

Universidade Federal de Lavras
Departamento de Ciências Exatas

Dependência via Cópulas

Devanil Jaques de Souza
Lucas Monteiro Chaves

27/11/2009

Variável aleatória bi-dimensional

$$(X, Y) \sim F_{X, Y}(x, y)$$

As marginais: $F_X(x)$ e $F_Y(y)$

Variável aleatória bi-dimensional

$$(X, Y) \sim F_{X, Y}(x, y)$$

As marginais: $F_X(x)$ e $F_Y(y)$

A estrutura de dependência entre X e Y

Variável aleatória bi-dimensional

$$(X, Y) \sim F_{X,Y}(x, y)$$

As marginais: $F_X(x)$ e $F_Y(y)$

A estrutura de dependência entre X e Y

O coeficiente de correlação de Pearson

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_X)^2]} \sqrt{E[(Y - \mu_Y)^2]}}$$

$$\rho_{X,Y} = \rho_{aX+b, cY+d}$$

Problema: Capta dependência linear

$$Y = aX + b \Rightarrow \rho = \pm 1$$

Como modelar outras dependências:

Tau de Kendall amostral

Uma amostra bivariada de tamanho n :

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

c – número de pares concordantes $(x_i - x_j)(y_i - y_j) > 0$

d – número de pares discordantes $(x_i - x_j)(y_i - y_j) < 0$

O coeficiente ***Tau de Kendall amostral***:

$$t = \frac{c - d}{c + d}$$

Tau de Kendall populacional

(X_1, Y_1) e (X_2, Y_2) duas cópias independentes de uma população $F_{X,Y}$.

Probabilidade de concordância do tipo 1

$$\begin{aligned}\mathcal{P}_{c1} &= \mathbb{P}[(X_2 - X_1)(Y_2 - Y_1) > 0] \\ &= 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x, y) f_{X,Y}(x, y) dx dy ,\end{aligned}$$

Probabilidade de discordância do tipo 1

$$\mathcal{P}_{d1} = \mathbb{P}[(X_2 - X_1)(Y_2 - Y_1) < 0] = 1 - \mathcal{P}_{c1}$$

Tau de Kendall populacional $\tau = \mathcal{P}_{c1} - \mathcal{P}_{d1}$

Se X e Y têm distribuição conjunta normal bivariada então

$$\tau = \frac{2}{\pi} \arcsin \rho$$

$$A_{ij} = \text{sgn}(X_j - X_i) \text{sgn}(Y_j - Y_i) \quad (j \neq i)$$

$$\text{sgn}(u) = \begin{cases} 1 & \text{caso } u > 0 \\ -1 & \text{caso } u < 0 \end{cases}$$

$$T = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij}$$

T é um estimador não viesado e consistente de τ .

O RHO de Spearman amostral

Amostra $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ de uma população bivariada (X, Y) . Sejam $(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)$ os respectivos postos.

$$S = \sum_{i=1}^n (p_i - q_i)^2.$$

O Rho de Spearman amostral

$$R = 1 - \frac{6S}{n(n^2 - 1)} = \frac{\sum_{i=1}^n (P_i - \bar{P})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2 \sum_{i=1}^n (Q_i - \bar{Q})^2}}$$

O RHO de Spearman populacional

(X_i, Y_i) , (X_j, Y_j) e (X_k, Y_k) três cópias independentes de uma população bivariada e contínua (X, Y) .

Probabilidade de concordância do tipo 2:

$$\mathcal{P}_{c_2} = \mathbb{P}\left[\left(X_j - X_i\right)\left(Y_k - Y_i\right) > 0\right]$$

Probabilidade de discordância do tipo 2:

$$\mathcal{P}_{d_2} = \mathbb{P}\left[\left(X_j - X_i\right)\left(Y_k - Y_i\right) < 0\right]$$

RHO de Spearman populacional

$$\rho_S = 3\left(\mathcal{P}_{c_2} - \mathcal{P}_{d_2}\right) = 6\mathcal{P}_{c_2} - 3 = 3 - 6\mathcal{P}_{d_2}$$

Relação entre o RHO de Spearman e o coeficiente de correlação produto-momento de Pearson

X e Y com distribuição conjunta normal bivariada:

$$\rho_S = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right)$$

Variável aleatória bidimensional (X, Y)

