UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

LINA LUCIA HERNANDEZ VELASCO

Ajuste Bayesiano para Cópulas Bivariadas

Campinas

2016

Lina Lucia Hernandez Velasco

Ajuste Bayesiano para Cópulas Bivariadas

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Estatística.

Orientadora: Laura Leticia Ramos Rifo

Este exemplar corresponde à versão final da Dissertação defendida pela aluna Lina Lucia Hernandez Velasco e orientada pela Profa. Dra. Laura Leticia Ramos Rifo.

Campinas

2016

Agência(s) de fomento e nº(s) de processo(s): CAPES

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

H43a

Hernandez Velasco, Lina Lucia, 1991-
Ajuste bayesiano para cópulas bivariadas / Lina Lucia Hernandez Velasco. – Campinas, SP : [s.n.], 2016.

Orientador: Laura Leticia Ramos Rifo.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Cópulas (Estatística matemática). 2. Inferência bayesiana. 3. Dependência (Estatística). I. Rifo, Laura Leticia Ramos,1970-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Bayesian analysis for bivariate copulas
Palavras-chave em inglês:
Copulas (Mathematical statistics)
Bayesian inference
Dependence (Statistics)
Área de concentração: Estatística
Titulação: Mestra em Estatística
Banca examinadora:
Laura Leticia Ramos Rifo [Orientador]
Verónica Andrea González-Lopéz
Marcio Alves Diniz
Data de defesa: 25-02-2016
Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 25 de fevereiro de 2016 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). LAURA LETICIA RAMOS RIFO

Prof(a). Dr(a). VERÓNICA ANDREA GONZÁLEZ-LÓPEZ

Prof(a). Dr(a). MARCIO ALVES DINIZ

A Ata da defesa com as respectivas assinaturas dos membros
encontra-se no processo de vida acadêmica do aluno.

Para meu sobrinho Matthew Vega, como um lembrete das recompensas que são obtidas com grande esforço

Agradecimentos

Em primeiro lugar, gostaria de agradecer a minha orientadora, a Prof. Dra. Laura Letícia Ramos Rifo, pelo seu apoio incondicional durante todo o meu processo de formação como mestre, e por contribuir com sua experiência tanto na elaboração deste trabalho quanto na minha formação como profissional e como pessoa.

Ao meu namorado e amor da minha vida, Jhon Andersson Rosero Gil, por me ajudar a resistir cada obstáculo que aparece no meu caminho e estar sempre ao meu lado para me lembrar que estamos vivendo esta aventura para construir nosso futuro.

Aos meus pais, por me motivar a tomar riscos e apoiar todas as minhas decisões, e ao meu irmão e sua família, por não permitir que a distância seja um problema na hora de permanecer unidos.

Aos meus colegas Cristian Garcia e a Vanessa Souza dos Santos, por me ajudar a ultimar os detalhes deste documento.

Também gostaria de agradecer à Universidade Estadual de Campinas por me permitir continuar com um estudo de pós-graduação, e à agencia CAPES por financiar esta nova fase da minha formação acadêmica.

Finalmente, mas não menos importante, gostaria de agradecer a todas as pessoas maravilhosas que conheci no meu passo por Campinas, e que fizeram desta experiência ainda mais memorável.

Resumo

Este trabalho de dissertação apresenta uma metodologia que permite analisar a dependência entre duas variáveis aleatórias usando a teória de cópulas como ferramenta. O objetivo principal é conseguir explicar ao leitor de forma prática, através de um exemplo aplicado, como implementar a análise Bayesiana para estimar uma cópula num contexto onde o objetivo é, dado um conjunto de cópulas candidatas, selecionar aquela que é a mais adequada para o nosso problema de interesse. Para isto, apresentamos inicialmente dois capítulos dedicados à teoria e os conceitos de cópula e da análise bayesiana para depois descrever a metodologia que vai nos permitir determinar uma cópula ótima em qualquer cenário. Finalmente fazemos uso de dados do Vestibular da Unicamp para mostrar passo a passo como implementar tal metodologia.

Palavras-chave: Cópulas, Inferência Bayesiana, Dependência, Vestibular UNICAMP.

Abstract

This dissertation is focused on presenting a methodology to analyze the dependence between two random variables using the copula theory. The main purpose is to explain in a practical way using an example application, how to implement Bayesian analysis to fit a copula in a context where the goal is, given a set of candidates, select the one that is most appropriate for our problem of interest. For this, initially we present two chapters devoted to theoretical contextualization of copula theory and concepts of Bayesian analysis. Then describe the methodology that will allow us to determine a great copula in any scenario. Finally we use of Unicamp Vestibular data to show step by step how to implement this methodology.

Keywords: latex. Copulas, Bayesian Inference, Dependence, Vestibular UNICAMP.

Lista de ilustrações

Lista de tabelas

Sumário

Introdução

O estudo da dependência entre duas ou mais variáveis é um dos aspectos mais interessantes da análise estatística, pois nos permite estabelecer relações e realizar inferências que contribuam para, por exemplo, melhorar um processo.

Para realizar uma análise desse tipo, existe uma ferramenta que surge a partir de um teorema enunciado por [Sklar, 1959], onde usa-se pela primeira vez a palavra *cópula* para definir uma classe de funções a partir das quais pode-se escrever a expressão da função de distribuição conjunta de certas variáveis em termos de suas marginais. Estas cópulas têm sido amplamente usadas na análise empírica de dados multivariados em diferentes áreas, incluindo a análise de sobrevivencia ([Clayton, 1978]; [Oakes, 1989]), ciências atuariais ([Frees and Valdez, 1998]), marketing ([Danaher and Smith, 2011]), estatística médica ([Lambert and Vandenhende, 2002] ; [Nikoloulopoulos and Karlis, 2008]) e econometria ([Smith, 2000]; [Patton, 2006]), entre outras.

Uma das maiores vantagens desta função cópula é que, dada a relação que tem com a função de distribuição conjunta, o estudo das probabilidades conjuntas, de duas variáveis pode-se reduzir ao estudo da cópula associada a elas. Isto é uma vantagem porque, na prática, às vezes achar a função de distribuição conjunta é um problema de alta complexidade, e se o nosso objetivo é apenas analisar dependência, as cópulas vão nos dar toda a informação que precisamos sem necessidade de conhecer a forma da distribuição. É por isso que são conhecidas também como *funções de dependência*.

Como já foi acima mencionado, as cópulas são funções, mas às vezes quando falamos de cópula na verdade estamos nos referindo a uma famílias de cópulas, que são conjuntos de funções que têm a mesma expressão matemática geral, e no caso das cópulas paramétricas, são indexadas por parâmetros. Para encontrar uma cópula adequada para um par de variáveis dentro de uma família paramétrica de cópulas, o processo se reduz a um problema de estimação, e uma das alternativas para abordar este problema é usar inferência Bayesiana, como fazem [Huard et al., 2006], que sugerem um método para selecionar entre diferentes cópulas bivariadas; [dos Santos Silva and Lopes, 2008], que usam métodos Markov Chain Monte Carlo (MCMC) para estimar funções cópula de poucos parâmetros; [Smith, 2011], que apresenta abordagens Bayesianas para diferentes cópulas; [Fernández et al., 2014], que trabalham o problema das distribuições conjugadas, apenas para citar alguns exemplos.

O objetivo principal nesta dissertação, é estudar o ajuste bayesiano para cópulas bivariadas como ferramenta de análise de dependência. Para isso, vamos estabelecer uma metodologia para selecionar a melhor cópula para um par específico de variáveis e

posteriormente aplicá-la a um conjunto de dados.

Para nossa aplicação, vamos trabalhar um problema relacionado com o processo de seleção através do qual ingressam os estudantes de graduação à Universidade Estadual de Campinas (UNICAMP). Este processo está baseado na prova de Vestibular, que é composto de duas fases de qualificação. A primeira consiste em uma prova de alternativas de conhecimentos gerais em diferentes disciplinas e a partir dela se decide se o candidato continua para a próxima fase. A segunda fase consiste em várias provas escritas, uma por área, onde o aluno deve demonstrar seu conhecimento específico em cada área. A partir de todas as provas, tanto da primeira quanto da segunda fase, é contruído um único indicador a partir do qual os candidatos são selecionados.

O nosso interesse com este problema, é estudar a relação existente entre o rendimento dos alunos que ingressam nos cursos pertencentes à área de Exatas e Engenharia em cada fase do Vestibular e dois indicadores de rendimento no primeiro semestre: o Coeficiente de Rendimento (CR) que detalharemos no Capítulo 4, e a nota obtida na disciplina MA111-Cálculo I, que é uma das mais importantes para a maioria dos cursos analisados. Com este estudo, pretendemos identificar aquelas provas do Vestibular que estão mais relacionadas com o desempenho do aluno no primeiro semestre e especificamente com a disciplina MA111-Cálculo I.

Esta dissertação é composta por cinco capítulos. O Capítulo 1 está dedicado ao estudo das cópulas, isto é, definições, teoremas, propriedades, métodos de construção. No Capítulo 2, encontramos todos os conceitos importantes sobre inferência Bayesiana, como estimação, teste de hipóteses, distribuição preditiva e simulação. Os Capítulos 3 e 4 estão dedicados à metodologia e à aplicação da mesma. Finalmente no Capítulo 5, encontramos as conclusões finais e sugestões para trabalhos futuros.

1 Cópulas e variáveis aleatórias

Incialmente podemos definir uma cópula como uma função de distribuição multivariada cujas marginais unidimensionais são uniformes no intervalo $[0, 1]$. Esta definição é natural, se a cópula é obtida a partir de uma distribuição multivariada contínua, onde a cópula é a distribuição multivariada original com marginais univariadas transformadas.

A ideia de uma função que caracterize a estrutura de dependência entre diferentes variáveis aleatórias vem dos trabalhos de Hoeffding de 1940-1948, [Hoeffding, 2012], que definiu uma classe de distribuições bivariadas padronizadas cujo suporte é o quadrado $[-1/2, 1/2]$ e cujas marginais são uniformes também nesse intervalo. Segundo [Schweizer, 1991], se Hoeffding tivesse usado como domínio para sua definição o quadrado $[0, 1]$ seria ele o precursor da teoria das cópulas. Mas foi Abe Sklar quem, em 1959, usou o termo *cópula* para definir funções que ligam funções de distribuição multivariadas com suas marginais unidimensionais. O antecessor mais direto de Sklar foi o trabalho de Ferón en 1956, que realizou um estudo sobre distribuições tridimensionais onde definia funções auxiliares, de domínio no quadrado $[0, 1]$, que permitiam ligar tais distribuições com suas marginais univariadas. A partir daí Sklar estabelece o teorema que leva seu nome e que constitui a pedra angular de uma teoria que se torna amplamente trabalhada.

1.1 Conceitos básicos

O objetivo principal desta primeira seção é definir cópula. Para isto, precisamos de alguns conceitos e propriedades adicionais, assim como estabelecer a notação que utilizaremos nas seções posteriores. Esta seção está baseada no livro de [Nelsen, 2013].

Seja $\mathbb{R}$ a reta real, $(-\infty, \infty)$, $\bar{\mathbb{R}}$ a reta real estendida, $[-\infty, \infty]$, e $\bar{\mathbb{R}}^2$ o plano estendido, $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$. Um retângulo em $\bar{\mathbb{R}}^2$ é o produto cartesiano de dois intervalos fechados: $B = [x_1, x_2] \times [y_1, y_2]$. Os vértices de um retângulo em B são os pontos $(x_1, y_1), (x_1, y_2), (x_2, y_1)$ e (x_2, y_2). O quadrado unitario I^2 é o produto $I \times I$, com $I = [0, 1]$. Consideramos funções reais bivariadas H cujo domínio, $DomH$, é um subconjunto de $\bar{\mathbb{R}}^2$, e cuja imagem, $RanH$, é um subconjunto de $\mathbb{R}$.

Definição 1.1. *Sejam S_1 e S_2 subconjuntos não vazios de $\bar{\mathbb{R}}$ e seja H uma função real bivariada tal que $DomH = S_1 \times S_2$. Seja $B = [x_1, x_2] \times [y_1, y_2]$ um retângulo tal que todos os seus vértices pertencem ao domínio $DomH$. Então o H-volume de B é dado por,*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \tag{1.1}$$

Definição 1.2. *Uma função real bivariada H é 2-crescente se $V_H(B) \geqslant 0$ para todos os retângulos B cujos vértices estão dentro do domínio $DomH$.*

Note que dizer que *H é 2-crescente* não implica necessariamente que *H é não decrescente em cada argumento*, como vemos nos seguintes exemplos.

