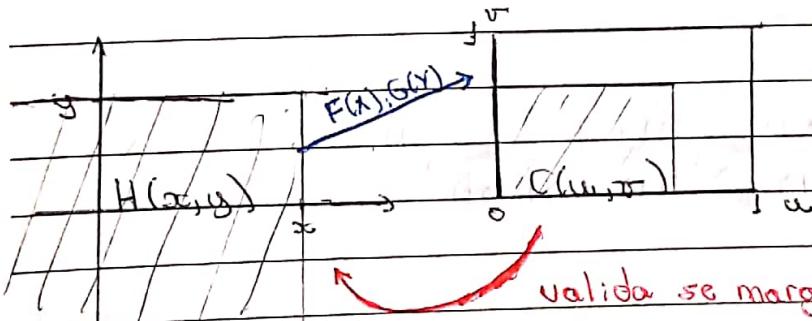
09/04 → 94-99 Extreme given value copulas

09/04 * Detecting dependence with Kendall plots (artigo) → American Statistician, 2003, 57, 275-284, Genest + Boies



$H(x,y) = C(F(x), G(y))$ serve bem pois $F(x)$ e $G(y)$ não têm influência na estrutura de dependência.

$C(\cdot, \cdot)$ é única se F e G são contínuas



valida se marginais são com a mesma distribuição para marginais de $C(\cdot, \cdot) \in U(0,1)$
i.e., distribuições permutáveis

$$H(x,y) = H(y,x)$$

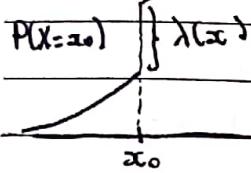
$$\lim_{x \rightarrow \infty} H(x,y) = \lim_{y \rightarrow \infty} H(x,y)$$

Secundaria

$$F_\Delta(X, \Delta) = F(X-) + \Delta [F(X) - F(X-)] \\ = \underbrace{\Delta F(X)}_U + (1-\Delta) \underbrace{F(X-)}_U$$

↓ soma ponderada de uniformes é uniforme!

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D



Construção mais ideal e lógica:

$$F_2(x, \lambda(x)) = P(X \leq x) + \lambda(x) P(X=x)$$

Open problem: $F_2(x, \Delta(x)) \rightarrow ?$

$$\Delta(\cdot) \sim [0,1]$$

Pode ser $U(0,1)$

Beta

Normal truncada

$$p(x,y) = \frac{E[(X-E(X))(Y-E(Y))]}{\sqrt{E(X-E(X))^2} \sqrt{E(Y-E(Y))^2}}$$



Versão local do coeficiente de correlação

$$p(x_0, y_0) = \frac{E[(X-E(X|Y=y_0))(Y-E(Y|X=x_0))]}{\sqrt{E(X-E(X|Y=y_0))^2} \sqrt{E(Y-E(Y|X=x_0))^2}}$$

$$p(x_1, y_1) = z_1$$

$$p(x_2, y_2) = z_2$$

Substituindo argumento

$$p(X, Y) = Z, \in [-1, 1]$$

$$p(x_n, y_n) = z_n$$



Introduzido por Bayzumoglu + Kotz (2003)

$H_1(x, y)$ then é f.d. com marginais $F_X(x)$ e $F_Y(y)$, mas não é f.d. de (X, Y)

$$F_X(x) = P(X \leq x) \leftarrow H_2(x, y) = P(X \leq y, Y \leq y)$$

$$G_Y(y) = P(Y \leq y) \quad \text{f.d. de } (X, Y) \rightarrow \text{contínuas}$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

$$H_1(x, y) = C_1(F_x(x), G_y(y))$$

$$H_2(x, y) = C_2(F_x(x), G_y(y))$$

$$P(H_1(X, Y) \leq t) \stackrel{\text{def.}}{=} H_1 / H_2(t) = \mu_{H_2} \{ (x, y) \in \mathbb{R}^2 \mid H_1(x, y) \leq t \}$$

para todo $t \in (0, 1)$

medida induzida por $H_2(\cdot, \cdot)$

$$= \mu_{H_2} \{ (x, y) \in \mathbb{R}^2 \mid C_1(F_x(x), G_y(y)) \leq t \} \stackrel{F_x(x)=u}{\stackrel{G_y(y)=v}{=}}$$

$$= \mu_{C_2} \{ (u, v) \in [0, 1]^2 \mid C_1(u, v) \leq t \} = C_1 / C_2(t) = P(C_1(U, V) \leq t)$$

$$\text{i.e. } H_1(X, Y) \stackrel{d}{=} C_1(U, V) \quad (*)$$

Encontrar (*) para algum dos casos:

$$M(u, v) = \min(u, v)$$

$$C_1 = W(u, v) = \max(u + v - 1, 0)$$

$$C_2 = \Pi(u, v) = u \cdot v$$

distribuição de Kendall é caso particular de $H_1(X, Y)$ quando $H_1 \equiv H_2$.

$$\bullet C_1 = M, C_2 = M$$

$$M / M(t) = P(M(U, V) \leq t) = \mu_M \{ (u, v) \in [0, 1]^2 \mid C_1 = M(u, v) \leq t \} =$$

" "

$$= P(U \leq t) = t$$

$$M(U, V) \sim U(0, 1) \text{ se } C_1 = C_2 = M$$

(Só nesse caso é uniforme)

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

• $C_1 = \Pi(u, v)$, $C_2 = \min(u, v)$

$$\begin{aligned} \Pi | C_1(t) &= P(\Pi(u, v) \leq t) = \mu_{\{(u, v) \in [0,1]^2 \mid C_1 = \Pi(u, v) = uv \leq t\}} = \\ &\stackrel{M=C_2}{=} P(U^2 \leq t) = P(U \leq \sqrt{t}) = \sqrt{t} \sim \text{Beta}(\gamma_2, 1) \end{aligned}$$

r.aguilera10@gmail.com

17/04/2013

X+Y

Según $(X, Y) \sim C(u, v)$

$$H(x, y) = P(X \leq x, Y \leq y)$$

$$F(x) = P(X \leq x)$$

$$G(y) = P(Y \leq y)$$

$$F^{-1}(w) = \inf \{x : F(x) \geq w\}$$

$$C(v|u) = \frac{\partial}{\partial u} C(u, v)$$

$$A_{X+Y}(s, t) = P(X \leq s, X+Y \leq t) = \int_{-\infty}^s P(X+Y \leq t \mid X=x) dF(x) =$$

$$= \int_{-\infty}^s P(Y \leq t-x \mid X=x) dF(x) = \int_{-\infty}^s \frac{\partial}{\partial x} C(F(x), G(t-x)) dF(x) =$$

$$\frac{\partial}{\partial x} (H(x, t-x))$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

substituir

$$w = F(x) \in (0,1)$$

$$x \rightarrow -\infty \Rightarrow w \rightarrow 0$$

$$x \rightarrow s \Rightarrow w \rightarrow F(s)$$

$$\int_0^{F(s)} \frac{\partial}{\partial w} G(w, G(t - F^{-1}(w))) dw$$

$$B_x(t) = \lim_{x+y} \lim_{s \rightarrow \infty} A_{x,x+y}(s,t) =$$

$$= \int_0^t \frac{\partial}{\partial w} G(w, G(t - F^{-1}(w))) dw$$

$F * G$

$$B_{x+y}(t) = \int_0^t \frac{\partial}{\partial v} G(F(t - G^{-1}(v)), v) dv$$

Exemplo 1.

$$G(w, v) = w, v \Rightarrow \frac{\partial}{\partial w} G(w, v) = v$$

dw

$$\frac{\partial}{\partial w} G(w, G(t - F^{-1}(w))) = G(t - F^{-1}(w))$$

dw

$$w = F(x) \Rightarrow F^{-1}(w) = x$$

$$F * G = \int_{-\infty}^{\infty} G(t - x) dF(x)$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

Exemplo 2.

$$C(u, v) = \min(u, v)$$

$$\frac{\partial C(u, v)}{\partial u} = \begin{cases} 1 & \text{se } u \leq v \\ 0 & \text{se } u > v \end{cases}$$

~~scribble~~

$$\Rightarrow \frac{\partial C(u, v)}{\partial u} = 1 \quad \text{se } u \in G(t - F^{-1}(v))$$

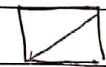
$$F^{-1}(u) + G^{-1}(u) \leq t$$

$$F^c G = \int_0^t I_{\{u : F^{-1}(u) + G^{-1}(u) \leq t\}} du =$$

$$= \sup \{u \in (0, 1) : F^{-1}(u) + G^{-1}(u) \leq t\}$$

$$\bar{E}_{x+y}(t) = \int_0^t \underline{\int}_0^u \bar{C}(u, \bar{G}(t - F^{-1}(v))) dv du$$

$$C(u, v) \stackrel{\text{Def.}}{=} \varphi(F^{-1}(u) + F^{-1}(v)) = C(v, u)$$



família de cônjuges com
suporte simétrico
em relação à diagonal
principal

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Natureza - às vezes calma (permutável)

$$P(U_1 \dots U_n \leq t) = \dots$$

$$P(U_1 \leq t) = t$$

Tentar por indução!

$$P(U_1 U_{n+1} \leq t) = \dots$$

Cherubini (2008)

(ODD)

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \Leftrightarrow$$

$$\Phi(x, y) = C(F(x), G(y))$$

Suposição → dependência que não é influenciada de marginais

1960.

Sibuya:

$$\Lambda(F(x), G(y)) = H(x, y) \quad \text{7} \} - \text{PQD}$$

$$F(x)G(y) \uparrow \downarrow - \text{DQD} \quad (1 + e^{-x} + e^{-y})$$

\Rightarrow NQD

$$\Gamma(x, y) = H(x, y)$$

$$\Sigma(x)G(y)$$

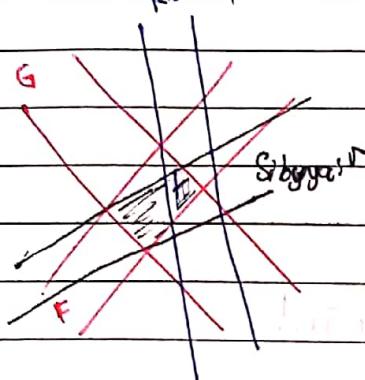
$x \rightarrow X$

$y \rightarrow Y$

$$\Gamma(X, Y) = \frac{H(X, Y)}{F(X)G(Y)} = \frac{H(X, Y)}{\bar{U} \cdot \bar{V}} = \Gamma(X, Y) \quad \text{N. de Sibuya.}$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA



$$\ln H(x,y) = \underbrace{\ln U}_{\text{Exponential}} + \underbrace{\ln Y}_{\text{Exponential}} = \ln \Gamma(x,y)$$

$$\frac{\ln \Gamma(x,y)}{H(x,y)} = E_1 + E_2$$

① Aplicação Cherubini F*G

23/04/2013

$$X \leq Y$$

$$P(X \leq t) = P(Y \leq t) \quad \forall t \geq 0$$

$$E X \leq E Y$$

$$(x_i, y_i) \in \underset{\text{corr}}{(X_i, Y_i)}$$

$$C(x_i, y_i) = \theta \Rightarrow$$

$$\min(x_i + y_i - 1, 0) \leq \theta \leq \min(x_i, y_i)$$

$$\Delta(x, y) = H(x, y)$$

$$F(x) G(y)$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

$$\ln \Delta(x,y) = \ln H(x,y) - \ln F(x) - \ln G(y)$$

Kendall $E_{LNExp}(1)$ $E_{LNExp}(1)$

$$\ln \Delta(x,y) - \ln H(x,y) = E_1 + E_2$$

dependentes.

$$P(F(x), G(y)) = H(x,y) \xrightarrow{\text{subuya } 60} (1 + e^{-x} + e^{-y})^{-1}$$

$F(x) G(y)$

$$\Gamma(x,y) = \frac{H(x,y)}{F(x) G(y)}$$

Let X be arbitrary real r.v. with $F_x(x) = P(X \leq t)$ and $\Lambda \in (0,1)$, independent of X ;

$$F_x(x-) = \lim_{w \uparrow x} F_x(w)$$

$$F(x, \lambda) \stackrel{\text{def}}{=} P(X \leq x) + \lambda P(X = x), \text{ for } \lambda \in (0,1) \text{ and } x \in \mathbb{R}.$$

Obs: If X is abs. continuous

$$P(X = x) = 0 \Rightarrow F(x, \lambda) = F_x(x)$$

fixed

Distributional transform

$$F(x, \Delta) = F_x(x-) + \Delta [F_x(x) - F_x(x-)] =$$

$$= \Delta F_x(x) + (1 - \Delta) F_x(x-)$$

	S	T	Q	Q	S	S	D
SEMANA	L	M	M	J	V	S	D

$$\text{Se } F_X(x) \stackrel{d}{=} U \sim U(0,1)$$

$$\nexists F(X, \Delta) \sim U(0,1)?$$

$$X = F_X^{-1}(v) ? \text{ a.s.}$$

Proof:

