



Universidade de Brasília
IE - Departamento de Estatística
Estágio Supervisionado 2

Modelos multidimensionais: Cópulas D-Vine

Lucas Lourenço Cunha Bragança

Brasília

Julho de 2015

Lucas Lourenço Cunha Bragança

Modelos Multidimensionais: D-Vine Cópulas

Relatório Final do Trabalho de Conclusão de Curso
para obtenção do título de Bacharel em Estatística

Orientadora: Prof.^a Dra. Cira Etheowalda Guevara Otiniano

Brasília

Julho de 2015

Agradecimentos

Agradeço em primeiro lugar a Deus, que é o supremo benfeitor e criador de todas as coisas. É apenas por sua imensa graça que posso desfrutar de momentos como esse. Agradeço também à minha família e amigos que sempre me apoiaram, me ajudando nos momentos de dificuldade e ansiedade, pra que eu pudesse prosseguir e chegar até aqui. Também sou muito grato à Professora Cira, pois melhor orientadora seria impossível de encontrar! Com toda a paciência em ensinar mesmo já tendo explicado várias vezes e eu ainda confuso, com todo o tempo dedicado e todas as ideias e correções sempre muito pertinentes.

Enfim, só tenho a agradecer por todas batalhas vencidas e agora me preparar para as próximas.

Resumo

Distribuições multivariadas flexíveis são necessárias para modelar sistemas em muitas áreas. A distribuição gaussiana multivariada é muito restritiva e não consegue capturar características como assimetria e caudas pesadas. Uma ferramenta alternativa no tratamento de tais distribuições são as cópulas, porém ainda hoje o estudo de adequação e estimação das cópulas clássicas em dimensões altas é um desafio. As cópulas D-Vine surgiram para tratar esse tipo de problema. Neste trabalho estudamos as cópulas D-Vine, bem como a inferência estatística dessas cópulas.

Palavras Chave: Cópulas, Cópula D-Vine, Função multivariada.

Abstract

Flexible multivariate distributions to model systems are needed in many areas. The multivariate Gaussian distribution is very restrictive and fails to capture characteristics such as asymmetry and heavy tails. An alternative tool in the treatment of such distributions is the copula, however nowadays the study of suitability and estimation of classic copulas in higher dimensions is still a challenge. The copula D-Vine emerged to solve this problem. In this work we study the D-Vine Copulas and the statistical inference of these copulas.

Key-words: Copulas, D-Vine copula, Multivariate Distributions.

Sumário

1	INTRODUÇÃO	7
2	Base teórica de Cópulas	9
2.1	Definição	10
2.2	Medidas de Associação	11
2.2.1	O Tau de Kendall	12
2.2.2	Rho de Spearman	13
2.3	Coeficiente Caudal	13
2.4	Exemplos de Famílias de Cópulas	14
2.4.1	Os Limitantes de Fréchet-Hoeffding	14
2.4.2	Cópulas Gaussianas	15
2.4.3	Cópulas Arquimedianas	16
3	Inferência Estatística para Cópulas	20
3.1	Método de Máxima Verossimilhança Exata	20
3.2	Método de Máxima verossimilhança com inferência nas marginais (Método IFM)	21
4	D-Vine Cópulas	23
4.1	Pair-copulas	23
4.2	Vine-copulas	26
4.3	Ilustrações numéricas para a cópula D-Vine	27
4.3.1	Representação Gráfica e ordem das variáveis	27

4.3.2	Estimação	29
5	Aplicação	33
5.1	O Conjunto de Dados	33
5.2	Análise exploratória dos dados	33
5.3	Estimação	36
6	Conclusão	38

1 INTRODUÇÃO

O trabalho com cópulas começou com Abe Sklar em 1959. A partir daí os estudos foram se tornando cada vez mais abrangentes em torno desse assunto, pois a modelagem por cópulas se tornou algo bastante popular em várias áreas de estudo em busca de funções de distribuição multivariadas mais flexíveis para modelar dados com certa dependência. A literatura tem crescido bastante, o número de artigos é extenso e o estudo tem levado a um grande crescimento quando se trata de medidas de dependência e construção de modelos d -variados menos intrincados usando cópulas. Uma cópula nada mais é que uma distribuição multivariada restrita a $[0, 1]^d$ cujas distribuições marginais são uniformes $[0, 1]$. Referências importantes para a teoria e aplicação de cópulas são os livros de Nelsen (1999) e de Cherubini, et al. (2004). É conhecida uma grande quantidade de famílias de cópulas, entre elas temos as cópulas elípticas, arquimedianas, extremas, pair-cópulas e outras. As três primeiras famílias são muito utilizadas em diversas áreas para a modelagem de dados bivariados.

Em 1996 Joe intensificou o estudo de famílias de cópulas bivariadas, bem como o de pair-cópulas. Posteriormente, esse estudo foi estendido e organizado por Bedford e Cooke (2001,2002) através do modelo vine-cópula. Esse modelo tem algumas vantagens em relação ao modelo de cópulas multivariadas clássicas tais como as arquimedianas, elípticas e extremas que não têm a flexibilidade de modelar vetores aleatórios com mais de dois componentes. Mais tarde, introduzido por Kurowicka e Cooke (2004), apareceu o modelo das cópulas D-vine e das cópulas C-vine que são cópulas multivariadas construídas através de cópulas bivariadas, chamadas de vine-

copulas, que são representadas através de uma estrutura de construção de um conjunto aninhado de árvores, o que possibilita visualizar facilmente a interdependência das marginais decompostas da função de densidade conjunta que representam.

Um estudo sobre tal assunto é importante pelo fato de ser algo extremamente atual e que requer uma grande quantidade de testes e simulações. Muitas aplicações já estão sendo feitas em cima desses estudos, principalmente quando se tratam de dados de retorno financeiro. Um exemplo é quando é necessária a modelagem de uma carteira de ativos financeiros diversificada com o objetivo do cálculo de alguma medida de risco.

Neste trabalho estudamos as cópulas D-vine e sua inferência estatística. Para isto, dividimos o trabalho em cinco capítulos. No Capítulo 1, apresentamos a definição de cópulas, bem como alguns exemplos das famílias clássicas e as medidas mais importantes que são abordadas ao se tratar de cópulas. No Capítulo 2, escrevemos os principais métodos de estimação de cópulas. No Capítulo 3, descrevemos as D-vine cópulas, sua origem e peculiaridades. No capítulo quatro apresentamos uma aplicação sobre o estudo e finalmente, no capítulo cinco, uma conclusão sobre todo o trabalho realizado.

2 Base teórica de Cópulas

Iniciamos este capítulo com a definição de cópula, em seguida apresentamos as principais medidas de associação como a correlação de Pearson, correlação de Kendall, correlação de Spearman e coeficiente caudal. Finalizamos este capítulo descrevendo as principais famílias de cópulas.

Antes de definirmos formalmente o conceito matemático de cópulas, podemos citar várias características desta função e sua utilidade na modelagem de funções multivariadas. De acordo com Fisher (1997):

Cópulas são de interesse para estatísticos por duas razões: Primeiramente, é uma forma de se estudar medidas de dependência em livre escala, e segundo, é o ponto inicial para construir famílias de distribuições bivariadas, às vezes, tendo em vista a simulação.

O autor Roger B. Nelsen diz basicamente o que são cópulas através de dois pontos de vista: No primeiro diz que cópulas são funções que juntam ou “acoplam” funções de distribuição multivariadas à suas funções de distribuição marginais unidimensionais. Sob um segundo ponto de vista diz que alternativamente, cópulas são funções de distribuição multivariadas a qual suas marginais unidimensionais são uniformes no intervalo $(0,1)$.

Por último, também podemos dizer que cópulas são funções que permitem construir um modelo multivariado que separa o comportamento marginal das variáveis aleatórias da sua estrutura de dependência.