Exemplo 1.1. *Seja H uma função definida em I^2 por $H(x, y) = \max(x, y)$. Então H é uma função não decrescente de x e y, no entanto, dado que $V_H(I^2) = -1$, H é 2-crescente.*

Exemplo 1.2. *Seja H a função definida em I^2 por $H(x, y) = (2x - 1)(2y - 1)$. Então H é 2-crescente, no entanto, é uma função decrescente de x para cada y em $(0, 0.5)$ e uma função decrescente de y para cada x em $(0, 0.5)$.*

Os seguintes dois lemas serão úteis para estabelecer a continuidade de subcópulas e cópulas em seções posteriores.

Lema 1.1. *Sejam S_1 e S_2 subconjuntos não vazios de $\bar{\mathbb{R}}$, e seja H uma função 2-crescente com domínio $S_1 \times S_2$. Sejam $x_1, x_2 \in S_1$, com $x_1 \leqslant x_2$, e sejam $y_1, y_2 \in S_2$, com $y_1 \leqslant y_2$. Então a função $t \to H(t, y_2) - H(t, y_1)$ é não decrescente em S_1 e a função $t \to H(x_2, t) - H(x_1, t)$ é não decrescente em S_2.*

Se supomos que S_1 e S_2 têm menor elemento a_1 e a_2 respectivamente, dizemos que uma função $H : S_1 \times S_2 \to \mathbb{R}$ é aplainada se $H(x, a_2) = 0 = H(a_1, y)$ para todo (x, y) em $S_1 \times S_2$.

Lema 1.2. *Seja S_1 um subconjunto não vazio de $\bar{\mathbb{R}}$ e seja H uma função 2-crescente aplainada com domínio em $S_1 \times S_2$. Então H é não decrescente em cada argumento.*

Se supomos que S_1 e S_2 têm maior elemento b_1 e b_2 respectivamente, então dizemos que uma função $H : S_1 \times S_2 \to \mathbb{R}$ tem marginais, denotadas como F e G, dadas por:

$$DomF = S_1, \text{e } F(x) = H(x, b_2) \text{ para todo} x \in S_1,$$
$$DomG = S_2, \text{e } G(y) = H(b_1, y) \text{ para todo} y \in S_2.$$

Exemplo 1.3. *Seja H a função com domínio $[-1, 1] \times [0, \infty]$ dada por:*

$$H(x, y) = \frac{(x + 1)(e^y - 1)}{x + 2e^y - 1}.$$

Então H é aplainada porque $H(x, 0) = 0$ e $H(-1, y) = 0$, com marginais $F(x) = H(x, \infty) = (x + 1)/2$ e $G(y) = H(1, y) = 1 - e^{-y}$.

Lema 1.3. *Sejam S_1 e S_2 subconjuntos não vazios de $\bar{\mathbb{R}}$ e seja H uma função aplainada 2-crescente com marginais cujo domínio é $S_1 \times S_2$. Sejam (x_1, y_1) e (x_2, y_2) dois pontos quaisquer em $S_1 \times S_2$. Então:*

$$|H(x_2, y_2) - H(x_1, y_1)| \leqslant |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| \tag{1.2}$$

Demonstração. Da desigualdade triangular, temos

$$|H(x_2, y_2) - H(x_1, y_1)| \leqslant |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|$$

Agora suponha que $x_1 \leqslant x_2$. Dado que H é 2-crescente e tem marginais, pelos Lemas 1.1 e 1.3 podemos escrever $0 \leqslant H(x_2, y_2) - H(x_1, y_2) \leqslant F(x_2) - F(x_1)$. Uma desigualdade análoga se consegue quando $x_2 \leqslant x_1$. Daí segue que para qualquer $x_1, x_2 \in S_1$, $|H(x_2, y_2) - H(x_1, y_2)| \leqslant |F(x_2) - F(x_1)|$. Similarmente para cada $y_1, y_2 \in S_2$, $|H(x_1, y_2) - H(x - 1, y_1)| \leqslant |G(y_2) - G(y_1)|$. □

Tendo já definidos os conceitos preliminares mais importantes, vamos nos focar na definição de cópula.

Definição 1.3. *Uma subcópula bidimensional é uma função C' com as seguintes propriedades:*

1. *$DomC' = S_1 \times S_2$ onde S_1 e S_2 são subconjuntos de I contendo 0 e 1.*
2. *C' é aplainada e 2-crescente.*
3. *Para todo $u \in S_1$ e todo $v \in S_2$, $C'(u, 1) = u$ e $C'(1, v) = v$.*

Note que para todo $(u, v) \in DomC'$ vale que $0 \leqslant C'(u, v) \leqslant 1$, e que $RanC'$ é também subconjunto de $I = [0, 1]$.

Definição 1.4. *Uma cópula bidimensional é uma função $C : I^2 \to I$ com as seguintes propriedades:*

1. *Para todos $u, v \in I$, $C(u, 0) = 0 = C(0, v)$ e $C(u, 1) = u$ e $C(1, v) = v$.*
2. *Para todos $u_1, u_2, v_1, v_2 \in I$ tais que $u_1 \leqslant u_2$ e $v_1 \leqslant v_2$,*

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geqslant 0 \tag{1.3}$$

Assim, muitas das propriedades importantes das cópulas são na verdade propriedades das subcópulas. Inicialmente esta diferença parece menor, mas vai ser muito importante na hora de falar sobre o Teorema de Sklar, que permite caraterizar a relação entre uma função cópula e uma função de distribuição.

Teorema 1.1. *Seja C' uma subcópula. Então para cada $(u, v) \in DomC'$:*

$$\max(u + v - 1, 0) \leqslant C'(u, v) \leqslant \min(u, v). \tag{1.4}$$

Demonstração. Seja (u, v) um ponto arbitrário em $DomC'$. Agora $C'(u, v) \leqslant C'(u, 1) = u$ e $C'(u, v) \leqslant C'(1, v) = v$ pelo que $C'(u, v) \leqslant min(u, v)$. Além disso, $V_{C'}([u, 1] \times [v, 1]) \leqslant 0$ implica que $C'(u, v) \geqslant 0$ pelo que $C'(u, v) \geqslant max(u + v - 1, 0)$. □

Dado que toda cópula é uma subcópula, a desigualdade do Teorema 1.1 é valida também para as cópulas. De fato, os limitantes dessa desigualdade são eles mesmos cópulas e são denotados normalmente $M(u, v) = \min(u, v)$ e $W(u, v) = \max(u + v - 1, 0)$. Assim, para toda cópula C e cada $(u, v) \in I^2$:

$$W(u, v) \leqslant C(u, v) \leqslant M(u, v). \tag{1.5}$$

Essa desigualdade é a versão para cópulas da desigualdade de Fréchet-Hoeffding ([Frechét, 1951] e [Hoeffding, 1940]) e por causa disso $W(u, v)$ e $M(u, v)$ são chamadas de cota inferior e superior de Fréchet-Hoeffding, respectivamente.

Uma terceira cópula importante que frequentemente encontramos é a cópula produto $\Pi(u, v) = uv$. A Figura 1 exibe os gráficos de superfície destas cópulas importantes, W, M e Π.

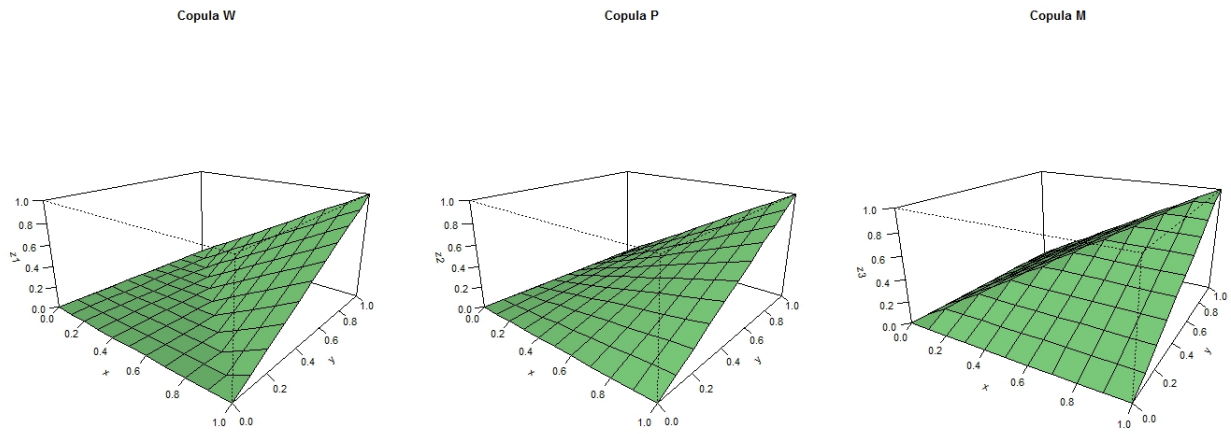

Figura 1 – Gráficos de superfície das cópulas W, Π e M.

O seguinte Teorema, que é consequência imediata do Lema 1.3, estabelece a continuidade das subcópulas e portanto das cópulas, usando uma condição de Lipschitz.

Teorema 1.2. *Seja C' uma subcópula, então para todo (u_1, u_2), $(v_1, v_2) \in Dom C'$,*

$$| C'(u_2, v_2) - C'(u_1, v_1) | \leqslant | u_2 - u_1 | + | v_2 - v_1 | . \tag{1.6}$$

Com isto, C' é uniformemente contínua no seu domínio.

Definição 1.5. *Sejam C uma cópula e $a \in I = [0, 1]$. A seção horizontal de C em a é a função $C_{\cdot,a} : I \to I$ dada por $t \to C(t, a)$; a seção vertical de C em a é a função $C_{a,\cdot} : I \to I$, dada por $t \to C(a, t)$ e a seção diagonal de C é a função $\delta_C : I \to I$ definida por $\delta_C = C(t, t)$.*

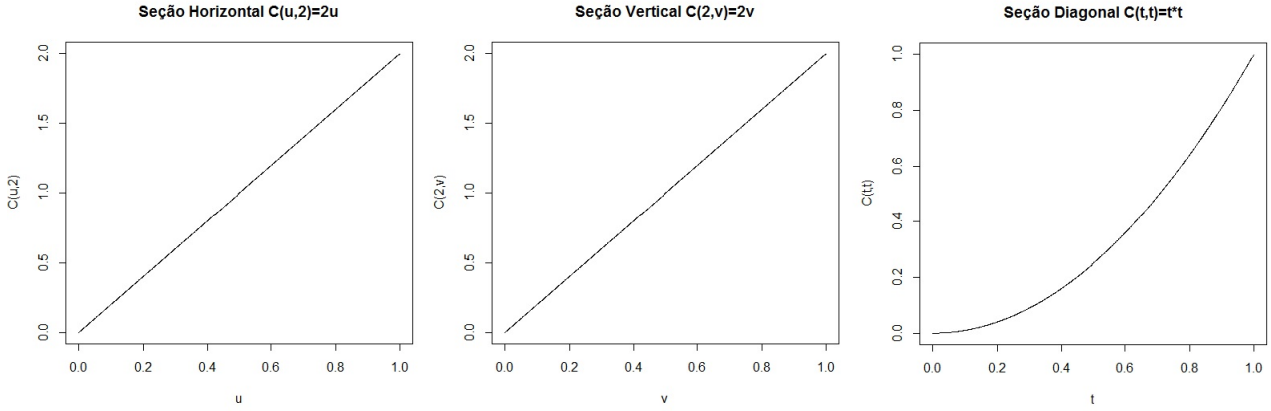

Figura 2 – Gráficos de seções horizontal, vertical e diagonal da cópula $C(u, v) = \Pi(u, v)$.

Corolário 1.1. *As seções horizontal, vertical e diagonal de uma cópula C são não-decrescentes e uniformemente contínuas em I.*

As seções de uma cópula vão ser muito úteis para construir cópulas com certas caraterísticas desejadas, como algumas propriedades de dependência condicional.

Para terminar esta seção, vamos apresentar dois resultados referentes às derivadas parciais de uma função cópula.

Teorema 1.3. *Seja C uma cópula. Para qualquer $v \in I = [0, 1]$, a derivada parcial $\dfrac{\partial C}{\partial u}(u, v)$ existe para quase todo u. Para cada (u, v),*

$$0 \leqslant \frac{\partial C}{\partial u}(u, v) \leqslant 1. \tag{1.7}$$

Similarmente, para cada $u \in I$, a derivada parcial $\dfrac{\partial C}{\partial v}(u, v)$ existe para quase todo v. Para cada (u, v),

$$0 \leqslant \frac{\partial C}{\partial v}(u, v) \leqslant 1. \tag{1.8}$$

Além disso, as funções $u \mapsto \dfrac{\partial C}{\partial v}(u, v)$ e $v \mapsto \dfrac{\partial C}{\partial u}(u, v)$ são definidas e não decrescentes em quase todo I.

