

Estudo comparativo de famílias de cópulas via simulações

Ricardo T. A. Oliveira^{1†}, Thaíze F. O. Assis², Paulo Renato A. Firmino³, Tiago A. E. Ferreira⁴

¹ Professor do Instituto Federal de Mato Grosso do Sul (IFMS). E-mail: ricardo.oliveira@ifms.edu.br.

² Professora Efetiva da Diretoria Acadêmica de Ciências (IFRN-CNAT-DIAC). E-mail: thaize.fernandes@ifrn.edu.br.

³ Professor Adjunto do Centro de Ciência e Tecnologia (UFRCA). E-mail: paulo.firmino@ufca.edu.br.

⁴ Professor Adjunto do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco (UFRPE-DEINFO). E-mail: tiago.espinola@ufrpe.br.

Resumo: Este trabalho apresenta uma introdução sobre modelos de cópulas, especificamente as cópulas empíricas e semi-empíricas, dedicando a avaliar as vantagens e desvantagens de se adotar cópulas empíricas, baseadas em distribuições marginais não-paramétricas das variáveis envolvidas, e as semi-empíricas que recorrem a ajustes paramétricos para as distribuições marginais. Simulando dados de variáveis dependentes linear e não-linear, assim como variáveis independentes que seguem distribuições exponenciais, ilustram-se as semelhanças entre estas alternativas de modelagem de cópulas. Evidencia-se a partir do estudo, os bons resultados da cópula empírica em relação à semi-empírica, bem como a necessidade de estudos mais profundos sobre as contribuições dos modelos de cópulas.

Palavras-chave: Cópulas empíricas; Cópulas semi-empíricas; Distribuições marginais.

Abstract: This paper provides an introduction to copulas models, specifically the empirical and semi-empirical copulas, dedicated to evaluate the advantages and disadvantages of adopting empirical copulas, based on nonparametric marginal distributions of the involved variables, and the semiempirical which resort to parametric adjustments for the marginals distributions. Simulation data from linear and non-linear dependent variables, as well as independent variables that follow exponential distributions, illustrate the similarities between these alternatives of copulas models. It is evident from the study, the good results of the empirical copula in relation to the semi-empirical, as well as the need for more in-depth studies on the contributions of copulas models.

Keywords: Empirical Copulas; Semi-empirical Copulas; Marginal Distributions.

Introdução

Os modelos de Cópulas são uma ferramenta aplicável na modelagem multivariada onde o estudo da dependência é evidente (MARCHI, 2008). Tais modelos são destaques na literatura devido a sua variabilidade de estruturas de dependência (OLIVEIRA, 2013b; VIOLA, 2009). O conceito de cópulas foi introduzido por Sklar (1959) *apud* Nelsen (2006) e frisa que a distribuição conjunta de um conjunto de variáveis pode ser função das distribuições marginais destas variáveis. Em outras palavras, a função cópula une funções de distribuição marginais em uma distribuição conjunta, sendo desta forma possível modelar uma estrutura de dependência das próprias variáveis envolvidas fazendo uso das respectivas distribuições marginais (LEAL, 2013; OLIVEIRA, 2013a; OLIVEIRA, 2016a).

† Autor correspondente: ricardotavares.gm@gmail.com.

Muitas são as famílias de cópulas a serem adotadas para modelar a dependência entre variáveis, desde as empíricas até as diversas alternativas paramétricas e semi-paramétricas (AZAM, 2013). A título de ilustração, Kao e Govindaraju (2010) abordam o problema de secas causadoras de prejuízos a agricultores norte-americanos. Os autores propõem, a partir de uma cópula empírica, implementar um algoritmo capaz de modelar o estado geral da seca por meio de cópulas. Eles afirmam o papel importante que as cópulas desempenham na análise de secas. No trabalho de Marchant *et al.* (2011) é abordado o problema de previsão de variação das propriedades do solo, possibilitando auxiliar os políticos e gestores ambientais. Neste estudo, os autores recorreram a cópula Gaussiana, demonstrando que o modelo Gaussiano não foi apropriado para representar o problema. Apesar disso, os autores mencionaram que metodologias baseadas em cópulas podem ser capazes de representar mais características gerais do solo e, assim, aprimorar previsões. Já Grimaldi e Serinaldi (2006) apresentam uma análise estatística multivariada sobre inundação. Utilizam cópulas paramétricas da classe Arquimedianas para estudar a distribuição conjunta de três variáveis específicas: o volume, o pico e a duração de inundações. Os autores declaram parecer condição rigorosa utilizar distribuições marginais de mesma família para todas as variáveis analisadas.

Cópulas são funções que relacionam duas ou mais distribuições marginais de forma a construir uma distribuição conjunta (GENEST, 2007; SABINO DA SILVA, 2017). A vantagem deste formalismo é que o analista pode se dedicar, em um primeiro momento, ao ajuste da distribuição marginal de cada variável e, apenas em seguida, modelar a dependência entre estas variáveis (OLIVEIRA, 2013b). O presente trabalho presta-se a revisar as alternativas de inferência propostas pela literatura de cópulas e as implicações práticas de algumas de suas propriedades. A partir de casos de estudo baseados em simulações, o artigo ilustra porque recorrer a marginais não-paramétricas mostra-se ao menos tão interessante quanto às paramétricas durante a etapa de inferência sobre os parâmetros de dependência da cópula, devido a propriedade de invariância da cópula a transformações monótonas sobre as variáveis. Destacando o fato de que qualquer variável contínua pode ser transformada em uma uniforme no intervalo $[0,1]$ a partir da sua distribuição acumulada marginal, cópulas podem ser usadas para fornecer uma estrutura de dependência multivariada separadamente das distribuições marginais (YAN, 2007; YAN, 2010). Como reflexo dessa propriedade de invariância, será ilustrada a capacidade de cópulas para modelar apenas a tendência da dependência entre duas variáveis (se positiva ou negativa), negligenciando aspectos mais específicos da forma funcional de tal dependência (se linear ou não-linear).

O artigo está dividido tal como segue. Na próxima seção, conceitos básicos de cópulas são introduzidos, assim como as alternativas difundidas pela literatura para inferência sobre os parâmetros da família de cópulas adotada. Em seguida, na seção 3, resultados e discussões baseados em casos de estudo gerados via simulação são apresentados. O artigo apresenta suas conclusões na seção 4.

Material e Métodos

Cópulas

Cópulas são funções que medem a dependência entre variáveis (GENEST, 2007; OLIVEIRA, 2018). Mais especificamente na modelagem estatística, considerando um vetor

envolvendo k variáveis aleatórias, $\mathbf{X}=(X_1,\dots,X_j,\dots,X_k)$, uma cópula $C(\cdot)$ é uma função de distribuição acumulada conjunta que opera sobre as distribuições acumuladas marginais das variáveis em \mathbf{X} , $\mathbf{V}=(V_1,\dots,V_j,\dots,V_k)$, onde para uma dada instância de \mathbf{X} , $\mathbf{x}=(x_1,\dots,x_j,\dots,x_k)$,

$$v_j=P(X_j\leq x_j), j=1,2,\dots,k, \quad (1)$$

e $C(V_1\leq v_1,\dots,V_j\leq v_j,\dots,V_k\leq v_k)$ é a cópula. As distribuições marginais da $C(\cdot)$ são uniformes no intervalo $[0,1]$ e sua imagem é o intervalo $[0,1]$.

O conceito de cópulas surgiu quando Sklar (1959) *apud* Nelsen (2006) percebeu que a função de distribuição acumulada conjunta $F_{\mathbf{X}}(x_1,\dots,x_j,\dots,x_k)$ para qualquer conjunto de variáveis aleatórias \mathbf{X} pode ser escrita da seguinte forma:

$$F_{\mathbf{X}}(x_1,\dots,x_j,\dots,x_k) = C(v_1,\dots,v_j,\dots,v_k), \quad (2)$$

onde, $\mathbf{v}=(v_1,\dots,v_j,\dots,v_k)$, dados pela Eq. (1), são instâncias das distribuições acumuladas marginais das variáveis em \mathbf{X} no ponto $\mathbf{x}=(x_1,\dots,x_j,\dots,x_k)$. A Eq. (2) acima trata-se do teorema de Sklar, que fundamenta cópulas e, de forma geral, afirma que:

Teorema 1 *Seja $\mathbf{X}=(X_1,\dots,X_j,\dots,X_k)$ um vetor de variáveis aleatórias com função de distribuição acumulada conjunta F e acumuladas das marginais $F_{X_j}(x_j)$, $j=1,2,\dots,k$. Então a cópula C é uma função acumulada k -dimensional, sendo escrita da seguinte forma:*

$$F_{\mathbf{X}}(x_1,x_2,\dots,x_k)=C(F_{X_1}(x_1),\dots,F_{X_j}(x_j),\dots,F_{X_k}(x_k)), \quad (3)$$

de forma que se $F_{X_1}(x_1),\dots,F_{X_j}(x_j),\dots,F_{X_k}(x_k)$ forem contínuos então C é única. ■