2.1 Definição

Definição 2.1. Cópula é uma função de distribuição de probabilidade n-dimensional $C : [0, 1]^n \rightarrow [0, 1]$, cujas marginais são uniformes $[0, 1]$, e satisfaz as propriedades abaixo:

1. $C(u_1, \dots, u_n)$ é crescente e contínua em cada um de seus componentes u_i , para todo $i = 1, \dots, n$
2. Para todo $u \in [0, 1]^n$, $C(u) = 0$ se pelo menos uma coordenada de u tende a zero, e $C(u) = 1$ se todas as coordenadas de u tenderem para 1.
3. Para todo $a, b \in [0, 1]^n$, tal que $a_i \leq b_i$, para todo i , $V_c([a, b]) \geq 0$.

Para $n=2$,

$$\begin{aligned} Vc([a, b]) &= P(a_1 \leq U_1 \leq b_1, a_2 \leq U_2 \leq b_2) \\ &= D_2 D_{1[a, b]} C(u_1, u_2) \\ &= C(b_1, b_2) + C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2), \end{aligned}$$

quando $[a, b] = [a_1, b_1] \times [a_2, b_2]$, sendo D o operador diferença definido por $D_{[a_1, b_1]} g(x, y) = g(b_1, y) - g(a_1, y)$.

Qualquer função de distribuição multivariada contém cópulas que podem ser usadas para construir funções de distribuição multivariadas a partir das funções de distribuição univariadas, ou seja, essas cópulas são funções de dependência que "acoplam" as distribuições marginais univariadas, formando uma distribuição multivariada. Podemos ver isso através do teorema de Sklar (1959) que diz:

Teorema 2.1. *Seja $F \in \mathfrak{S}(F_1, \dots, F_n)$ uma função de distribuição n -dimensional com marginais F_1, \dots, F_n , então existe uma cópula $C \in \mathfrak{S}(U, \dots, U)$ com marginais uniformes de tal forma que:*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \quad (1)$$

Caso F_1, \dots, F_n sejam contínuas, então a cópula C é única (caso contrário, existe mais de uma cópula que pode ser usada para agregar as marginais). Se as funções de distribuição marginal de F são contínuas e C é função de distribuição de $(F_1(x_1), \dots, F_n(x_n))$ desde que $F_i(x_i) \sim U(0, 1)$, então C é uma cópula que tem a seguinte representação:

$$C(u_1, \dots, u_n) = F_x(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad (2)$$

sendo F_i^{-1} a inversa generalizada de F_i , definida por:

$$F_i^{-1}(t) = \inf\{x \in \mathbb{R}^1; F_i(x) \geq t\}. \quad (3)$$

2.2 Medidas de Associação

Naturalmente, ao se estudar cópulas, se faz necessário falar em associação ou dependência de variáveis aleatórias, pois as cópulas são uma forma de representar a relação de dependência das variáveis.

Sendo assim, falaremos de algumas medidas de correlação, sendo a mais conhecida delas a correlação de Pearson, ou coeficiente de correlação linear.

Definição 2.2. Dados X e Y variáveis aleatórias, o coeficiente de correlação entre elas é dado por:

$$\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sigma[X]\sigma[Y]}, \quad (4)$$

na qual $\text{Cov}[X, Y]$ é a covariância entre X e Y dada por $E(XY) - E(X)E(Y)$ e $\sigma[X]$ e $\sigma[Y]$ são os desvios-padrão de X e Y .

Essa medida de dependência detecta unicamente a correlação linear das variáveis, portanto não é a mais adequada para medir qualquer modelo que fuja do modelo linear. Como sabemos, esse coeficiente assume valores em $[-1, 1]$, sendo que, quanto mais distante do zero o valor assumido, tanto pra mais quanto para menos, a correlação é mais forte (positivamente ou negativamente) e quanto mais próximo de zero a correlação é mais fraca. Mas em certos casos essa definição pode refletir em uma inverdade, pois as variáveis podem ter alguma correlação não-linear a qual esta medida não consegue detectar.

Olhemos então para outras medidas, bastante usuais, que detectam outros tipos de correlação.

2.2.1 O Tau de Kendall

O Coeficiente de correlação de Kendall, chamado mais comumente de $\text{Tau}(\tau)$ de Kendall é calculado pela diferença entre a probabilidade de concordância e discordância das variáveis aleatórias. É geralmente definido matematicamente desta forma:

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}, \quad (5)$$

em que (X_1, Y_1) e (X_2, Y_2) são vetores aleatórios independentes obtidos através das observações provenientes do vetor (X, Y) . Temos também que a relação do τ de

Kendall e a cópula que relaciona X e Y é

$$\tau(X, Y) = 4 \int \int C(u, v) dC(u, v) - 1. \quad (6)$$

É importante notar que $\tau(X, Y) = 4E(C(u, v)) - 1$, ou seja, a integral mostrada em (6) é nada menos que o valor esperado da variável aleatória $C(u, v)$.

2.2.2 Rho de Spearman

Tal qual o τ de Kendall, o Rho de Spearman também é calculado usando as probabilidades de concordância e discordância. Sua medida é definida por:

$$\rho_s(X, Y) = 3[P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}], \quad (7)$$

com (X_2, Y_3) independentes.

Em Embrechts et al. (2001) temos o seguinte teorema:

Teorema 2.2. *Seja (X, Y) um vetor de variáveis aleatórias contínuas com uma cópula C . Então o ρ de Spearman para o vetor (X, Y) é dado por:*

$$\rho_s(X, Y) = 12 \int \int_{[0,1]} C(u, v) dudv - 3$$

Esse teorema mostra a relação do ρ de Spearman com a cópula que relaciona X e Y.

2.3 Coeficiente Caudal

A dependência Caudal é um conceito interessante para o estudo de dependência entre valores extremos. Verifica-se que a dependência caudal entre duas variáveis aleatórias contínuas X e Y é uma propriedade da cópula e, conseqüentemente, a

quantidade de dependência na cauda é invariante sob transformações estritamente de X e Y (Embrechts et al. (2001)).

Definição 2.3. Seja (X, Y) um vetor de variáveis aleatórias contínuas com funções de distribuição marginais F e G . Então, os coeficientes de dependência caudal superior e inferior, λ_U e λ_L para o vetor (X, Y) são dados por:

$$\lambda_U = \lim_{u \rightarrow 1^-} P\{Y > G^{-1}(u) | X > F^{-1}(u)\} \quad (8)$$

$$\lambda_L = \lim_{u \rightarrow 0^+} P\{Y < G^{-1}(u) | X < F^{-1}(u)\} \quad (9)$$

se os limites de $\lambda_U \in [0, 1]$ e $\lambda_L \in [0, 1]$ existem. Se $\lambda_U \in (0, 1]$, X e Y são chamados assintoticamente dependentes na parte extrema da cauda superior, e se $\lambda_L \in (0, 1]$, são assintoticamente dependentes na cauda inferior. Caso contrário, se $\lambda_U = 0$ são assintoticamente independentes na cauda superior ou, se $\lambda_L = 0$, na cauda inferior.

Escrito através de Cópulas, os coeficientes de dependência caudal λ_U e λ_L ficam:

$$\lambda_U = \lim_{u \rightarrow 1^-} \frac{C(u, u)}{1 - u}$$

$$\lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}$$

2.4 Exemplos de Famílias de Cópulas

2.4.1 Os Limitantes de Fréchet-Hoeffding

Para qualquer cópula existem limites de variação que são dados pela versão cópula da desigualdade de Fréchet-Hoeffding, denominados limite inferior W^n e li-

mite superior M^n , isto é:

Teorema 2.3. *Se C é uma cópula qualquer, então, para $u \in [0, 1]^n$:*

$$W^n(u) \leq C(u) \leq M^n(u) \quad (10)$$

e também

Teorema 2.4. *Seja X_1, X_2, \dots, X_n variáveis aleatórias contínuas com cópula C e distribuições marginais F_1, \dots, F_n , então X_1, \dots, X_n são independentes se, e somente se, $C = F_1(X_1)F_2(X_2)\dots F_n(X_n) = \Pi^n$*

A partir desses dois teoremas, dizemos que M^n e Π^n são cópulas n -dimensionais para todo $n \geq 2$, com Π^n sendo a cópula de independência, e W^n não é cópula para todo $n \geq 3$. Embora W^n não seja cópula com n a partir de 3, ele é o melhor limite inferior possível.