Demonstração. Sabemos que as funções monótonas são diferenciáveis, portanto, as derivadas existem. A desigualdade 1.7 segue da desigualdade 1.6 definindo $v_1 = v_2$ e $u_1 = u_2$ respectivamente. Se $v_1 \leqslant v_2$ então do Lema 1.1 temos que a função $u \to C(u, v_2) - C(u, v_1)$ é não decrescente. Portanto, $\partial(C(u, v_2) - C(u, v_1))/\partial u$ é definida e não negativa em quase todo I. Daí segue que $v \to \partial C(u, v)/\partial u$ é definida e não decrescente para quase todo I. Da mesma forma pode ser feito para $\partial C(u, v)/\partial v$. □

Teorema 1.4. *Seja C uma cópula. Se $\dfrac{\partial C}{\partial v}(u, v)$ e $\dfrac{\partial^2 C}{\partial u \partial v}(u, v)$ são contínuas em I^2, e $\dfrac{\partial C}{\partial u}(u, v)$ existe para todo $u \in I = (0, 1)$, quando $v = 0$, então $\dfrac{\partial^2 C}{\partial v \partial u}(u, v)$ e $\dfrac{\partial^2 C}{\partial u \partial v}(u, v)$ existem em $(0, 1)^2$ e $\dfrac{\partial^2 C}{\partial u \partial v}(u, v) = \dfrac{\partial^2 C}{\partial v \partial u}(u, v)$.*

Para seções posteriores, a densidade de uma cópula fará referência à segunda derivada parcial dela, isto é, $\partial^2 C(u, v)/\partial u \partial v$.

1.2 Teorema de Sklar e a interpretação probabilística das cópulas

Para a abordagem utilizada neste trabalho, um dos aspectos mais interessantes das cópulas é sua relação com distribuições de variáveis aleatórias e portanto a sua interpretação probabilística. Esta relação é estabelecida a partir do Teorema de [Sklar, 1959] que afirma tanto que as cópulas são funções de distribuição conjunta (em nosso caso, bivariadas) quanto a recíproca, ou seja, que as funções de distribuição conjunta podem se reescrever em termos das marginais e uma única subcópula, que por sua vez pode se estender (em geral de forma não única) a uma cópula. Isto implica, que em geral o estudo das funções de distribuição conjunta pode se reduzir ao estudo das cópulas associadas a elas.

Antes de começar, relembremos o que é uma função de distribuição, inicialmente só nos casos univariado e bivariado.

Definição 1.6. *Uma função de distribuição real é uma função F com domínio $\bar{\mathbb{R}}$ tal que:*

1. *F é não decrescente.*
2. *$F(-\infty) = 0$ e $F(\infty) = 1$.*

Exemplo 1.4. *Para qualquer par de números $a, b \in \mathbb{R}$, com $a < b$, a distribuição uniforme em $[a, b]$ é a função de distribuição U_{ab} dada por*

$$U_{ab}(x) = \begin{cases} 0 & x \in (-\infty, a) \\ \dfrac{x-a}{b-a} & x \in [a, b] \\ 1 & x \in (b, \infty) \end{cases} \tag{1.9}$$

Definição 1.7. *Uma função de distribuição conjunta bivariada é uma função H com domínio em $\bar{\mathbb{R}}^2$ tal que:*

1. *H é 2-crescente.*
2. *$H(x, -\infty) = H(-\infty, y) = 0$ e $H(-\infty, \infty) = 1$.*

Desta definição, temos que H é aplainada e, dado que $DomH = \bar{\mathbb{R}}^2$, H tem marginais F e G dadas por $F(x) = H(x, -\infty)$ e $G(y) = H(\infty, y)$. Pelo Corolário 1.1, concluímos que F e G são funções de distribuição.

Exemplo 1.5. *Seja H uma função com domínio $\bar{\mathbb{R}}^2$ dada por:*

$$H(x,y) = \begin{cases} \dfrac{(x+1)(e^y-1)}{x+2e^y-1} & (x,y) \in [-1,1] \times [0,\infty) \\ 1-e^{-y} & x \in [1,\infty) \times [0,\infty) \\ 0 & \textit{caso contrário} \end{cases} \tag{1.10}$$

Podemos verificar que H é 2-crescente e aplainada, e que $H(\infty, \infty) = 1$. Portanto H é uma função de distribuição conjunta com marginais dadas por F e G:

$$F(x) = U_{-1,1}(x) \quad e \quad G(y) = \begin{cases} 0 & y \in (-\infty, 0) \\ 1-e^{-y} & y \in [0,\infty) \end{cases}. \tag{1.11}$$

Antes de seguir com o Teorema de Sklar, enunciemos e demostremos dois lemas necessários para a demonstração deste teorema.

Lema 1.4. *Seja H uma função de distribuição conjunta com marginais F e G. Então existe uma única subcopula C' tal que,*

- *$DomC' = RanF \times RanG$,*
- *Para todo $x, y \in \bar{\mathbb{R}}$, $H(x,y) = C'(F(x), G(y))$.*

Demonstração. A distribuição conjunta H satisfaz a hipótese do Lema 1.3 com $S_1 = S_2 = \bar{\mathbb{R}}$. Portanto, para cada par de pontos $(x_1, y_1), (x_2, y_2) \in \bar{R}^2$,

$$|H(x_2, y_2) - H(x_1, y_1)| \leqslant |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

Daí resulta que se $F(x_1) = F(x_2)$ e $G(y_1) = G(y_2)$, então $H(x_1, y_1) = H(x_2, y_2)$. Assim, o conjunto de pares ordenados

$$\{((F(x), G(y)), H(x,y)) | x, y \in \bar{\mathbb{R}}\}$$

define uma função real bivariada C' cujo domínio é $RanF \times RanG$. Além disso, esta função C' é uma subcópula, afirmação que segue diretamente das propriedades de H. Por exemplo,

para verificar a condição 3 da Definição 1.3 basta notar que para cada $u \in RanF$, existe um $x \in \bar{R}$ tal que $F(x) = u$. Assim, $C'(u,1) = C'(F(x), G(\infty)) = H(x,\infty) = F(x) = u$. Desta forma é possível verificar as demais condições da Definição 1.3. □

Lema 1.5. *Seja C' uma subcópula. Então existe uma cópula C tal que $C(u,v) = C'(u,v)$ para todo $(u,v) \in DomC'$; isto é, qualquer subcópula pode ser estendida a uma cópula. Esta extensão geralmente não é única.*

Demonstração. Seja $DomC' = S_1 \times S_2$. Usando o Teorema 1.2 e o fato de C' ser não decrescente em todo o seu domínio, podemos estender C' por continuidade a uma função C'' com domínio $\bar{S_1} \times \bar{S_2}$, onde $\bar{S_1}$ é o fecho de S_1 e $\bar{S_2}$ é o fecho de S_2. É claro que C'' é também uma subcópula. Agora estendemos C'' a uma função C com domínio em $I^2 = (0,1)^2$. Para esta última parte, consideremos (a,b) sendo qualquer ponto de I^2. Sejam a_1 e a_2 os elementos menor e maior de $\bar{S_1}$, respectivamente satisfazendo $a_1 < a < a_2$; analogamente, sejam b_1 e b_2 os elementos menor e maior de $\bar{S_2}$, satisfazendo $b_1 < b < b_2$. Note que se $a \in \bar{S_1}$, então $a_1 = a = a_2$; e se $b \in \bar{S_2}$, então $b_1 = b = b_2$. Agora sejam

$$\lambda_1 = \begin{cases} (a-a_1)/(a_2-a_1) & \text{se } a_1 < a_2 \\ 1 & \text{se } a_1 = a_2 \end{cases}.$$

$$\mu_1 = \begin{cases} (b-b_1)/(b_2-b_1) & \text{se } b_1 < b_2 \\ 1 & \text{se } b_1 = b_2 \end{cases}.$$

e definamos

$$\begin{aligned} C(a,b) = {} & (1-\lambda_1)(1-\mu_1)C''(a_1,b_1) + (1-\lambda_1)\mu_1 C''(a_1,b_2) + \\ & \lambda_1(1-\mu_1)C''(a_2,b_1) + \lambda_1\mu_1 C''(a_2,b_2). \end{aligned} \tag{1.12}$$

Note que a interpolação acima é linear para todo domínio, pois a que λ_1 e μ_1 são lineares em a e b, respectivamente.

É fácil notar que $DomC = I^2$, $C(a,b) = C''(a,b)$ para qualquer $(a,b) \in DomC''$ e C satisfaz a condição 1 da Definição 1.4, e portanto, temos apenas que mostrar que C satisfaz a equação (1.3). Para isto, seja $(c,d) \in I^2$ outro ponto tal que $c > a$ e $d > b$ e sejam $c_1, d_1, c_2, d_2, \lambda_2, \mu_2$ relacionados com c e d da mesma forma que $a_1, b_1, a_2, b_2, \lambda_1, \mu_1$ estão relacionados com a e b. Agora, calculamos o volume $V_C(B)$ onde $B = [a,c] \times [b,d]$ e consideramos os casos, se tem ou não um ponto em $\bar{S_1}$ estritamente entre a e c, e se tem um ponto em $\bar{S_2}$ estritamente entre b e d. Substituindo (1.12) e os respectivos valores de $C(a,d), C(c,b)$ e $C(c,d)$ na equação (1.1) obtemos o seguinte depois de simplificar os cálculos

$$V_C(B) = V_C([a,c] \times [b,d]) = (\lambda_2 - \lambda_1)(\mu_2 - \mu_1)V_C([a_1,a_2] \times [b_1,b_2]) \tag{1.13}$$

do que segue que $V_C(B) \geqslant 0$ neste caso, pois $c \geqslant a$ e $d \geqslant b$ implica que $\lambda_2 \geqslant \lambda_1$ e $\mu_2 \geqslant \mu_1$.

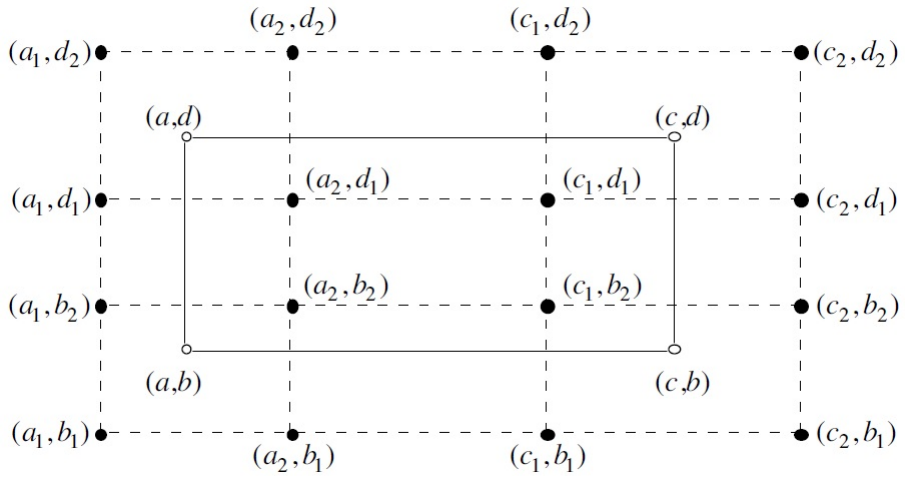

Figura 3 – Caso menos simples do Lema 1.5.

No outro extremo, o caso menos simples acontece quando há pelo menos um ponto em $\bar{S}_1$ estritamente entre a e c, e pelo menos um ponto em $\bar{S}_2$ estritamente entre b e d, ou seja, $a < a_2 < c_1 < c$ e $b < b_2 < d_1 < d$ (Figura 3). Neste caso, substituindo (1.12) e os respectivos $C(a,d), C(c,b)$ e $C(c,d)$ na expressão do volume $V_C(B)$, como no caso mais simples, obtemos

$$\begin{aligned} V_C(B) \;=\; & (1-\lambda_1)\mu_2 V_C([a_1,a_2]\times[d_1,d_2]) + \mu_2 V_C([a_2,c_1]\times[d_1,d_2]) \\ & +\lambda_2\mu_2 V_C([c_1,c_2]\times[d_1,d_2]) + (1-\lambda_1) V_C([a_1,a_2]\times[b_1,d_1]) \\ & +V_C([a_2,c_1]\times[b_2,d_1]) + \lambda_2 V_C([c_1,c_2]\times[b_2,d_1]) \\ & +(1-\lambda_1)(1-\mu_1) V_C([a_1,a_2]\times[b_1,b_2]) \\ & +(1-\mu_1) V_C([a_2,c_1]\times[b_1,b_2]) + \mu_2(1-\mu_1) V_C([c_1,c_2]\times[b_1,b_2]) \end{aligned}$$

onde a parte direita da expressão acima é uma combinação linear de nove quantidades não negativas (correspondentes aos C-volumes dos nove retângulos da Figura 3) com coeficientes não negativos e portanto a expressão é não negativa. Para os outros casos a demonstração é similar. □

Agora enunciemos a versão bivariada do Teorema de Sklar.