Considerando v_j como na Eq. (1), se tem do teorema 1 e das igualdades (2) e (3), que

$$C(v_1,\dots,v_k) = P(V_1 \leq v_1,\dots,V_k \leq v_k). \quad (4)$$

Sobre a função de densidade de probabilidade conjunta de \mathbf{X} , $p_{\mathbf{X}}(x_1,\dots,x_j,\dots,x_k)$, pode-se demonstrar que:

$$p_{\mathbf{X}}(x_1,\dots,x_k) = \frac{\partial^k F_{\mathbf{X}}(x_1,\dots,x_k)}{\partial x_1 \dots \partial x_k} = c(F_{X_1}(x_1),\dots,F_{X_k}(x_k)) \prod_{j=1}^k p_{X_j}(x_j), \quad (5)$$

onde $p_{X_j}(x_j)$ é a função de densidade de probabilidade marginal de X_j e a função $c(\cdot)$ representa a densidade de probabilidade de cópula:

$$c(v_1,\dots,v_k) = \frac{\partial^k}{\partial v_1 \dots \partial v_k} C(v_1,\dots,v_k). \quad (6)$$

Assim, a densidade de cópula é também a função densidade de probabilidade conjunta para as variáveis de \mathbf{V} .

Como destacado anteriormente, há diversas famílias paramétricas de cópulas a serem consideradas do lado direito da Eq. (4). A título de ilustração, a cópula $C(\cdot)$ é dita Arquimediana se admite a representação,

$$C(v_1, \dots, v_k) = \varphi^{-1}[\varphi(v_1) + \dots + \varphi(v_k)], \quad (7)$$

onde $\varphi(t)$ é conhecida como a função geradora da cópula com inversa $\varphi^{-1}(t)$ se $\varphi(0) = \infty$ (NELSEN, 2006).

A classe das cópulas Arquimedianas abrange uma grande variedade de estruturas de dependência. Em particular, as cópulas Arquimedianas podem assumir dependência caudal assimétrica, sendo uma propriedade a favor de sua aplicação à modelagem de dados com estrutura de dependência assimétrica (OLIVEIRA, 2013^a). Quando a função geradora é dada por

$$\varphi(t) = (-\ln t)^\theta, \quad \forall \theta \geq 1, \quad (8)$$

$\varphi^{-1}(t) = \exp\left(-t^{\frac{1}{\theta}}\right)$ e a Eq. (7) resulta na chamada cópula de Gumbel-Hougaard (SILVA

FILHO, 2013). Neste caso, a distribuição acumulada conjunta é dada por

$$F_X(x_1, \dots, x_k) = C(v_1, \dots, v_k) = \exp\left\{-\left[\sum_{j=1}^k (-\ln v_j)^\theta\right]^{\frac{1}{\theta}}\right\}, \quad (9)$$

onde v_j é dado pela Eq. (1).

Na cópula de Gumbel-Hougaard (GENEST, 2007; OLIVEIRA, 2013b; OLIVIERA, 2016a; OLIVEIRA, 2016b), o grau de dependência entre as variáveis em \mathbf{X} é medido unicamente a partir do parâmetro θ . Para $\theta = 1$ obtêm-se a cópula de marginais independentes e se $\theta \rightarrow \infty$ obtêm-se a cópula dedicada à dependência perfeita positiva. Desta forma, a cópula de Gumbel-Hougaard permite modelar desde situações de independência até de dependência perfeita positiva. Ela trata de uma cópula de valores extremos e apresenta dependência na cauda superior (EUGÊNIO FILHO, 2017; LEAL, 2013).

Por outro lado, é também possível se trabalhar com cópulas empíricas. As cópulas empíricas são baseadas no conjunto de n amostras (instâncias) independentes do vetor \mathbf{X} disponíveis. Nelas, $C(\cdot)$ é aproximada pela distribuição não paramétrica $C_n(\cdot)$:

$$C_n(v_1, \dots, v_k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\left(\frac{R_1(x_{i,1})}{n+1} \leq v_1, \dots, \frac{R_j(x_{i,j})}{n+1} \leq v_j, \dots, \frac{R_k(x_{i,k})}{n+1} \leq v_k\right), \quad (10)$$

onde $\mathbb{1}(A)$ denota a função indicadora do conjunto A , $x_{i,j}$ representa a i -ésima observação da variável j e $R_j(x_{i,j})$ a posição (ordem ou *ranking*) do valor $x_{i,j}$ quando as observações da variável X_j são postas em ordem crescente. Um dos principais argumentos em favor da cópula apresentada na Eq. (10) reside na sua simplicidade e propriedades de convergência. À medida que n cresce, o estimador $C_n(v_1, \dots, v_k)$ tende a uma distribuição normal com média $C(v_1, \dots, v_k)$

(GENEST, 2007). Das equações (9) e (10), podem-se destacar alternativas de modelagem de cópulas tanto paramétricas quanto empíricas (não-paramétricas).

De modo a simbolizar o funcionamento da cópula empírica baseada em ordem (*ranking*), foram elaboradas duas tabelas com dados fictícios. A Tabela 1 apresenta instâncias de duas variáveis X_1 e X_2 , os vetores $\mathbf{x}_i=(x_{i,1}, x_{i,2})$ (com $i=1, 2, \dots, n=5$). A Tabela 2 mostra os *rankings* associados aos elementos de \mathbf{x}_i . Desta forma, $R_j(x_{i,j})$, ($j=1,2$), representará a ordem de $x_{i,j}$ quando as instâncias da variável X_1 , forem ordenadas crescentemente. Pode-se observar que o menor valor apresentado pelas instâncias de X_1 é $x_{1,1} = 0.0765$, logo $R_1(x_{1,1}) = 1$, ou seja, a menor ordem possível (dado que R_1 representa os *rankings* das instâncias da variável X_1 e R_2 de X_2). O maior valor em X_1 é $x_{2,1}=0.876$, logo $R_1(x_{2,1})=5$ (o número que representa a maior ordem para X_1). Os demais casos seguem o mesmo raciocínio.

Tabela 1. Dados fictícios das variáveis X_1 e X_2 para ilustração do funcionamento do *ranking* das cópulas empíricas.

I	1	2	3	4	5
$x_{i,1}$	0.0765	0.876	0.161	0.210	0.428
$x_{i,2}$	2.729	0.200	0.548	0.246	0.551

Tabela 2. *Ranking* fictício gerado a partir das variáveis X e Y .

I	1	2	3	4	5
$R_1(x_{i,1})$	1	5	2	3	4
$R_2(x_{i,2})$	5	1	3	2	4

Outra consequência muito importante de cópulas reside no fato de que se $Y_j = \rho(X_j)$ é uma função monótona crescente de X_j , então

$$F_X(x_1, \dots, x_k) = C(v_1, \dots, v_k) = F_Y(y_1, \dots, y_k). \quad (11)$$

Em palavras, esta consequência implica no fato de que cópulas são capazes apenas de modelar a tendência da dependência entre duas ou mais variáveis (se positiva ou negativa), negligenciando aspectos mais específicos da forma funcional de tal dependência (GENEST, 2007). Assim, para duas variáveis X_1 e X_2 , seja $X_2 = X_1$ ou $X_2 = X_1^2$, a cópula associando ambas as variáveis será a mesma. Esta propriedade de cópulas será ilustrada no presente trabalho através de vários casos de estudo baseados em simulações. Além disso, será estudada a influência da estrutura das distribuições acumuladas das marginais $F_{X_j}(x_j)$ sobre a cópula. Mais especificamente, serão estudadas as cópulas empíricas onde $F_{X_j}(x_j)$ é aproximada tanto por estimativas de estatísticas de ordem baseadas nos dados (Eq. (10)) quanto por distribuições paramétricas.

Métodos de estimação dos parâmetros de cópulas

Na literatura atual, existem diversos métodos de estimação e inferência sobre a cópula mais adequada aos dados empíricos (NELSEN, 2006; SILVA FILHO, 2010; SILVA FILHO, 2013; LEAL, 2013; JOE, 2013). Em muitos trabalhos são citados tanto métodos não-

paramétricos como as cópulas empíricas, quanto paramétricos de máxima verossimilhança (MV) e de maximização a partir das distribuições marginais, também conhecida como (IFM), acrônimo para “*Inference function for margins*” (GENEST, 2007; JOE, 2013). Há ainda as cópulas semi-paramétricas, onde as marginais não-paramétricas são consideradas para o ajuste da cópula paramétrica (NELSEN, 2006). Todas estas alternativas serão brevemente introduzidas a seguir.