2.4.2 Cópulas Gaussianas

Um exemplo de família de cópula é a cópula gaussiana. Uma de suas principais características é que não possui coeficiente de dependência na cauda. Um exemplo do uso prático desse tipo de cópula é calcular o risco de um conjunto de financiamentos garantidos por hipotecas. Essa cópula é usada frequentemente ao se simular dados financeiros pois tem normalidade multivariada, uma suposição comum para estes fins.

A cópula Gaussiana apresenta forma de uma cópula com distribuição normal multivariada n -dimensional, com matriz de correlação R expressa da seguinte ma-

neira:

$$C_R^{Ga}(u) = \Phi_R^n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)),$$

em que:

1. Φ_R^n denota a função de distribuição conjunta de n funções de distribuição normal padrão com matriz de correlação linear R;
2. Φ^{-1} representa a inversa da função de distribuição de uma normal padrão univariada.

Note que:

$$\lambda_L = \lambda_U = 2 \lim_{x \rightarrow -\infty} \Phi\left(x \frac{\sqrt{1-p}}{\sqrt{1+p}}\right).$$

E assim, podemos ver que a cópula gaussiana realmente não possui coeficiente de dependência caudal, pois este é igual a zero, como foi dito anteriormente.

Na figura 1 seguem exemplos gráficos da cópula gaussiana, a sua densidade e suas curvas de nível onde podemos ver claramente a inexistência da dependência caudal.

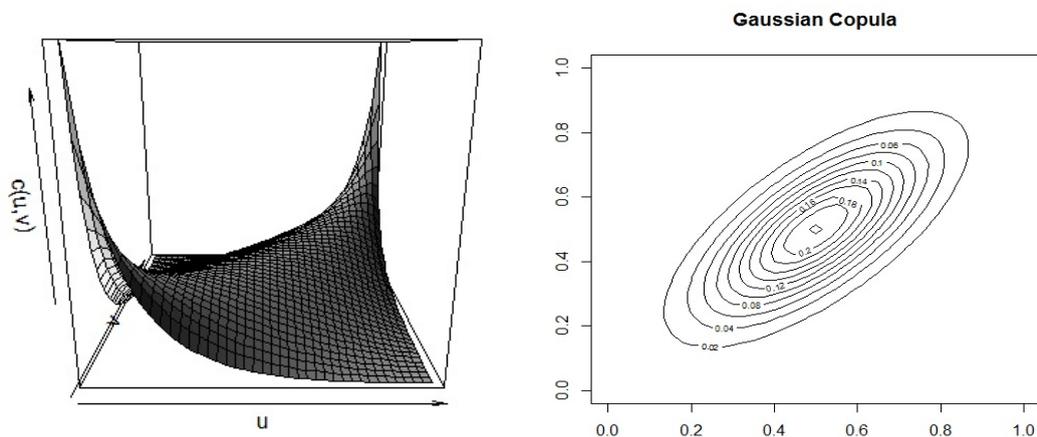


Figura 1: Densidade e Curvas de Nível da Cópula Gaussiana

2.4.3 Cópulas Arquimedianas

As cópulas arquimedianas tem muito destaque na literatura por causa de sua aplicação na modelagem de dados atuariais, econômicos e financeiros, e além disso apresentam expressões analíticas relativamente simples.

Para se definir a cópula arquimediana, primeiro precisa-se definir uma função pseudo-inversa. Segue:

Definição 2.4. Seja φ uma função contínua $[0, 1] \rightarrow [0, \infty]$, estritamente decrescente, tal que $\varphi(1) = 0$. Chama-se pseudo-inversa de φ a função $\varphi^{[-1]} : [0, \infty] \rightarrow [0, \infty]$ dada por:

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases} \quad (11)$$

Então, a partir dessa informação, verificamos que $\varphi^{[-1]}$ é contínua, não-crescente em $[0, \infty]$ e estritamente decrescente em $[0, \varphi(0)]$. Ademais, $\varphi^{[-1]}(\varphi(u)) = u$ em $[0, 1]$ e,

$$\varphi(\varphi^{[-1]}(t)) = \begin{cases} t, & 0 \leq t \leq \varphi(0) \\ \varphi(0), & \varphi(0) \leq t \leq \infty. \end{cases} \quad (12)$$

Sendo assim, se $\varphi(0) = \infty$ então $\varphi^{[-1]} = \varphi^{-1}$, isto é, a função inversa é igual à função pseudo-inversa.

Teorema 2.5. *Seja φ uma função contínua, estritamente decrescente em $[0, 1] \rightarrow [0, \infty]$ tal que $\varphi(1) = 0$, e $\varphi^{[-1]}$ a pseudo-inversa de φ . Seja também C uma função $[0, 1]^2 \rightarrow [0, 1]$ dada por:*

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) \quad (13)$$

Então C é uma cópula arquimediana se, e somente se, φ for convexa.

A função φ também é conhecida por função geradora. A prova deste teorema se encontra em Nelsen (1999).

Algumas propriedades das cópulas arquimedianas são:

1. Dada uma cópula arquimediana $C(u, v)$, esta é simétrica para todo $u, v \in [0, 1]$, ou seja, $C(u, v) = C(v, u)$
2. Seja C uma cópula arquimediana, então ela também é chamada associativa se

$$C(u, C(v, w)) = C(C(u, v), w), \forall u, v, w \in [0, 1].$$

Alguns exemplos de cópulas arquimedianas são: Cópula Clayton, Gumbel, Frank, BB1, BB6, BB7, BB8 e outras. Para este estudo mostraremos alguns exemplos gráficos das citadas. Segue uma tabela 1 das cópulas arquimedianas com suas formas matemáticas:

Tabela 1: Cópulas Arquimedianas

Cópula	Parâmetros	$\Psi'(\Psi^{-1}(u))$
Frank(B3)	$0 \leq \delta < \infty$	$(1/\delta)(1 - e^{\delta u})$
MTCJ(B4)	$0 \leq \delta < \infty$	$-(1/\delta)u^{1+\delta}$
Joe(B5)	$1 \leq \delta < \infty$	$-(1/\delta)(1 - u)^{1-\delta}[1 - (1 - u)^\delta]$
Gumbel(B6)	$1 \leq \delta < \infty$	$-(1/\delta)u(-\log u)^{1-\delta}$
BB1	$1 \leq \delta < \infty; 0 < \theta < \infty$	$-(1/(\theta\delta))u^{\theta+1}(u^{-\theta} - 1)^{1-\delta}$
BB2	$0 < \delta; \theta < \infty$	$-(1/(\theta\delta))u^{1+\theta}e^{-\delta(u^{-\theta}-1)}$
BB3	$0 < \delta < \infty; 1 \leq \theta < \infty$	$-(1/(\theta\delta))u(-\log u)^{1-\theta}e^{-\delta(-\log u)^\theta}$
BB6	$1 \leq \delta; \theta < \infty$	$-(1/(\theta\delta))(1 - u)^{1-\theta}[1 - (1 - u)^\theta] [-\log[1 - (1 - u)^\theta]]^{1-\delta}$
BB7	$1 \leq \delta < \infty; 0 < \theta < \infty$	$-(1/(\theta\delta))(1 - u)^{1-\delta}[1 - (1 - u)^\delta]^{1+\theta}$
BB8	$0 < \delta \leq 1; 1 \leq \theta < \infty$	$-(1/(\theta\delta))(1 - \delta u)^{1-\theta} + (1/(\theta\delta))(1 - \delta u)$
BB9	$0 \leq \alpha < \infty; 1 \leq \theta < \infty$	$-(1/\theta)u(\alpha - \log u)^{1-\theta}$
BB10	$0 < \alpha < \infty; 0 \leq \theta \leq 1$	$-\alpha u - \alpha\theta(1 - \theta)^{-1}u^{1+1/\alpha}$