Teorema 1.5. *[Sklar 1959]*

Seja H uma função de distribuição bivariada conjunta com marginais F e G. Então existe uma cópula C tal que, para todo $x, y \in \bar{\mathbb{R}}$:

$$H(x,y) = C\left(F(x), G(y)\right). \tag{1.14}$$

Se F e G são contínuas, então C é única. Nos demais casos, C é unicamente determinada em $RanF \times RanG$. Inversamente, se C é uma cópula e F e G são funções de distribuição reais, então $H(x, y)$ definida pela equação acima é uma função de distribuição conjunta com marginais F e G.

Em palavras simples, através do Teorema 1.5 podemos representar uma probabilidade conjunta usando as marginais e uma cópula onde esta última representa de forma única a associação entre X e Y. É por isto que as cópulas são funções de dependência.

Segue a demonstração deste resultado baseada no argumento de [Schweizer and Sklar, 1974], e apresentada por [Nelsen, 2013].

Demonstração. A existência de uma cópula C tal que $H(x, y) = C(F(x), G(y))$ para todo $x, y \in \bar{\mathbb{R}}$ segue dos Lemas 1.4 e 1.5. Se F e G são contínuas, então $RanF = RanG = I$, e portanto a única subcópula do Lema 1.4 é uma cópula. Demonstrar a inversa é simplesmente verificar que $H(x, y)$ é uma função de distribuição e que suas marginais são exatamente F e G. □

Definição 1.8. *Seja F uma função de distribuição. Uma quase inversa de F é qualquer função $F^{(-1)}$ com domínio em I tal que:*

1. *Se $t \in RanF$ então $F^{(-1)}(t)$ é qualquer número x em $\bar{\mathbb{R}}$ tal que $F(x) = t$, ou seja $F\left(F^{(-1)}(t)\right) = t$.*
2. *Se $t \in RangF$ então $F^{(-1)}(t) = \inf\{x \mid F(x) \geqslant t\} = \sup\{x \mid F(x) \leqslant t\}$.*

Se F é estritamente crescente, então tem uma única quase-inversa, que é a inversa usual que vamos denotar por F^{-1}.

Usando a Definição 1.8 é possível relacionar uma cópula com uma função de distribuição com o seguinte corolário.

Corolário 1.2. *Seja H uma função de distribuição bivariada conjunta com marginais contínuas F, G e cópula C. Então para todos $u, v \in I$,*

$$C(u, v) = H\left(F^{-1}(u), G^{-1}(v)\right). \tag{1.15}$$

Este corolário, junto com o Teorema 1.5, permite construir distribuições bivariadas de forma simples, dado que só precisamos unir as distribuições marginais das variáveis de interesse em uma função que satisfaça a Definição 1.4, o que é muito prático dado que não precisamos que as variáveis sejam do mesmo tipo, ou seja, transformações lineares afins uma da outra, como nos métodos tradicionais de construção.

Para facilitar o entendimento, estamos enunciando unicamente os resultados para o caso bivariado, mas todos estes podem se estender à dimensão $n \geqslant 2$.

Exemplo 1.6. *Consideremos novamente a função H descrita no Exemplo 1.5. As quase-inversas de F e G são dadas por $F^{(-1)}(u) = 2u - 1$ e $G^{(-1)}(v) = -ln(1-v)$ para $u, v \in I$. Como $RanF = RanG = I$, usamos a Equação 1.15 para deduzir a expressão da cópula C associada a H que vem dada por $C(u,v) = \dfrac{uv}{u+v-uv}$.*

1.3 Propriedades das cópulas

Nesta seção vamos estudar algumas propriedades das cópulas que serão úteis adiante. Para facilitar a notação vamos denotar a cópula de X e Y como $C_{XY}(u,v)$ e a cópula de $\alpha_1(X)$ e $\alpha_2(Y)$ como $C_{\alpha_1(X)\alpha_2(Y)}(u,v)$.

Na Seção 1.1 falamos sobre as cópulas $W(u,v), \Pi(u,v)$ e $M(u,v)$ e de como toda cópula se relaciona com as cotas de Frechét-Hoeffding a partir do Teorema 1.1. Nesse sentido poderíamos perguntar o que significa que a cópula associada a duas variáveis aleatórias coincida com alguma destas três cópulas. Para responder esta pergunta consideremos as seguintes proposições.

Proposição 1.1. *Duas variaveis aleatórias X e Y são comonotônicas se e somente se $(X,Y) \overset{d}{=} (\alpha_1(Z), \alpha_2(Z))$, para alguma variável Z e funções monótonas crescentes α_1 e α_2, onde $\overset{d}{=}$ denota igualdade em distribuição.*

Proposição 1.2. *Duas variaveis aleatórias X e Y são contramonotônicas se e somente se $(X,Y) \overset{d}{=} (\alpha_1(Z), \alpha_2(Z))$, para alguma variável Z, α_1 função monótona crescente e α_2 função monótona decrescente, ou vice-versa.*

Desta forma, vamos dizer que

1. Se a cópula associada a duas variáveis aleatórias X e Y é a cópula $M(u,v)$, então há dependência perfeita positiva entre as variáveis já que esta cópula sugere que valores grandes (ou pequenos) das variáveis aleatórias ocorrem simultaneamente. Dizemos então que X e Y são comonotônicas.

2. Se a distribuição está caraterizada pela cópula $W(u,v)$, então a relação entre as variáveis é perfeita e negativa já que esta cópula sugere que valores grandes de uma das variáveis tendem a ocorrer quando a outra variável toma valores pequenos. Dizemos nesse caso que X e Y são contramonotônicas.

Isso a respeito dos tipo de dependência perfeita, para a ausencia de dependência ou indepenência temos o seguinte resultado.

Teorema 1.6. *Sejam X e Y variáveis aleatórias contínuas. Então X e Y são independentes se e somente se $C_{XY}(u,v) = \Pi(u,v)$.*

Demonstração. Sejam X e Y duas variáveis aleatórias contínuas independentes com função de distribuição conjunta $H(x,y)$ e marginais $F(x)$ e $G(y)$. Dado que X e Y são independentes temos que $H(x,y)$ pode ser escrita como $H(x,y) = F(x)G(y)$. Agora, usando o Corolário 1.2 temos que a cópula associada a X e Y é

$$\begin{aligned} C(u,v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= F(F^{-1}(u))G(G^{-1}(v)) \\ &= uv \\ &= \Pi. \end{aligned}$$

No outro sentido temos que, se a cópula associada a duas variáveis aleatórias contínuas é Π então $C(u,v) = uv$ e, usando o Teorema de Sklar,

$$\begin{aligned} H(x,y) &= C(F(x), G(y)) \\ &= F(x)G(y). \end{aligned}$$

Então concluímos que X e Y são independentes. □

Outra boa propriedade das cópulas é que são invariantes sob transformações estritamente monótonas das variáveis aleatórias (v.a) envolvidas na análise. Isto é facilmente observável quando temos uma v.a X com função de distribuição contínua e uma função estritamente monótona α cujo domínio contém $RanX$. Nesse caso a função de distribuição da v.a $\alpha(X)$ também é contínua. Para o caso em que α é estritamente crescente temos o seguinte resultado.

Teorema 1.7. *Sejam (X,Y) v.a. contínuas com cópula $C_{XY}(u,v)$. Se (α_1, α_2) são funções estritamente crescentes sobre $RanX$ e $RanY$ respectivamente, então $(\alpha_1(X), \alpha_2(Y))$ também tem cópula $C_{XY}(u,v)$, i.e, C_{XY} é invariante sob transformações estritamente crescentes de X e Y.*

Demonstração. Sejam F_1, G_1, F_2, G_2 as funções de distribuição de $X, Y, \alpha_1(X)$ e $\alpha_2(Y)$ respectivamente. Devido a que α_1 e α_2 são estritamente crescentes, $F_2 = P[\alpha_1(X) \leqslant x] = P[X \leqslant \alpha_1^{-1}(x)] = F_1\alpha_1^{-1}(x))$, e da mesma forma $G_2 = P[\alpha_2(Y) \leqslant y] = P[Y \leqslant \alpha_2^{-1}(y)] = G_1\alpha_2^{-1}(y))$. Então, para todo $x, y \in \mathbb{R}$,

$$\begin{aligned} C_{\alpha_1(X)\alpha_2(Y)}(F_2(x), G_2(y)) &= P[\alpha_1(X) \leqslant x, \alpha_2(Y) \leqslant y] \\ &= P[X \leqslant \alpha_1^{-1}(x), Y \leqslant \alpha_2^{-1}(y)] \\ &= C_{XY}(F_1(\alpha_1^{-1}(x)), G_1(\alpha_2^{-1}(y)] \\ &= C_{XY}(F_2(x), G_2(y)) \end{aligned}$$

Dado que X e Y são contínuas, $RanF_2 = RanG_2 = I$, daí segue que $C_{\alpha_1(X)\alpha_2(Y)} = C_{XY}$ em I^2.

□

1.4 Representação gráfica de uma cópula

Os gráficos são uma ferramenta muito útil na hora de representar funções em geral e são uma boa forma de resumir muita informação de forma amigável para o leitor. No caso particular das cópulas bidimensionais da forma $z = C(u, v)$, o gráfico é uma superfície contínua no cubo unitario $[0, 1]^3$, delimitada pelo quadrilátero de vértices $(0, 0, 0)$, $(1, 0, 0)$, $(1, 1, 1)$ e $(0, 1, 0)$. Como já mencionado, dada qualquer cópula C, se verifica que $W(u, v) \leqslant C(u, v) \leqslant M(u, v)$.

Assim, o gráfico de C fica entre os gráficos das cotas de Fréchet-Hoeffding (Figura 1), ou seja, as superfícies $z = W(u, v)$ e $z = M(u, v)$ limitam mais uma vez a superfície definida por qualquer cópula, pelo que é difícil diferenciar entre uma cópula e outra unicamente a partir de um gráfico de superfície como aquele exibido pela Figura (1).

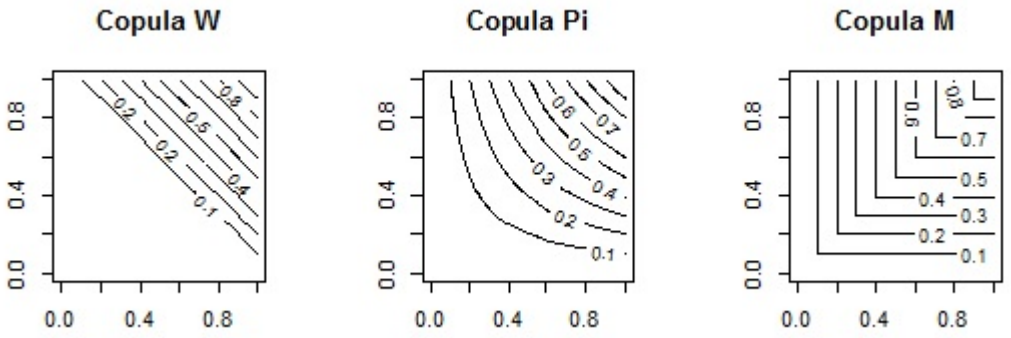

Figura 4 – Gráfico de contorno das cópulas W, Π e M.

Para comparar cópulas preferimos usar as curvas de nível ou contornos, que são conjuntos em $[0, 1]$ dados por $C(u, v) = k$ onde k é constante. A Figura 4 apresenta, por exemplo, os contornos das cópulas M, W e Π.

Fazer uso desse recurso melhora a percepção de diferenças entre duas cópulas, mas o que verdadeiramente vai nos ajudar a estabelecer graficamente quando uma cópula cumpre com certa caraterística de interesse, é a análise da sua densidade, isto é, tanto da superfície da forma $\dfrac{\partial^2 C}{\partial u \partial v}(u, v) = z$ quanto das curvas de nível da forma $\dfrac{\partial^2 C}{\partial u \partial v}(u, v) = k$.