Método de Máxima Verossimilhança

O método de máxima verossimilhança (MMV) é usualmente adotado para estimar tanto os parâmetros das distribuições marginais $F_{X_j}(x_j)$ quanto da cópula paramétrica $C(\cdot)$. O MMV estima simultaneamente os parâmetros das distribuições marginais de \mathbf{X} e de $C(\cdot)$. Os estimadores obtidos são aqueles que maximizam a função de log-verossimilhança resultante da Eq. (5) diante dos dados evidenciados \mathbf{x}_i (BOLFARINE, 2001; YAN, 2010):

$$\begin{aligned} l &= \sum_{i=1}^n \ln(p_{X_1, \dots, X_k}(x_{i,1}, \dots, x_{i,j}, \dots, x_{i,k} | \alpha_1, \dots, \alpha_j, \dots, \alpha_k, \theta)) \\ &= \sum_{i=1}^n \ln(c(F_{X_1}(x_{i,1} | \alpha_1), \dots, F_{X_k}(x_{i,k} | \alpha_k) | \theta)) + \sum_{i=1}^n \sum_{j=1}^k \ln(p_{X_j}(x_{i,j} | \alpha_j)) \end{aligned} \quad (12)$$

onde α_j representa o conjunto de parâmetros de $F_{X_j}(x_j)$ e θ o conjunto de parâmetros de dependência da densidade de cópula $c(\cdot)$, genericamente apresentada na Eq. (6). Por sua vez, o vetor $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,j}, \dots, x_{i,k})$ representa a i -ésima amostra (instância ou evidência) do vetor de variáveis \mathbf{X} . Assim, via MMV, os parâmetros $\theta, \alpha_1, \dots, \alpha_k$ são todos estimados simultaneamente, de forma a maximizar a função l . Esta estimação conjunta torna a obtenção de estimativas para os parâmetros bastante complexa (LEAL, 2013).

Método IFM (*inference function for margins*)

O método IFM permite separar o processo de estimação em duas etapas. Primeiramente, são estimados os parâmetros das marginais (α_j) para, em seguida, serem estimados os parâmetros da cópula (θ). Neste sentido, é comum que se adote o MMV em ambas as etapas de estimação, separando-se a soma da Eq. (12) em duas funções a serem otimizadas em sequência (YAN, 2007). Desta forma, o problema de inferência das marginais é tratado primeiramente de maneira usual, e pode-se recorrer a testes de ajuste tais como o de Kolmogorov-Smirnov (veja Simard (2011)) ou a critérios de qualidade de ajuste como o AIC (sigla para *Akaike information criterion*) (AKPA, 2011). Apenas em um segundo momento é que o modelo de cópulas é, de fato, tratado.

Esta alternativa de inferência mostra-se interessante por diversos motivos. Em termos práticos, pode-se estar interessado em estudar a dependência entre variáveis cujo comportamento marginal tenha sido previamente modelado, tal como ocorre em problemas de incerteza de modelos, por exemplo (FIRMINO, 2010). Por outro lado, embora que as estimativas obtidas via IFM não sejam necessariamente equivalentes às do MMV, as

diferenças são desprezíveis na grande parte dos casos (LEAL, 2013). Por fim, destaque-se ainda o fato de que ao decompor o problema de inferência, IFM o torna computacionalmente atraente, já que se reduz a sua complexidade em relação ao MMV. Para maiores detalhes sobre IFM e sua comparação com MMV, sugere-se literatura mais específica (SILVA FILHO, 2010; JOE, 2013; LEAL, 2013; MELO, 2008; SANTOS, 2012).

Métodos Bayesianos

A literatura também apresenta outras alternativas paramétricas além do MMV e IFM para inferir sobre os parâmetros da distribuição conjunta uma vez obtidos os dados. Por exemplo, Rossi (2012) implementa o algoritmo de Metropolis-Hastings utilizando uma abordagem Bayesiana, na qual os parâmetros são estimados em conjunto e separados, a exemplo de MMV e IFM, utilizando métodos com base nas simulações de Monte Carlos via cadeias de Markov.

Métodos Semi e Não-Paramétricos

Nos métodos não-paramétricos, as cópulas são construídas somente a partir da distribuição de frequência dos dados e modelos matemáticos paramétricos são negligenciados (EUGÊNIO FILHO, 2017). Desta abordagem emergem as cópulas empíricas descritas na Eq. (10). Alguns autores têm se dedicado a esclarecer a importância desta abordagem, principalmente devido à sua simplicidade de aplicação e propriedades estatísticas (GENEST, 2007).

Como já destacado anteriormente, as cópulas empíricas não fazem uso de modelos paramétricos nem sobre as distribuições marginais das variáveis em X nem sobre a própria cópula. Nelas, a distribuição acumulada da variável X_j é aproximada pela distribuição de frequências acumuladas de acordo com os dados, $R_j(x_{i,j})$, por sua vez baseada na ordem das observações em $x_{i,j}$ ($i=1,\dots,n$; $j=1,\dots,k$). As cópulas empíricas foram introduzidas primeiramente por Deheuvels (1979) *apud* Nelsen (2006), chamando-a na ocasião de “função de dependência empírica”. Elas tendem a ser computacionalmente eficientes e são estimadores centrados para a verdadeira cópula à medida que o tamanho da amostra n cresce (IYENGAR, 2009; KAO, 2010).

Sobre as cópulas semi-paramétricas, as marginais são usualmente aproximadas pela distribuição de frequências acumuladas $R_j(x_{i,j})$. Contudo, a cópula $C(\cdot)$, diferentemente do método não-paramétrico, se dá por um modelo matemático (NELSEN, 2006).

Embora pouco mencionada na literatura, é ainda possível que a modelagem envolva distribuições marginais paramétricas ajustadas aos dados e uma cópula não-paramétrica. Para tanto bastaria substituir na Eq. (10) $R_j(x_{i,j})/(n+1)$ pela função $F_j(x_{i,j})$, ajustada via MMV, por exemplo. No presente artigo, esta modalidade de cópula será denominada de "semi-empírica".

Na próxima subseção, as diferenças entre cópulas empíricas e semi-empíricas serão ilustradas a partir de casos de simulação. O intuito é verificar as semelhanças e diferenças entre os modelos de cópulas obtidos.

Casos de estudo

As simulações executadas neste trabalho foram geradas a partir do *software* R (TEAM, 2021), sempre envolvendo duas variáveis $\mathbf{X}=(X_1=X, X_2=Y)$ e diversas relações entre elas. O intuito era constatar se há algum ganho em termos de acurácia do modelo de cópula semi-empírica em relação à empírica, uma vez que a cópula semi-empírica tem um custo computacional maior a depender da forma das distribuições marginais $F_{X_j}(x_j)$. Destaca-se que esta cópula é raramente citada na literatura, embora factível. Além da análise gráfica das curvas de nível das cópulas obtidas, foi adotado como critério de comparação a capacidade da cópula de captar a tendência do relacionamento entre as variáveis em \mathbf{X} e, quando pertinente, o momento em que tal relação mudava de sentido.

Desta forma, cada relação foi simulada a partir da geração de 100 pontos aleatórios. Os pontos aleatórios foram gerados probabilisticamente a partir de duas distribuições, sempre considerando Y como sendo uma função determinística de X no caso de haver dependência. A ausência de ruído nas relações entre X e Y foi adotada de forma a dar um maior destaque à acurácia de cada método de modelagem.

Especificamente, foi gerada uma amostra de tamanho $n=100$ para X , denominada \mathbf{x} , de forma que X segue uma distribuição Exponencial ($\lambda=0,5$), onde $\lambda=1/E[X]$ sendo $E[\cdot]$ a função média.

Sobre Y , foram consideradas duas amostras a serem pareadas com \mathbf{x} ; em um primeiro momento, uma amostra independente de X (denominada y_I) e outra funcionalmente dependente de \mathbf{x} (denominada y_{II}), de forma a se avaliar as cópulas nos casos onde X e Y são ou não independentes. Para o primeiro cenário, se considerou $Y_I \sim$ Exponencial ($\lambda=0,5$).

Para o segundo momento, diversas relações de dependência entre X e Y foram simuladas, observe o Quadro 1 a seguir.

Quadro 1. Relações de dependência entre X e Y .

(a)	(b)	(c)
$y_{II} = x+2, y_{II} = x^3+2$	$y_{II} = -x+2, y_{II} = -x^3+2$	$y_{II} = -(x+5)^2, y_{II} = (x+5)^2$

Nota-se que nos casos em (a) ilustram-se relações de dependência positiva entre X e Y (isto é, Y é uma função crescente de X), enquanto em (b) trata-se de casos de dependência negativa (isto é, Y é uma função decrescente de X). Já em (c), estudam-se casos em que o sentido da dependência muda, envolvendo regiões de dependência positiva e negativa. Observe-se que quaisquer outras relações de dependência que permitissem ilustrar dependências lineares e não-lineares crescentes ou decrescentes ou com mudança de sentido poderiam ter sido usadas no estudo.