Agora na figura 2 os gráficos das curvas de nível que representam cada uma dessas cópulas:

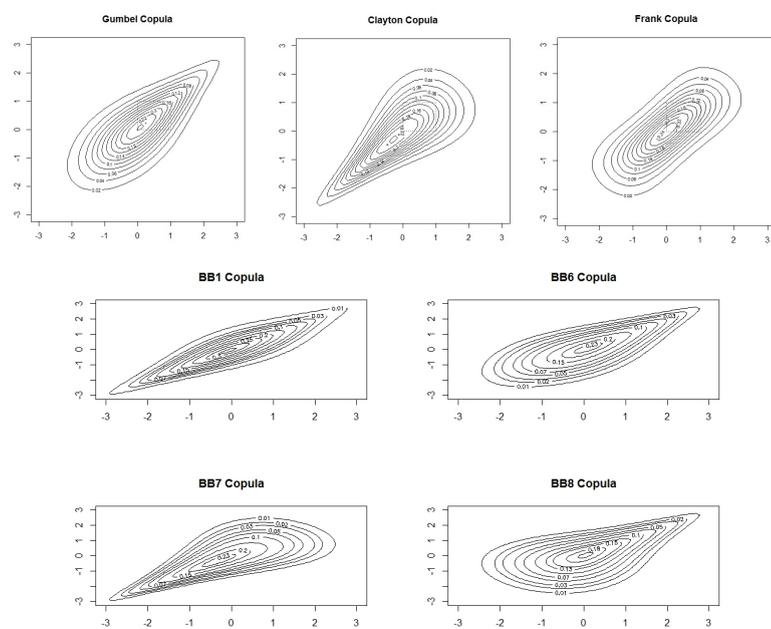


Figura 2: Curvas de Nível das Cópulas Arquimedianas: Gumbel, Clayton, Frank, BB1, BB6, BB7 e BB8

3 Inferência Estatística para Cópulas

Existem diversos métodos que podem ser utilizados para estimar os parâmetros das cópulas descritas no Capítulo 2. Por exemplo, há métodos não paramétricos para as cópulas Arquimedianas e Extremais. Porém, o método que é válido para todas as famílias de cópulas é o de máxima verossimilhança, o qual é descrito neste Capítulo.

3.1 Método de Máxima Verossimilhança Exata

Para utilização desse método, assumimos que C é uma cópula e (X_1, \dots, X_n) é um vetor associado a ela, e as marginais desse vetor são dadas por F_j , onde $j = 1, \dots, n$.

Sendo assim, temos:

$$f(x_1, x_2, \dots, x_n) = c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i), \quad (14)$$

em que:

$$c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = \frac{\partial^n (C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)))}{\partial F_1(x_1) \partial F_2(x_2) \dots \partial F_n(x_n)}, \quad (15)$$

é a enésima derivada parcial mista da cópula C , e c é a densidade da cópula.

A equação (14) é chamada a representação canônica da cópula que, para a função de densidade multivariada, nos permite afirmar que, geralmente, um problema de modelagem estatística para a cópula, pode ser dividido em duas etapas:

- Identificar as distribuições marginais.
- Definir a função cópula apropriada.

Esse é um ponto importante e primordial ao se tratar de questões de estimação, como veremos a seguir.

Seja $X \sim F_1(x_1, \dots, x_m; \alpha_1)$ e $Y \sim F_2(y_1, \dots, y_m; \alpha_2)$ com cópula $C = C_\theta$. Então a expressão que representa a função de verossimilhança é dada por:

$$\begin{aligned} L(\theta) &= f((x_1, \alpha_1), (y_1, \alpha_2); \beta) \cdots f((x_m, \alpha_1), (y_m, \alpha_2); \beta) \\ &= c(F_1(x_1, \alpha_1), F_2(y_1, \alpha_2); \theta) \cdots c(F_1(x_m, \alpha_1), F_2(y_m, \alpha_2); \theta) \cdot \\ &\quad \cdot f_1(x_1, \alpha_1) f_2(y_1, \alpha_2) \cdots f_1(x_m, \alpha_1) f_2(y_m, \alpha_2) \end{aligned} \quad (16)$$

A função log de verossimilhança fica assim:

$$l(\theta) = \sum_{i=1}^m \log c(F_1(x_i, \alpha_1), F_2(y_i, \alpha_2); \beta) + \sum_{j=1}^m (\log f_1(x_i, \alpha_1) + \log f_2(y_i, \alpha_2)) \quad (17)$$

onde θ é o conjunto de parâmetros, tanto da cópula quanto das marginais, ou seja, $\theta = (\alpha_1, \alpha_2, \beta)$.

Consequentemente, dada a cópula e um conjunto de funções de densidade das marginais, a função log de verossimilhança pode ser escrita, e, por maximização desta, obtém-se o estimador de máxima verossimilhança:

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} l(\theta) \quad (18)$$

Pode-se dizer que o estimador de máxima verossimilhança, sob condições de regularidade, existe e é consistente e assintoticamente eficiente.

3.2 Método de Máxima verossimilhança com inferência nas marginais (Método IFM)

O método de máxima verossimilhança, mostrado anteriormente, pode ter um alto custo computacional porque é necessário estimar conjuntamente os parâmetros

das distribuições marginais e os parâmetros das estruturas de dependência representada pela cópula. Entretanto, é possível ver que a função log de verossimilhança é composta por dois termos positivos: um que envolve a densidade da cópula e seu parâmetros, e outro que envolve as marginais e todos os parâmetros da densidade da cópula. Joe (1996), propôs então o método de máxima verossimilhança com inferência nas marginais (Método IFM), mostrando que esse conjunto de parâmetros deveria ser estimado em duas etapas:

1. Estimar os parâmetros $\alpha = (\alpha_1, \alpha_2)$ das marginais, fazendo a estimação das distribuições marginais univariadas:

$$\hat{\alpha} = \arg \max_{\alpha} \left(\sum_{i=1}^m \ln f_1(x_i; \alpha_1) + \ln f_2(y_i; \alpha_2) \right),$$

2. Dado $\hat{\alpha}$, fazemos a estimação do parâmetro β da cópula:

$$\hat{\beta} = \arg \max_{\beta} \sum_{i=1}^m \ln c(F_1(x_i, \hat{\alpha}_1), F_2(y_i, \hat{\alpha}_2); \beta).$$

Esse método é chamado Inferência pelas Marginais (Inference for the Margins - IFM) e ele é descrito como um vetor:

$$\hat{\theta}_{IFM} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta})'.$$

O estimador IFM fornece um ótimo ponto de partida para se obter um estimador de máxima verossimilhança exato (EMV). Joe (1997) também provou que o estimador IFM, sob condições regulares, atende à propriedade de normalidade assintótica, tal qual o EMV.