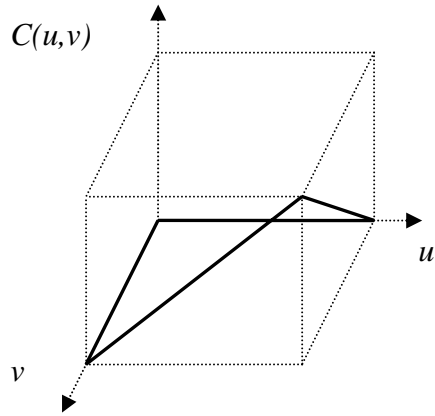
$$(x, y) \rightarrow \begin{cases} F_{X,Y}(x, y) \\ F_X(x) \\ F_Y(y) \end{cases}$$

$(F_X(x), F_Y(y)) \rightarrow F_{X,Y}(x, y)$ **Uma cópula**

Sejam $I = [0,1]$ $(u,v) \in I^2$

Cópula bi-dimensional: uma função $C : I^2 \rightarrow I$

P1 - $C(0,v) = C(u,0) = 0$, $C(u,1) = u$ e $C(1,v) = v$



P2 - $C(b,d) - C(a,d) - C(b,c) + C(a,c) \geq 0$

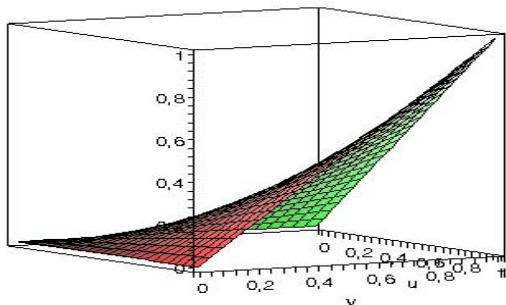
em que

$$a < b$$

$$c < d$$

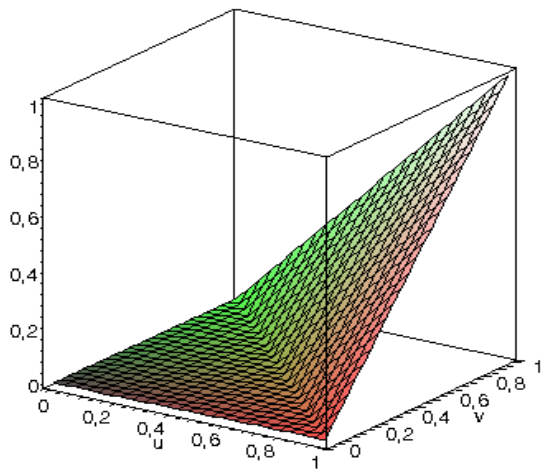
Exemplos de Cópulas

E1 - $C_I(u, v) = uv$



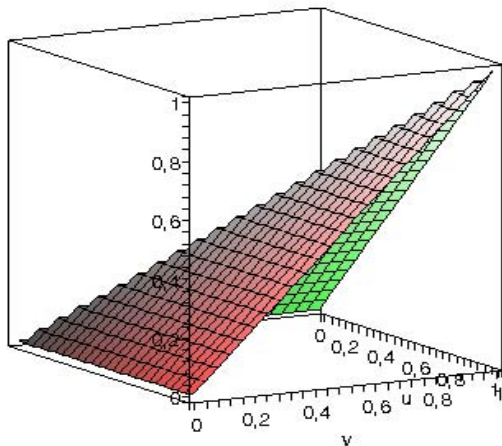
E2 -

$$C_0(u, v) = \max\{u + v - 1, 0\}$$



E3 -

$$C_1(u, v) = \min\{u, v\}$$



Desigualdade de Fréchet-Hoeffding:

Se $C_0(u, v) = \max\{u + v - 1, 0\}$

e $C_1(u, v) = \min\{u, v\}$

então

$$C_0(u, v) \leq C(u, v) \leq C_1(u, v)$$

Teorema de Sklar:

Sejam X e Y variáveis aleatórias com distribuições marginais

$$F_X(x) = P[X \leq x] \qquad G_Y(y) = P[Y \leq y]$$

e distribuição conjunta

$$H_{X,Y}(x, y) = P[X \leq x, Y \leq y].$$

Então, tomando-se $u = F_X(x)$ e $v = G_Y(y)$,

$$C(u, v) = H_{X,Y}(F_X^{-1}(u), G_Y^{-1}(v)) \quad \text{é uma cópula.}$$

Além disso:

- para quaisquer $F_X(x)$ e $G_Y(y)$ e qualquer cópula $C(u,v)$, $H_{X,Y}(x,y) = C(F_X(x), G_Y(y))$ é uma conjunta admissível.
- se $F_X(x)$ e $G_Y(y)$ são contínuas, para cada conjunta $H_{X,Y}(x,y)$ admissível existe uma e somente uma cópula $C(u,v)$ tal que $H_{X,Y}(x,y) = C(F_X(x), G_Y(y))$.