Destaca-se que como os modelos de cópulas são invariantes as transformações monótonas crescentes da variável \mathbf{X} (SILVA FILHO, 2013), eles são capazes de capturar apenas o sentido da dependência, se positiva ou negativa, entre as variáveis. Como será ilustrado a seguir, por exemplo, a cópula associada às relações presentes em (a) será a mesma.

Resultados e discussão

Para o primeiro momento do estudo, a Figura 1 ilustra o comportamento das cópulas empírica e semi-empírica quando as variáveis são independentes. Se percebe que a Figura 1 (a) apresenta os dados gerados a partir de variáveis que seguem distribuições exponenciais independentes entre si, já em (b) mostra a relação entre as distribuições acumuladas marginais baseadas em ordem (*rankings*), (R_X vs R_Y), apresentados na forma dos quadrados em preto, e entre as distribuições marginais paramétricas exponenciais (F_X vs F_Y), representados pelos triângulos em cinza.

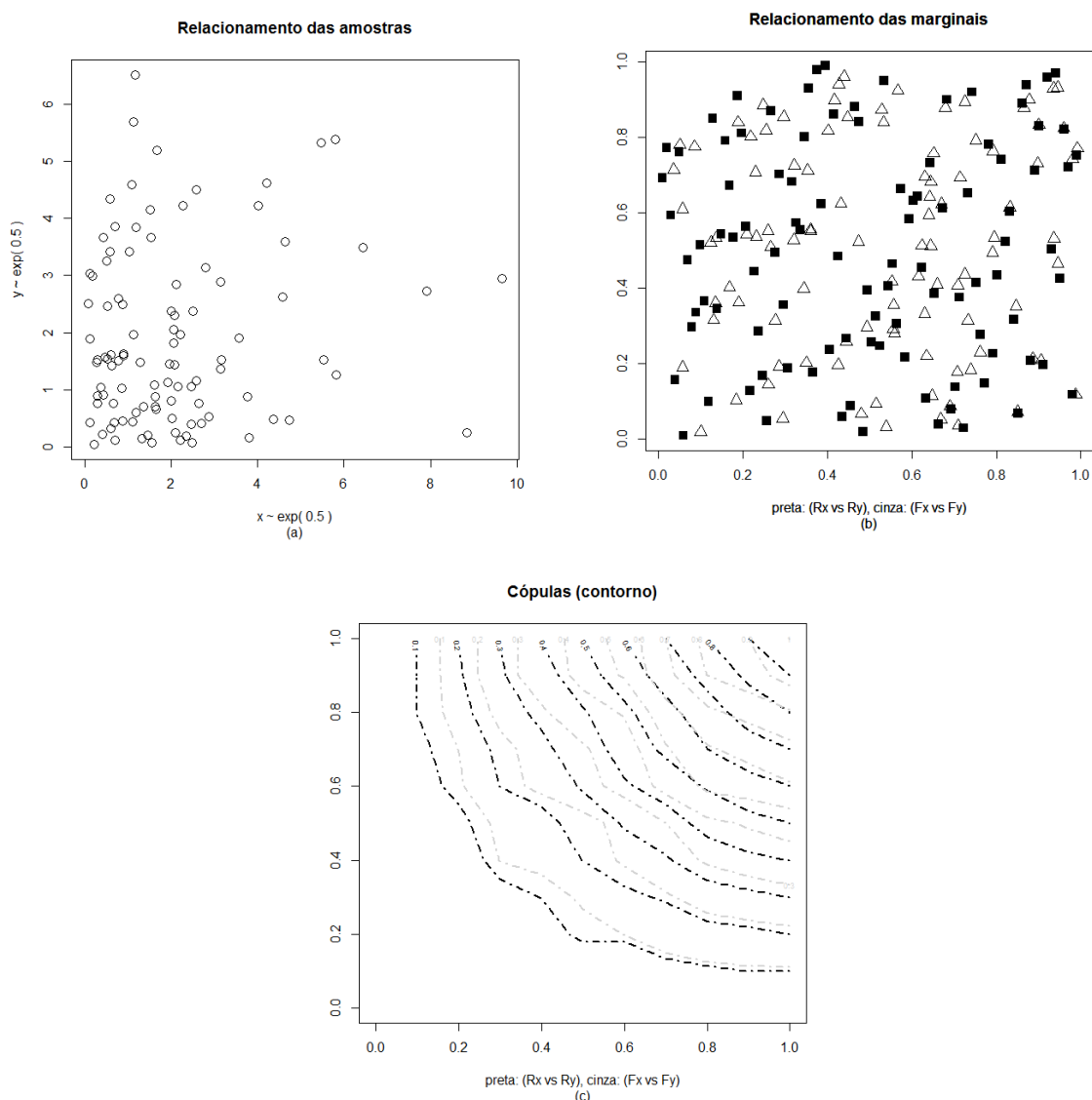


Figura 1. Gráficos de pares de 100 dados simulados a partir de variáveis exponenciais independentes (a), respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (R_X vs R_Y) e distribuições paramétricas exponenciais (F_X vs F_Y) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

Já em (c), as curvas de nível em preto representam a cópula empírica e as em cinza referem-se à cópula semi-empírica. Como pode-se observar, nas Figura 1 (a) e (b), as variáveis não apresentam qualquer comportamento correlato, desta forma, os dados plotados estão completamente desorganizados, sendo difícil visualizar qualquer relação entre os pontos. A Figura 1 (c), gerada pelas cópulas em questão, mostram como as curvas de nível se comportam com o formato de ondas nos casos em que as distribuições marginais são independentes. Não haveria, a partir de uma análise visual, razões para destacar maiores diferenças entre ambas as curvas de nível.

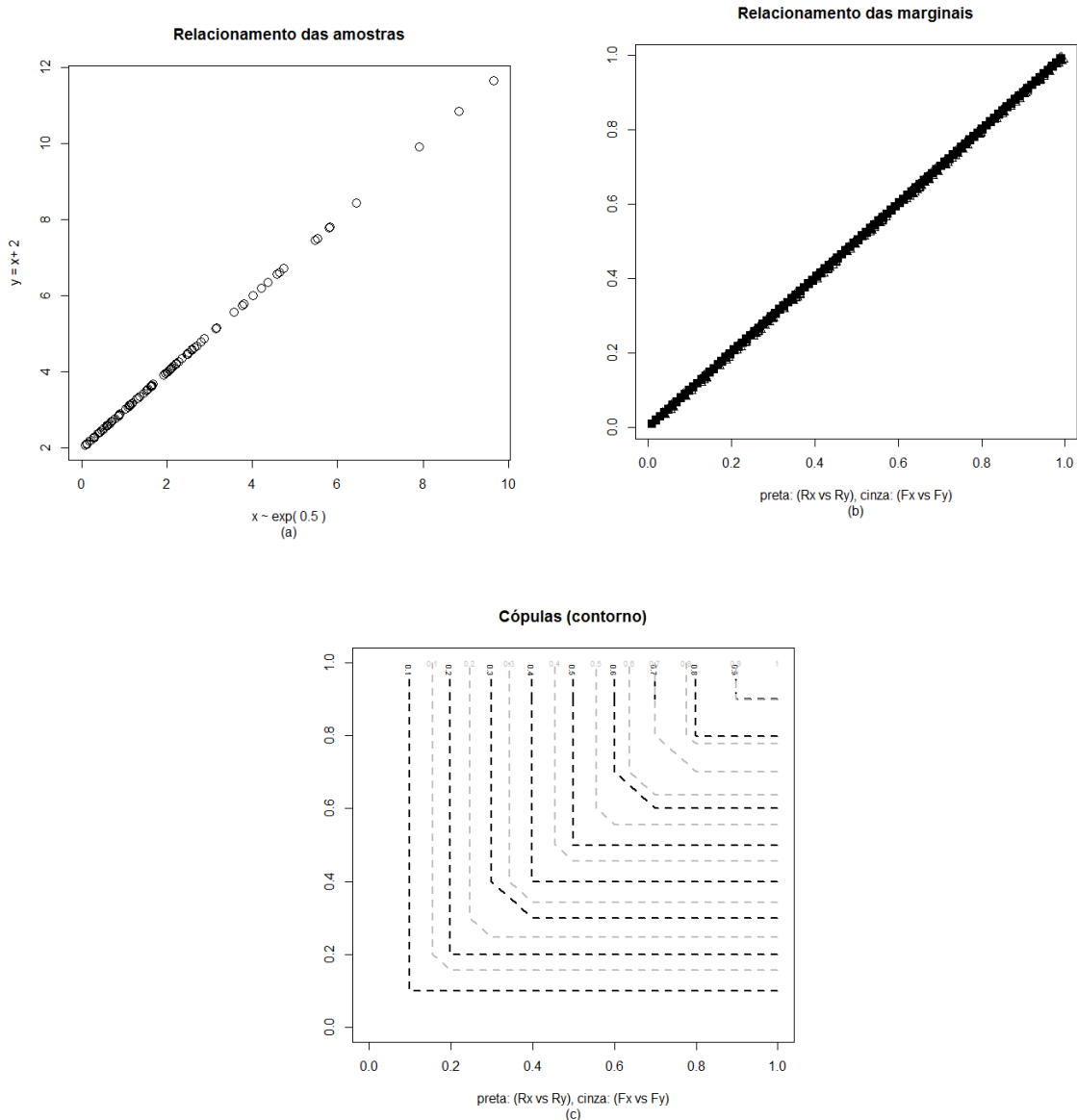


Figura 2. Gráficos de pares de 100 dados simulados a partir de X exponencial, $Y = X+2$, respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (Rx vs Ry) e distribuições paramétricas exponenciais (Fx vs Fy) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