4 D-Vine Cópulas

Todo o estudo apresentado no capítulo anterior já tem sido bastante trabalhado ao longo dos anos e se mostrou de imensa utilidade para modelagem de funções multivariadas. Os exemplos e toda teoria mostrada, se mostra muito produtivo em âmbitos bivariados, mas quando se trata de trabalhar em âmbitos maiores, uma série de dificuldades aparece e o trabalho com cópulas já não se torna tão fluido. Sendo assim, vários matemáticos e estatísticos tiveram que buscar métodos que melhorassem, ou, otimizassem o uso de funções cópulas em dimensões maiores, como por exemplo a facilitação da estimação de cópulas ou até uma simples análise visual das funções. Joe (1996) propôs um modelo inicial para resolver esse problema e dele surgiram as vine-cópulas e por conseguinte a D-vine cópula. Aas et al. (2009) escreveu mais detalhadamente, explicando muito bem sobre esse assunto.

Este é o capítulo principal deste trabalho. Aqui descrevemos detalhadamente as D-vine cópulas e sua inferência. Iniciamos com as Pair cópulas.

4.1 Pair-copulas

Podemos chamar o modelo Pair-copula de uma construção probabilística que representa modelos multivariados altamente dependentes e complexos de uma forma que se assemelha a modelagem hierárquica clássica. Essa forma é bem simples. Modela-se a dependência através de construção de blocos baseados na independência condicional, o que quer dizer que esse esquema de modelagem decompõe uma densidade multivariada através das variáveis originais e de suas funções de distribuição condicionais e não condicionadas.

Podemos mostrar a decomposição Pair-copula de uma distribuição multivariada geral desta forma:

Considere um vetor $X = (X_1, \dots, X_n)$ de variáveis aleatórias com função de densidade conjunta $f(x_1, \dots, x_n)$. Essa densidade pode ser fatorada como:

$$f(x_1, \dots, x_n) = f_n(x_n)f(x_{n-1}|x_n)f(x_{n-2}|x_{n-1}, x_n)\dots f(x_1|x_2, \dots, x_n), \quad (19)$$

e essa decomposição é única até a reclassificação das variáveis. Utilizando o Teorema de Sklar (1), podemos manipular sua expressão passando para a função de densidade conjunta f , por uma F absolutamente contínua com densidades marginais contínuas e estritamente crescentes F_1, \dots, F_n , e então, usando a regra da cadeia temos:

$$f(x_1, \dots, x_n) = c_{1\dots n}\{F_1(x_1), \dots, F_n(x_n)\}f_1(x_1) \cdots f_n(x_n) \quad (20)$$

para algumas densidades de cópulas n -variadas $c_{1\dots n}(\cdot)$. No caso Bivariado essa expressão é simplificada para:

$$f(x_1, x_2) = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_1(x_1) \cdot f_2(x_2), \quad (21)$$

em que $c_{12}(\cdot, \cdot)$ é a densidade pair-copula apropriada para o par de variáveis transformadas $F_1(x_1)$ e $F_2(x_2)$. Para se obter uma densidade condicional, é fácil ver que:

$$f(x_1|x_2) = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_1(x_1), \quad (22)$$

para a mesma pair-copula. Para três variáveis aleatórias X_1, X_2 e X_3 temos:

$$f(x_1|x_2, x_3) = c_{12|3}\{F(x_1|x_2), F(x_2|x_3)\} \cdot f(x_1|x_3), \quad (23)$$

para a pair-copula $c_{12|3}$ apropriada, aplicada às variáveis transformadas $F(x_1|x_3)$ e $F(x_2|x_3)$. Uma decomposição alternativa seria:

$$f(x_1|x_2, x_3) = c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot f(x_1|x_2), \quad (24)$$

em que $c_{13|2}$ é diferente da pair-copula $c_{12|3}$. Decompondo mais $f(x_1|x_2)$ na pair-copula $c_{13|2}$, nos leva a:

$$f(x_1|x_2, x_3) = c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot c_{12}\{F(x_1), F(x_2)\} \cdot f_1(x_1), \quad (25)$$

onde vemos duas pair-copulas presentes.

Agora, que já foram mostradas as densidades de uma pair-copula em dimensionalidades menores, podemos ver que a fórmula geral fica dessa forma:

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v}_{-j}}\{F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})\} \cdot f(x|\mathbf{v}_{-j}), \quad (26)$$

para um vetor \mathbf{v} d -dimensional. Note que v_j é um componente escolhido arbitrariamente de \mathbf{v} e que \mathbf{v}_{-j} é um vetor \mathbf{v} sem o componente v_j . Concluindo, sob condições regulares apropriadas, uma densidade multivariada pode ser expressa como um produto de pair-copulas. Também é claro que é uma construção iterativa por natureza e que, dada a fatoração específica, tem-se várias re-parametrizações diferentes.

A construção pair-copula envolve a distribuição marginal condicional da forma $F(x|\mathbf{v})$. Para cada j , Joe (1996) mostrou que:

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}\{F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})\}}{\partial F(v_j|\mathbf{v}_{-j})}, \quad (27)$$

em que $C_{ij|k}$ é uma função de distribuição de uma cópula bivariada.

4.2 Vine-copulas

Existem diferentes tipos de construções pair-copulas para cada nível de dimensionalidade, e quanto maior a dimensão, maior será o número de diferentes construções pair-copulas. Bedford e Cooke (2001,2002) organizaram o trabalho de Joe (1996), e introduziram o modelo gráfico chamado de *The Regular Vine*. Existem vários tipos de decomposições de pair-copulas na classe das Regular Vine copulas, mas neste trabalho vamos focar apenas na cópula D-vine, ou Drawable-Vine Copula (Kurowicka e Cooke, 2004) tridimensional.

A expressão geral da estrutura D-Vine no caso tridimensional é:

$$\begin{aligned}
 f(x_1, x_2, x_3) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\
 &\times c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \\
 &\times c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\}.
 \end{aligned} \tag{28}$$

Cada modelo de Regular Vine dá um caminho próprio para decompor a densidade e a D-Vine copula é representada através de um conjunto aninhado de árvores como mostra a figura 3:

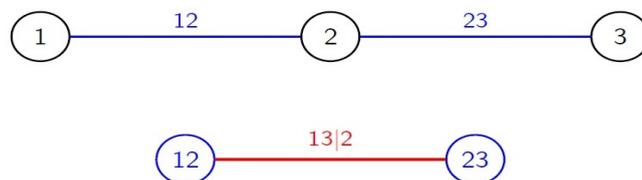


Figura 3: Decomposição da cópula D-vine tridimensional

Nesta imagem vemos a decomposição de um cópula D-Vine com 3 componentes.

Ela consiste em 2 árvores T_j , onde $j = 1, 2$, e tem $4 - j$ nós por árvore T_j e $3 - j$ ligações por árvore T_j . Cada ligação corresponde à densidade da pair-copula e o nome de cada ligação corresponde ao subscrito da densidade da pair-copula, ou seja, a ligação $13|2$ corresponde à densidade da cópula $c_{13|2}(\cdot)$. Podemos perceber que toda a decomposição é composta por $\frac{n(n-1)}{2}$ ligações e pelas densidades marginais de cada variável. Os nós na árvore T_j só são necessários para rotular as ligações na árvore T_{j+1} como podemos ver na imagem acima.

4.3 Ilustrações numéricas para a cópula D-Vine

Nesta seção exploramos a parte inferencial, apresentando o método de estimação das D-Vine cópulas conforme o artigo de Brechman e Schepsmeier (2013).