For $p \in (0,1)$ let $F_X^{-1}(p) = \inf_{\text{fixed}} \{x; F_X(x) \geq p\}$

$$F(X, \Delta) \leq p \Leftrightarrow (X, \Delta) \in \{(x, \Delta); P(X \leq x) + P(X > x) \leq p\}$$

① Let $\beta = P(X = F_X^{-1}(p)) > 0$ (corresponds to "flat" parts of X) with

$$q = P(X \leq F_X^{-1}(p))$$

Then,

$$\{V = F(X, \Delta) \leq p\} \quad \text{if } q + \Delta \beta = p$$

$$\Downarrow \quad \{X \leq F_X^{-1}(p)\} \cup \{X = F_X^{-1}(p), q + \Delta \beta \leq p\}$$

$X \in \Delta$ ~~are~~ independentes

2004

$$P(V \leq p) = P(F(X, \Delta) \leq p) =$$

$$= P(X \leq F_X^{-1}(p)) + P(X = F_X^{-1}(p)) P(q + \Delta \beta \leq p) =$$

$$= q + \frac{\beta}{1-\beta} p - q = p$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

② Let $\beta = P(X = F_X^{-1}(p)) = 0$ (corresponds to abs. cont. case) \Rightarrow

$$P(F(x, \Delta) \leq p) = P(X \in F_X^{-1}(p)) = E_x(F_X^{-1}(p)) = p$$

Rosenblatt Transform (s2)

Let (X, Y) be a random vector on (Ω, \mathcal{F}, P) .

$$Z = (z_1, z_2) = T(x, y)$$

$$\begin{aligned} T: & \quad z_1 = P(X \leq x) \Rightarrow d_{x_1}(F_X(x)) = dz_1 \\ & \quad z_2 = P(Y \leq y | X=x) = F_{Y|X}(y|x) \Rightarrow d_y(F_{Y|X}(y|x)) = dz_2 \end{aligned}$$

$(z_1, z_2) \sim \text{Uniform Distribution on unit cube } [0,1]^2$

$$P(z_1 \leq z_1, z_2 \leq z_2) = \int_0^{z_1} \int_0^{z_2} dy (F_{Y|X}(y|x)) dx (F_X(x)) =$$

$$= \int_0^{z_1} \int_0^{z_2} dz_1 dz_2 = z_1 z_2 = P(z_1 \leq z_1) P(z_2 \leq z_2)$$

$$\text{i.e. } z_1 = F_X(x) \sim U(0,1)$$

$$z_2 = F_{Y|X}(y|x) \sim U(0,1)$$

$$\text{Se } Y \sim N(0,1) \Rightarrow \phi(y) = P(Y \leq y) \Rightarrow \phi^{-1}(F_X(x)) \sim N(0,1)$$

Ihen, let $A_1, A_2 \sim N(0,1) \Rightarrow A_1^2 + A_2^2 \sim \chi^2(2)$
Ind.

Therefore, the random variable

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

$$S(X, Y) \stackrel{d}{=} [\underbrace{\phi^{-1}(F_X(x))}_{z_1}]^2 + [\underbrace{\phi^{-1}(F_{Y|X}(y|x))}_{z_2}]^2 \sim \chi^2(2) \quad (1)$$

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a random sample from (X, Y) .
 $S(X_1, Y_1), S(X_2, Y_2), \dots, S(X_n, Y_n)$ is a random sample from $\chi^2(2)$.

Test:

$H_0: (X, Y)$ has cont. distr. $H(x, y) = P(X \leq x, Y \leq y)$ vs $H_A: \neq$
 U (não vale)

$H_0^*: S(X, Y) \sim \chi^2(2)$ vs. $H_A: S(X, Y) \neq \chi^2(2)$

Um caminho para verificar se (X, Y) tem f.d. $H(x, y)$. Pela volta
 não ser válida, outras ferramentas devem ser utilizadas para
 que se tenha certeza sobre a afirmação.

$$C(v/u) = \mathbb{E}[C(u, v)]$$

$$\Rightarrow \begin{cases} z_1 = U = F_X(x) \sim U(0,1) \\ z_2 = C(v/u) = C(G_Y(y)/F_X(x)) \sim U(0,1) \end{cases}$$

$$S(X, Y) \stackrel{d}{=} [\phi^{-1}(F_X(x))]^2 + [\phi^{-1}(C(G_Y(y)/F_X(x)))]^2 \sim \chi^2(2)$$

Test:

$H_0: (X, Y)$ has a copula $C(u, v)$

$H_0^*: S(X, Y) \sim \chi^2(2)$ vs $H_A: \neq \chi^2$

A volta só é válida em distribuições simétricas. Ainda é necessário
 uso de outras ferramentas.

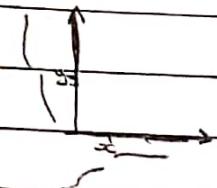
S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

$$F(x, \lambda) = P(X \leq x) + \lambda P(X = x) \Rightarrow U_1 = F(X, \lambda) \sim U(0,1) \rightarrow U_1 \in U_2 \text{ have copula } C_2(u_1, u_2)$$

$$G(y, \lambda) = P(Y \leq y) + \lambda P(Y = y) \Rightarrow U_2 = G(Y, \lambda) \sim U(0,1)$$

$$H_2(x, y) = P(X \leq x, Y \leq y) = C_2(F(x, \lambda), G(y, \lambda))$$



$H_2(x, y) = C_2(F(x, \lambda), G(y, \lambda))$ given that λ_1, λ_2 são indep. on X and Y.

E se

$$F(x, \lambda) = P(X \leq x) + \Delta(x) P(X = x)$$

1. Condition on $\lambda(x)$? $0 \leq \lambda(x) \leq 1$

2. $F(x, \Delta(x)) = ?$ Se $\Delta(x) \sim \text{Beta}$, o que aconteceria com

$$F(x, \Delta(x)) / \Delta(x) = \lambda \quad F(\Delta(x)) = \lambda$$

$$P(X \leq x) + \Delta(x) P(X = x)$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

105/2013

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y)) \quad (1)$$

\nwarrow Se " $=$ " \nearrow

$$X = \beta(Y)$$

$\beta \rightarrow$ decrescente

$$P(X=Y) = 1$$

$X = \alpha(Y)$, α -crescente



Open problema: versão de (1) em termos de densidades

$$? \leq h(x, y) \leq ?$$

$$P(X_1 > x_1, X_2 > x_2) = P(\min(T_1, T_2) > x_1, \min(T_2, T_3) > x_2) =$$

$$= P(T_1 > x_1, T_3 > x_1, T_2 > x_2, T_3 > x_2) =$$

$$= P(T_1 > x_1, T_2 > x_2, T_3 > \max(x_1, x_2))$$

$$\underbrace{P(X_1 < X_2)}_{\text{continua}} + P(X_1 > X_2) + P(X_1 = X_2) = 1$$

\nwarrow singular

$(1-\alpha) \qquad \qquad \alpha$

$$S(x_1, x_2) = \alpha S_S(x_1, x_2) + (1-\alpha) S_\phi(x_1, x_2)$$

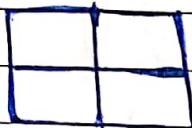
$$\frac{\partial^2}{\partial x_1 \partial x_2} S(x_1, x_2) \geq 0 \quad e \quad \frac{\partial^2}{\partial x_1 \partial x_2} S_S(x_1, x_2) = 0$$

14.05.13

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Ex 2 Mais de 20 medidas



Kendall's Tau.

 (X, Y) e (X_i, Y_i) com mesma distribuição
 \downarrow
 $i = 1, 2$
 $H(x_i, y_i)$
 \downarrow
 $C(u_i, v_i)$

contínua

def.

$$\tau(X, Y) = \underbrace{P[(X_1 - X_2)(Y_1 - Y_2) > 0]}_{1 - P[(X_1 - X_2)(Y_1 - Y_2) \leq 0]}$$

□	○
○	□

○ concordância
□ discordância

Logo,

$$\tau(X, Y) = 2 \underbrace{P[(X_1 - X_2)(Y_1 - Y_2) > 0]}_{P(X_1 > X_2, Y_1 > Y_2) + P(X_1 < X_2, Y_1 < Y_2)} - 1$$

$$P(X_1 > X_2, Y_1 > Y_2) + P(X_1 < X_2, Y_1 < Y_2)$$

e

$$P(X_1 > X_2, Y_1 > Y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(X_2 \leq x_1, Y_2 \leq y_1) dH(x_1, y_1) =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(F(x), G(y)) dC(F(x), G(y)) =$$

	S	T	Q	Q	S	S	D
SEMANA	<input type="checkbox"/>						
	L	M	M	J	V	S	D

$$= \int_0^1 \int_0^1 C(u, v) dC(u, v) = \int_0^1 \int_0^1 C(u, v) c(u, v) du dv$$

$$P(X_1 \leq X_2, Y_1 \leq Y_2) = \int_0^1 \int_0^1 [1 - u - v + C(u, v)] dC(u, v)$$

$$\int_0^1 \int_0^1 1 dC(u, v) = \int_0^1 \int_0^1 c(u, v) du dv = 1$$

$$\int_0^1 \int_0^1 u dC(u, v) = \int_0^1 u \int_0^1 c(u, v) du dv = E(u) = \frac{1}{2}$$

$$\int_0^1 \int_0^1 v dC(u, v) = \int_0^1 v \int_0^1 c(u, v) du dv = E(v) = \frac{1}{2}$$

$$P(X_1 \leq X_2, Y_1 \leq Y_2) = \int_0^1 \int_0^1 C(u, v) dC(u, v)$$

Logo

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

$$= 4 E[C(u, v)] - 1$$

\downarrow
z - variável aleatória de Kendall

$$= 4 E(z) - 1$$

Sejam $(x_1, y_1), \dots, (x_n, y_n) \rightarrow$ amostra ind. de (X, Y)

Sejam R_1, R_2, \dots, R_n - postos de (x_1, \dots, x_n)

S_1, S_2, \dots, S_n - postos de (y_1, \dots, y_n)

S	T	Q	Q	S	S	D

L M M J V S D SEMANA

$$T_n = \frac{2}{n^2 - n} \sum_{1 \leq i < j \leq n} \text{sign}(R_i - R_j) \text{ sign}(S_i - S_j) \quad \text{versão comstral}$$

Versão mais simples

$$T_n = \frac{c - d}{\binom{n}{2}}, \quad c = \# \text{ pares concordantes}$$

$$d = \# \text{ .. discordantes}$$

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2} = \frac{n^2 - n}{2}$$

~~Se~~ é absurdo utilizar T_n quando temos uma variável ordinal e outra variável não ordinal.

Antes de usar T_n , verificar se as variáveis são ordinais.

Invariante sobre transformações crescentes.

$$(x(x), \beta(y)) \text{ então } G_{(x(u), \beta(v))}(u, v) = G_{x(u)}(u, v) \Rightarrow$$

crescente crescente

Kendall's tau é invariante sob transformações crescentes.

Spearman's rho

$$(x_i, y_i) \quad i=1, 2, 3$$

versão populacional

def

$$\rho(x, y) =$$

versão amostral

$$p_n = \frac{12}{n^3 - n} \sum_{i=1}^n R_i S_i - \frac{3(n+1)}{n-1}$$

$$= \frac{3}{n^2} \left[P[(x_1 - x_2)(y_1 - y_2) \geq 0] - P[(x_1 - x_2)(y_1 - y_2) \leq 0] \right]$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

(X_i, Y_i) , $i = 1, 2, 3$ são cópias de (X, Y)

$$\rho(X, Y) = \rho(U, V) = \frac{E(UV) - E(U)E(V)}{\sqrt{\text{Var}(U)} \sqrt{\text{Var}(V)}} = \frac{E(UV) - \frac{1}{2} \cdot \frac{1}{2}}{\sqrt{\frac{1}{12}} \sqrt{\frac{1}{12}}} \Rightarrow$$

coef. de
correlação
de U e V

$$\Rightarrow \rho(U, V) = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3$$

$$\rho(U, V) = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3$$

$$-1 \leq 3\tau - 2\rho \leq 1$$

Se estamos trabalhando com o mesmo tipo de variável, altura e altura, faz até sentido usar posterior. Mas se utilizarmos medidas de altura e peso, já não faria tanto.