A Figura 2 ilustra o comportamento da cópula quando a relação entre X e Y é linear e positiva. Nesta Figura apresenta-se a simulação a partir da distribuição exponencial para X e da relação $Y = X+2$. Assim como na Figura 1 (b), esta Figura apresenta a função acumulada das marginais baseadas em ordem (*rankings*) de (R_X vs R_Y), apresentados na forma dos quadrados pretos e as distribuições paramétricas exponenciais de (F_X vs F_Y), representadas pelos triângulos cinzas. Enquanto (c) ilustra o comportamento das cópulas quanto as distribuições exponenciais são positivas. Nota-se em (b), que as relações entre as acumuladas empíricas e paramétricas se apresentam da mesma forma, o que conduz às cópulas em (c) apresentaram comportamentos semelhantes.

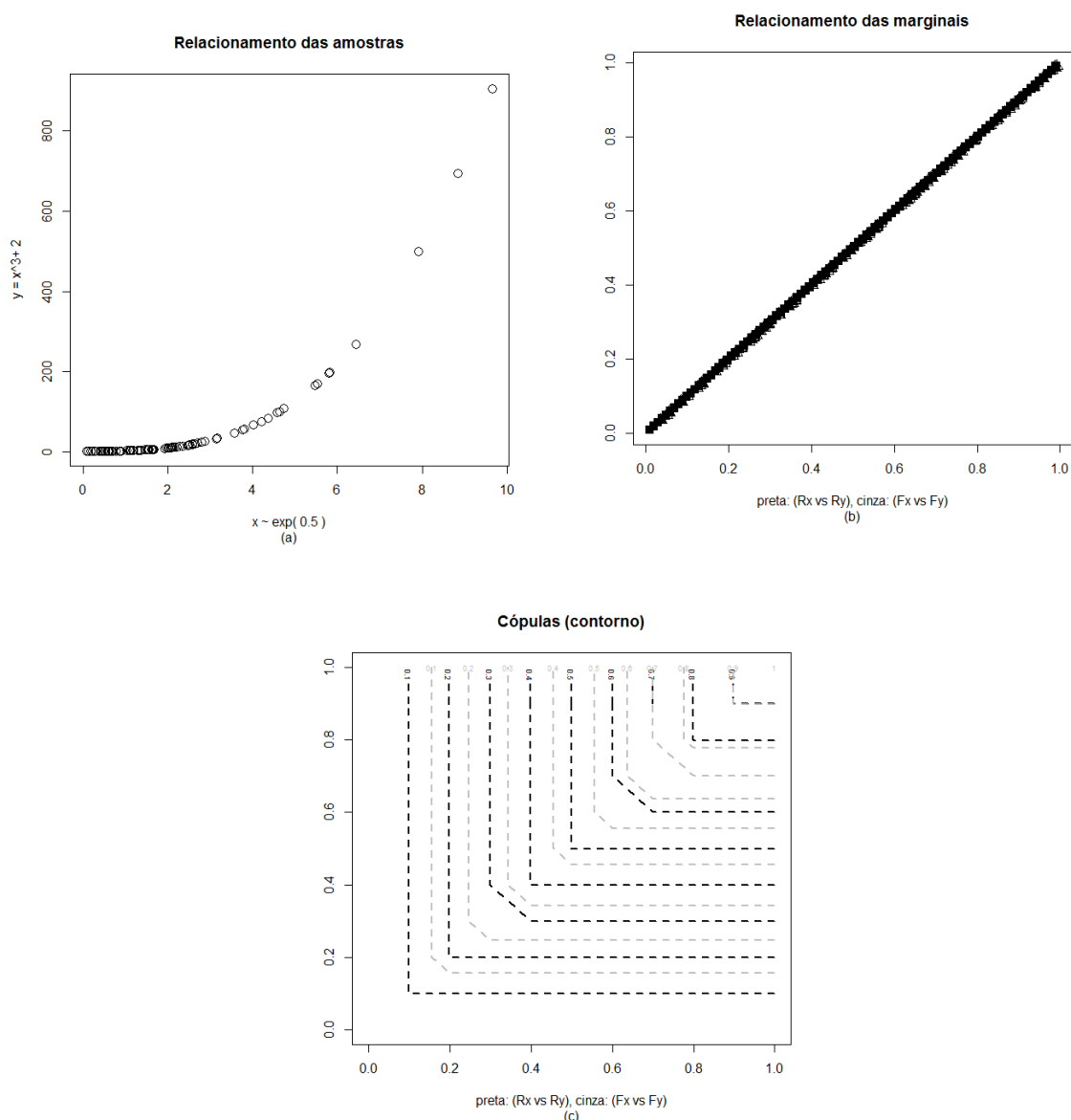


Figura 3. Gráficos de pares de 100 dados simulados a partir de X exponencial, $Y = X^3+2$, respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (R_X vs R_Y) e distribuições paramétricas exponenciais (F_X vs F_Y) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

A Figura 3, assim como a Figura 2, foi simulada a partir de uma distribuição exponencial para X e da relação $Y = X^3 + 2$. Fica evidente nas Figuras 2 (b) e 3 (b) que as relações entre as acumuladas empíricas e paramétricas são semelhantes, mais uma vez conduzindo a cópulas semelhantes (Figuras 2(c) e 3(c)). Logo, as cópulas empíricas e semi-empírica não foram capazes de captar as diferenças apresentadas nas simulações quando a dependência é linear ($Y = X+2$) e não-linear ($Y = X^3 + 2$).

A Figura 4 apresenta o comportamento da cópula empírica e semi-empírica quando X segue uma distribuição exponencial e Y é uma função com dependência linear e negativa de X ($Y = -X+2$).

Nas Figuras 4 (b) e (c), percebe-se pouca diferença entre as relações envolvendo marginais empíricas e paramétricas. As curvas de nível são, mais uma vez, semelhantes. Comparando as Figura 2 (c) e 3 (c) com a Figura 4 (c) fica claro que as curvas de nível se comportam diferente quando a dependência assumida for positiva ou negativa.