4.3.1 Representação Gráfica e ordem das variáveis

Antes de falar da estimação em si, é preciso falar da ordem de seleção das variáveis em uma D-Vine copula e sua construção. A construção da D-Vine se dá pela escolha de uma ordem específica de variáveis, então, na primeira árvore temos a dependência da 1ª com a 2ª variáveis, da 2ª com a 3ª, da 3ª com a 4ª e assim sucessivamente. Na segunda árvore temos a dependência condicional da 1ª e 3ª dada a 2ª variável, ou seja, o par $1, 3|2$. Depois temos a 2ª e 4ª dada a 3ª ($2, 4|3$) e assim por diante. Isso se segue em todas as árvores em sequência (por exemplo na terceira árvore teremos a dependência condicional mostrado no par $1, 4|2, 3$ e na quarta árvore o par $1, 5|2, 3, 4$ e assim por diante). Vemos então que a ordem das variáveis muda completamente as dependências condicionais, e assim muda também a estrutura da D-Vine copula. A construção de uma D-Vine copula se dá da seguinte maneira:

- 1^a Árvore— $(1, 2), (2, 3), (3, 4), \dots, (d - 1, d)$
 2^a Árvore— $(1, 3|2), (2, 4|3), \dots, (d - 2, d|d - 1)$
 3^a Árvore— $(1, 4|2, 3), (2, 5|3, 4), \dots, (d - 3, d|d - 2, d - 1)$
 \vdots
 $\text{Árvore}(d - 1)$ — $(1, d|2, \dots, d - 1)$.

Para melhor vizualização, fica dessa forma mostrada na figura 4:

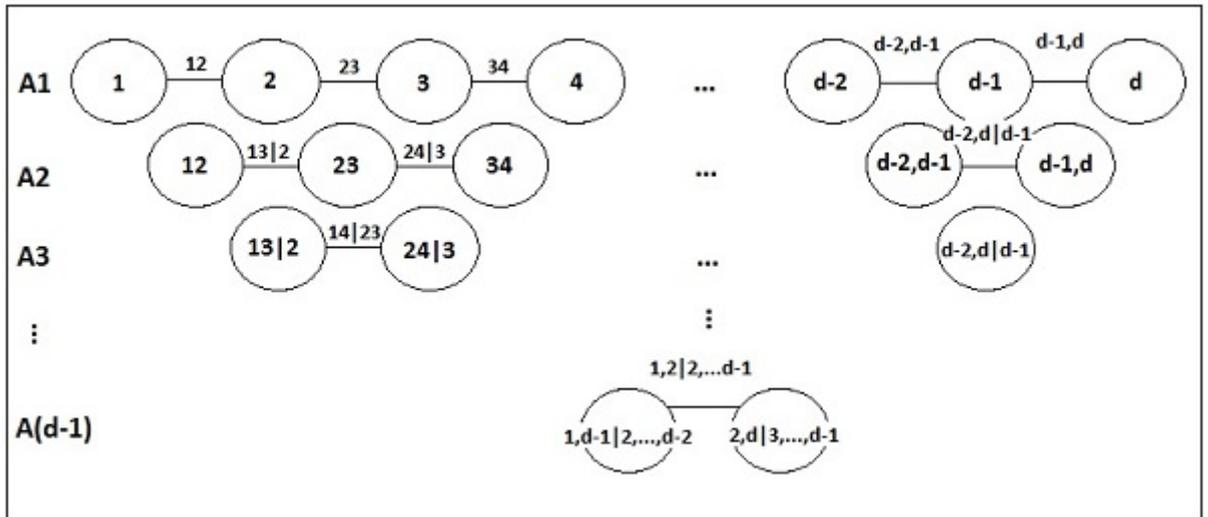


Figura 4: Representação Gráfica Geral da Cópula D-Vine

A partir disso temos a densidade da D-Vine no campo n-dimensional:

$$\begin{aligned}
 f(x) &= \prod_{k=1}^n f(x_k) \\
 &\times \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\dots,i+j-1} \{F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})|\theta_{i,i+j|(i+1):(i+j-1)}\},
 \end{aligned} \tag{29}$$

em que o índice j representa as árvores e o índice i representa as ligações (ou

dependências) entre os nós (variáveis). Na representação gráfica da D-Vine, um nó é conectado a no máximo 2 ligações.

4.3.2 Estimação

Tendo decidido a estrutura da D-Vine a ser usada, deve-se selecionar famílias pair-copula para cada par de variáveis (condicionais) e para isso, segundo Brechmann e Schepsmeier (2013) usa-se a função `CDVineCopSelect` do pacote `CDVine` do R. O processo de seleção de pair-copulas segue árvore por árvore, desde os pares condicionais da árvore dois até a árvore d-1, sempre dependendo das especificações das árvores anteriores através das funções-h. Para explicar melhor essas funções-h vamos a expressão geral da estrutura da D-Vine (no exemplo, o caso tridimensional) mais detalhadamente. Re-escrevemos a expressão (28), como:

$$\begin{aligned}
f(x_1, x_2, x_3) &= f(x_1) \underbrace{c_{12}(F(x_1), F(x_2))}_{f(x_2|x_1)} f(x_2) \\
&\times \underbrace{c_{32|1}(F(x_3|x_1), F(x_2|x_1)) c_{31}(F(x_3), F(x_1))}_{f(x_3|x_2, x_1)} f(x_3) \\
&= \underbrace{f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3)}_{\text{Marginais}} \\
&\times \underbrace{c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\}}_{\text{Cópulas Não Condicionais}} \\
&\times \underbrace{c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\}}_{\text{Cópulas Condicionais}}.
\end{aligned} \tag{30}$$

Em (30) não é possível obter de forma direta as distribuições acumuladas condicionais $F(x_1|x_2)$ e $F(x_3|x_2)$, assim como qualquer outra distribuição condicional da estrutura geral de uma cópula D-Vine. Entretanto, é possível se tomarmos a derivada da cópula com relação a uma das variáveis. Para mostrar isso tomamos uma cópula $C(u_1, u_2)$,

$$\begin{aligned}
C(u_1, u_2) &= C(F_1(x_1), F_2(x_2)), \quad \text{onde } u_1 = F_1(x_1) \text{ e } u_2 = F_2(x_2) \\
&= F(\underbrace{F_1^{-1}(u_1)}_{x_1}, \underbrace{F_2^{-1}(u_2)}_{x_2})
\end{aligned}$$

Sabendo isso, podemos agora obter a distribuição condicional dessa forma:

$$\begin{aligned}
P(U_2 \leq u_2 | U_1 = u_1) &= \frac{\partial}{\partial u_1} C(u_1, u_2) \\
&= \frac{\partial}{\partial u_1} C(\underbrace{F_1(x_1)}_{u_1}, \underbrace{F_2(x_2)}_{u_2}) \\
&= \frac{\partial}{\partial u_1} C(F_1(x_1), F_2(x_2)) f_1(x_1) \tag{31} \\
&= \frac{\partial}{\partial u_1} F(x_1, x_2) f(x_1), \quad \text{dividindo tudo por } f(x_1) \\
&= \frac{\partial}{\partial u_1} C(F_1(x_1), F_2(x_2)),
\end{aligned}$$

ou seja,

$$F_{2|1}(x_2|x_1) = \frac{\partial}{\partial x_1} F_{12}(x_1, x_2) / f(x_1) = \frac{\partial}{\partial u_1} C_{12}(F_1(x_1), F_2(x_2)).$$

Então, mostramos a forma de se obter as distribuições acumuladas condicionais no caso bidimensional. Para o caso tridimensional, utiliza-se o mesmo processo:

$$F_{3|21}(x_3|x_2, x_1) = \frac{\partial F_{321}(x_3, x_2|x_1)}{\partial x_2} / f(x_2|x_1) = \frac{\partial}{\partial u_2} C_{321}(F_{2|1}, F_{3|1}),$$

$F_{2|1}$ e $F_{3|1}$ podem ser obtidas como anteriormente para o caso bivariado, mostrado em (31).