A Figura 6 exibe o gráfico da densidade de $C(u, v) = uv + uv(1-v)(1-u)$, que é a mesma cópula do Gráfico 5, e ajuda a exemplificar que é melhor analisar a densidade da cópula no lugar da função cópula. Isto devido, a simples visão dos contornos da cópula

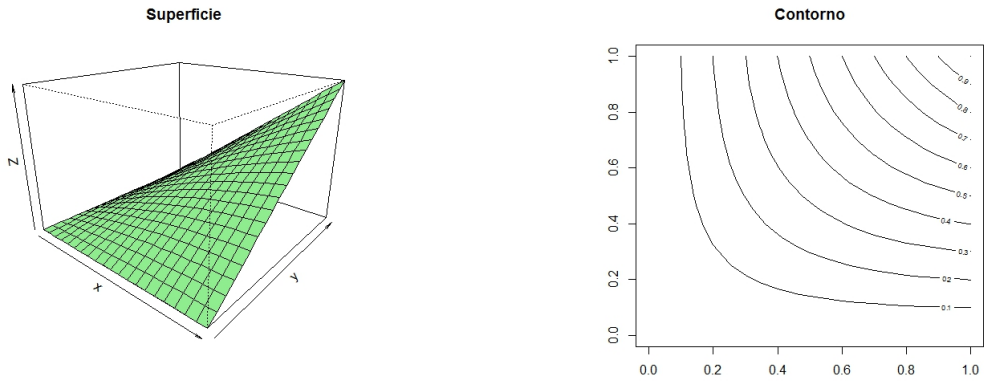

Figura 5 – Exemplo de superfície e contorno de uma cópula $C(u,v) = uv + uv(1-v)(1-u)$

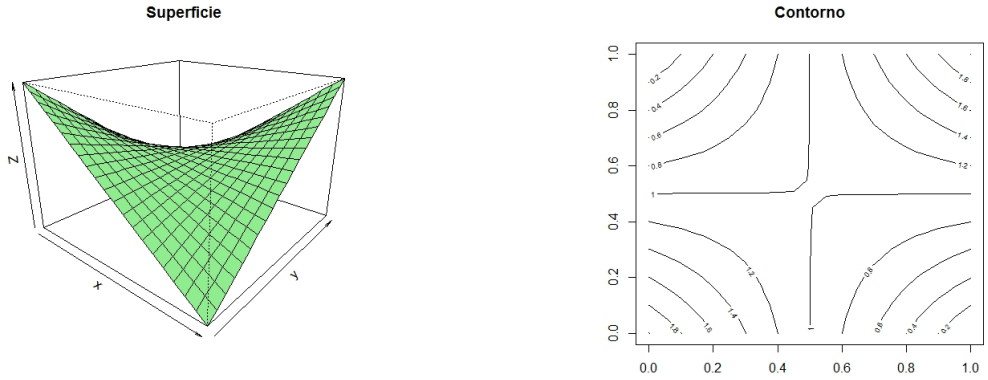

Figura 6 – Exemplo de densidade e contorno de uma cópula $C(u,v) = uv + uv(1-v)(1-u)$.

II da Figura 4 e a cópula da Figura 5, são iguais mesmo sendo de funções cópula diferentes. A maioria dos contornos das cópulas exibem um comportamento gráfico similar, pelo que essa única ferramenta não é suficiente. Por outro lado, os gráficos da Figura 6 são claramente diferentes dos análogos da cópula II, que correspondem ao plano $z = 1$.

1.5 Cópulas e dependência

Neste trabalho, o propósito do estudo das cópulas é usá-las como ferramenta para descrever a relação de dependência entre duas variáveis, digamos X e Y. Essa relação pode ser quantificada a partir de várias medidas de resumo, sendo a mais conhecida o coeficiente de correlação de Pearson $\rho(x,y)$. Embora $\rho(x,y)$ seja o mais usado, é o mais limitado de todos dado que unicamente reflete um tipo de dependência, linear.
Existem outros tipos de medidas de dependência chamadas medidas de concordância. Dizemos que duas variáveis são *concordantes* se valores grandes de uma delas estão associados a valores grandes da outra, e da mesma forma para valores pequenos, valores

pequenos de uma implicam valores pequenos na outra. No caso contrário, quando valores grandes de uma variável estão associados a valores pequenos da outra, dizemos que essas duas variáveis são *discordantes*.

Assim, podemos dizer que este tipo de medidas generalizam a relação de dependência do coeficiente de correlação de Pearson, considerando relações não lineares. Duas destas medidas, o Tau de Kendall ([Kendall, 1938]) e o Rho de Spearman ([Spearman, 1904]) são definidas a seguir.

Tau de Kendall

Sejam (X_1, Y_1) e (X_2, Y_2) dois vetores aleatórios *iid* com funções de distribuição conjunta H_1 e H_2, respectivamente, com marginais F (para X_1 e X_2) e G (para Y_1 e Y_2). Sejam C_1 e C_2 as cópulas associadas a (X_1, Y_1) e (X_2, Y_2) respectivamente. Definimos a medida τ de Kendall como a probabilidade de concordância (entre os vetores (X_1, Y_1) e (X_2, Y_2)) menos a probabilidade de discordância,

$$\tau = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right].$$

[Nelsen, 2013] demonstra que τ pode expressar-se em termos das cópulas como

$$\tau_C = 4 \int\int_{I^2} C_2(u, v) dC_1(u, v) - 1. \tag{1.16}$$

Rho de Spearman

Sejam $(X_1, Y_1), (X_2, Y_2)$ e (X_3, Y_3) três vetores aleatórios independentes com função distribuição conjunta comum H (cujas marginais são de novo F e G) e C a cópula associada a H. A versão populacional do $\rho_{X,Y}$ de Spearman é definida para ser proporcional à probabilidade de concordância menos a probabilidade de discordância de dois vetores (X_1, Y_1) e (X_2, Y_3), isto é, um par de vetores com mesmas marginais, mas um deles tem função de distribuição H, enquanto as componentes do outro vetor são independentes pelo que a sua função de distribuição é $F(x)G(y)$ ([Kruskal, 1958] e [Lehmann, 1966]).

$$\rho_{X,Y} = P\left[(X_1 - X_2)(Y_1 - Y_3) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_3) < 0\right].$$

Note que a cópula associada ao vetor (X_2, Y_3) é Π. Similarmente como para o caso do τ de Kendall, [Nelsen, 2013] prova que $\rho_{X,Y}$ pode ser escrito como

$$\rho_{X,Y} = 12 \int\int_{I^2} \left[C(u, v) - uv\right] dudv = 12 \int\int_{I^2} C(u, v) dudv - 3, \tag{1.17}$$

que vamos denotar por ρ_C.

Se consideramos duas variáveis aleatórias X e Y, estas duas medidas, τ e ρ_C, vão atingir o valor 1 se a cópula associada a elas é cota superior de Fréchet-Hoeffding (ou seja que as variáveis são comonotônicas). Contrariamente se tais medidas atinguem o valor (-1) a cópula de X e Y é a cota inferior de Fréchet e dizemos que estas variáveis são contramontônicas.

Outra medida de dependência muito usada nas aplicações em economia é o coeficiente de dependência de caudas (TDC Tail Dependence Coeficient) que, em geral, é importante no estudo de dependência de valores extremos.

Definição 1.9. *Seja C uma cópula bivariada tal que $\lim_{u\to 1} -\dfrac{1-2u-C(u,u)}{(1-u)} = \lambda_U$ existe. Dizemos que C tem dependência na cauda superior se $\lambda_U \in (0,1]$ e independência se $\lambda_U = 0$.*

Definição 1.10. *Seja C uma cópula bivariada tal que $\lim_{u\to 0^+} \dfrac{C(u,u)}{u} = \lambda_L$ existe. Dizemos que C tem dependência na cauda inferior se $\lambda_L \in (0,1]$ e independência se $\lambda_L = 0$.*

Esta dependência nas caudas entre duas variáveis aleatória contínuas X e Y pode-se estabelecer como uma propriedade da cópula e não das distribuições marginais, e portanto sua quantificação é invariante sob transformações estritamente crescentes de X e Y, como diz o Teorema 1.7.

1.6 Cópula empírica

A seguir, explicamos como estimar uma cópula de forma não paramétrica. Este procedimento foi introduzido por [Deheuvels, 1979].

Definição 1.11. *Considere as variáveis aleatorias (X,Y). Seja $(x_k,y_k)_{k=1}^n$ uma amostra observada de tamanho n obtida a partir da distribuição bivariada de (X,Y). A cópula empírica, C_n, associada a estas variáveis está definida por,*

$$\begin{aligned} C_n\left(\frac{i}{n},\frac{j}{n}\right) &= \frac{\#\text{ de pares } (x,y) \text{ da amostra tais que } x \leqslant x_{(i)} \text{ e } y \leqslant y_{(j)}}{n} \\ &= \frac{1}{n}\sum_{k=1}^{n} I\left(x \leqslant x_{(i)},\, y \leqslant y_{(j)}\right) \quad i,j = 1,..,n\ , \end{aligned}$$

em que $x_{(i)}$ e $y_{(j)}$ são estatísticas de ordem da amostra.

Definição 1.12. *A frequência da cópula empírica, c_n, vem dada por*

$$c_n\left(\frac{i}{n},\frac{j}{n}\right) = \begin{cases} \dfrac{1}{n} & \text{se } (x_{(i)},y_{(j)}) \text{ é um elemento da amostra,} \\ 0 & \text{em outro caso.} \end{cases}$$

[Deheuvels, 1979] também demostrou que a cópula empírica converge para a verdadeira cópula quando n cresce. A cópula empírica permite construir contrastes não paramétricos de independência como mostram [Deheuvels, 1980] e [Deheulves, 1981], assim como [Genest and Rémillard, 2004]. Além disso, estas cópulas são uma ferramenta muito útil para a análise exploratória, já que é uma primeira aproximação aos dados.

1.7 Tipos de cópulas

Nesta oportunidade vamos estabelecer dois critérios de classificação de cópulas, com alguns exemplos.

1.7.1 Tipos de cópulas dado o conhecimento explícito da sua forma

Por conhecimento explícito da sua forma, vamos nos referir à expressão matemática que descreve a cópula. Neste sentido, podemos dividir as cópulas em dois grandes grupos: cópulas paramétricas e não paramétricas.

Vamos chamar de paramétricas a todas aquelas cópulas pertencentes a uma família cuja expressão matemática está indexada por um parâmetro. A relação entre o parâmetro e a família é de um a um, pelo que selecionar uma cópula dentro de uma família específica é equivalente a selecionar o conjunto de parâmetros que esta envolve. As cópulas não paramétricas vão ser todas aquelas que não estão indexadas por um parâmetro, como a cópula empírica.

Em particular, neste trabalho, todas as cópulas que consideramos, exceto a empírica, são paramétricas.

1.7.2 Tipos de cópulas dada a relação de dependência que refletem

Outra forma de classificar as cópulas é a partir da relação de dependência que são capazes de modelar. Algumas das classes de maior interesse são:

Cópulas de dependência extrema:

Como seu nome diz, estas cópulas modelam relações perfeitas entre as variáveis, isto é, dependência positiva perfeita (cópula M), dependência negativa perfeita (cópula W) ou independência (cópula Π).

Cópulas elípticas:

Estas são todas as cópulas associadas às distribuições elípticas, como a normal e se caracterizam por representar relações de dependência simétricas. A simulação a partir

destas distribuições é bem simples e, como consequência do Teorema de Sklar, também é fácil simular este tipo de cópulas.

Os exemplos mais comuns destas cópulas, são a cópula gaussiana que vem da distribuição normal e a t-cópula (ou cópula de Student) que vem da distribuição t-Student, das quais falaremos em seções posteriores.

Cópulas de valor extremo:

Estas cópulas representam relações que dão maior peso ao que acontece nas caudas (extremos) das distribuições marginais. [Segers, 2004] diz que as cópulas de valor extremo são os possíveis limites (se existirem) de cópulas associadas a máximos de amostras aleatórias identicamente distribuídas.

1.7.3 Cópulas arquimedianas

Esta classe encerra um grande número de famílias de cópulas com diferentes e variadas caraterísticas. Isto torna difícil classificá-las num tipo específico de dependência, como no caso das cópulas elípticas (que refletem simetria) ou as cópulas de valor extremo (que dão mais importância à dependência nas caudas). Mas então, por que é que falamos das arquimedianas como uma classe específica de cópulas? Antes de responder consideremos a seguinte definição.

Definição 1.13. *Seja Φ o conjunto de funções contínuas, estritamente decrescentes e convexas da forma $\varphi : [0,1] \to [0,\infty]$ onde $\varphi(1) = 0$.*

Schweizer e Sklar demonstraram que cada elemento de Φ gera uma cópula C a partir da seguinte relação [Joe, 1997],

$$C(u,v) \quad = \quad \varphi^{-1}\left(\varphi(u) + \varphi(v)\right) \quad \text{com } 0 \leqslant u, v \leqslant 1.$$

A função φ é conhecida como o gerador da cópula. Quando $\varphi(0) = \infty$ se diz que φ é um gerador estrito.

