

Valores Extremos Multivariados mediante R-vines



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

Leonardo Moreno

II Jornadas LPE-MAREN

La Paloma, Rocha, 2014

Esquema:

Eventos extremos

El Problema

Una idea de Cópulas

Regular Vines

Estimación y Predicciones

Eventos extremos climáticos.



Inundaciones río amarillo 1931
China

Eventos extremos climáticos.



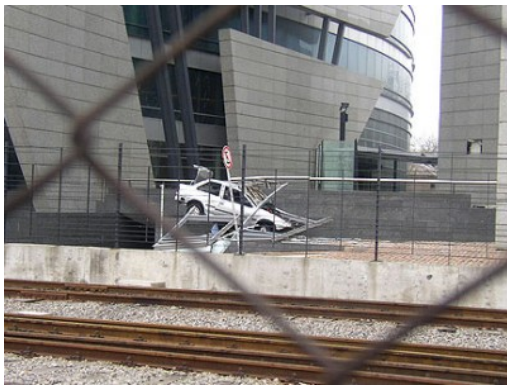
Inundaciones en León
México

Eventos extremos climáticos.



Deslaves 1999
Venezuela

Eventos extremos climáticos.



Temporal 2005
Uruguay

Eventos extremos climáticos.



Temporal 2005
Uruguay

Eventos extremos

¿Es posible predecir determinados eventos extremos?

Poca información....

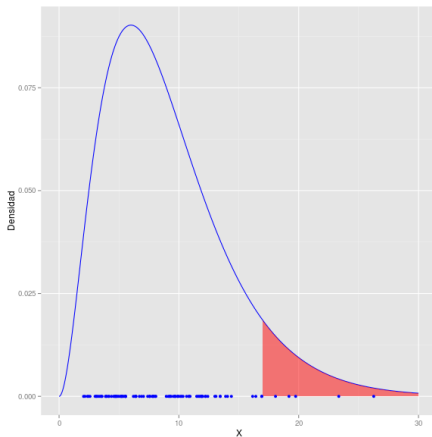


Figura : Estimación de la cola de la distribución

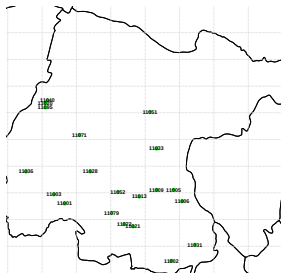
Una Aplicación: LLuvias Extremas en Guanajuato



Una Aplicación: LLuvias Extremas en Guanajuato



Estaciones seleccionadas



- Se decide trabajar con 20 estaciones que presentan datos simultáneamente en 37 años.

Cóputas: Teorema de Sklar

Para cóputas bidimensionales

Sea H la función de distribución conjunta de una v.a vectorial (X, Y) con F y G las distribuciones marginales (supongamos continuas). Entonces existe y es única la cóputa C tal que $\forall (x, y) \in \overline{R}^2$ cumple que,

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

Propiedades

- Sean $M(u, v) \equiv \min(u, v)$ y $W(u, v) \equiv \max(u + v - 1, 0)$. Entonces para toda cópula C se cumple que,

$$W(u, v) \leq C(u, v) \leq M(u, v). \quad (2)$$

W y M son llamadas las cotas inferior y superior de *Fréchet-Hoeffding* para Cópulas

- Sean X e Y v.a absolutamente continuas. X e Y son independientes si y sólo si $C_{XY}(u, v) = \Pi(u, v) \equiv u.v$.

Densidad de la Función de Cópula

Se llama densidad asociada a la cópula $C(u_1, u_2)$ a,

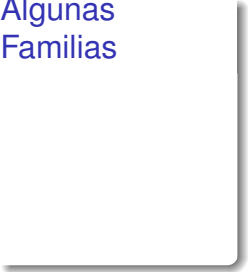
$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}. \quad (3)$$

En v.a absolutamente continuas se cumple que,

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2), \quad (4)$$

Cóputas paramétricas

Algunas
Familias



Cóputas paramétricas

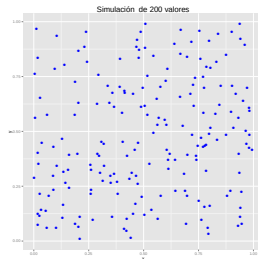
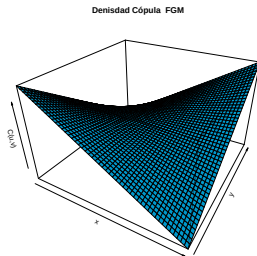
Algunas
Familias