Medida de Blest (2000)

versão amostral

$$v_n = \frac{\text{def}}{n-1} \frac{2n+1}{n^2-n} \sum_{i=1}^n \text{Si} \left(1 - \frac{R_i}{(n+1)} \right)^2 \in [-1, 1]$$

$$\text{populacional. } V(X, Y) = 2 - 12 \int_0^1 \int_0^1 (1-n)^2 v dC(u, v)$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

e

$$Y(X, Y) = p(X, Y) - 6 \mathbb{E}[U^2 V - (1-U)^2 (1-V)]$$

$$Y(-X, Y) = Y(X, Y) - 2p(X, Y)$$

Todas estas medidas estão entre -1 e 1.

Todas " " são iguais a 0 quando as v.a. são ind.

$$\max(f(x) + g(y) - 1, 0) \leq h(x, y) \leq \min(f(x), g(y))$$

$$\tau_L \leq \tau \leq \tau_U$$

$$\rho_L \leq \rho \leq \rho_U$$

15.05.13

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

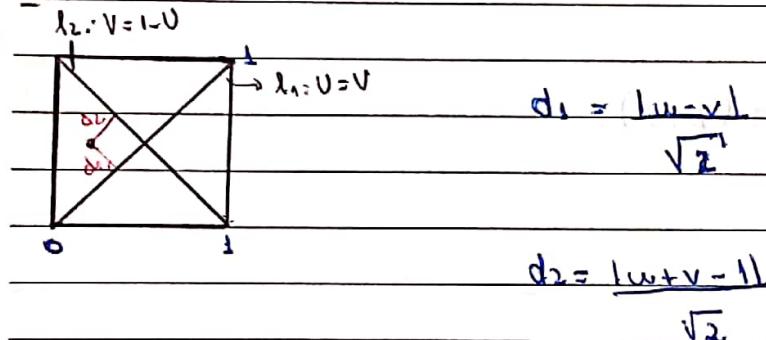
SEMANA

Análise de Dependência Por Meio de Cúpulas

Nikolai

"Se você quer mudar um sistema, você deve conhecer muito bem esse sistema".

"Ninguém sabe como vai se portar em uma situação que ainda não conhece."



$$\lambda(u,v) = d_2^2 - d_1^2 = 4uv - 2(w+v) + 1$$

2)

1) $\lambda(u,v) > 0$ se (u,v) é mais próximo de l_1 do que de l_2 .

2) $\lambda(u,v) < 0$ se (u,v) é mais próximo de l_2 do que de l_1 .

3) $\lambda(u,v) = 0$ para pontos (u,v) "equidistantes": $(u,0.5)$ e $(0.5,v)$

4) $\lambda_{\min}(u,v) = -0.5$ para $(0,1)$ e $(1,0)$

5) $\lambda_{\max}(u,v) = 0.5$

para $(0,0)$ e $(1,1)$

S	T	Q	Q	S	S	D
□	□	□	□	□	□	□

SEMANA L M M J V S D

$U, V \sim U(0,1)$ e $(U, V) \sim C(u, v)$

$$E(\Delta(U, V)) = \int_0^1 \int_0^1 \Delta(u, v) d_{UV} C(u, v) =$$

$$= \frac{1}{2} E \left[4UV - 2(U+V) + 1 \right] = 2E(UV) - 1/2 \Rightarrow$$

$$\delta E[\Delta(U, V)] = \rho(x, y)$$

já que $\rho(x, y) = 12E(UV) - 3$

"Dados são quem falam"

Versão Local

$$\rho(x, y) = \text{corr}(x, y) = \frac{E[(U - E(U|V=v_0))(V - E(V|U=u_0))]}{\sqrt{E(U - E(U|V=v_0))^2} \sqrt{E(V - E(V|U=u_0))^2}}$$

$$(u_1, v_1) \rightarrow \rho(u_1, v_1) = z_1$$

se 1000 observações

reduzir os dados mais

uma vez ...

ideia

$$(u_2, v_2) \rightarrow \rho(u_2, v_2) = z_2$$

reduzir os dados mais

uma vez ...

ideia

$$(u_n, v_n) \rightarrow \rho(u_n, v_n) = z_n$$

Curvas de níveis

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA



$$0,7 \leq p \leq 0,9$$

$$0,3 \leq p \leq 0,7$$

Medidas de dependência local \rightarrow várias.

"Gaiola - povão".

Marginal não tem influência \propto chefe & totalitarismo

(X, Y) vetor a. abs. contínuo com dist. $H(x, y) \Rightarrow h_{xy}(x, y) \geq H(x, y)$

$P(X \leq x) = F(x)$ $P(Y \leq y) = G(y)$ $dx dy$

ordenada nominal

$$\text{Cov}(Y, F(X)) = E[(Y - E(Y))(F(X) - E(F(X)))] =$$

$$= E[(Y - E(Y))(F(X) - 1/2)] =$$

$$= E(YF(X) - Y/2 - E(Y)F(X) + E(Y)/2) =$$

$$= E(YF(X)) - \frac{E(Y)}{2} - \frac{E(Y)F(X)}{2} + \frac{E(Y)}{2}$$

$$= E(YF(X)) - \frac{E(Y)}{2}$$

$$\text{com } E(YF(X)) = \iint y F(x) h_{XY}(x, y) dx dy$$

Analogamente,

$$\text{Cov}(X, G(Y)) = E(XG(Y)) - \frac{EX}{2}$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

Nas próximas aulas...

$$\text{Cov}(Y, F(X)) \stackrel{?}{=} E[(Y_1 - Y_2) \text{ sign}(X_1 - X_2)]$$

Se (X_1, Y_1) e (X_2, Y_2) são cópias independentes de (X, Y) .

Gini Measure:

$$\Gamma(Y, X) = \frac{\text{Cov}(Y, F(X))}{\text{Cov}(Y, G(Y))} = 0 \Leftrightarrow X \text{ e } Y \text{ são independentes}$$

Note que:

$$\Gamma(Y, X) \neq \Gamma(X, Y) = \frac{\text{Cov}(X, G(Y))}{\text{Cov}(X, F(X))}$$

O que significa que não é uma medida de simetria, enquanto que

$$\rho_{XY} = \rho_{YX}$$

21.05.13

S	T	Q	Q	S	S	D
	☒					

L M M J V S D
SEMANA

Modelos de Regressão com Cúpulas

Copula-Based Regression Model

Antes, sobre cúpulas no caso discreto

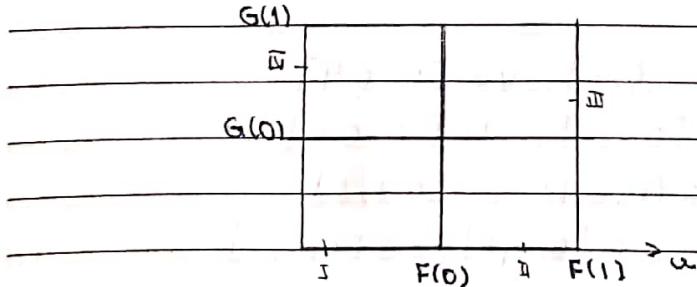
X/Y	0	1	$g(i_2)$
0	0.25	0.15	0.4
1	0.15	0.45	0.6
$f(i_1)$	0.4	0.6	1.0

$$H(i_1, i_2) = \begin{cases} 0.25 & \text{se } i_1=0, i_2=0 \\ 0.4 & \text{se } i_1=0, i_2=1 \\ 0.4 & \text{se } i_1=1, i_2=0 \\ 1 & \text{se } i_1=1, i_2=1 \\ 0 & \text{se } i_1 < 0, i_2 > 0 \end{cases}$$

Agora, temos que encontrar C tal que:

$$H(i_1, i_2) = C(F_x(i_1), F_y(i_2))$$

Para tal, vejamos o esquema gráfico abaixo



Agora, vamos definir probabilidades para cada uma das regiões definidas acima.

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

$$C_D(u, v) = \frac{p(i_1, i_2)}{f(i_1) g(i_2)}$$

$$\sum_{i_1} \sum_{i_2} \frac{p(i_1, i_2)}{f(i_1) g(i_2)} = 1$$

$$C_D = \begin{cases} (1) \frac{25}{16} & u \in [F(-1), F(0)], v \in [G(-1), G(0)] \\ (2) \frac{15}{16} & u \in [F(0), F(1)], v \in [G(0), G(1)] \\ (3) \frac{15}{16} & u \in [F(-1), F(0)], v \in [G(0), G(1)] \\ (4) \frac{45}{36} & u \in [F(0), F(1)], v \in [G(0), G(1)] \end{cases}$$

Agora,

$$C_D f(u, v) = \int_0^u \int_0^v C_D(s, t) ds dt$$

$$(1) \Rightarrow C_D(u, v) = \int_0^u \int_0^v \frac{25}{16} ds dt = \frac{25}{16} uv, \quad u \in [0, 0.4], v \in [0, 0.4]$$

$$(2) \Rightarrow C_D(u, v) = \int_0^{0.4} \int_0^v C_D(u, v) ds dt = \int_{0.4}^{0.4} \int_{0 \leq s \leq u} C_D(u, v) ds dt$$

$$+ \int_{0.4=s}^u \int_{s=t}^v C_D(u, v) ds dt \quad u \in [0.4, 1], v \in [0, 0.4]$$

$$C_D(u, v) = \begin{cases} \frac{25}{16} uv & u \in [0, 0.4], v \in [0, 0.4] \\ \frac{3}{8} v + \frac{5}{8} uv & u \in [0.4, 1], v \in [0, 0.4] \\ \frac{3}{8} u + \frac{5}{8} uv & u \in [0, 0.4], v \in [0.4, 1] \\ \frac{9}{16}(1-u-v) + \frac{25}{16} uv & u \in [0.4, 1], v \in [0.4, 1] \end{cases}$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

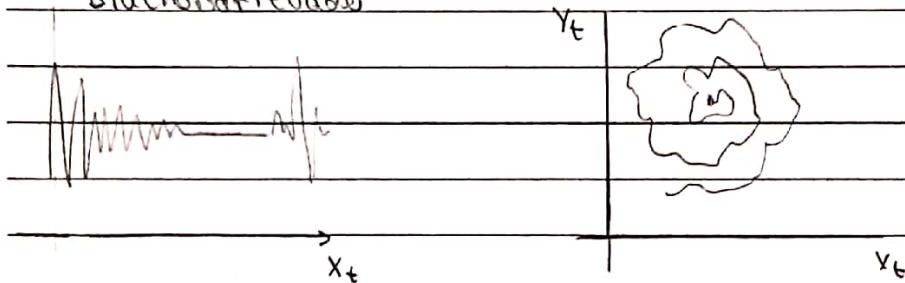
Niewiadomska-Bugaj and Kowalczyk (2005).

Grade transformation. Extension of the Probability Integral transform

Pitt et. al (2006). Gaussian copula regression

Orsi and Greenberg (1998) multivariate probit model

Estacionariedad



22.05.13

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Análise - Clusterização + Variáveis Latentes

Considera X uma v.a. contínua (também valeria p/ o caso discreto)

Então, temos (X é não negativa)

$$P(X \leq t) = F(t) \Rightarrow P(X > t) = 1 - F(t) = \bar{F}(t)$$

$$\text{com } f(t) = \frac{d}{dt} F(t) = F'(t).$$

Taxa de Mortalidade.

$$\tau(t) = \frac{f(t)}{F(t)} = P(t \leq X \leq t + \Delta t / X > t)$$

Taxa de Mortalidade Acumulada

$$R(t) = \int_0^t \tau(x) dx$$

Das relações, segue que:

$$\bar{F}(t) = \exp \{-R(t)\}$$

Logo. De modo geral,

$$X \rightarrow f(x), F(x), \tau(x), R(x)$$

ou seja, a distrib. de X é unicamente determinada por qualquer uma das 4 funções acima. (No caso univariado)

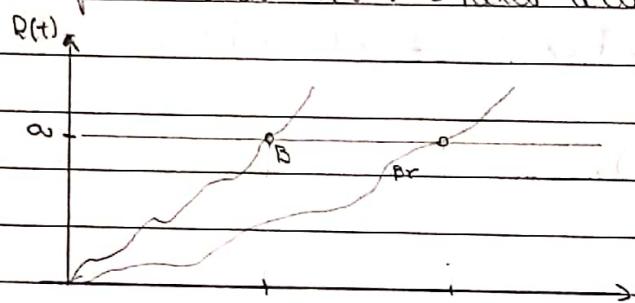
S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

Uma vez que

$$(*) \quad \bar{F}(t) = \exp\{-R(t)\} \Rightarrow$$

$$P(X > t) = P(T > \omega | R(t)) \text{ com } T \sim \text{Exp}(1)$$

X - tempo de vida $R(t)$ - taxa acumulada.



No física, conhecido como Hazard Potential.

No estatística, introduzida em 2006 por Nader Singpurwalla.