Respondendo à pergunta formulada anteriormente, as cópulas arquimedianas são todas aquelas que podem ser geradas da forma descrita na Definição 1.13. Muitas das famílias de cópulas mais conhecidas pertencem à esta classe, e por isso é interessante escrever alguns dos resultados apresentados anteriormente em termos do gerador da cópula arquimediana. De fato, isto permite simplificar os cálculos. Por exemplo, o τ de Kendall

pode-se expressar por,

$$\tau_C = 1 + 4\int_0^1 \frac{\varphi(u)}{\varphi'(u)}du.$$

E para a dependência de caudas, temos o seguinte teorema.

Teorema 1.8. *Seja C uma cópula arquimediana com gerador $\varphi \in \Omega$. Então,*

$$\lambda_U = 2 - \lim_{s\to 0^+}\left[\frac{1-\varphi^{(-1)}(2s)}{1-\varphi^{(-1)}(s)}\right]$$

e

$$\lambda_L = \lim_{s\to\infty}\left[\frac{\varphi^{(-1)}(2s)}{\varphi^{(-1)}(s)}\right].$$

Demonstração. Da Definição 1.10 temos que $\lambda_L = \lim_{t\to 0^+} C(t,t)/t$. Se escrevemos esta expressão em termos do gerador $\varphi(t)$, temos que

$$\lambda_L = \lim_{t\to 0^+}\frac{\varphi^{(-1)}(2\varphi(t))}{t}.$$

Se definimos $s = \varphi(t)$ então

$$\lambda_L = \lim_{s\to\infty}\frac{\varphi^{(-1)}(2s)}{\varphi^{(-1)}(s)}.$$

Para λ_U, a análise é semelhante.

□

1.8 Algumas cópulas comuns

Geralmente quando falamos dos diferentes tipos de cópulas que existem, na verdade fazemos referência a diferentes tipos de famílias. Todas as cópulas que pertencem a uma mesma família apresentam a mesma estrutura matemática que pode depender de parâmetros. Portanto, para cada um dos valores dos parâmetros obtemos um membro de tal família. Nesta seção vamos definir algumas das famílias de cópulas paramétricas mais comuns.

Para uma ilustrão gráfica destas cópulas, ver os gráficos do Anexo A.

1.8.1 Cópula gaussiana

O nome gaussiana se deve a que a sua expressão coincide com a função de distribuição de uma normal padrão bivariada, que é uma distribuição elíptica, logo esta

é uma cópula da classe elíptica. Considerando o coeficiente de correlação de Pearson, $-1 \leqslant \rho \leqslant 1$, temos a seguinte expressão,

$$C(u,v) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{2(1-\rho^2)}\right) dt_1 dt_2$$

com $x_1 = \Phi^{-1}(u)$, $x_2 = \Phi^{-1}(v)$, onde Φ denota a função distribuição de uma N(0,1). Por definição, as funções de distribuição marginais coincidem com a normal padrão. A densidade da cópula gaussiana vem dada pela seguinte expressão:

$$c(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right).$$

Esta cópula não tem expressões fechadas para o ρ de Spearman nem para o τ de Kendall, e por isso é preciso aproximá-los usando os resultados de seções anteriores. Com respeito aos coeficientes de dependência de caudas, tanto superior quanto inferior, obtemos que $\lambda_U = \lambda_L = 0$.

1.8.2 Cópula T-Student

Como no caso anterior, temos outra cópula elíptica cuja expressão coincide com uma função distribuição, neste caso, a t-Student bivariada e coeficiente de correlação $-1 \leqslant \rho \leqslant 1$,

$$C(u,v) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2} dt_1 dt_2,$$

onde ν corresponde aos graus de liberdade da t-Student considerada. As distribuições marginais desta cópula coincidem com a t-Student padrão ($x_1 = t_\nu^{-1}(u)$ e $x_2 = t_\nu^{-1}(v)$). A função densidade desta cópula tem a seguinte forma

$$c(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2}$$

Para a cópula t-Stundent, é preciso usar as expressões descritas na Seção 1.5 para o $\rho_{X,Y}$ de Spearman (1.16) e o τ de Kendall (1.17) para achar estas medidas, dado que elas não têm expressões fechadas. Com respeito aos coeficientes de dependência caudal, temos que $\lambda_U = \lambda_L = 2t_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}\right) \geqslant 0$ e portanto esta cópula apresenta dependência em ambas caudas.

1.8.3 Cópula de Gumbel

Esta cópula é uma cópula arquimediana com gerador $\varphi(t) = (-\ln(t))^\theta$, portanto a função cópula vem dada por:

$$C(u,v) = \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right),$$

onde $\theta \in [1, +\infty)$ e sua densidade é dada pela equação:

$$c(u,v) = \frac{(-ln(u))^{\theta-1}\left[-1+\theta+(-ln(u))^{\theta}+(-ln(v))^{\theta}\right]^{\frac{1}{\theta}}\left[(-ln(u))^{\theta}+(-ln(v))^{\theta}\right]^{\frac{1}{\theta}-2}(-ln(v))^{\theta-1}}{exp\left[(-ln(u))^{\theta}+(-ln(v))^{\theta}\right]^{1/\theta}uv}.$$

Na cópula de Gumbel, a dependência é positiva perfeita (tem forma equivalente à cópula M) quando $\theta \to \infty$ e coincide com a cópula Π quando $\theta = 1$. A expressão fechada para o τ de Kendall é bem simples:

$$\tau(\theta) = 1 - \frac{1}{\theta}.$$

Esta cópula inclui o caso de dependência de cauda superior, que se obtém considerando $\lambda_U = 2 - 2^{1/\theta}$e $\lambda_L = 0$.

1.8.4 Cópula de Clayton

Esta também é uma cópula arquimediana, neste caso o gerador da cópula (descrito na Seção 1.8) é $\varphi = (1/\theta)\left(t^{-\theta} - 1\right)$ e a cópula é:

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

onde $\theta > 0$. Para a densidade, temos esta expressão:

$$c(u,v) = (1+\theta)\, u^{(-1-\theta)} v^{(-1-\theta)} \left(u^{-\theta} + v^{-\theta} - 1\right)^{(-2-\frac{1}{\theta})}.$$

A cópula de Clayton, quando $\theta \to \infty$, coincide com a cópula M (dependência positiva perfeita) e coincide com a cópula independência quando $\theta \to 0$. Para esta cópula também temos uma forma explícita para calcular o valor do τ de Kendall:

$$\tau(\theta) = \frac{\theta}{\theta + 2}.$$

Diferentemente da cópula de Gumbel, a de Clayton possui dependência na cauda inferior se $\lambda_L = 2^{-1/\theta}$ e para cauda superior se $\lambda_U = 0$.

1.8.5 Cópula de Frank

Seguindo com as cópulas arquimedianas, definimos a família de Frank como as cópulas geradas pela função $\varphi = -\ln\left[\dfrac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right]$ cuja expressão é

$$C(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{\left(e^{-\theta} - 1\right)}\right),$$

onde $\theta \in (-\infty, \infty) \setminus \{0\}$. Enquanto que a densidade tem a seguinte expressão,

$$c(u,v) = \frac{\theta e^{\theta(1+u+v)}\left(e^{\theta} - 1\right)}{\left(e^{\theta(u+v)} - e^{\theta}\left(e^{\theta u} + e^{\theta v} - 1\right)\right)^2}.$$

Assim como as cópulas gaussiana e t-Student, a cópula de Frank admite dependência tanto positiva (quando $\theta \to \infty$) quanto negativa (quando $\theta \to -\infty$) e reflete independência quando $\theta \to 0$. O τ de Kendall pode-se calcular usando

$$\tau(\theta) \quad = \quad 1 - \frac{4}{\theta} + \frac{4}{\theta^2} \int_0^\theta \frac{t}{e^t - 1} dt.$$

1.9 Métodos de construção de cópulas

Como consequência do Teorema de Sklar, consegue-se obter distribuições bivariadas ou multivariadas a partir de uma função cópula e das marginais que desejamos fixar. Este fato é uma vantagem importante para a modelagem e simulação de variáveis aleatórias e torna a construção e escolha de cópulas questões de interesse para quem trabalha com esses objetivos. Ao longo do tempo, muitos autores têm desenvolvido métodos que servem para construir cópulas que possuam certas caraterísticas desejáveis orientadas a identificar algum tipo particular de relação de dependência. Nesta oportunidade vamos apresentar um resumo dos métodos mais comuns encontrados na literatura.

1.9.1 Método de inversão

Este método permite obter funções cópula a partir de inversas de funções de distribuição, baseando-se apenas no Corolário 1.2 de onde temos que :

$$C(u, v) \quad = \quad H(F^{-1}(u), G^{-1}(v)), \tag{1.18}$$

para F, G e H satisfazendo as condições do Corolário 1.2, Nelsen generaliza este resultado usando as quase-inversas de $F^{(-1)}$ e $G^{(-1)}$ e por isso não é necessário que F e G sejam estritamente crescentes nem contínuas. Note que a função cópula C da expressão (1.18) serve para, dadas duas funções de distribuição univariadas F' e G' (diferentes de F e G), obter uma outra conjunta H' (também diferente de H) com marginais F' e G'. De fato, pelo Teorema de Sklar $H'(x, y) = C(F'(x), G'(y))$, é uma função de distribuição bivariada com marginais F' e G'. Dois exemplos deste método são as cópulas gaussiana e t-Student, que podem se escritas respectivamente como $C_G(u, v) = \Phi_2(\Phi_1^{-1}(u), \Phi_1^{-1}(v)$ e $C_T(u, v) = t_{\nu,2}\left(t_{\nu,1}^{-1}(u), t_{\nu,1}^{-1}(v)\right)$, onde Φ_n é uma distribuição normal n-variada e $t_{\nu,n}$ é uma distribuição t-Student n-variada com ν graus de liberdade.

1.9.2 Métodos geométricos.

Estes métodos são baseados na própria definição de cópula, isto é, no lugar de fazer referência às variáveis aleatórias ou às suas funções de distribuição, usamos as características geométricas das funções cópula para achar restrições sobre algumas funções

para que cumpram as características da Definição 1.4. Entre algumas destas características, temos o suporte ou a forma dos gráficos das suas seções.

1.9.2.1 Soma ordinal de cópulas

A partir deste método, consegue-se construir uma cópula C como combinação de um número finito e enumerável de cópulas C_i. O suporte de C obtém-se ao rescalar o suporte de cada C_i a um quadrado $J_i^2 = [a_i, b_i] \times [a_i, b_i] \subseteq I^2$, onde esses intervalos $[a_i, b_i]$ sao fechados, disjuntos e não degenerados. Ou seja, o suporte de C é a combinação dos suportes de uma sequência de cópulas. A Definição 1.14 formaliza este tipo de construção.

Definição 1.14. *Seja $\{J_k\}_{k\in K}$ uma partição enumerável de I, isto é, um conjunto de intervalos fechados e disjuntos $[a_k, b_k]$ cuja união é o quadrado unitário. Seja $\{C_k\}_{k\in K}$ um conjunto de cópulas associados à partição $\{J_k\}_{k\in K}$. Define-se uma soma ordinal de $\{C_i\}_{i\in I}$ com respeito a $\{J_k\}_{k\in K}$ como a cópula:*

$$C(u,v) = \begin{cases} a_k + (b_k - a_k)C_k\left(\dfrac{u-a_k}{b_k-a_k}, \dfrac{v-a_k}{b_k-a_k}\right) & (u,v) \in J_k^2 \\ M(u,v) & \textit{Caso contrário} \end{cases}$$

Exemplo 1.7. *Consideremos a soma ordinal de (W, W) com respeito a $([0,\theta],[\theta,1])$. Temos então, para $(u,v) \in [0,\theta]^2$,*

$$C(u,v) = \theta W(u/\theta, v/\theta) = \theta \max(\frac{u}{\theta} + \frac{v}{\theta} - 1, 0),$$

se $(u,v) \in [\theta,1]^2$

$$C(u,v) = \theta + (1-\theta)W(\frac{u-\theta}{1-\theta}, \frac{v-\theta}{1-\theta}) = \theta + (1-\theta)\max(\frac{u+v}{1-\theta} - 1, 0)$$

Então a expressão para a cópula $C(u,v)$ é

$$C(u,v) = \begin{cases} \max(u+v-\theta, 0) & \textit{se } (u,v) \in [0,\theta]^2 \\ \max(u+v-1, \theta) & \textit{se } (u,v) \in [\theta,1]^2 \\ min(u,v) & \textit{se nao} \end{cases}$$

[Mesiar and Sempi, 2010] demostram que uma soma ordinal de n cópulas é de fato uma cópula, e também mostram que toda cópula pode ser escrita como uma soma ordinal de cópulas dadas algumas condições.