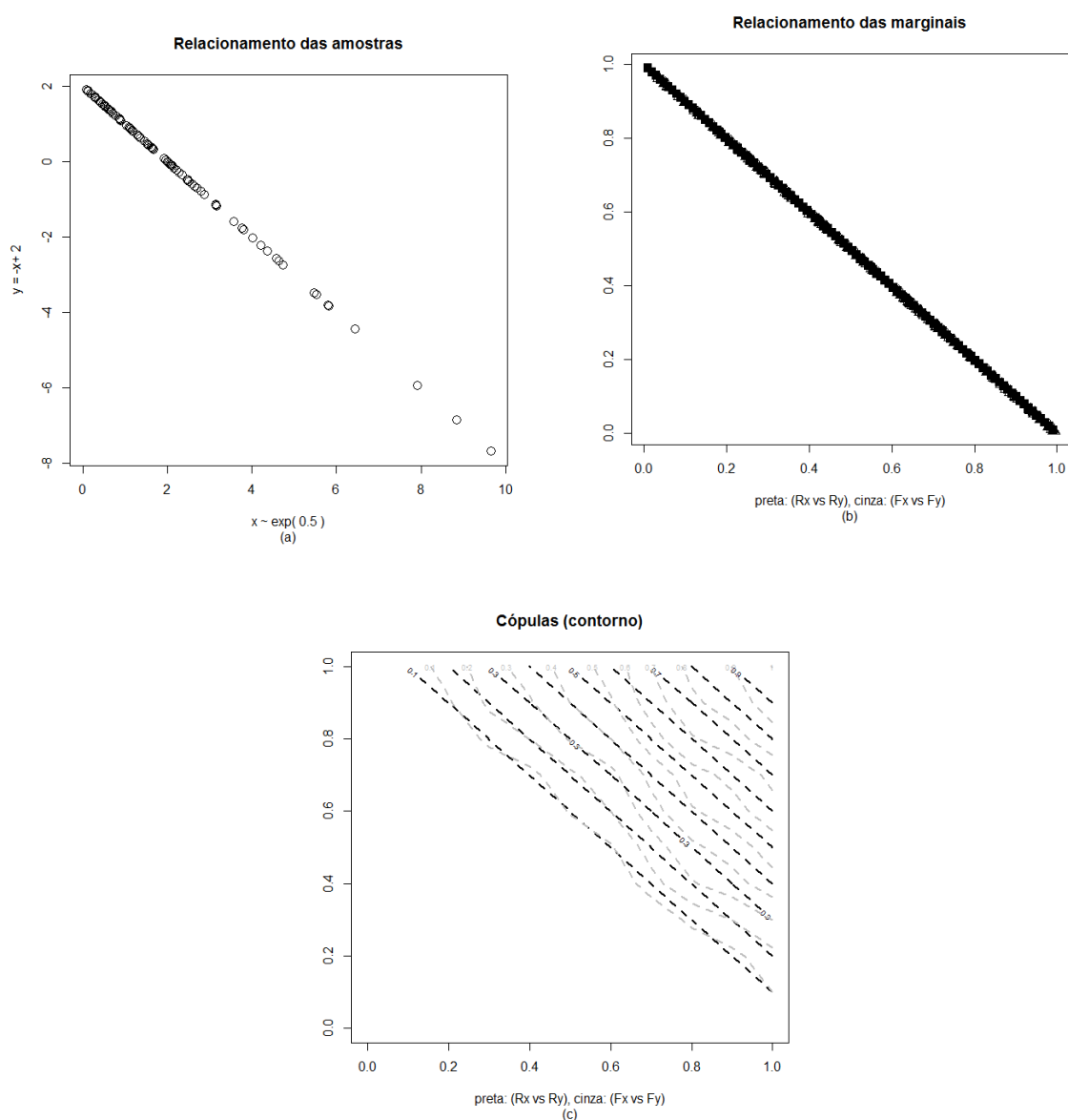


Figura 4. Gráficos de pares de 100 dados simulados a partir de X exponencial, $Y = -X + 2$, respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (R_x

vs R_y) e distribuições paramétricas exponenciais (F_x vs F_y) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

A Figura 5 apresenta a simulação a partir da distribuição exponencial para X e da relação $Y = -X^3 + 2$. Destaca-se as semelhantes nas interpretações das Figuras 5 (b) e (c) em relação aos resultados apresentados nas Figura 4 (b) e (c), ou seja, as cópulas para este caso também não conseguiram captar os detalhes quanto à natureza da dependência negativa entre as variáveis. Analisando a Figura 4 (c) e 5 (c) é notável que a relação de dependência das distribuições marginais são negativas e ambas as cópulas não foram capazes de minuciar a relação de dependência entre as marginais dadas por $Y = -X + 2$ e $Y = -X^3 + 2$.

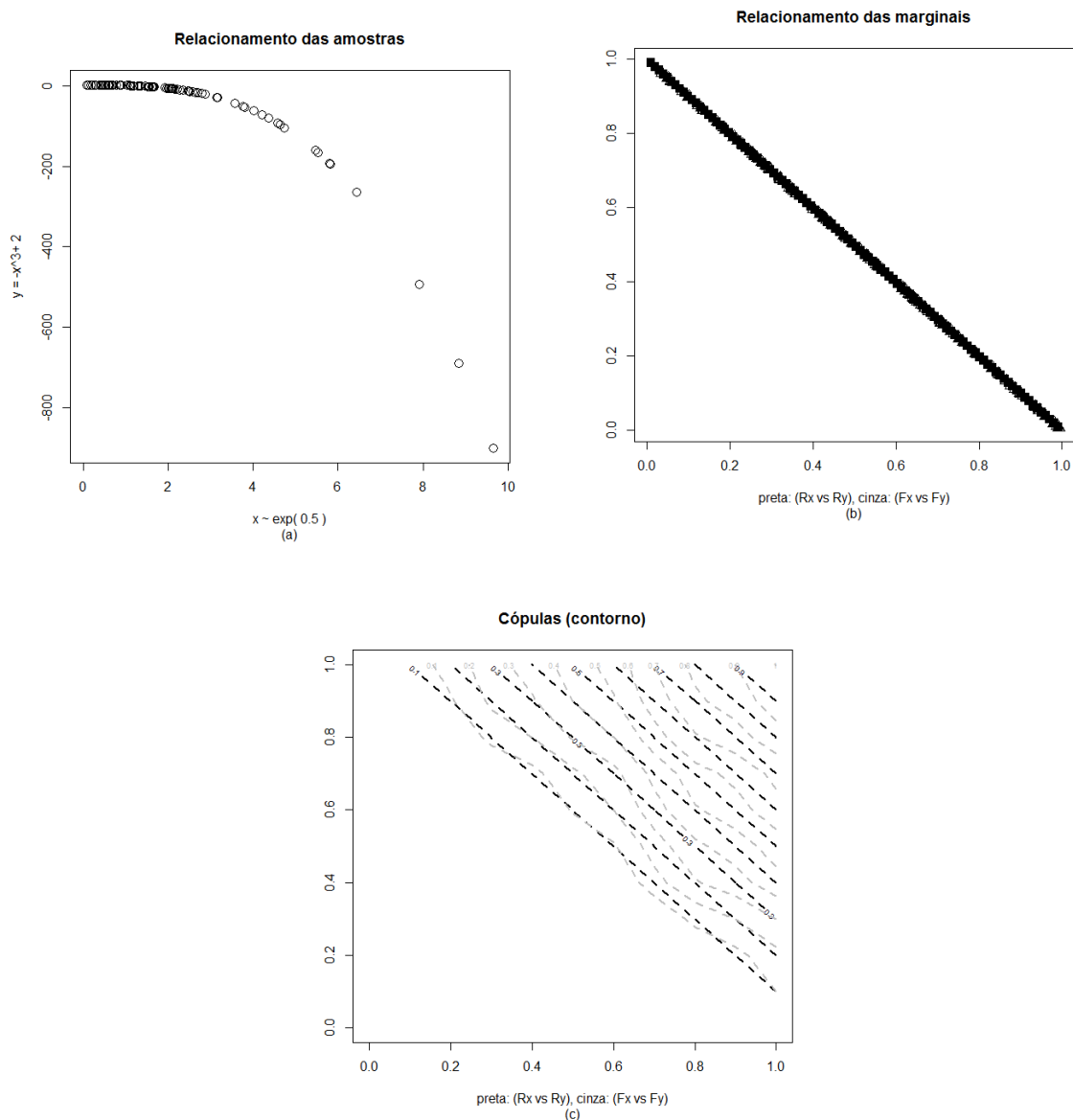


Figura 5. Gráficos de pares de 100 dados simulados a partir de X exponencial, $Y = -X^3 + 2$, respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (R_x vs R_y) e distribuições paramétricas exponenciais (F_x vs F_y) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

A Figura 6 ilustra o comportamento da cópula empírica e semi-empírica quando X segue uma distribuição exponencial e Y é uma função de X cujo sinal da dependência é positivo em uma região do espaço de possibilidades de X e negativo em outra. Para realizar tal simulação, se utilizou a estrutura de dependência dada por $Y = -(X-5)^2$.

Com base nos resultados apresentados nas figuras anteriores pode-se constatar que a cópula empírica e semi-empírica não foram capazes de captar o comportamento específico das relações entre X e Y , atendo-se apenas a captar a tendência destas relações. Já nas Figuras 6 e 7, pode-se perceber que as relações entre as marginais empíricas e as marginais paramétricas, assim como as respectivas curvas de nível, são capazes de captar as tendências e são também precisas quanto ao momento em que tais tendências mudam de sinal.