Cóputas paramétricas

Algunas Familias

- Farlie-Gumbel-Morgenstern

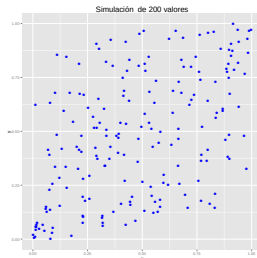
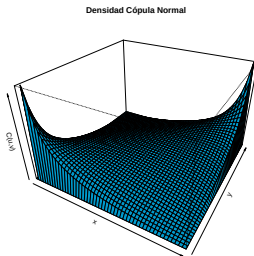


$$\theta = 0,7$$

Cóputas paramétricas

Algunas Familias

- Farlie-Gumbel-Morgenstern
- Gaussiana

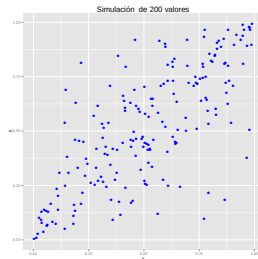
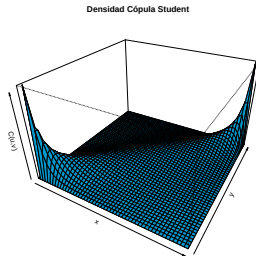


$$\theta = 0,25$$

Cóputas paramétricas

Algunas Familias

- Farlie-Gumbel-Morgenstern
- Gaussiana
- Student

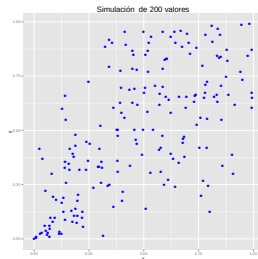
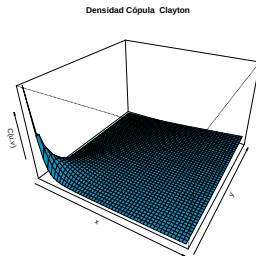