No caso bivariado,

$$\bar{F}(t_1, t_2) = P(X_1 > t_1, X_2 > t_2)$$

$$\bar{F}_i(t_i) = P(X_i > t_i)$$

$$\bar{F}(t_1, t_2) = \frac{\bar{F}_1(t_1) \bar{F}_2(t_2)}{\bar{F}_1(t_1) \bar{F}_2(t_2)}$$

$$= \bar{F}_1(t_1) \bar{F}_2(t_2) \exp \left\{ \ln \left[\frac{\bar{F}(t_1, t_2)}{\bar{F}_1(t_1) \bar{F}_2(t_2)} \right] \right\}$$

Logo, usando a relação (*), temos que

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

$$\bar{F}(t_1, t_2) = \exp \{-R_1(t_1) - R_2(t_2) + A(t_1, t_2)\}$$

onde

$$A(t_1, t_2) = \frac{\ln \bar{F}(t_1, t_2)}{\bar{F}_1(t_1) \bar{F}_2(t_2)} = \frac{\ln \bar{F}(t_1, t_2)}{F_1(t_1) F_2(t_2)}$$



> 1 - Posit.

< 1 - Neg.

$= 1$ - Indep.

Ou seja,

$$\bar{F}(t_1, t_2) = \exp \{-H(t_1, t_2)\}$$

com

$$-H(t_1, t_2) = R_1(t_1) + R_2(t_2) - A(t_1, t_2)$$

No caso bivariado, taxa de mortalidade é definida por

$$Z(t_1, t_2) = \frac{f(t_1, t_2)}{\bar{F}(t_1, t_2)}$$



Não caracteriza distribuição bivariada unicamente.

Para caracterizar ~~distribuição~~ precisa, características condicionais adicionais

$$\tau_i(t_1, t_2) = \frac{\partial}{\partial t_i} \left[-\ln \bar{F}(t_1, t_2) \right], i=1,2$$

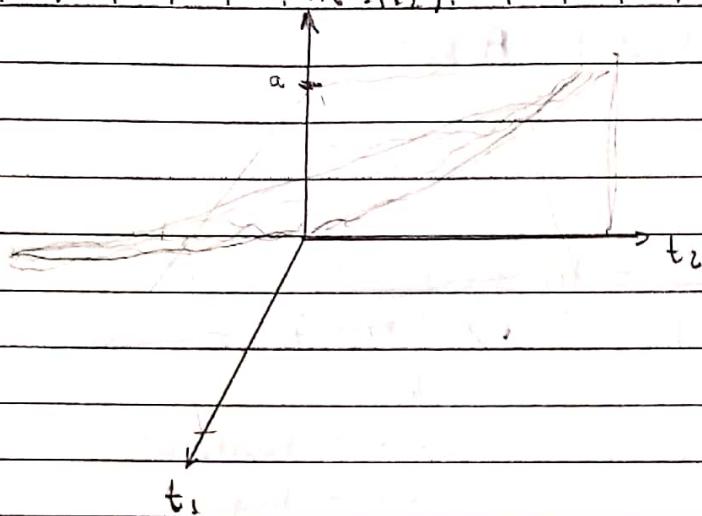
$$(\tau_1(t_1, t_2), \tau_2(t_1, t_2)) \neq Z(t_1, t_2)$$

hazard gradient function

$Z(t_1, t_2)$ e $(\tau_1(t_1, t_2), \tau_2(t_1, t_2))$ → determina unicamente a distribuição bivariada.

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

$H(t_1, t_2)$.

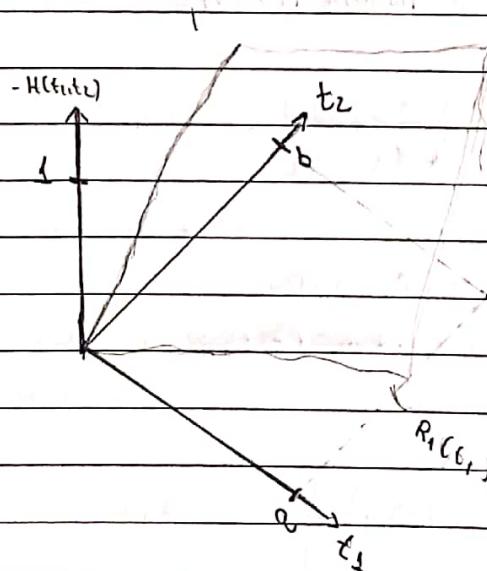


* Também podemos fazer a construção considerando

$$F(x_1, x_2) = F_1(x_1) F_2(x_2) \exp \left[\ln F(x_1, x_2) \right] / F(x_1) F(x_2)$$

O que resultaria na mesma expressão vista anteriormente.

Ideal para censuras à direita. →



28/05/2013

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

Opulas.

M.O. (67)

$$(x_1, x_2) = [\min(T_1, T_2), \min(T_2, T_3)] \quad T_1, T_2, T_3 - \text{ind. exp.}$$

$$S_{x_1, x_2}(x_1, x_2) = P(X_1 \geq x_1, X_2 \geq x_2) = \begin{cases} S_1(x_1 - x_2) \exp(-\theta x_1) & \text{se } x_1 \geq x_2 \\ S_2(x_2 - x_1) \exp(-\theta x_2) & \text{se } x_2 \geq x_1 \\ \exp(-\theta x_1) & \text{se } x_1 = x_2 \end{cases}$$

$$\frac{\partial^2 S_{x_1, x_2}(x_1, x_2)}{\partial x_1 \partial x_2} = f_{x_1, x_2}(x_1, x_2)$$

$$S_{x_1, x_2}(x_1, x_2) = (1-\alpha) S_{x_1, x_2}^C(x_1, x_2) + \alpha S_{x_1, x_2}^{sin}(x_1, x_2)$$

$$\Downarrow \quad \Downarrow$$

$$P(X_1 \geq x_2) + P(X_2 \geq x_1) \quad P(X_1 = x_2)$$

$$C(u, v) = \begin{cases} u^{1-\beta} v^\beta & \text{se } u^\beta \geq v^\beta, \quad u^\beta = v^\beta \Rightarrow u = v^{\beta/2} \\ u^\beta v^{1-\beta} & \text{se } u^\beta \leq v^\beta \end{cases}$$



$$X \sim \text{Exp}(\lambda) \Rightarrow S(x) = \exp(-\lambda x)$$

$$S(x+t) = S(x) \delta(t) \quad \forall x, t \geq 0$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

$$S(x) = S(x+t) = P(X \geq x+t) = x+ = [x-t / X \geq t]$$

$$\downarrow \quad S(t) \quad \quad P(X \geq t)$$

↓
residual lifetime

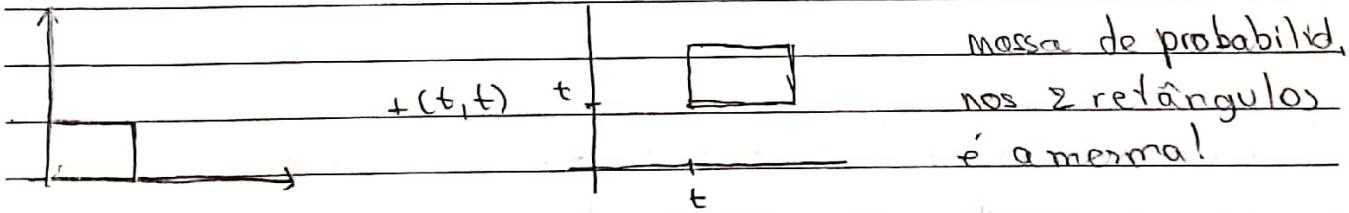
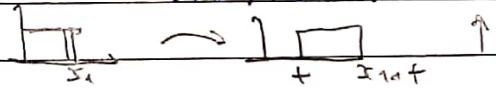
$$X \stackrel{d}{=} X_t$$

No caso bivariado,

$$S_{x_1, x_2}(x_1, x_2) = \exp\{-\lambda_1 x_1 - \lambda_2 x_2 - \theta \max(x_1, x_2)\}$$

$$S_{x_1, x_2}(x_1+t, x_2+t) = S_{x_1, x_2}(x_1, x_2) S_{x_1, x_2}(t, t)$$

$$\frac{P(X_1 \geq x_1+t, X_2 \geq x_2+t)}{P(X_1 \geq t, X_2 \geq t)} = \frac{S_{x_1, x_2}(x_1+t, x_2+t)}{S_{x_1, x_2}(t, t)} = \underbrace{S_{x_1, x_2}(t, t)}_{(x_1, x_2)}$$



$$\begin{aligned} \frac{\partial^2}{\partial x_1 \partial x_2} S_{x_1, x_2}(x_1, x_2) &= \frac{\partial^2}{\partial x_1 \partial x_2} [S_{x_1, x_2} \exp(-\theta x_1)] = \\ &= \frac{\partial}{\partial x_2} [-f_{x_1}(x_1-x_2) \exp(-\theta x_1) + S(x_1-x_2)(-\theta) \exp(-\theta x_1)] \\ &= \exp(-\theta) [f'_{x_1}(x_1-x_2) - \theta f_{x_1}(x_1-x_2) \exp(-\theta x_1)] \end{aligned}$$

Devemos ter

$$0 \leq \exp(-\theta) [f'_{x_1}(x_1-x_2) - \theta f_{x_1}(x_1-x_2)] \leq$$

$$\frac{d}{dx_1} \ln f_{x_1}(x_1-x_2) \geq 0$$

S	T	Q	Q	S	S	D
L	M	M	J	V	S	D

SEMANA

$$S_{x_1, x_2}(x_1, x_2) = \exp(-\kappa(x_1, x_2))$$

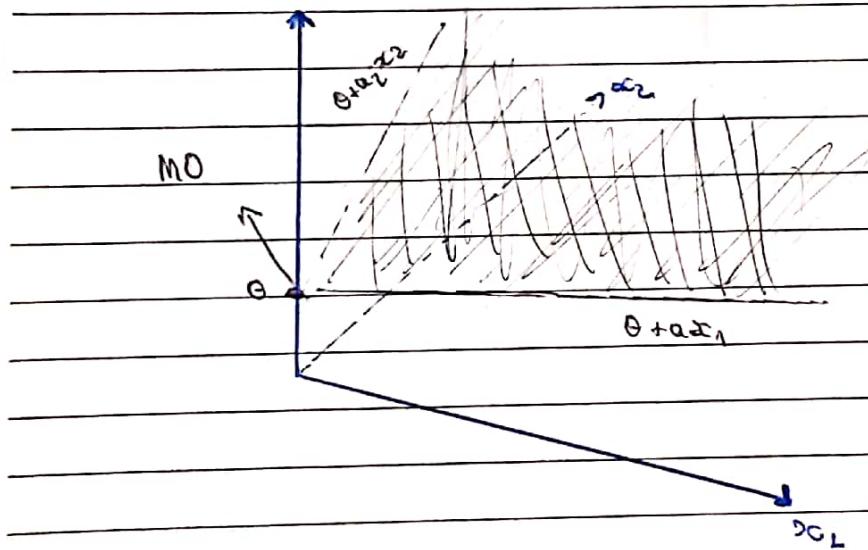
$$S_{x_1}(x_1) S_{x_2}(x_2)$$

$$\underline{S_{x_1}(x_1, x_2)}$$

$$S_{x_1}(x_1, 0) S_{x_2}(0, x_2)$$

→ falso domínio

$$\tau_1(x_1, x_2) + \tau_2(x_1, x_2) = \Theta + \alpha_1 x_1 + \alpha_2 x_2$$



BLMP;

$$(x_L / x_L > t) - LMP$$

e

$$(x_L / x_1 > t) - LMP$$

↓

Gumbel → abs. cont.

$$\exp(-\lambda_1 x_1 - \lambda_2 x_2 - \Theta \lambda_1 \lambda_2 x_1 x_2)$$

$$\tau_1(x_1, x_2) = \exp(-\Theta \lambda_1 x_1)$$

$$\tau_2(x_1, x_2) = \Theta + \alpha_2 x_2$$

30/04

Rosenthal (52)

24/04/2013

Seja $\underline{X} = (X_1, \dots, X_n)$ com $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$

$$\underline{Z} = (Z_1, \dots, Z_n) = T_n \underline{X}$$

$$Z_1 = P(X_1 \leq x_1) = F_1(x_1)$$

$$Z_2 = P(X_2 \leq x_2 | X_1 = x_1) = F_{z_1}(x_2 | x_1)$$

$$\vdots$$

$$Z_n = P(X_n \leq x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = F_{z_{n-1}}(x_n | x_1, \dots, x_{n-1})$$

$$\begin{aligned} P(Z_1 \leq z_1, \dots, Z_n \leq z_n) &= \int_0^{z_n} \dots \int_0^{z_1} dZ_n (F_{n|1,2,\dots,n-1}) \dots dZ_1 (F_1(x_1)) \\ &= \int_0^{z_n} \dots \int_0^{z_1} dz_1 \dots dz_n = \prod_{i=1}^n z_i \end{aligned}$$

$\Rightarrow Z_1, \dots, Z_n$ são independentes, uniformemente distribuídos

em $[0, 1]^n$.