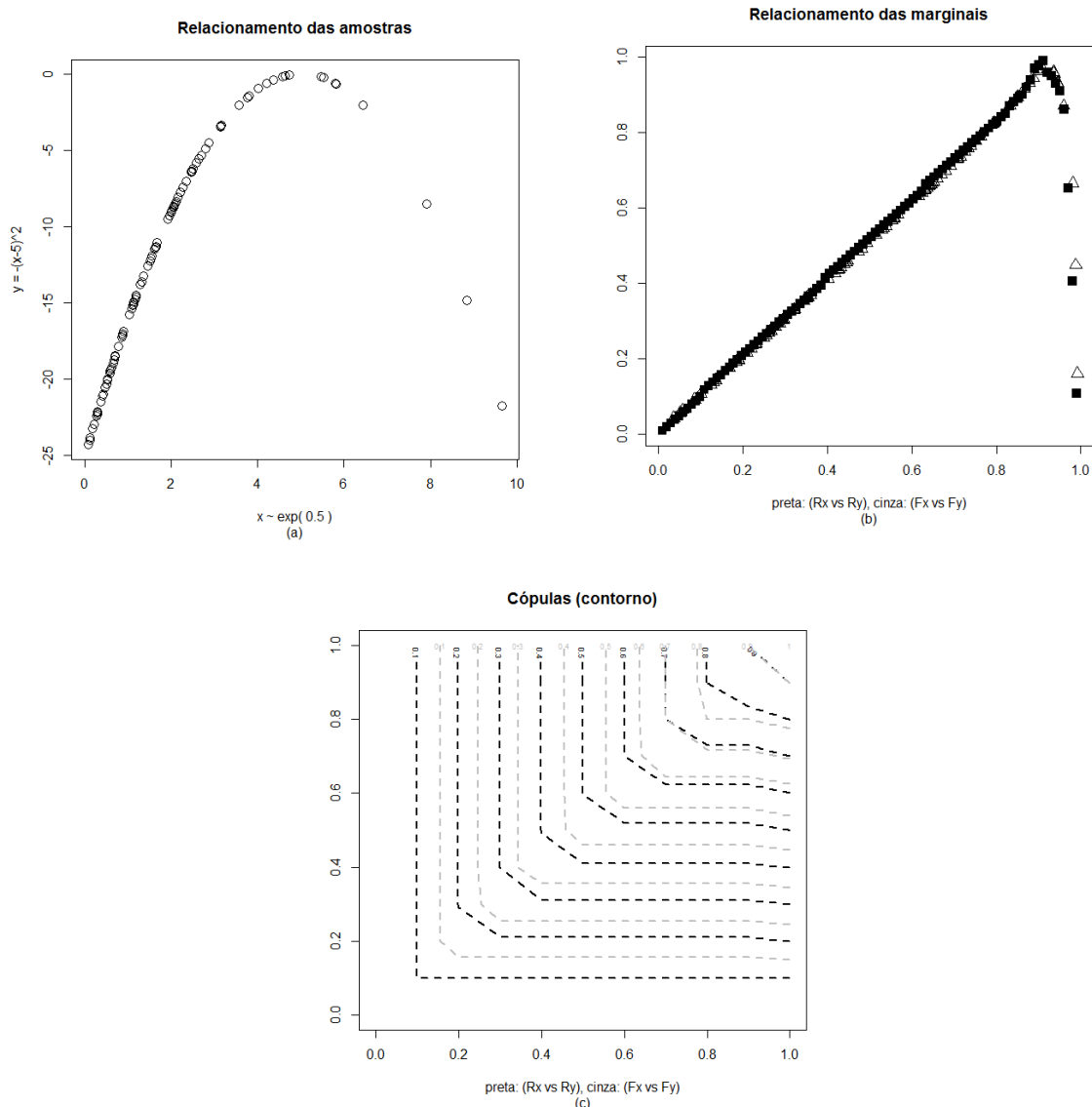


Figura 6. Gráficos de pares de 100 dados simulados a partir de X exponencial, $Y = -(X-5)^2$, respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (R_x vs R_y) e distribuições paramétricas exponenciais (F_x vs F_y) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

A Figura 7, apresenta uma estrutura de dependência dada por $Y = (X-5)^2$, ou seja, a estrutura de dependência inversa ilustrada na figura anterior. Ao observar cuidadosamente as Figura 6 (b) e 7 (b), nota-se, que a relação das distribuições marginais acompanha a relação entre os dados de X e Y.

Se pode notar ainda uma razoável diferença entre os resultados obtidos nas Figuras 6 (c) e 7 (c) com as demais figura apresentadas anteriormente que mostram como as curvas de nível se comportam neste trabalho, com dependência positiva ou negativa. Ao se comparar os resultados obtidos na Figura 6 (b) e Figura 7 (b) em relação as figuras anteriores referentes ao “Relacionamento as marginais”, se pode observar que a cópula empírica e semi-empírica conseguem captar o comportamento da dependência das distribuições marginais, quando estas tendem a oscilar de sentido, ou seja, quando a relação da distribuição marginal passa do sentido positivo para o negativo por exemplo.

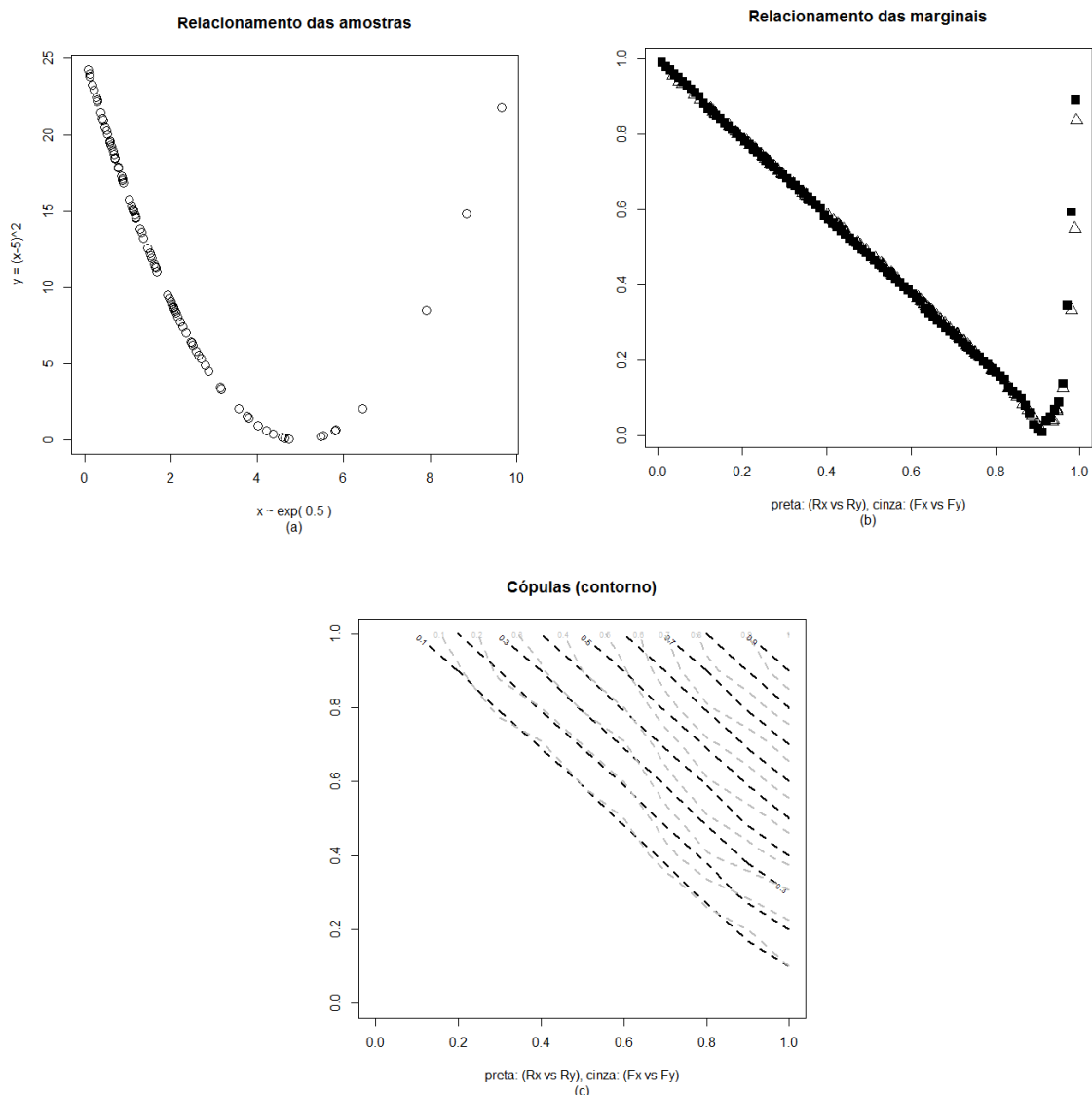


Figura 7. Gráficos de pares de 100 dados simulados a partir de X exponencial, $Y = (X-5)^2$, respectiva relação entre funções acumuladas marginais baseadas em estatísticas de ordem (R_x vs R_y) e distribuições paramétricas exponenciais (F_x vs F_y) (b) e curvas de nível das relativas cópulas, a empírica apresentada na cor preta e a semi-empírica na cor cinza (c).