$$\rho = 0,7 \text{ y } 4 \text{ g.l.}$$

Cóputas paramétricas

Algunas Familias

- Farlie-Gumbel-Morgenstern
- Gaussiana
- Student
- Clayton



$$\theta = 1,5$$

Cópuas Extremas

Algumas
Famílias



Cópuas Extremas

Algumas
Famílias

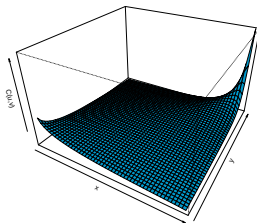


Cóputas Extremas

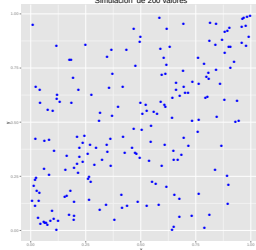
Algunas Familias

- Gumbel-Hougaard

Densidad Cóputa Gumbel



Simulación de 200 valores



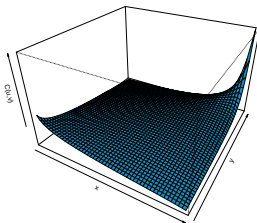
$$\alpha = 2/3$$

Cóputas Extremas

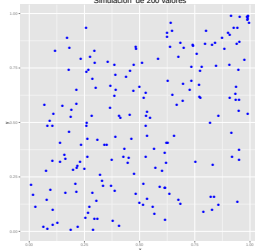
Algunas Familias

- Gumbel-Hougaard
- Galambos

Densidad Cópula Galambos



Simulación de 200 valores



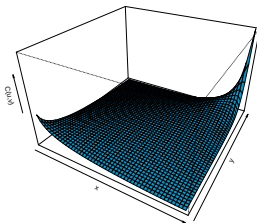
$$\alpha = 2/3$$

Cóputas Extremas

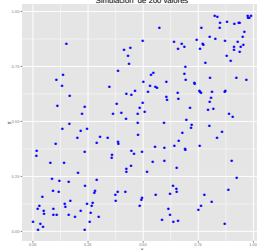
Algunas Familias

- Gumbel-Hougaard
- Galambos
- Hüsler-Reiss

Densidad Cópula Husler-Reiss



Simulación de 200 valores



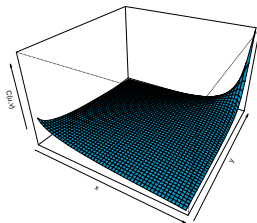
$$a = 3/2$$

Cóputas Extremas

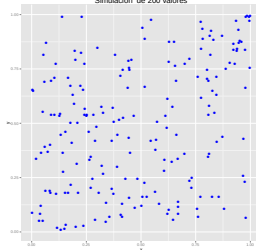
Algunas Familias

- Gumbel-Hougaard
- Galambos
- Hüsler-Reiss
- *t*-Extremal

Densidad Cópula t-extremal



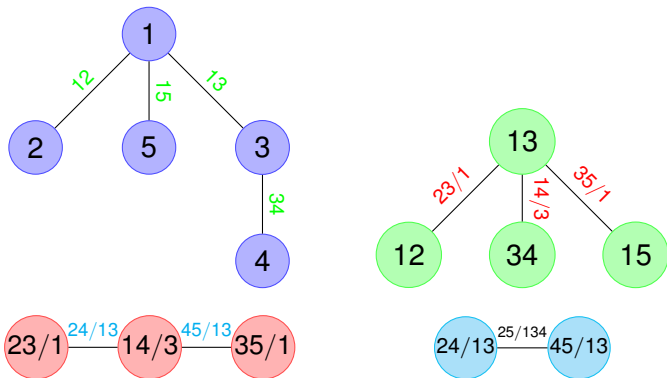
Simulación de 200 valores



$$\rho = 2/3 \text{ y } 4/3$$

Estructura de R-vines

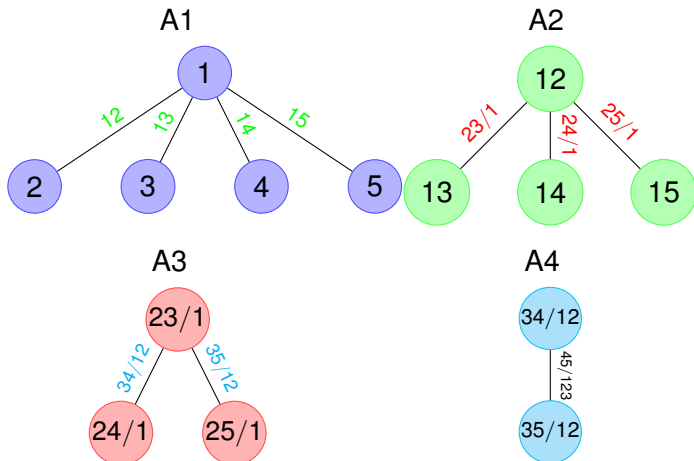
$$f_{1,2,3,4,5} = f_1 f_2 f_3 f_4 f_5 c_{12} c_{13} c_{34} c_{15} c_{23/1} c_{14/3} c_{35/1} c_{24/13} c_{45/13} c_{25/134},$$



El número posible de R-vines en $\mathbb{R}^d$ es $\frac{d!}{2} 2^{\frac{(d-2)(d-3)}{2}}$.

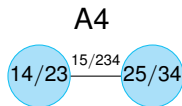
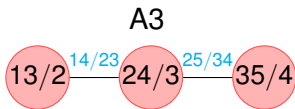
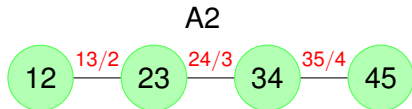
Ejemplo. C-vines

$$f_{1,2,3,4,5} = f_1 f_2 f_3 f_4 f_5 c_{12} c_{13} c_{14} c_{15} c_{23/1} c_{24/1} c_{25/1} c_{34/12} c_{35/12} c_{45/123}$$



Ejemplos: D-vines

$$f_{1,2,3,4,5} = f_1 f_2 f_3 f_4 f_5 c_{12} c_{23} c_{34} c_{45} c_{13/2} c_{24/3} c_{35/4} c_{14/23} c_{25/34} c_{15/234},$$



Selección del Modelo

La selección de la cópula R-vines consta de 3 pasos.

1. Seleccionar la estructura R-vines entre todas la posibles.
2. Elegir las $d(d - 1)/2$ familias paramétricas de cópulas de a pares.
3. Estimar los parámetros de las cópulas seleccionadas.

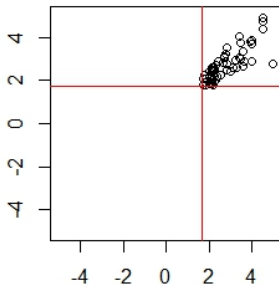
Selección de la estructura R-vines

- Procedimiento secuencial introducido por Dímann, Brechmann, Czado y Kurowicka (2013), [5].
- Se elige una **medida de dependencia** $\delta_{i,j}$ que le asigne pesos a las aristas.
- Bajo un algoritmo (usamos Prim) adecuado se selecciona un árbol, el primero del conjunto, que maximice,

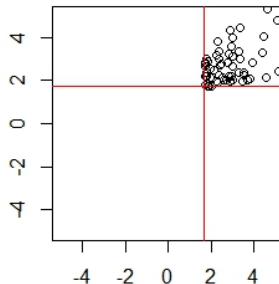
$$\sum_{\text{aristas } e=\{i,j\}} |\delta_{i,j}|$$