PROBLEMA: ① Tem $n!$ transformações tipo T1.

② caso particulares

$$E_{\underline{X}} \underline{X} = (X_1, X_2) \sim N(m_1, m_2, \beta_1^2, \beta_2^2, \rho) \quad \text{com } (X_1, X_2)$$

$$F_{X_1}(x_1) = \phi \left(\frac{x_1 - m_1}{\sigma_1} \right)$$

$$\begin{aligned}
H(F^{-1}(u), G^{-1}(v)) &= 1 - \exp \left[\ln \cdot (1-u) \right] - \exp \left[\ln (1-v) \right] \\
&+ \exp \left[- \left(-\ln(1-u) - \ln(1-v) + \theta \ln(1-u) \ln(1-v) \cdot \right) \right] \\
&= \frac{-1 - (1-u) \cdot (1-v) + (1-u)(1-v) \exp \left\{ -\theta \ln(1-u) \ln(1-v) \right\}}{\underline{1 - (1-u) \cdot (1-v)}} \\
&= \underline{I(u, v)}
\end{aligned}$$

$$\begin{aligned}
\bar{I}(u, v) &= \cancel{u} \cancel{+ v - 1 + 1 - u - v + uv \exp \left\{ -\theta \ln(u) \ln(v) \right\}} \\
\underline{(\bar{I}(u, v) = uv \exp \left\{ -\theta \ln u \ln v \right\})} &
\end{aligned}$$

$$\begin{cases} \bar{F}(u) = e^{-x} = u \\ \bar{G}(y) = e^y = v \end{cases} \quad \begin{cases} x = \bar{F}^{-1}(u) = -\ln u \\ y = \bar{G}^{-1}(v) = -\ln v \end{cases}$$

$$\begin{aligned}
\bar{H}(x, y) &= 1 - F(x) - G(y) + H(x, y) = \\
&= \underline{1 - (1 - e^{-x}) - (1 - e^{-y}) + 1 - e^{-x} - e^{-y} + e^{-x+y + \theta xy}}
\end{aligned}$$

$$\begin{aligned}
H(x, y) &= e^{-x+y+\theta xy}, \Rightarrow \bar{H}(\bar{F}^{-1}(u), \bar{G}^{-1}(v)) = \\
&\exp \left\{ - \left[-\ln(u) - \ln(v) + \theta \ln u \ln v \right] \right\} = uv \exp \left\{ -\theta \frac{\ln u}{\ln v} \right\}
\end{aligned}$$

$$H(x,y) = \bar{F}(x) + \bar{G}(y) - 1 + C(1-\bar{F}(x), 1-\bar{G}(y))$$

$$\boxed{\bar{E}(u,v) \stackrel{\text{def}}{=} u+v-1+C(1-u, 1-v)}$$

survival copula

$$\underline{E}(u,v) = P(U>u, V>v) = 1-u-v + C(u,v) = \underline{\bar{E}}(1-u, 1-v)$$

$$\bar{H}(x,y) = \bar{E}(\bar{F}(x), \bar{G}(y)) \quad] \quad \text{some thanks}$$

$$\ell(F(x), G(y)) = H(x,y) \quad]$$

$$\text{define } x = \bar{F}^{-1}(u), y = \bar{G}^{-1}(v)$$

$$\boxed{\bar{H}(\bar{F}^{-1}(u), \bar{G}^{-1}(v)) = \bar{\bar{E}}(u, v)} \quad]$$

Ex. (Gumbel bivariate)

$$H(x,y) = 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)}$$

$$\bar{F}(x) = \lim_{y \rightarrow \infty} H(x,y) = 1 - e^{-x} = u \quad \left\{ F^{-1}(u) = -\ln(1-u) \right.$$

$$\bar{G}(y) = \lim_{x \rightarrow \infty} H(x,y) = 1 - e^{-y} = v \quad \left\{ G^{-1}(v) = -\ln(1-v) \right.$$

$$c(F(x), G(y)) f(x) g(y) = h(x, y)$$

$$\Rightarrow c(F(x), G(y)) = \frac{h(x, y)}{f(x) g(y)}$$

$$u = F(x) \Rightarrow x = F^{-1}(u)$$

$$v = G(y) \Rightarrow y = G^{-1}(v)$$

$$c(u, v) = \frac{h(F^{-1}(u), G^{-1}(v))}{f(F^{-1}(u)) g(G^{-1}(v))}$$

} if you have
h, f, g $\Rightarrow c$

$$c(u, v) = \int_0^u \int_0^v c(w, z) dw dz$$

survival copulas

$$P(X > x, Y > y) = \bar{H}(x, y) = 1 - F(x) - G(y) + H(x, y)$$

$$\left\{ \begin{array}{l} \bar{F}(u) = 1 - F(u) \Rightarrow F(u) = 1 - \bar{F}(u) \\ \bar{G}(v) = 1 - G(v) \Rightarrow G(v) = 1 - \bar{G}(v) \end{array} \right\} \left\{ \begin{array}{l} \bar{F}(u) = u \\ \bar{G}(v) = v \end{array} \right.$$

$$\bar{H}(x, y) = \bar{F}(x) + \bar{G}(y) - 1 + H(x, y)$$

$$\underline{F_{X,Y}(t)} = P(X \leq t) = \lim_{s \rightarrow \infty} F_{X,Y}(s, t) \stackrel{F(s) \rightarrow 1}{=} 1$$

$$= \int_0^1 DIC\left(u, G\left(\frac{t}{F^{-1}(u)}\right)\right) du$$

If substitute $C(u, v) = uv \Rightarrow DIC(u, v) = v$

$$\boxed{F_{X,Y}(t) = \int_{-\infty}^{\infty} G\left(\frac{t}{x}\right) dF(x)}$$

Classical probability

$$h(x, y) = \frac{\partial^2}{\partial x \partial y} H(x, y)$$

$$\frac{\partial^2 P(X+Y \leq t)}{\partial t} = \int_{-\infty}^{\infty} G(t-x) dF(x) \quad (\text{density function})$$

$$= \int_{-\infty}^t h(x, t-x) dx$$

regular density

$$h(x, y) = \frac{\partial^2}{\partial x \partial y} [H(x, y)] = \frac{\partial^2}{\partial x \partial y} [\mathcal{E}(F(x), G(y))] = \frac{\partial}{\partial x} (F(x), G(y)) \cdot \frac{\partial}{\partial x} F(x), \frac{\partial}{\partial y} G(y)$$

$$P(X+Y \leq t) = \int_0^1 D_1 C(u, G(t - F^{-1}(u))) du$$

$$(X, Y) \rightarrow \mathcal{E} \quad D_1 = \frac{\partial}{\partial u} C(u, v) \quad C(u, v) = uv$$

$P(XY \leq t)$ It turns useful in cases in which

we have products of discount factors

both in temporal settings - (ie forward prices)

or in spatial setting like risk free
discount factor - firms survival prob.)

$$H(x, y) = P(X \leq x, Y \leq y)$$

$$= C(F(x), G(y))$$

$$F_{X, XY}(x, t) = P(X \leq x, XY \leq t) = \int_{-\infty}^x P(XY \leq t | X=x) dF(x)$$

$$H(F(x), \mathbb{E}(t/x))$$

$$= \int_0^x P(Y \leq \frac{t}{x} | X=x) dF(x) = \int_0^1 \frac{\partial}{\partial x} H(F(x), \mathbb{E}(t/x)) dF(x)$$

$$\underbrace{\frac{\partial}{\partial x} H(x, t/x)}$$

substitute - $u = F(x) \rightarrow x = F^{-1}(u) \in (0, 1)$

$$x \rightarrow 0 \Rightarrow u \rightarrow 0$$

$$x \rightarrow 1 \Rightarrow u \rightarrow F(1)$$

$$\int_0^{F(1)} \frac{\partial}{\partial u} H(u, G(\frac{t}{F^{-1}(u)})) du$$

Análise de Dependência por Meio de Cúplas

30/04/2013

Seja $\underline{X} = (X_1, \dots, X_n)$ com $F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$

$$\underline{Z} = (Z_1, \dots, Z_n) = T_{\underline{X}} \underline{X}$$

$$Z_1 = P(X_1 \leq x_1) = F_1(x_1)$$

$$Z_2 = P(X_2 \leq x_2) = F_2(x_2 | x_1)$$

⋮

$$Z_n = P(X_n \leq x_n) = F_n(x_n | x_1, \dots, x_{n-1})$$

$$P(Z_1 \leq z_1, \dots, Z_n \leq z_n) = \int_{-\infty}^{z_n} \cdots \int_{-\infty}^{z_1} dx_n (F_{n|1,2,\dots,n-1}) \cdots dx_1 F_1(x_1) =$$

$$dz_1 = dx_1 F_1(x_1)$$

$$= \int_{-\infty}^{z_n} \cdots \int_{-\infty}^{z_1} dx_n (F_{n|1,2,\dots,n-1}) \cdots dx_1 F_1(x_1) =$$

$$= \int_{-\infty}^{z_n} \cdots \int_{-\infty}^{z_1} dz_1 \cdots dz_n = \prod_{i=1}^n z_i \quad \Rightarrow \text{Z}_i \text{ sаo independentes com } z_i \sim U(0,1), i=1, \dots, n.$$

Problema:

Rosenblatt caso multivariado

① tem $n!$ transformações

② Casos particulares

Ex:

$$\underline{X} = (X_1, X_2) \sim N_2(m_1, m_2, \sigma_1^2, \sigma_2^2, \rho)$$

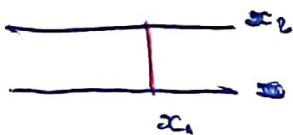
$$F_{X_1}(x_1) = \Phi\left(\frac{x_1 - m_1}{\sigma_1}\right)$$

$$F_{X_2|X_1}(x_2) = \Phi\left(\frac{x_2 - m_2 - \frac{\rho \sigma_1}{\sigma_2} (x_1 - m_1)}{\sigma_2 \sqrt{1 - \rho^2}}\right)$$

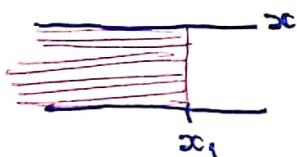
$$X_1 \sim N(m, \sigma_1^2)$$

$$X_2 | X_1 \sim N\left(m_2 + \frac{\rho \sigma_1}{\sigma_2} (x_1 - m_1), \sigma_2^2(1 - \rho^2)\right).$$

Problema Aberto! (Dá pra resolver, eu acho)



$$P(X_2 \leq x_2 | X_1 = x_1) = 0 \text{ Não traz informação.}$$



$$P(X_2 \leq x_2 | X_1 \leq x_1) \neq 0 \text{ - Traz informação.}$$

Problema aberto. Qual seria versão de Rosenblatt

$$T(Y) = (Y_1, Y_2)$$

$$\begin{cases} y_1 = P(X_1 \leq x_1) = F_1(x_1) \\ y_2 = P(X_2 \leq x_2 | X_1 \leq x_1) = \frac{P(X_1 \leq x_1, X_2 \leq x_2)}{P(X_1 \leq x_1)} = \frac{F(x_1, x_2)}{F(x_1)} \end{cases}$$

$$G(x, \lambda) = P(X < x) + \lambda P(X = \infty) \stackrel{\text{cont.}}{\rightarrow} P(X \leq x)$$

$$U = G(X, \Lambda) = G(X-) + \Lambda [G(X) - G(X-)]$$

Se $\Lambda \sim U(0,1)$ e ind. de X

$\Rightarrow U \sim U(0,1)$.

$$X = (X_1, \dots, X_n) \xrightarrow{T_2} T_2 X = (Z_1, \dots, Z_n) + \Delta_1, \dots, \Delta_n \sim^{\text{ind.}} U(0,1)$$

\downarrow independente de X_1, \dots, X_n .

$$F_1(x_1-) = P(X_1 \leq x_1)$$

$$F_2(x_2 | x_1) = P(X_2 \leq x_2 | X_1 = x_1), \text{ etc.}$$

$$Z_1 = (1 - \Delta_1) F_1(x_1-) + \Delta_1 F_1(x_1)$$

$$Z_2 = (1 - \Delta_2) F_2(x_2 | x_1) + \Delta_2 F_2(x_2 | x_1)$$

⋮

$$Z_n = (1 - \Delta_n) F_n(x_n | x_1, \dots, x_{n-1}) + \Delta_n F_n(x_n | x_1, \dots, x_{n-1})$$

Teorema. $P(Z_1 \leq z_1, \dots, Z_n \leq z_n) = \prod Z_i$ se Z_i 's independentes e $Z_i \sim U(0,1)$