1.9.2.2 Soma convexa de cópulas

Nelsen propõe como exercício (ver exercício 2.3(a) [Nelsen, 2013]) o seguinte resultado,

Proposição 1.3. *Sejam duas cópulas C_1 e C_2 e um valor $\theta \in I = [0,1]$. Então $C(u,v) = (1-\theta)C_1(u,v) + \theta C_2(u,v)$ é também cópula.*

Demonstração. Sejam C_1 e C_2 duas funções que cumprem as condições da Definição 1.4, isto é, duas funções cópula quaisquer. Definamos a soma convexa destas duas funções como a função $C(u,v) = (1-\theta)C_1(u,v) + \theta C_2(u,v)$, onde $\theta \in I = [0,1]$. Assim, a função $C(u,v)$ vai ser cópula se verificamos para ela as duas condições da Definição 1.4. Para a primeira condição, temos que dado que C_1 e C_2 são cópulas,

$$C(0,v) \;=\; (1-\theta)C_1(0,v) + \theta C_2(0,v) = (1-\theta)(0) + \theta(0) = 0,$$

e analogamente demostramos que $C(u,0) = 0$. Também temos que

$$C(1,v) \;=\; (1-\theta)C_1(1,v) + \theta C_2(1,v) = (1-\theta)(v) + \theta(v) = v$$

e da mesma forma podemos obter que $C(u,1) = u$, e assim demostramos a condição 1 da Definição 1.4. Para a segunda condição, devemos verificar que o C-volume é não negativo, isto é, $C(u_2,v_2) - C(u_2,v1) - C(u_1,v_2) + C(u_1,v_1) \geqslant 0$ para todos $u_1, u_2, v_1, v_2 \in I$ tais que $u_1 \leqslant u_2$ e $v_1 \leqslant v_2$. Para isto, consideremos o seguinte: dado que C_1 e C_2 cumprem esta condição e que por sua vez C se escreve em termos destas duas, podemos dividir a expressão do C-volume nas expressões correspondentes aos C_1-volume e C_2-volume, assim

$$\begin{aligned}
V_C([u_1,u_2]\times[v_1,v_2]) &= C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \\
&= (1-\theta)C_1(u_2,v_2) + \theta C_2(u_2,v_2) - (1-\theta)C_1(u_2,v_1) - \theta C_2(u_2,v_1) - \\
&\quad -(1-\theta)C_1(u_1,v_2) - \theta C_2(u_1,v_2) + (1-\theta)C_1(u_1,v_1) + \theta C_2(u_1,v_1) \\
&= (1-\theta)\left[C_1(u_2,v_2) - C_1(u_2,v_1) - C_1(u_1,v_2) + C_1(u_1,v_1)\right] + \\
&\quad +\theta\left[C_2(u_2,v_2) - C_2(u_2,v_1) - C_2(u_1,v_2) + C_2(u_1,v_1)\right] \\
&= (1-\theta)V_{C_1}([u_1,u_2]\times[v_1,v_2]) + \theta V_{C_2}([u_1,u_2]\times[v_1,v_2]).
\end{aligned}$$

Dado que $\theta \in I$ e que tanto $V_{C_1} \geqslant 0$ quanto $V_{C_2} \geqslant 0$ conluimos que $V_C \geqslant 0$ o que encerra a demostração. □

Este resultado pode-se estender para considerar a combinação de uma família arbitrária de cópulas, ou seja, uma soma convexa de $\{C_\theta\}_{\theta\in\Theta}$, onde consideramos o parâmetro θ como uma observação de uma variável aleatória contínua Θ com função de densidade Λ.

Definição 1.15. *Define-se a soma convexa de uma sequência de cópulas $\{C_\theta\}_\Theta$ com respeito a Λ, função que recebe o nome de função de mistura, como $\int_R C_\theta(u,v)d\Lambda(\theta)$.*

1.9.2.3 Método baseado no conhecimento das seções da cópula.

No Seção 1.1, foram apresentados os conceitos de seções de uma cópula (horizontais, verticais e diagonais) na Definição 1.5. Relembrando, estas seções correspondem à imagem bidimensional da cópula que resulta ao fixar o valor de uma das variáveis (u no caso horizontal e v no caso vertical) ou de fazer ambas iguais ao mesmo valor (no caso diagonal).

Este método permite construir cópulas a partir do conhecimento da expressão das seções horizontal e vertical principalmente. Por exemplo, a cópula cuja expressão corresponde a um polinômio de grau n em u vai ter a forma $u \to C(u,v) = a_n(v)u^n + a_{n-1}(v)u^{n-1} + ... + a_1(v)u + a_o(v)$ e equivalentemente em v será do tipo $v \to C(u,v) = b_n(u)v^n + b_{n-1}(u)v^{n-1} + ... + b_1(u)v + b_o(u)$, ou seja, as seções horizontal e vertical respctivamente.

O método supõe o conhecimento das seções de uma cópula e a partir delas perguntar-se pelas condições que devem satisfazer os polinômios $a_i(v)$ e $b_i(u)$ para que a função $C(u,v)$ seja mesmo uma cópula. [Nelsen, 2013] faz uma explicação detalhada para os casos mais simples dos polinômios, isto é, quando as seções se supõem lineares, quadráticas ou cúbicas.

Cópulas com seções lineares

Neste caso escrevemos por exemplo $C(u,v) = a_1(v)u + a_0(v)$ se desejamos uma cópula com seção linear em u. As funções a_1 e a_0 podem ser achadas usando a condição 1 da Definição 1.4, isto é,

$$0 = C(0,v) = a_0(v) \quad \text{e} \quad v = C(1,v) = a_1(v),$$

pelo que existe unicamente uma cópula com seção horizontal (ou vertical) linear, que é a cópula Π.

Cópulas com seções quadráticas.

De novo, suponhamos que queremos uma cópula com seção quadrática em, por exemplo, u. Temos então $C(u,v) = a_2(v)u^2 + a_1(v)u + a_0(v)$, onde as funções a_i são tais que:

$$0 = C(0,v) = a_0(v) \quad \text{e} \quad v = C(1,v) = a_2(v) + a_1(v).$$

Se definimos $a_2(v) = -\psi(v)$, então $a_1(v) = v - a_1(v) = v + \psi(v)$ e podemos escrever:

$$C(u,v) = uv + \psi(v)u(1-u), \tag{1.19}$$

onde ψ é uma função tal que C é 2-crescente e $\psi(0) = \psi(1) = 0$ (para que $C(u,0) = 0$ e $C(u,1) = u$). Para nos ajudar a encontrar funções ψ temos o seguinte teorema.

Teorema 1.9. *Seja ψ uma função com domínio em I e seja C uma cópula dada por (1.19) para $u, v \in I$. Então C é uma cópula se e somente se:*

1. $\psi(0) = \psi(1) = 0$.
2. *ψ satifaz a condição de Lipschitz*

$$|\psi(v_2) - \psi(v_1)| \leqslant |v_2 - v_1|,$$

para todo v_1, v_2 em I. Além disso, C é absolutamente contínua.

Um exemplo clássico neste contexto é a família de cópulas conhecida como a família de Farlie-Gumbel-Morgenstern (FGM). Suponha que C é simétrica e tem seção quadrática em v. Isto implica que C satisfaz $C(u,v) = uv + \psi(u)v(1-v)$. Consequentemente, se $\psi(u) = \theta u(1-u)$ para algum parâmetro θ, então

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v).$$

O C_θ-volume do retângulo $[u_1, u_2] \times [v_1, v_2]$ pode-se simplificar até obter:

$$V_{C_\theta}([u_1,u_2] \times [v_1,v_2]) = (u_2 - u_1)\left[1 + \theta(1-u_1-u_2)(1-v_1-v_2)\right]$$

Dado que $(1-u_1-u_2)(1-v_1-v_2)$ pertence a $[-1,1]$ em I, segue que C_θ é 2-crescente, e portanto é uma cópula se e somente se $\theta \in [-1,1]$. Note que os membros desta família são cópulas com seções quadráticas tanto verticais quanto horizontais. Principalmente por sua forma analítica simples, a família FGM tem sido amplamente usada em modelagem, para testes de associação e para estudar a eficiência de procedimentos não paramétricos. Para mais aplicações desta família consultar [Conway, 1983] e [Hutchinson and Lai, 1990]

Cópulas com seções cúbicas.

Seguindo a linha dos dois casos anteriores, podemos estender essas idéias para construir cópulas com seções cúbicas. Seja uma cópula com seções cúbicas em u. Logo, C vem dada por $C(u,v) = a_3(v)u^3 + a_2(v)u^2 + a_1(v)u + a_0(v)$ onde as a_i são funções apropriadas. De novo usando as condições da Definição 1.4,

$$0 = C(0,v) = a_0(v) \quad \text{e} \quad v = C(1,v) = a_3(v) + a_2(v) + a_1(v).$$

Se definimos $\alpha(v) = -a_3(v) - a_2(v)$ e $\beta(v) = -2a_3(v) - a_2(v)$ podemos reescrever:

$$C(u,v) = uv + u(1-u)\left[\alpha(v)(1-u) + \beta(v)u\right], \tag{1.20}$$

onde α e β são funções tal que $\alpha(0) = \beta(0) = \beta(0) = \beta(1) = 0$. As condições requeridas para que C seja uma cópula vêm dadas pelo seguinte teorema.

Teorema 1.10. *Sejam α e β duas funções de $I \to R$ satisfazendo $\alpha(0) = \beta(0) = \beta(0) = \beta(1) = 0$, e seja C uma função definida por (1.20). Então C é uma cópula se e somente se para todo u_1, u_2, v_1, v_2 em I tais que $u_1 < u_2, v_1 < v_2$, temos:*

$$\left[(1-u_1)^2+(1-u_2)^2+u_1u_2-1\right]\frac{\alpha(v_2)-\alpha(v_1)}{v_2-v_1}-\left[u_1^2+u_2^2+(1-u_1)(1-u_2)-1\right]\frac{\beta(v_2)-\beta(v_1)}{v_2-v_1} \quad \leqslant \quad -1.$$

Usar este teorema diretamente pode ser muito complexo. Na sequência, enunciamos um teorema que surge do anterior e que simplifica um pouco as contas.

Teorema 1.11. *Sejam α, β e C definidos como no Teorema 1.10. Então C é uma cópula se e somente se:*

1. *$\alpha(v)e\ \beta(v)$ são absolutamente contínuas.*
2. *Para quase todo v em I o ponto nas derivadas $(\alpha'(v), \beta'(v))$ é tal que, $-1 \leqslant \alpha'(v) \leqslant 2$ e $-2 \leqslant \beta'(v) \leqslant 1$ ou $[\alpha'(v)]^2 - \alpha'(v)\beta'(v) + [\beta'(v)]^2 - 3\alpha'(v) + 3\beta'(v) \leqslant 0$. Além disso, C é absolutamente contínua.*

Vamos chamar a região descrita na parte 2 do Teorema 1.11 de S (na Figura 7), que é a união do conjunto de pontos do quadrado $[-1, 2] \times [-2, 1]$ e o conjunto de pontos dentro da elipse em $\mathbb{R}^2$ cuja equação é $x^2 - xy + y^2 - 3x - 3y = 0$.

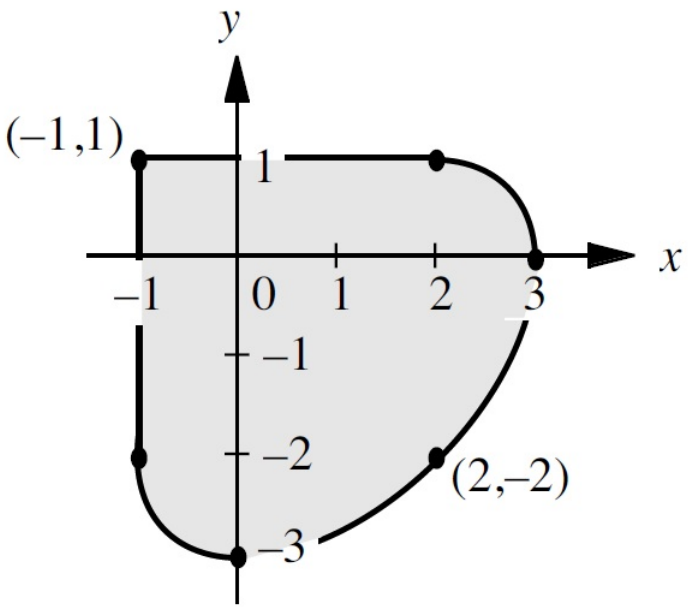

Figura 7 – Região S descrita no Teorema 1.11.