Relaciones de dependencia ($\delta_{i,j}$). Correlación

Corr =

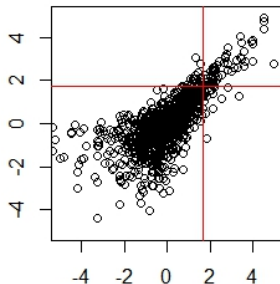


Corr =

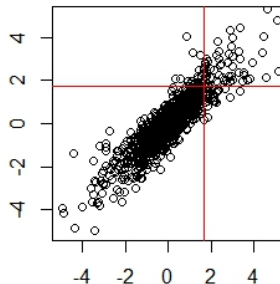


Relaciones de dependencia. Correlación

Corr = 0.66



Corr = 0.91



Otras medidas de Dependencia

- Rango Correlación de Kendall, (“Kendall’s tau”).

$$\rho_{\tau}(X, Y) = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0],$$

donde (X_1, Y_1) y (X_2, Y_2) pares de v.a iid con distribución F .

- Coeficiente Dependencia en las Colas.

$$\lambda_u = \lim_{u \rightarrow 1^-} P(Y > F_2^{-1}(u) | X > F_1^{-1}(u)),$$

- El F -madograma (para un campo aleatorio).

$$\nu^F(h) = \frac{1}{2} E |F(Z(x+h)) - F(Z(x))|, \quad (5)$$

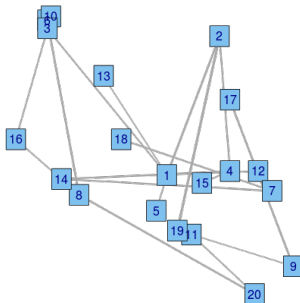
- Distancia covarianza (Szekely 2007) y MIC (Reshef 2011).

Selección de Familias.

- Luego de elegida la estructura se seleccionan las familias de cópulas bidimensionales y se estiman los parámetros de cada cópula por máxima verosimilitud.
- El método de simplificación, que consiste en considerar cópulas de una sola familia, y truncamiento, suponer que a partir de cierto nivel las cópulas son independientes, fueron introducidos por Brechmann, Czado y Aas en el 2012, [1].
- Por último se calculan las distribuciones condicionales empíricas de forma recursiva pues. A partir de ellas se reitera el algoritmo.

Primer árbol de la R-vine (Iluvioso)

Tree 1



Algunas Predicciones

período (años)	<i>Junio – Agosto</i>				
	Mediana	Media	Desvío	P^* (supere 100 mm)	P^* (supere 150 mm)
10	105.2	110.6	23.4	0.625	0.063
20	116.0	122.5	25.3	0.87	0.121
30	123.4	129.1	25.4	0.948	0.161
40	127.2	133.5	27.4	0.98	0.194
50	133.0	139.2	28.7	0.994	0.268

GRACIAS!!



Referencias (1)



E.Brechmann; C. Czado; K. Aas.

Truncated regular vines in high dimensions with application to financial data.
Canadian Journal of Statistics, 40(1):68–85, 2012.



K. Aas; C. Czado; A. Frigessi; H. Bakken.

Pair-copula constructions of multiple dependence.
Insurance, Mathematics and Economics, 44:182–198, 2009.



J.F. Dissmann.

Statistical inference for regular vines and application.
Master's thesis, Technische Universit at Munchen., 2010.



D. Kurowicka and R. Cooke.

Uncertainty analysis with high dimensional dependence modelling.
Wiley, 2006.



J. Dißmann; E. C. Brechmann; C. Czado; D. Kurowicka.

Selecting and estimating regular vine copulae and application to financial returns.
Computational Statistics & Data Analysis, 59:52–69, 2013.



O. Morales-Nápoles.

Bayesian belief nets and vines in aviation safety and other applications.
PhD thesis, Technische Universiteit Delft, 2008.

Referencias (2)



R. Nelsen.

An introduction to copulas.
Springer-Verlag, 2006.



A. Sklar.

Fonctions de répartition a n dimensions et leurs marges.
Publ. Inst. Stat. Univ. Paris, 8:229–231, 1959.



R.M Cooke T. Bedford.

Probability density decomposition for conditionally dependent random variables modeled by vines.
Annals of Mathematics and Artificial Intelligence, 32:245–268, 2001.



H.Joe; J.J. Xu.

The estimation method of inference functions for margins for multivariate models.
Technical report, Department of Statistics, University of British Columbia, 1996.