Rüchendorf (2009).

$$(X, Y) \stackrel{H}{\sim} C(\omega, \nu)$$

$$\hat{F}_x(x) = U = F_X(x)$$

$$C(v/\omega) = \frac{\partial}{\partial \omega} C(\omega, \nu) \quad Z_2 = C(V/U) = C(G_Y(y)/F_X(x))$$

Nesse caso, via cópulas

Z_1 e Z_2 são independentes e $Z_i \sim N(0, 1)$, $i=1, 2$.

$$H_0: (X, Y) \sim C(\omega, \nu) \quad \text{vs} \quad H_A: (X, Y) \not\sim C$$

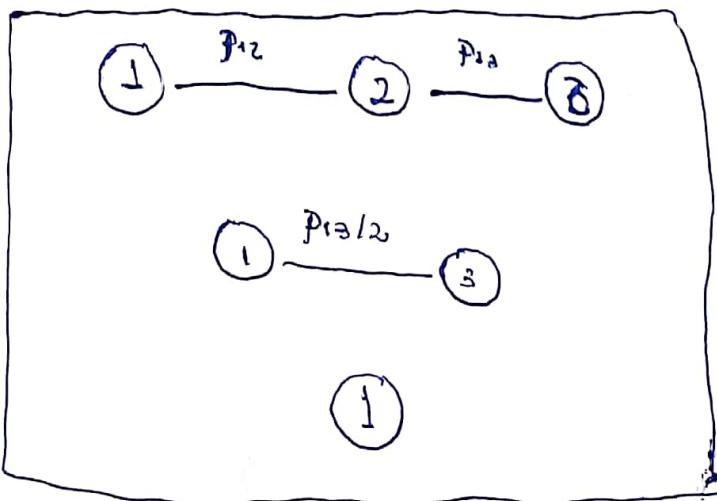
$$S(X, Y) = \left[P^2(F_X(x))^2 + \left[\Phi^{-1}(C(G_Y(y)/F_X(x))) \right]^2 \right] \sim \chi^2_{(2)}$$

↓ ∇ não é sempre verdadeira (Em dist. simétricas, isso é verdade)

$$H_0^*: S(X, Y) \sim \chi^2(2) \quad \text{vs.} \quad H_A: S(X, Y) \text{ não é } \chi^2(2).$$

state space Models (89)

Vine copula ou pair copula (estrutura vine)

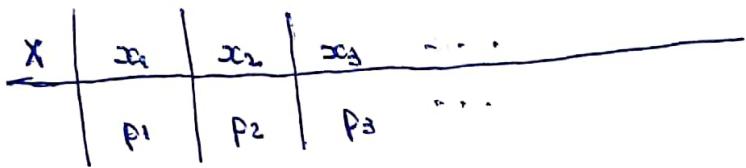


3 especialistas
tornando decisões.
No segundo passo,
apenas 2 prevaleceram,
mas com a influência
do que ficou pra trás.
Depois, segue a
última decisão influenciada
por todos mas tomada
por apenas 1.

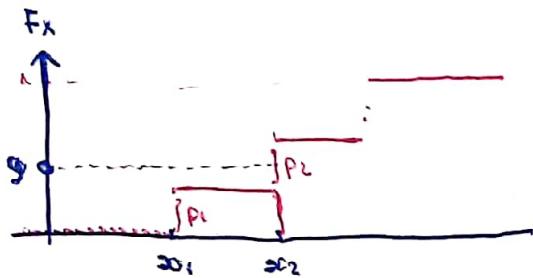
Estrutura
vine copula
(lembra urna)
uma

{ Roger Cooke (TUDelft)
Dorota Kurowicka

Harry Joe
Claudia Czado

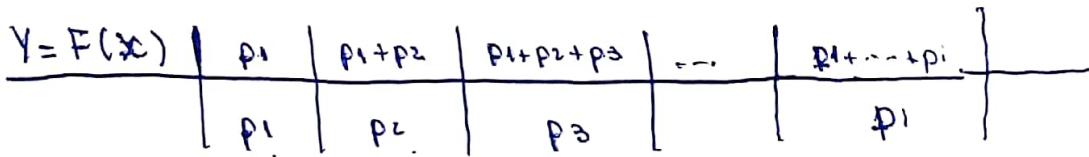


$$F_X(x) = P(X \leq x) = \sum_{j \leq i} p_j \quad \text{para } x_j \leq x < x_{j+1}$$



$$Y = F_X(x)$$

Obs. $P(Y = F_X(x) \leq y) \leq y$



$$P(Y = F_X(x) \leq y) = \sum_{j \leq i} p_j \quad \text{se } \sum_{j \leq i} p_j \leq y < \sum_{j \leq i+1} p_j$$

Em caso discreto: $F(x)$ não é $\sim U(0,1)$

$$\mathbb{E}(F_X(x)) = \frac{1}{2} + \left[\frac{1}{2} \sum_{i=1}^n p_i^2 \right] \quad F_X(x) \sim U(0,1)$$

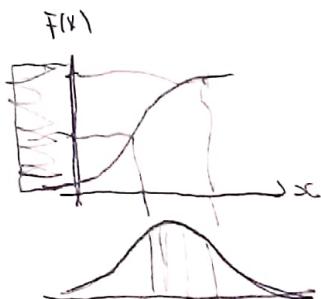
$$\text{Var}(F_X(x)) = \frac{1}{12} + \left[f \left(\sum_{i=1}^n p_i^2 \right) \right]$$

08/05/2013

Paul Embrechts - Risk Lab - ETH-Zürich
 Jef Teugels (1965) -

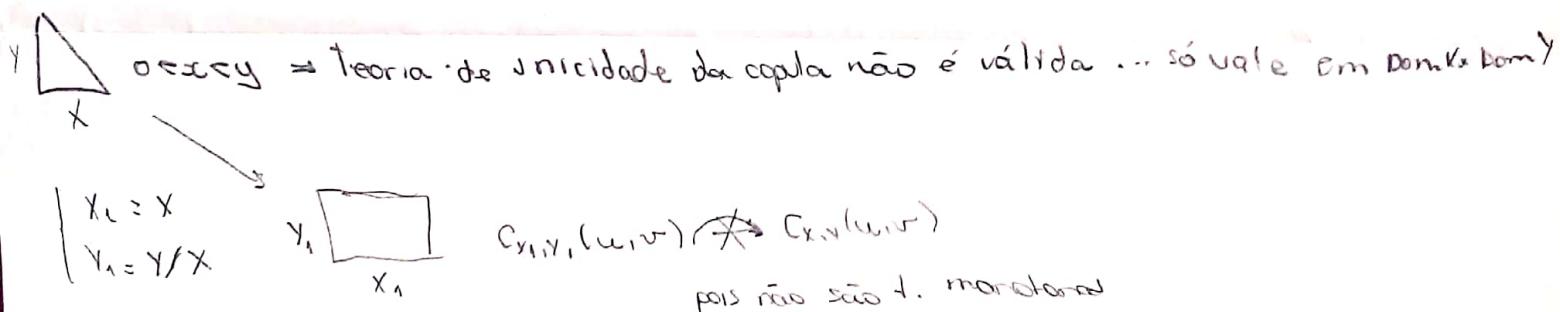
↓ Win Schoutens - livro sobre finanças

Credit Lyounis, Bouyé, Durrleman, Riboulet e Roncalli (2000), Bikelghbal, Lindskog e McNeil (2003) e Cherubini, Luciano e Vecchiato.



Idade de programa no R (não se sabe) - feito

Koja dinovich.



$E(\sum X_i)$ se torna uma v.a. univariada. Daí, usamos o limite de Frechet-Hoeffding

$$h(x, y) = c(F(x), G(y))$$

① C. Genest and A.-C. Favre (2007), Everything you
always wanted to know about copulae modeling
but were afraid to ask.

Journal of Hydrologic Engineering, 12, 347-368

② C. Genest and B. Remillard (2006). ^(*) Goodness-of-fit
procedures for copula models based on probability
integral transform.

Scandinavian Journal of Statistics,
33, 337 - 366

③ C. Genest, B. Remillard and D. Beaudoin (2009). ^(A)
Goodness-of-fit tests for copulas; A review and
power study.

Insurance: Mathematics & Economics, 44, 199-213

Bom dia!

Eu me chamo Maicon A. Pinheiro e este é o Rafael A. Mazzei.

Em conjunto, construímos esse trabalho cujo título é "Detectando Dependência por Meio de Kendall Plots".

De modo geral, o nosso objetivo durante a apresentação é falar sobre algumas ferramentas gráficas úteis para a detecção de dependência.

Ou seja, gráficos construídos a partir dos dados que nos permitem verificar se as v.a. no estudo são ou não independentes.

E para fazer isso de uma maneira gradual, passo-a-passo, decidimos dividir a apresentação em 6 seções:

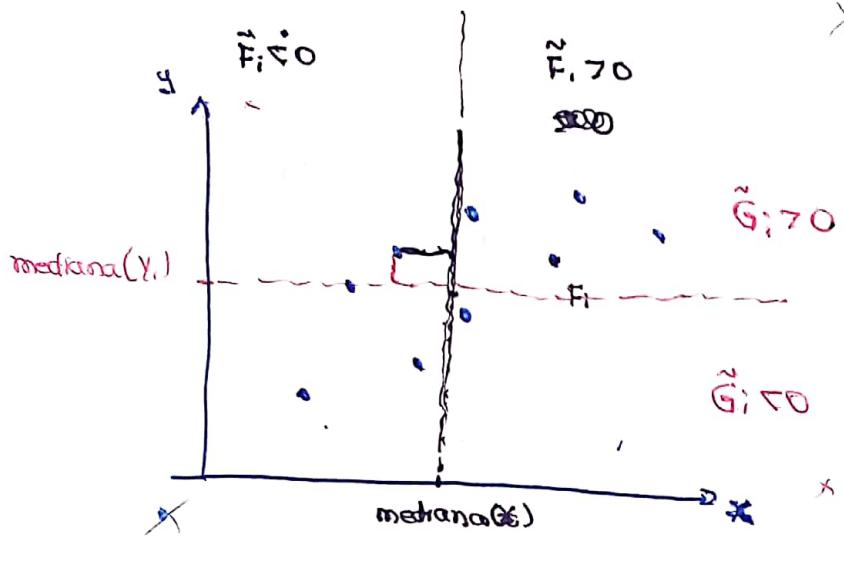
1. INTRODUÇÃO
2. CHI-PILOTS
3. K-PILOTS
4. R
5. Conclusão
6. Referências

→
— Rafael ...

→
Apesar destas boas características que esse gráfico possui, ele ainda não é uma boa opção para a detecção de dependência entre as v.a. Por quê?

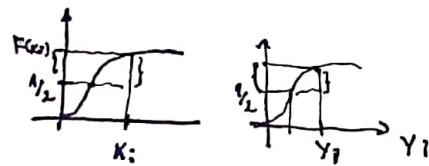
É que avaliar essa nuvem de pontos dentro do quadrado unitário nem sempre é uma tarefa clara, fácil de ser feita com precisão.

Em função da necessidade um gráfico ainda mais preciso ~~para~~ quanto à detecção de dependências, Fisher e Switzer (1985, 2001)

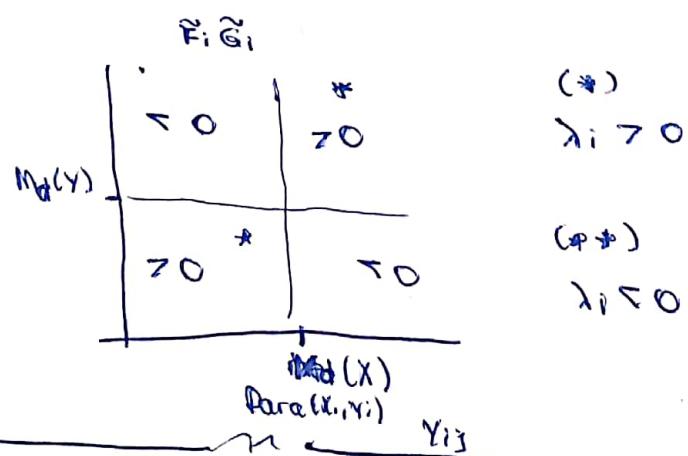


$$\tilde{F}_i(\text{mediana}(x)) = \frac{1}{2}$$

$\tilde{F}_i = F_i - \frac{1}{2} \Rightarrow$ distância entre x_i e a mediana das x_i 's em relação à f.d. empírica.