A função-h é simplesmente a forma geral (ou n-dimensional) dessas relações mostradas. Elas representam a função de distribuição condicional quando x e v são

uniformes. De forma geral ela é escrita dessa forma:

$$h(x, v) = F_{j|\mathbf{v}}(x_j|x_{\mathbf{v}}) = \frac{\partial C_{kj|\mathbf{v}_{-k}}(F_{k|\mathbf{v}_{-k}}(x_k|x_{\mathbf{v}_{-k}}), F_{j|\mathbf{v}_{-k}}(x_j|x_{\mathbf{v}_{-k}}))}{F_{k|\mathbf{v}_{-k}}(x_k|x_{\mathbf{v}_{-k}})}, \quad (32)$$

em que $\mathbf{v} = \{k, j_1, \dots, j_n\}$ em que $k < j$ e o subscrito k indica $\mathbf{v}_{-k} = \{j_1, \dots, j_n\}$.

Então, o modelo da cópula D-Vine é enquadrado sequencialmente procedendo de forma iterativa árvore por árvore e assim sua estimação só envolve a estimação de cada pair-copula individualmente. No programa, a estimação pode ser estabelecida utilizando a função CDVineSeqEst, a qual é realizada nos parâmetros de cada pair-copula através de EMV. Ainda que essas estimativas sequenciais muitas vezes proporcionem um bom ajuste, normalmente se está interessado em maximizar a log-verossimilhança da vine-copula. Assim a log-verossimilhança para cópula D-Vine é:

$$\sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \sum_{t=1}^T \log\{c_{i,i+j|i+1,\dots,i+j-1}[F(x_{i,t}|x_{i+1,t}, \dots, x_{i+j-1,t}), F(x_{i+j,t}|x_{i+1,t}, \dots, x_{i+j-1,t})]\},$$

em que $c_{i,i+j|i+1,\dots,i+j-1}[\cdot, \cdot]$ é a densidade da cópula.

No programa a log-verossimilhança de uma Vine-Copula para um conjunto de dados, famílias pair-copula e parâmetros pode ser obtido usando a função CDVineLogLik. Com os cálculos dessa log-verossimilhança, pode-se agora estimar os parâmetros em conjunto usando EMV em contraste com a estimativa sequencial de pares falado anteriormente.

No caso de inferência para um modelo de 3 variáveis, distribuídas uniformemente em $[0, 1]$ temos:

$$\sum_{t=1}^T \{\log c_{12}(x_{1,t}, x_{2,t}, \Theta_{11}) + \log c_{23}(x_{2,t}, x_{3,t}, \Theta_{12}) + \log c_{13|2}(v_{1,t}, v_{2,t}, \Theta_{21})\}$$

onde:

$$v_{1,t} = F(x_{1,t}, x_{2,t}) = h(x_{1,t}, x_{2,t}, \Theta_{11})$$

e

$$v_{2,t} = F(x_{3,t}, x_{2,t}) = h(x_{3,t}, x_{2,t}, \Theta_{12}).$$

Os parâmetros a serem estimados são $\Theta = (\Theta_{11}, \Theta_{12}, \Theta_{21})$, onde $\Theta_{j,i}$ é um conjunto de parâmetros da densidade da cópula correspondente $c_{i,i+j|i+1,\dots,i+j-1}(\cdot, \cdot)$.

5 Aplicação

Até agora apresentamos a teoria de cópulas, até chegarmos à D-Vine cópula, que é o cerne deste trabalho. Nesta seção, apresentaremos uma aplicação da teoria da cópula D-Vine com três conjuntos de dados, desde a análise exploratória dos dados até a estimação e sua representação gráfica. Para tal utilizamos o pacote CDVine do software R.

5.1 O Conjunto de Dados

Este conjunto de dados contém os resíduos padronizados transformados de log retornos diários dos principais índices de ações do mundo em 2009 e 2010. São 396 observações para cada um dos três países, sendo eles Estados Unidos da América, Alemanha e Grã-Bretanha. Cada série cronológica está devidamente filtrada através dos modelos ARMA(1,1)-GARCH(1,1) com inovações t-Student.

5.2 Análise exploratória dos dados

A análise exploratória dos dados não é feita da mesma forma que em um conjunto de dados comum. Estamos lidando com cópulas, portanto vamos analisá-las par a par. Começando com o cálculo da correlação de Kendall (tabela 2) e alguns gráficos (figuras 5 e 6):

```
cor('dados',method='kendall')
```

Tabela 2: Correlação de Kendall

País	Estados Unidos	Alemanha	Grã-Bretanha
Estados Unidos	1.000	0.503	0.486
Alemanha	0.503	1.000	0.732
Grã-Bretanha	0.486	0.732	1.000

```
plot(dados[,1],dados[,2], xlab="EUA", ylab="Alemanha", pch = ".", cex = 5.0)
```

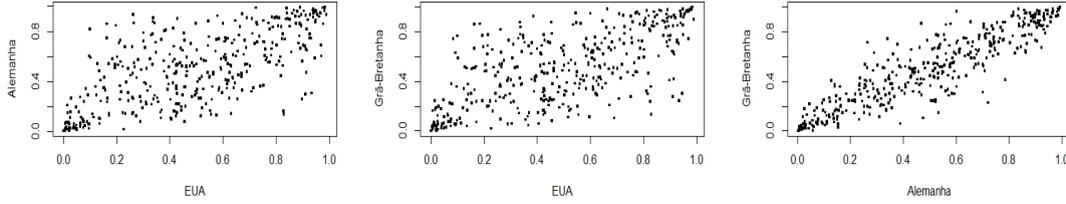


Figura 5: ScatterPlots: EUA-Alemanha, EUA-GRB e Alemanha-GRB, respectivamente.

```
BiCopMetaContour(dados[,1],dados[,2],bw=1.7,size=100,
```

```
levels=c(0.01,0.03,0.05,0.07,0.1,0.15,0.18),par=1,family="emp")
```

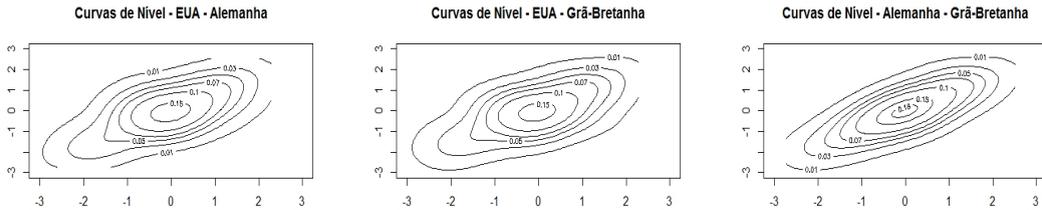


Figura 6: Curvas de Nível com distribuição empírica.

Pelos ScatterPlots vemos que existe uma relação entre os mercados de ações desses países. Entre os europeus a relação é mais forte e entre o americano com os europeus, existe uma relação mais fraca. Também podemos ver que existe uma dependência caudal simétrica, e as curvas de nível nos dão suporte ao afirmar isso.

Fizemos testes de independência utilizando a função $BiCopInd-Test(USA,GER)$p.value$ do programa. A estatística do teste é: $T = \sqrt{\frac{9N(N-1)}{2(2N+5)}} \times |\hat{\tau}|$ onde N é o número de observações e o $\hat{\tau}$ é o Tau de Kendall das cópulas. O teste foi realizado para cada par das variáveis e em todos eles o p-valor foi zero, o que nos informa que há uma dependência muito grande entre as variáveis.