Dos resultados apresentados neste experimento, é razoável afirmar que o modelo de cópulas empíricas é capaz de captar tendências da relação de dependência entre as distribuições marginais e, desta forma, das variáveis em questão, a um nível de precisão comparável ao das cópulas semi-empíricas.

Conclusões

O presente trabalho teve como principal objetivo apresentar uma breve introdução de conceitos importantes de cópulas e estudar o comportamento das cópulas empíricas e semi-empíricas através de simulações. Tais simulações foram direcionadas aos casos em que a estrutura de dependência entre as variáveis, é positiva, negativa, linear ou não-linear, além dos casos de independência. Para tal, foram aplicados os conceitos de cópulas empíricas, primeiramente introduzidos por Deheuvels (1979) *apud* Nelsen (2006), para avaliar o comportamento das variáveis quando estas seguem distribuições distintas, utilizando duas cópulas empíricas, a primeira baseada em ordem (*rankings*) e a segunda baseada na função acumulada das distribuições marginais ajustada aos dados.

Os resultados obtidos com aplicação referida sugerem que as cópulas empíricas são capazes de modelar tendências de forma ao menos tão acurada quanto as semi-empíricas para captar a tendência da relação entre variáveis. Ilustrou-se ainda a fragilidade de cópulas em captar formatos de dependência diferentes do linear entre as variáveis. Desta forma, a aplicação da cópula semi-empírica se mostra menos atraente, uma vez que, sua utilização exige mais esforço teórico e computacional comparada a aplicação da cópula empírica.

Como trabalhos futuros, pode-se explorar a utilização de técnicas como IFM para estimar os parâmetros das marginais da cópula semi-empírica e analisar se após a maximização é possível captar diferenças entre as distribuições com sentido único. Pode-se ainda, estudar formas alternativas à análise visual, tais como testes de hipóteses, dedicados à comparação de desempenho entre as cópulas.

Referências

AKPA, O.M.; E.I. Unuabonah. Small-Sample Corrected Akaike Information Criterion: An appropriate statistical tool for ranking of adsorption isotherm models. *Desalination*. v. 272, issue 1-3, p. 20-26, 2011. Acesso em: 17 de março de 2013. Disponível em: <http://dx.doi.org.ez19.periodicos.capes.gov.br/10.1016/j.desal.2010.12.057>.

AZAM, K. *Bayesian Inference for a Semi-Parametric Copula-based Markov Chain*. Acesso em: 17 de março de 2013. Disponível em: http://www2.warwick.ac.uk/fac/soc/economics/staff/phd_students/azam/job_market_paper.pdf.

BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*. Rio de Janeiro: SBM, 2001.

DEHEUVELS, P. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad Roy Belg Bul Cl Sci* (5). v. 65, issue p. 274-292, 1979.

EUGÊNIO FILHO, E. C. *Um teste para dependência de valores extremos utilizando cópulas*. 2017. 80 p. (Mestrado em Estatística), Universidade Federal de Lavras, Minas Gerais, 2017.

SILVA FILHO, O. C. *Cópulas Tempo-Variante em Finanças*. 2010. 92 p. (Doutorado em Economia), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

SILVA FILHO, O. C.; ZIEGELMANN, F.A.; DUEKER, M. J. Assessing dependence between financial market indexes using conditional time-varying copulas: applications to Value at Risk (VaR). *Quantitative Finance*. v. 0, issue p. 1-16, 2013. Disponível em: <http://dx.doi.org/10.1080/14697688.2012.739726>.

FIRMINO, P.R.; E.L. DROGUETT. Model Uncertainty via Mixed Bayesian Networks. *In: 10th International Probabilistic Safety Assessment & Management Conference*. p. 2010.

GENEST, C.; A.-C. FAVRE. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*. v. 12, issue 4, p. 347-368, 2007.

GRIMALDI, S.; F. SERINALDI. Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*. v. 29, issue 8, p. 1155-1167, 2006. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2011.03.005>.

IYENGAR, S.G.; T. DAMARLA. A parametric copula based framework for multimodal signal processing. *In: Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. Taipei. p. 1893-1896, 2009.

JOE, H.X.; J. J. *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Disponível em: <http://www.stat.ubc.ca/~harry/ifm.pdf>. Acesso em: 02 de março de 2013.

KAO, S.C.; R.S. GOVINDARAJU. A copula-based joint deficit index for droughts. *Journal of Hydrology*. v. 380, issue 1-2, p. 121-134, 2010. Disponível em: <http://dx.doi.org/10.1016/j.jhydrol.2009.10.029>.

LEAL, D.M.B. *Aplicação de cópulas ao ramo vida: Risco de resgate e risco de taxa de juro*. Disponível em: http://pascal.iseg.utl.pt/~alfredo/ftp/papers/leal_msc.pdf. Acesso em: 11 de março de 2013.

MARCHANT, B.P., et al. Spatial prediction of soil properties with copulas. *Geoderma*. v. 162, issue 3-4, p. 327-334, 2011. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2011.03.005>.

MARCHI, V.A.A. *Uma nova abordagem para análise de dependência bivariada*. 2010. 114 p. (Mestrado em Estatística), Universidade Federal de São Carlos, São Carlos, 2010.

MELO, E.F.L. Uma Aplicação de Cópulas de Lévy na Agregação de Processos Multivariados de Ruína. *Revista Brasileira de Risco e Seguro*. v. 4, n. 7, p. 47-64, 2008.

NELSEN, R.B. *An Introduction to Copulas*. Portland: Springer. 2006. 276 p.

OLIVEIRA, R.T.A.; ASSIS, T.F.O.; FIRMINO, P.R.A.; FERREIRA, T.A. E.; OLIVEIRA, A.L. I. Copulas-based ensemble of Artificial Neural Networks for forecasting real world time series. *In: International Joint Conference on Neural Networks (IJCNN)*, 2016a.

OLIVEIRA, R.T.A.; ASSIS, T.F.O.; FIRMINO, P.R.A.; FERREIRA, T.A.E. Copulas-based time series combined forecasters. *Information Sciences*, v. 376, p. 110-124, 2016b.

OLIVEIRA, R.T.A.; ASSIS, T. F.O; FIRMINO, P. R. A.; FERREIRA, T. A. E.; OLIVEIRA, A.L.I. Aggregation of Time Series Forecasts via Cacoullos Copula.. *In: International Joint Conference on Neural Networks (IJCNN)*, 2018.

OLIVEIRA, T. F.; OLIVEIRA, R. T. A.; FIRMINO, P. R. A.; MATTOS NETO, P. S. G.; FERREIRA, T. A. E. Combining Time Series Forecasting Models Via Gumbel-Hougaard Copulas. *In: 1st BRICS Countries congress on Computational iNtelligence (BRICS-CCI) and 11th Brazilian Congress on Computacional Intelligence (CBIS 2013)*. 2013a.

OLIVEIRA, T.F.; FIRMINO, P.R.A.; FERREIRA, T.A.E. Estudo da incerteza de modelos via cópulas de Gumbel. *In: Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBras)*, 2013, Campina Grande. Modelagem estatística em áreas multidisciplinares: impactos causados pelas mudanças climáticas da região nordeste, 2013b.

ROSSI, J.L. *Seleção de modelo cópula-GARCH: uma abordagem bayesiana*. 2012. 88 p. (Mestrado em Ciência da Computação e Matemática Computacional), Universidade de São Paulo, São Carlos, 2012.

SANTOS, M.F. *Aplicação de cópulas na modelação do número de sinistros de grupos de risco homogêneos no seguro automóvel*. Acesso em: 12 de novembro de 2012. Disponível em: <http://www.isp.pt/NR/rdonlyres/0B1306AD-9B60-4A8E-9B68-02A0D00BE1C2/0/C%C3%B3pulasTrabalhoMarianaFariadosSantos.pdf>.

SILVA, F.A.B.S. *Ensaio em cópulas e finanças empíricas*. 2017. 92 p. (Doutorado em Economia), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

SIMARD, R.; P. L' ECUYER. Computing the Two-Sided Kolmogorov-Smirnov Distribution. *Journal of Statistical Software*. v. 39, issue 11, p. 1-18, 2011. Disponível em: <http://www.jstatsoft.org/v39/i11>.

SKLAR, A. Fonctions de répartition à n dimensions et leurs marges. *l'Institut de Statistique de L'Université de Paris*, v. 8, issue p. 229-231, 1959.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. 2021.

VIOLA M. L. L. *Tipos de Dependência entre Variáveis Aleatórias e Teoria de Cópulas*. 2009. Disponível em: <http://www.ime.unicamp.br/~veronica/dependence/book.pdf>.

YAN, J. *Enjoy the Joy of Copulas: With a Package copula*. Journal of Statistical Software. v. 21, issue 4, p. 2007.

YAN, J. Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis*. Elsevier, p. 2010.