Propriedade: Se $f(\cdot)$ e $g(\cdot)$ são funções
estritamente crescentes então

$$C_{U,V} = C_{f(U),g(V)}$$

Cópuas e o TAU de Kendall

$$\begin{aligned}\tau &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{X,Y}(x,y) f_{X,Y}(x,y) dx dy - 1 \\ &= 4 \int_0^1 \int_0^1 C(u,v) \frac{\partial^2 C(u,v)}{\partial u \partial v} du dv - 1\end{aligned}$$

Cópuas e o RHO de Spearman

$$\begin{aligned}\rho_S &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_X(x) F_Y(y) f_{X,Y}(x,y) dx dy - 3 \\ &= 12 \int_0^1 \int_0^1 uv \frac{\partial^2 C(u,v)}{\partial u \partial v} du dv - 3\end{aligned}$$

Construção de Cópulas

- Método da mistura (compound)

Exemplo:

$$\begin{cases} \mathbf{P}[X \leq x | \gamma] = 1 - e^{-\gamma x} \\ \gamma \sim \text{Gama}(r, \lambda) \end{cases} \rightarrow \mathbf{P}[X \leq x] = E_{\gamma}[\mathbf{P}[X \leq x | \gamma]] = 1 - (1 + x/\lambda)^{-r}$$

X_1 e X_2 independentes, a menos de um risco γ .

$$F(x_1, x_2) = F_1(x_1) + F_2(x_2) - 1 \left[(1 - F_1(x_1))^{-1/r} + (1 - F_2(x_2))^{-1/r} - 1 \right]^{-r}$$

$$C(u, v) = u + v - 1 \left[(1 - u)^{-1/r} + (1 - v)^{-1/r} - 1 \right]^{-r}$$

Construção de Cópulas

- Cópulas Arquimedianas

$\phi: [0,1] \rightarrow [0,\infty]$ função convexa estritamente decrescente

$$\phi^{[-1]}(\phi(u) + \phi(v)) = C(u, v)$$

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & t \leq \phi(0) \\ 0 & t \geq \phi(0) \end{cases}$$

$\phi(\cdot)$ função geradora da cópula $C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v))$

Ajuste de Cópulas:
$$\tau = 4 \int_0^1 \int_0^1 C(u, v) \frac{\partial^2 C(u, v)}{\partial u \partial v} du dv - 1$$

Os dados: X – índice de estresse hídrico da região de Castro – Paraná

Y – produtividade de soja no município de Castro – Paraná

Período de 1980/1981 a 2007/2008

$$X = (1.156, 1.147, \dots, 1.2446)$$

$$Y = (2200, 2134, \dots, 3126)$$

$$\text{Corr}(X, Y) = 0.47$$

As cópulas e a relação entre os parâmetros e o Tau de Kendall

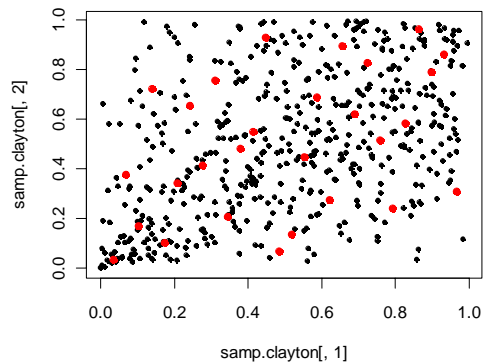
Família	Expressão $C(u, v)$	parâmetro	Tau de Kendall
Clayton	$\max\left(\left[u^{-\theta} + v^{-\theta} - 1\right]^{-1/\theta}, 0\right)$	$\theta \geq -1$	$\theta/(\theta+2)$
Gumbel	$\exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}$	$\theta \geq 1$	$1 - \theta^{-1}$
Normal	$N_{X,Y}\left(N_U^{-1}(u), N_V^{-1}(u)\right)$	$-1 \leq \rho \leq 1$	$\frac{2}{\pi} \arcsin(\rho)$
Morgenstern	$uv[1 + \theta(1-u)(1-v)]$	$-1 \leq \theta \leq 1$	$\frac{2}{9}\theta$

$$\tau = 0.3386 \rightarrow \begin{cases} \text{Clayton} & \theta = 1.024 \\ \text{Gumbel} & \theta = 1.512 \\ \text{Normal} & \rho = 0.518 \\ \text{Morgenstern} & NA \end{cases}$$