Por último, realizamos o teste de seleção, que busca indícios de qual família de cópulas pertencem as pair-cópulas. Para tal, fizemos o teste de seleção básico utilizando a função `CDVineCopSelect(dados,familyset=c(1:10,13,14,16:20),type=2,selectioncrit="AIC", indeptest=FALSE,level=0.05)`. Esse teste baseia nos critérios de informação AIC ou BIC. Primeiramente todas as cópulas são devidamente ajustadas usando estimação por máxima verossimilhança. Então os critérios são computados para todas as famílias disponíveis no programa e a família com o valor mínimo é escolhida. O critério AIC para uma cópula bivariada é:

$$AIC := -2 \sum_{i=1}^N \ln[c(u_{i,1}, u_{i,2}|\theta)] + 2k. \quad (33)$$

para observações $u_{i,j}$ com $i = 1, \dots, N$ e $j = 1, 2$ onde $k = 1$ para um parâmetro e $k = 2$ para dois parâmetros (isso nos casos das cópulas t-Student, BB1, BB6, BB7, BB8). O critério BIC é similar:

$$BIC := -2 \sum_{i=1}^N \ln[c(u_{i,1}, u_{i,2}|\theta)] + \ln(N)k. \quad (34)$$

Como resultado para a relação entre dados de retorno financeiro dos países europeus obteve-se a t-student como família selecionada e para a relação dos dados do país americano com os europeus obteve-se uma cópula arquimediana de nome BB7.

Sabemos que a montagem da cópula D-Vine é feita árvore por árvore, então na segunda árvore tem-se dois nós que são as pair-cópulas obtidas da árvore de cima. No nosso caso a Pair-cópula EUA,ALEMANHA e a outra é ALEMANHA,GRÃ-BRETANHA e precisamos selecionar a que família pair-copula se adequa a relação

entre essas duas outras pair-copulas. Para isso utilizamos a função do R `CDVineCopSelect`, que segue o mesmo princípio da função `BiCopSelect` e dos testes (33) e (34), mas faz para todas as árvores da cópula D-Vine. Então para essa última a família selecionada foi a da cópula arquimediana BB8.

5.3 Estimação

Após a análise exploratória dos dados vamos partir para inferência estatística, mas como dito no capítulo anterior, antes de fazermos a estimação precisamos determinar a ordem das variáveis que constituirão a primeira árvore da seleção da D-Vine. Sendo assim, o ideal é deixar como o nó central os índices alemães pois através da análise exploratória dos dados, tomando por base a tabela 2 percebemos que é o que tem relação mais forte com os outros dois índices, o americano e o da Grã-Bretanha. Sendo assim, utilizando essa ordem e as famílias de cópulas selecionadas pelo programa podemos começar com a parte da estimação.

Utilizando as funções `CDVineSeqEst` e `CDVineMLE`, estimamos os parâmetros das pair-copulas através da estimação por máxima verossimilhança. A primeira função estima os parâmetros por estimação sequencial e a segunda através de estimação conjunta. Vemos os resultados gerados pela tabela 3.

family < -c(9,2,20)

seqPar < -CDVineSeqEst(*dados*, *family* = *family*, *type* = 2, *method* = "mle")

mlePar < - CDVineMLE(*dados*, *family* = *family*, *start* = *seqPar*\$*par*, *start2* =
seqPar\$*par2*, *type* = 2)

Tabela 3: Parâmetros Estimados

Método/Parâmetros	Par1	Par2	Par3
Sequencial1	1.909	0.910	1.355
Sequencial2	1.209	11.887	0.914
Conjunta1	1.929	0.911	1.355
Conjunta2	1.214	11.887	0.914

E então calculamos a Log-Verossimilhança otimizada de ambos os tipos de estimação e vemos a pequena melhoria dos parâmetros estimados pela conjunta em relação a sequencial. Podemos constatar os resultados na tabela 4.

CDVineLogLik(*dados*, *family* = *family*, *par* = *mlePar*\$*par*, *par2* =
mlePar\$*par2*, *type* = 2)\$loglik

Tabela 4: Log-Verossimilhança

Método	Log-Verossimilhança
Sequencial	532.986
Conjunta	533.0092

Por fim, obtemos o gráfico da cópula D-Vine:

CDVineTreePlot(*dados*, *family* = *family*, *par* = *mlePar*\$*par*, *par2* = *mlePar*\$*par2*
, *names* = *colnames*(*dados*), *type* = 2, *tree* = 1, *edge.labels* = c("family", "theotau"))

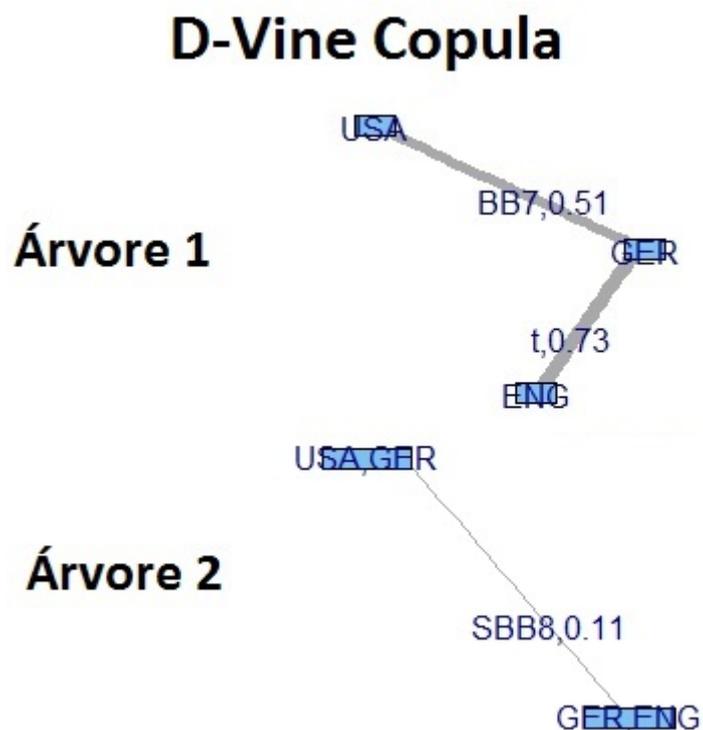


Figura 7: Gráfico D-Vine cópula obtido através dos dados de retorno financeiro da Alemanha, Estados Unidos e Grã-Bretanha

A figura 7 segue a mesma forma das figuras 3 e 4, mas com seu desenho gerado pelo programa. Nela podemos ver os “nós” das árvores representando cada país, as ligações dos nós com a família de cópula selecionado e valor da relação entre os nós, inclusive na árvore 2, a qual vemos o valor da relação geral entre todas as componentes gerado pelas funções condicionais através das funções-h.

6 Conclusão

O objetivo traçado para este trabalho foi atingido. Consegui estudar a teoria geral de cópulas, das cópulas D-vine, bem como trabalhar com os programas computacionais próprios para análise, estimação e representação gráfica das cópulas D-vine. Fica para um trabalho futuro a aplicação de cópulas D-vine no cálculo de medidas de risco de finanças e atuária.

Referências

- [1] Aas, K., C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, 2009.
- [2] Bedford, T.J., and Cooke, R.M., Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 2001.
- [3] Brechmann, E C., Schepsmeier, U., *Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine*, 2013.
- [4] Embrechts, McNeil, A., and Straumann, D. “Correlation and Dependency in Risk Management: Properties and Pitfalls”. *Value at Risk and Beyond*. Cambridge University Press, 2001.
- [5] Fisher, N. I. Copulas, *Encyclopedia of Statistical Sciences*, Update Vol. 1, 159-163. John Wiley Sons, New York, (1997).
- [6] Joe, H. “Families of m-variate distributions with given margins and $m(m-1) = 2$ bivariate dependence parameters”. In L. Rüchendorf and B. Schweizer and M. D. Taylor (Ed.), *Distributions with fixed marginals and Related Topics*, 1996.
- [7] Joe, H., *Multivariate Models and Dependence Concepts*. Chapman and Hall, London (1997).

- [8] Kurowicka, D. and R. M. Cooke. Distribution - free continuous bayesian belief nets. In *Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice*, Santa Fe, New Mexico, 2004.
- [9] Nelsen, R B., *An Introduction to Copulas*, Springer Science and Business Media, New York, 1999.
- [10] Sklar, A., Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 1959.
- [11] Cherubini U., Luciano E., Vecchiato W., *Copula Methods in Finance*, Wiley Finance, 2004.