$\tilde{G} = F_i - \frac{1}{2} \Rightarrow$ distância entre y_i e a mediana das y_i 's em relação às suas f.d. empíricas

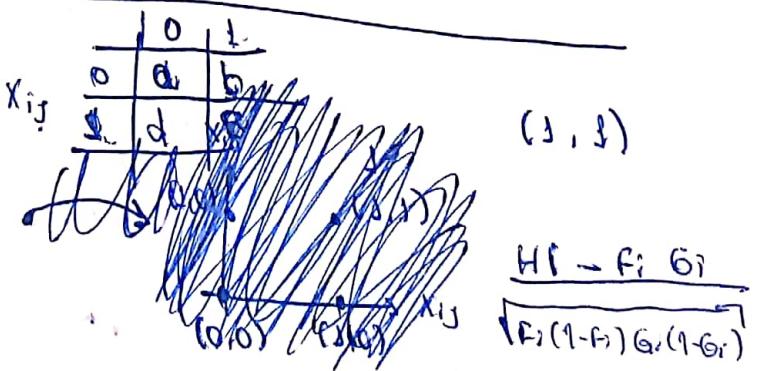
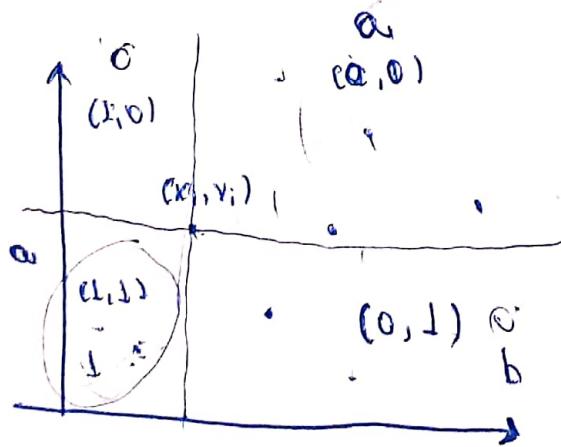


(*)

$$x_i > 0$$

(***)

$$x_i \leq 0$$



$$P_{X_{ij}, Y_{ij}} = \frac{E(X_{ij} Y_{ij}) - E(X_{ij}) E(Y_{ij})}{\sqrt{\text{Var}(X_{ij}) \text{Var}(Y_{ij})}}$$

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

e

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

$$\tilde{F}_i = F_i - 1/2, \quad \tilde{G}_i = G_i - 1/2 \quad \text{para } 1 \leq i \leq n.$$

• $\lambda_i \in [-1, 1]$

$$\lambda_i = 4 \underbrace{\operatorname{sgn}(\tilde{F}_i \tilde{G}_i)}_{\text{sgn}} \max(\tilde{F}_i^2, \tilde{G}_i^2)$$

$$0 \leq F_i \leq 1 \quad \text{e} \quad 0 \leq G_i \leq 1 \Leftrightarrow$$

$$-1/2 \leq \underbrace{F_i - 1/2}_{\tilde{F}_i} \leq 1/2 \quad \text{e} \quad -1/2 \leq \underbrace{G_i - 1/2}_{\tilde{G}_i} \leq 1/2 \Leftrightarrow$$

$$-1/2 \leq \tilde{F}_i \leq 1/2 \quad \text{e} \quad -1/2 \leq \tilde{G}_i \leq 1/2 \Rightarrow$$

$$0 \leq \tilde{F}_i^2 \leq 1/4 \quad \text{e} \quad 0 \leq \tilde{G}_i^2 \leq 1/4.$$

Logo,

$$0 \leq \max(\tilde{F}_i^2, \tilde{G}_i^2) \leq 1/4 \Leftrightarrow (\times 4)$$

$$0 \leq 4 \max(\tilde{F}_i^2, \tilde{G}_i^2) \leq 1$$

Se multiplicarmos por $\operatorname{sgn}(\tilde{F}_i \tilde{G}_i)$, teremos

$$-1 \leq 4 \operatorname{sgn}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i \tilde{G}_i) \leq 0 \quad \text{se } \operatorname{sgn}(\tilde{F}_i \tilde{G}_i) = -1$$

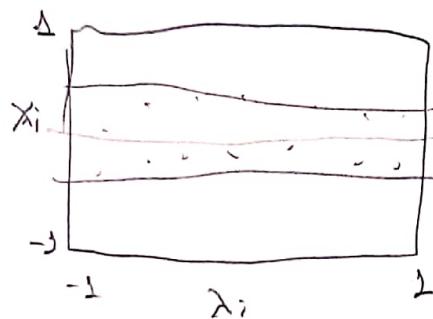
$$\text{e} \quad 0 \leq 4 \operatorname{sgn}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i \tilde{G}_i) \leq 1 \quad \text{se } \operatorname{sgn}(\tilde{F}_i \tilde{G}_i) = 1.$$

Por, $-1 \leq \lambda_i \leq 1 \Rightarrow \lambda_i \in [-1, 1]$

propor o chi-plot.

Ele é, na verdade, um outro gráfico de dispersão, só que agora de duas medidas ~~pares~~ obtidas por meio dos dados, denotadas por (λ_i, χ_i) .

Mas antes de conhecer estas medidas, vamos entender - de modo geral e geométrico - o que é esse chi-plot.



Essa é a área onde serão plotados os pares (λ_i, χ_i) . Se estes formarem uma faixa horizontal entre estas duas linhas, então não rejeitamos a hipótese de que as v.a. originais são independentes.

Dito isso, só nos falta conhecer quem:

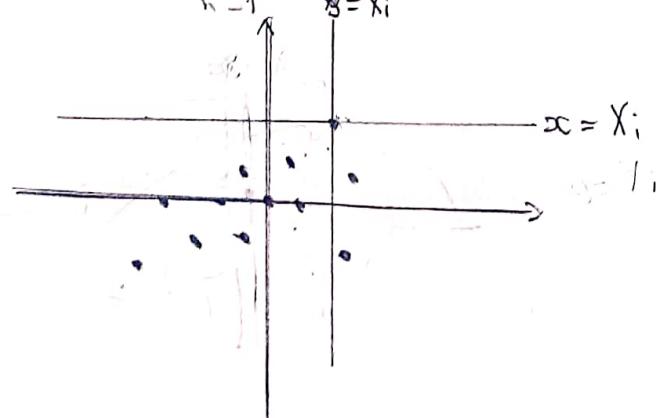
λ_i , χ_i e estas linhas.

Para isso, precisamos primeiro definir:

$$H_i = \frac{1}{n-1} \# \{ i \# j : X_j \leq X_i, Y_j \leq Y_i \} \quad \text{F.d. empírica associada ao par } (X_i, Y_i)$$

$$F_i = \frac{1}{n-1} \# \{ i \# j, X_j \leq X_i \} \quad \rightarrow \text{F.d. } X$$

$$G_i = \frac{1}{n-1} \# \{ i \# j, Y_j \leq Y_i \} \quad \rightarrow \text{F.d. } Y$$



$m_d(x) = \text{mediana } \{x_1, \dots, x_n\}$ e $m_d(y) = \text{mediana } \{y_1, \dots, y_n\}$

~~mediana é mediana de x_1, \dots, x_n~~

$$\hat{F}_n(m_d(x)) = \frac{\#\{i : x_i \leq m_d(x)\}}{n} = \frac{1}{2}$$

$$\hat{G}_n(m_d(y)) = \frac{\#\{i : y_i \leq m_d(y)\}}{n} = \frac{1}{2}$$

Dado um par (x_i, y_i) ,

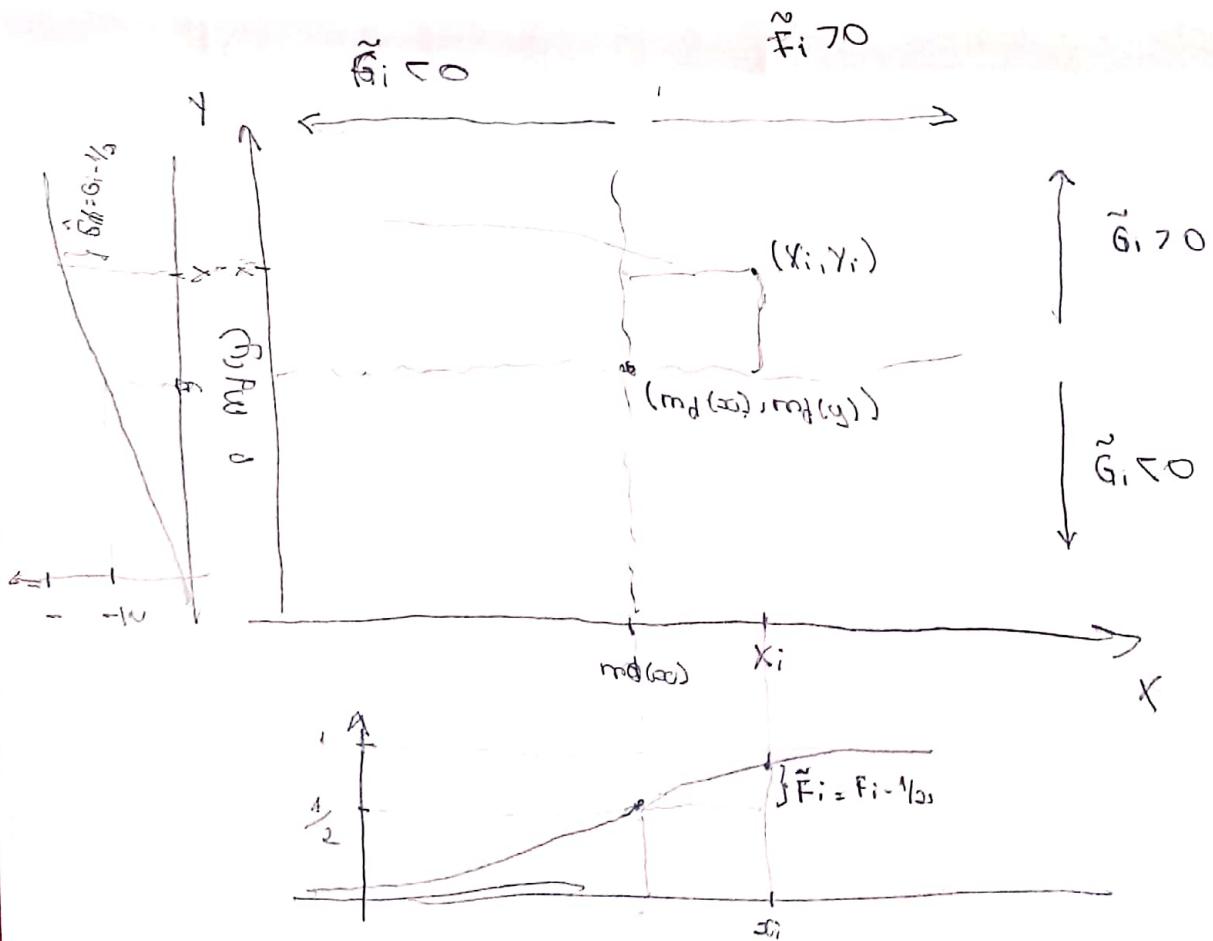
$$\tilde{F}_i = F_i - 1/2 \Leftrightarrow$$

$$\tilde{F}_i = F_i - \hat{F}_n(m_d(x))$$

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

$$\tilde{G}_i = G_i - \hat{G}_n(m_d(y))$$

centro dos dados série ^{par} $(m_d(x), m_d(y))$.



A vantagem do Kendall-Plot é que ele pode ser estendido pro caso multivariado. Ou seja,

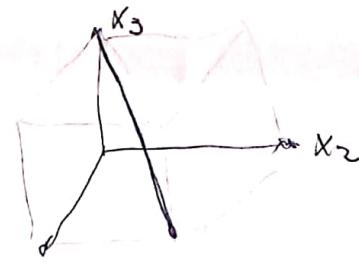
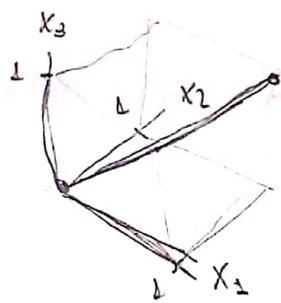
Dada uma amostra p-variada de tamanho $n \geq 2$,

com putarmos f.d.c. empíricas

ordenamos-as

e cruzamos com os valores esperados das estatísticas de ordem sob a hipótese de independência.