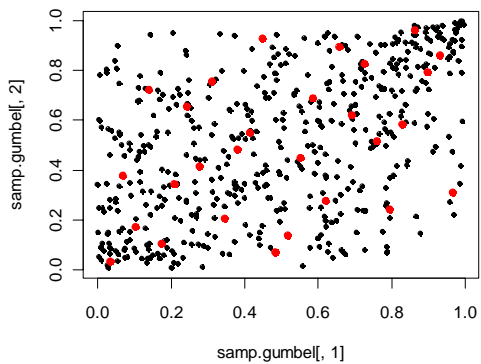
Qualidade do ajuste

Método gráfico: comparar o scatter plot de uma amostra grande da cópula ajustada e os pontos $(R_i/(n+1), S_i/(n+1))$ (suporte da cópula empírica)

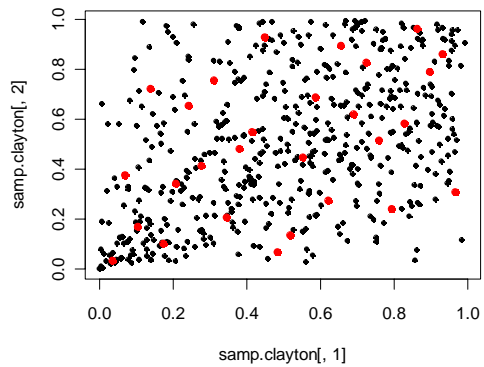
Clayton(1,024)



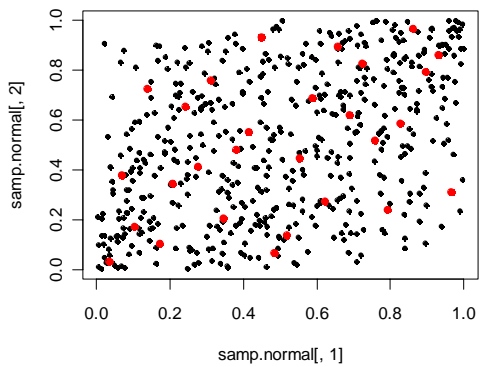
Gumbel-Hougaard(1,512)



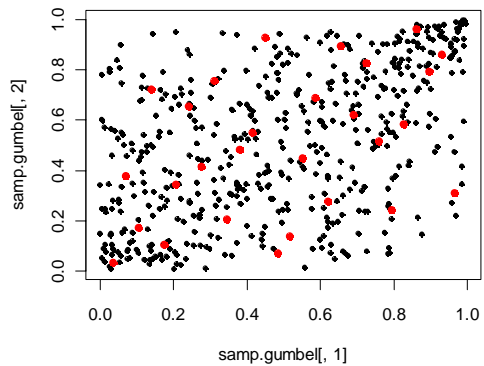
Clayton(1,024)



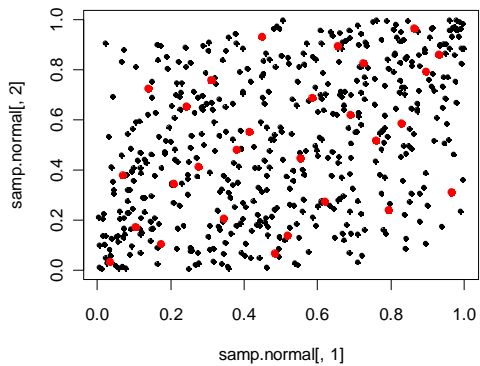
Normal(0,518)



Gumbel-Hougaard(1,512)



Normal(0,518)



Referências

- Anjos, U.U., Ferreira, F.H., Kolev, N.V., Mendes, B.V.M., 2004. *Modelando dependências via Cópulas*. ABE. 16º SINAPE.
- Joe, H., 1997, *Multivariate Models and Dependence Concepts*, Chapman&Hall-CRC
- Nelsen, R.B., 2005. *Dependence Modeling with Archimedean Copulas*. Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance. IME-USP. <http://www.lclark.edu/~mathsci/nelsen.html>
- Nelsen, R.B., 2002. *Concordance and copulas: A survey*. Kluwer Academic Publishers, Dordrecht, pp. 169-178. <http://www.lclark.edu/~mathsci/nelsen.html>
- Nelsen, R.B. 1999, *An Introduction to Copulas*, Springer.
- Yan, J. 2007, *Enjoy the joy of copulas: With a package copula*. Journal of Statistical Software, Volume 21, Issue 4. <http://www.jstatso.org/>

- Genest, C., Rémillard, B. e Beaudoin, D. 2007, ***Goodness-of-fit for copulas: A review and a power study.*** Insurance:Mathematics and Economics 44, 199-213, www.sciencedirect.com
- Genest, C. e Favre, A.C. , 2007. ***Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask.*** Journal of Hydrological Engineering .
- Frees, E.W. e Valdez, E.A., 1998, ***Understanding Relationships Using Copulas.*** North American Actuarial Journal, volume 2, number 1.1