Exemplo:

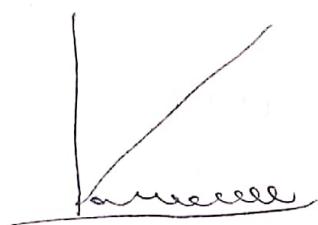
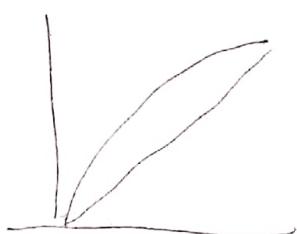


R. Nelsen

$$C(u, v) = 0$$

$$\text{corr}(U, V) = \rho_0$$

$$\text{Kendall Tau} = \tau_0$$



$$K_0(u) = u + u \sum_{k=1}^{p-1} \frac{1}{k!} \log^k(1/u), \quad 0 \leq u \leq 1.$$

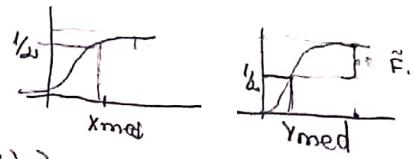
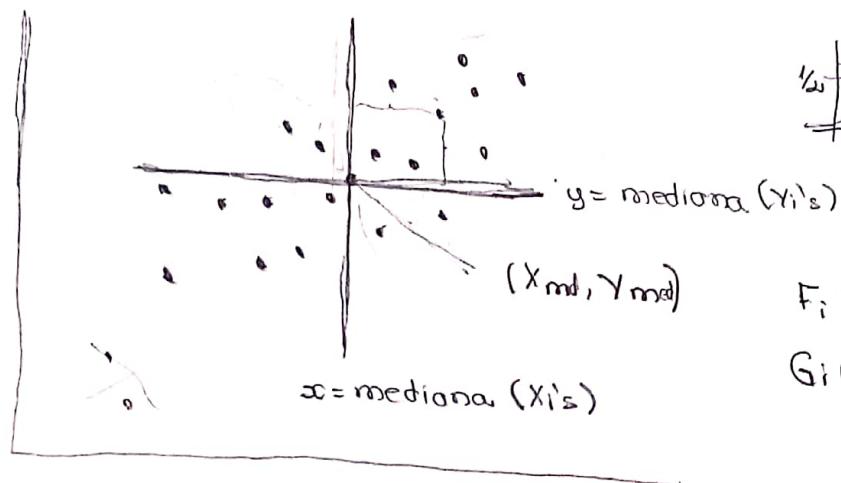
sob Independ.

$$K_{10}(H(x_1, \dots, x_p)) = P(H(x_1, \dots, x_p) \leq u) =$$

$$= P(F_1(x_1), \dots, F_p(x_p) \leq u) = P(U_1 U_2 \dots U_p \leq u) =$$

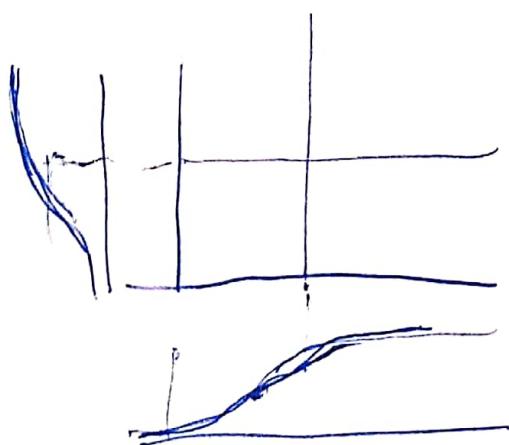
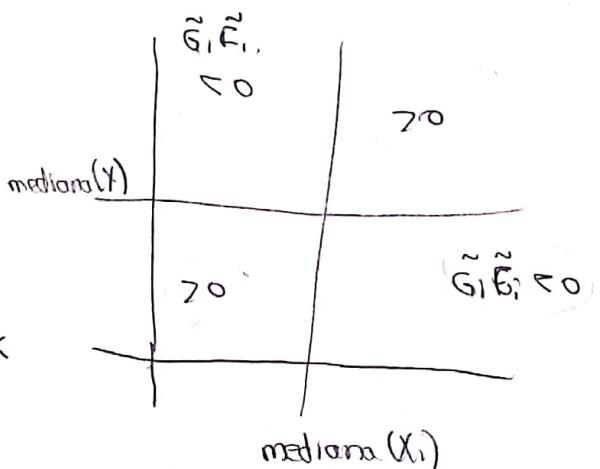
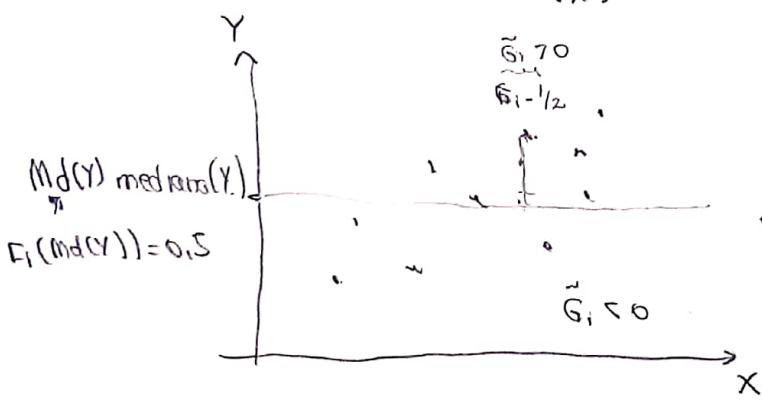
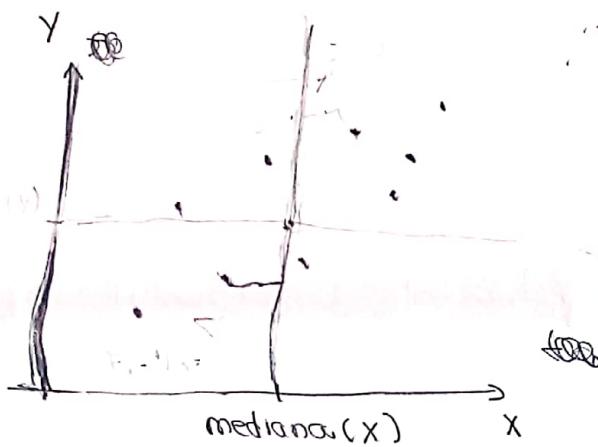
$$= P(\sum \log U_i \leq \log(u)) = P(\sum \log U_i \geq -\log(u))$$

Medida de distância do par (x_i, y_i) ao centro dos dados:



$$F_i(x_{\text{med}}) = 0,5$$

$$G_i(y_{\text{med}}) = 0,5$$





Detecting Dependence with Kendall Plots

Author(s): Christian Genest and Jean-Claude Boies

Reviewed work(s):

Source: *The American Statistician*, Vol. 57, No. 4 (Nov., 2003), pp. 275-284

Published by: American Statistical Association

Stable URL: <http://www.jstor.org/stable/30037296>

Accessed: 14/03/2013 18:26

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Statistician*.

<http://www.jstor.org>

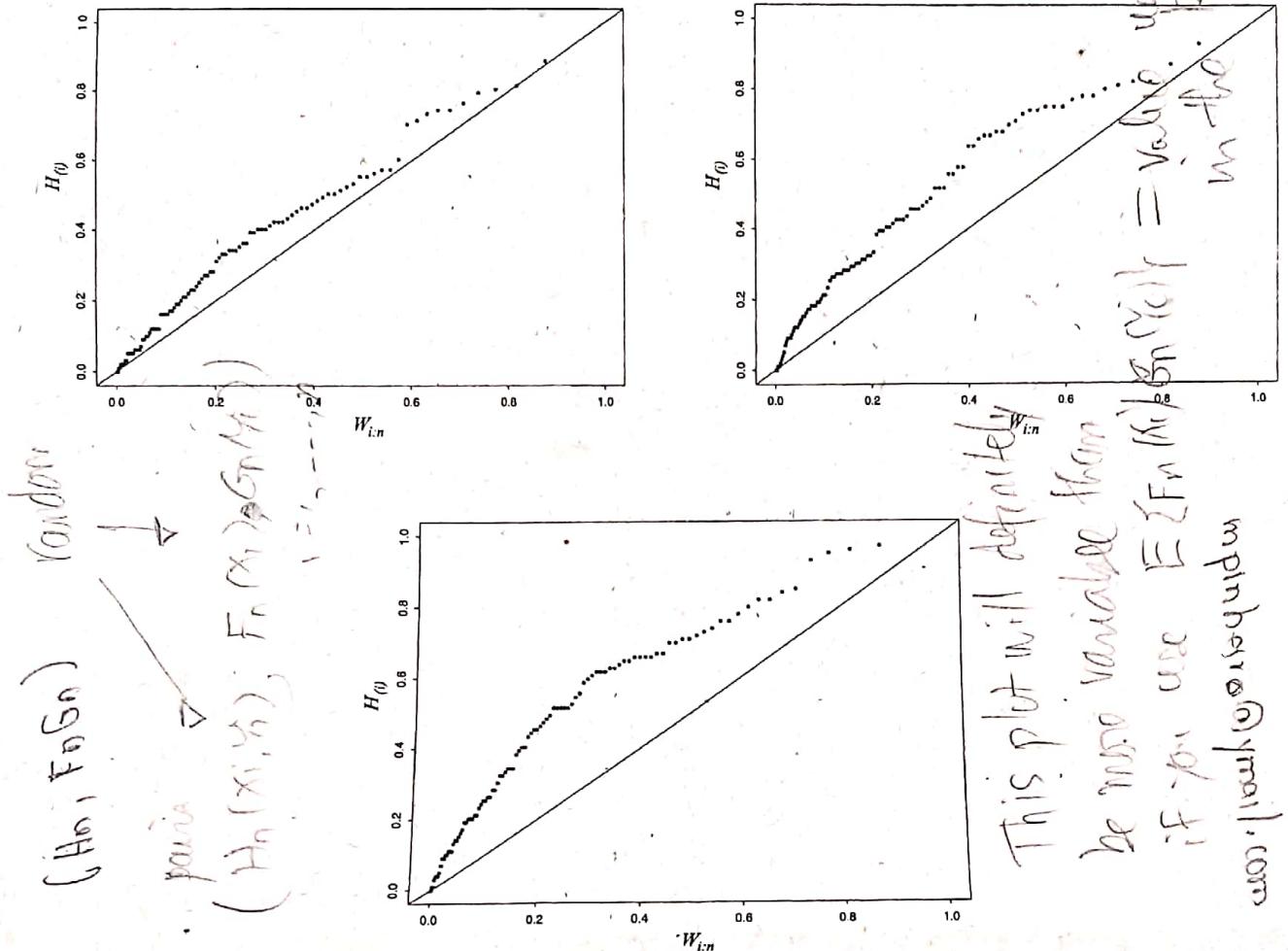


Figure 6. K-plots based on pseudo-random samples of size $n = 100$ from a bivariate vector (X, Y) with Clayton copula and Kendall's tau equal to $1/4$ (top left panel), $1/2$ (top right panel), and $3/4$ (bottom panel).

as can be readily checked using the closed-form expression

$$K_\alpha(w) = w + \frac{w}{\alpha} (1 - w^\alpha), \quad 0 \leq w \leq 1$$

which derives from the Archimedean character of C_α (Genest and MacKay 1986; Genest and Rivest 2001).

Figure 6 shows K-plots based on pseudo-random samples of size $n = 100$ from C_α with $\tau = 1/4, 1/2, 3/4$. The curves are rather different from those displayed in Figure 5. One should keep in mind, however, that bivariate distributions with distinct copulas C and C^* do not necessarily have different distributions K and K^* associated with their probability integral transformations.

That the implication $C \neq C^* \Rightarrow K \neq K^*$ is false may be verified easily using the class of bivariate extreme-value distributions, whose underlying copulas (see, e.g., Capéraà, Fougères, and Genest 1997b, 2000) are of the form

$$C_A(u, v) = \exp \left[\log(uv) A \left\{ \frac{\log(u)}{\log(uv)} \right\} \right]$$

for some convex function $A : [0, 1] \rightarrow [1/2, 1]$ such that $A(0) = A(1) = 0$ and $A(w) \geq \max(w, 1-w)$ for all $0 \leq w \leq 1$. As shown by Ghoudi, Khoudraji, and Rivest (1998), $W =$

$C_A(U, V)$ is distributed as

$$K_A(w) = w - (1 - \tau_A)w \log(w), \quad 0 \leq w \leq 1$$

where

$$\tau_A = \int_0^1 \frac{w(1-w)}{A(w)} dA'(w) \quad \text{a proposito}$$

is the population value of Kendall's tau (which, by the way, is always positive for an extreme-value distribution, since A is convex; for a stronger result along these lines, see Garralda-Guillem 2000).

Thus, if two extreme-value distributions with generators $A \neq A^*$ verify $\tau_A = \tau_{A^*}$, then clearly $K_A = K_{A^*}$. A fortiori, one could not hope to catalogue with complete precision the types of dependence implied by various K-plots. Of course, chi-plots suffer from the same limitation, though presumably to a lesser extent, given their bivariate nature. At the same time, however, the exact features of the copula function that chi-plots depict seem hard to pin down.

5.2 Real Data

Figure 7 displays a scatterplot, a chi-plot, and a K-plot of the ranks for 88 pairs of measurements used to investigate the