Usando os teoremas anteriores enunciamos o seguinte que vai simplificar mais a construção de cópulas de seções cúbicas.

Teorema 1.12. *Suponha que C tem seções cúbicas em ambas u e v. Então,*

$$C(u,v) = uv + uv(1-u)(1-v)\left[A_1 v(1-u) + A_2(1-v)(1-u) + B_1 uv + B_2 u(1-v)\right] \tag{1.21}$$

onde A_1, A_2, B_1, B_2 são constantes reais tais que os pontos $(A_2, A_1), (B_1, B_2), (B_1, A_1)$ e (A_2, B_2) pertencem a S.

Como consequência imediata do Teorema 1.12, temos o seguinte corolário.

Corolário 1.3. *Suponha que C tem seções cúbicas em u e v, isto é, C é dada pela equação (1.21) do Teorema 1.12. Então:*

1. *C é simétrica ($C(u,v) = C(v,u)$) se e somente se $A_1 = B_2$.*
2. *C apresenta simetria radial se e somente se $A_1 = B_2$ e $A_2 = B_1$.*
3. *Se $A_1 = B_2 = -A_2 = -B_1$ então C é tal que $C(u,v) = u - C(u, 1-v)$ e $C(u,v) = v - C(1-u, v)$.*

Um exemplo de cópulas de seções cúbicas são as da família Sarmanov, que tem a seguinte forma:

$$C_\theta(u,v) \;=\; uv + uv(1-u)(1-v)\left[3\theta + 5\theta^2(1-2u)(1-2v)\right],$$

onde $\theta \in [0,1]$. Aplicações desta familia encontram-se em [Ting Lee, 1996]. Algumas generalizações podem-se achar em [Bairamov et al., 2011].

Também seguindo o corolário podemos construir uma familia de cópulas fixando $A_1 = A_2 = a$ e $B_1 = B_2 = b$ na equação (1.21), obtendo uma expressão para cópulas assimétricas com seções cúbicas em u e quadráticas em v quando $a \neq b$:

$$C_{a,b}(u,v) = uv + uv(1-u)(1-v)\left[a(1-u) + bu\right], \tag{1.22}$$

onde $-1 \leqslant a, b \leqslant 1$. Se $a = b$ obtemos a família FGM.

Se queremos por exemplo, uma cópula com seções cúbicas tanto em u quanto em v, podemos fixar $A_1 = a$ e $A_2 = B_1 = B_2 = b$ e obtemos

$$C(u,v) = uv + uv(1-u)(1-v)\left[(a-b)v(1-u) + b\right] \tag{1.23}$$

onde $|b| \leqslant 1$ e $\left[b - 3 - (9 + 6b - 3b^2)^{1/2}\right]/2 \leqslant a \leqslant 1$. Esta familia de cópulas também é assimétrica como no caso anterior.

2 Preliminares sobre inferência Bayesiana

O principal objetivo deste trabalho é usar a análise Bayesiana para estimar uma cópula para um conjunto de dados, pelo que precisamos falar de estatística Bayesiana para contextualizar o leitor neste paradigma. Este capítulo está baseado principalmente no trabalho de [Schervish, 2012].

Um dos problemas fundamentais no estudo da estatística é a *inferência*. Considerando um conjunto de dados, estamos interessados em tirar conclusões ou *obter inferências* sobre caraterísticas desconhecidas relacionadas com o sistema objetivo do estudo ao que pertencem os referidos dados.

Exemplo 2.1. *Em época de eleições, é costume fazer pesquisas para medir a intenção de voto da população. Em particular, A pode estar interessada em saber que porcentagem da população votaria a esse candidato se nesse dia fossem as eleições. Neste caso, precisa-se de uma amostra (que se supõe representativa de toda a população), para responder a uma pergunta: considera votar no candidato A nas próximas eleições? A população é bem definida, são todas aquelas pessoas habilitadas para votar nas eleições, mas a caraterística θ: votar ao candidato A é desconhecida. O objetivo da inferência é usar as respostas da amostra para fazer afirmações sobre a quantidade θ para, no caso da equipe do candidato A, ajudar a melhorar a estratégia de campanha.*

O problema da inferência tem sido objetivo de estudo desde o século XVIII quando começou o estudo sistemático da teoria da probabilidade. A partir da interpretação de probabilidade é possível considerar dois enfoques: frequentista ou clássico e Bayesiano. [DeGroot, 2005]

Por exemplo, quando afirmamos que ao lançar uma moeda a probabilidade de obter cara é 1/2 podemos interpretar o significado desta afirmação de duas formas. Inicialmente desde o ponto de vista frequentista, pode significar que se jogamos uma moeda muitas vezes esperamos obter aproximadamente a mesma proporção de caras que de coroas (interpretação clássica de uma probabilidade). Por outro lado, segundo a interpretação Bayesiana, a probabilidade 1/2 é uma afirmação subjetiva, ou seja, é o que certo indivíduo espera ao lançar uma moeda, mas pode ser um valor diferente para outro indivíduo.

Estabeleçamos inicialmente pontos comuns entre estes enfoques. Para ambos casos se utilizam modelos com parâmetros desconhecidos para caracterizar certo fenômeno e para estimar tais parâmetros ambos métodos necessitam de dados. Com respeito às diferenças, a mais clara delas é a forma de tratar os parâmetros: para a escola frequentista,

os parâmetros são valores fixos, mesmo desconhecidos, e a sua estimação é baseada na escolha de um valor para o parâmetro tal que a probabilidade de se observar os dados seja a máxima possível; para a escola Bayesiana, a informação sobre estes parâmetros é modelada por uma distribuição, obtida a partir do Teorema de Bayes, isto é, eles são considerados variáveis aleatórias.

Teorema 2.1 (Teorema de Bayes)**.** *Sejam $B_1, B_2, ..., B_k$ eventos mutuamente exclusivos e exaustivos de um espaço amostral. Para qualquer evento novo A, temos*

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)} \tag{2.1}$$

O Teorema 2.1 é a versão probabilística geral do Teorema de Bayes, mas para a análise que estamos trabalhando é mais interessante a seguinte versão:

Teorema 2.2 (Teorema de Bayes para Variáveis Aleatórias)**.** *Sejam X e θ variáveis aleatórias com fdp (função de densidade de probabilidade) $f(x|\theta)$ e $\pi(\theta)$. Então:*

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_\Theta f(x|\theta)\pi(\theta)d\theta}. \tag{2.2}$$

A ideia é a seguinte, o pesquisador tem informação prévia sobre os parâmetros que quer estimar e antes de obter os dados, ela pode ser quantificada por uma distribuição de probabilidade. Esta informação inicial vai se modificar em função dos dados observados obtendo assim uma distribuição a posteriori (após observar) que resumirá todo o conhecimento do investigador. No contexto do Teorema 2.2 temos que:

- X: Variável aleatória que caracteriza os dados,
- θ: Parâmetro desconhecido,
- $f(x|\theta)$: Densidade da variável X dado o parâmetro (desconhecido) θ,
- $\pi(\theta)$: Densidade a priori de θ,
- $\pi(\theta|X)$: Densidade a posteriori de θ.

Segundo [Albert, 1997], isto é consistente com a implementação do método científico. A distribuição a priori representa as informações iniciais sobre o modelo. Feitas as observações, a distribuição a posteriori representa sua informação atualizada depois de olhar seus dados.

A equação do Teorema 2.2 tem $f(x|\theta)$ mas, para nosso cálculo da densidade a posteriori na verdade precisamos substituir essa função por uma que encerre toda a informação que podemos obter da amostra. Essa função é chamada de função de

verossimilhança que não é outra coisa que a função de densidade conjunta da amostra e é denotada por $L(\theta|x) = f(x_1, ...x_n|\theta)$. Assim seguindo o Teorema 2.2 temos:

$$\pi(\theta|x_1, ..., x_n) = \frac{f(x_1, ..., x_n|\theta)\pi(\theta)}{\int_\Theta f(x_1, ..., x_n|\theta)\pi(\theta)d\theta} \tag{2.3}$$

Esta expressão é a base da inferência Bayesiana. Na pratica, o denominador da expressão anterior não precisa ser calculado devido a que é constante com respecto a θ, pelo que em geral a regra de Bayes se escreve:

$$\pi(\theta|x_1, ..., x_n) \propto f(x_1, ..., x_n|\theta)\pi(\theta). \tag{2.4}$$

Precisamos apenas conhecer a densidade a posteriori a menos de uma constante de normalização. Muitas vezes somos capazes de identificar a distribuição a posteriori do parâmetro apenas olhando o numerador da expressão.

Exemplo 2.2. *De novo no contexto do Exemplo 2.1, suponha que a equipe de trabalho do candidato A tem informação suficiente para pensar que θ: probabilidade de uma pessoa votar a A, tem distribuição $Beta(2, 3)$. Se, além disso, supomos que X (Intenção de voto) é uma v.a com distribuição $Bernoulli(\theta)$, onde $x_i = 1$ se a pessoa responde que considera votar ao candidato A e $x_i = 0$ no caso contrário. A equipe selecionou uma amostra aleatória de tamanho 120 da variável X observando que 58 pessoas consideram votar ao candidato A.*

$$\begin{aligned}
\pi(\theta|x_1, ..., x_{120}) \propto f(x_1, ..., x_{120}|\theta)\pi(\theta) & \\
& \propto \left[\prod_{i=1}^{120} \theta^{x_i}(1-\theta)^{1-x_i}\right] \\
& \left[\frac{1}{B(2,3)}\theta(1-\theta)^2\right] \\
& \propto \left[\theta^{58}(1-\theta)^{62}\right]\left[\theta(1-\theta)^2\right] \\
& \propto \theta^{59}(1-\theta)^{64}
\end{aligned}$$

$\theta \in (0, 1)$. Olhando a expressão $\theta^{59}(1-\theta)^{64}$ podemos identificar que $\pi(\theta|x_1, ..., x_{120})$ tem forma da densidade de uma $Beta(60, 65)$ pelo que podemos afirmar que a distribuição a posterior de θ é Beta com esses parâmetros.

Esta é uma situação desejável para o investigador, pois é uma forma de simplificar os seus resultados a partir do fato de que, sob certas condições, deve existir uma família de distribuições padrão para o parâmetro θ tal que cumpra com a propriedade: se a distribuição a priori de θ pertence a esta família, então para qualquer tamanho de amostra e quaisquer valores observados dela, a distribuição a posteriori de θ também pertence a esta família. Vamos formalizar isto a partir da seguinte definição.

Definição 2.1. *Seja $\mathcal{F}$ a classe de f.d.p ou f.m.p $f(x|\theta)$. A classe Π de distribuições a priori é uma familia conjugada para $\mathcal{F}$ se a distribuição a posteriori está na classe Π $\forall$ $f \in \mathcal{F}$ e todas as prioris em Π, para todo x que pertence ao suporte de X.*

2.1 Estimação pontual

Voltando ao Exemplo 2.1, temos a equipe do candidato A interessada em conhecer a proporção θ de pessoas que consideram votar nesse candidato nas próximas eleições. Inicialmente eles poderiam definir uma função do tipo $\theta = \delta(X_1, ..., X_n)$ (onde$X_1, ..., X_n$ é uma amostra aleatória de X) e estabelecê-la como o estimador de θ, mas como selecionar essa função tal que a sua estimativa de θ seja a melhor em algum sentido? Se supomos que o verdadeiro valor de θ é θ_0 e a equipe decide que d é o valor que estima θ, então uma forma de determinar que tão apropriado é d para estimar θ é medir a distância entre θ_0 e d a partir de alguma função.

Esta abordagem faz parte da teoria da decisão onde, em geral, consideramos que o agente decisor (quem toma a decisão) deve escolher dentre um conjunto de decisões possíveis que têm consequências dependendo do estado da natureza (o que acontece na realidade), que é desconhecido. Definimos os seguinte conceitos:

- S : Espaço amostral ($w \in S$),
- X : Observação de uma variável aleatória $X(w)$,
- P : Modelo Probabilístico para X,
- Θ: Espaço de possíveis valores do parámetro θ desconhecido,
- D: Espaço de possíveis decisões,
- L: função perda de valor real.

Definindo o problema de decisão como (Θ, D, L), o problema de inferência estatística se reduz a selecionar o procedimento estatístico (às vezes chamado de função ou regra de decisão), que vai nos descrever a forma de tomar a decisão quando já temos a informação amostral.

Definição 2.2. *Uma função de decisão, procedimento estatístico ou regra de decisão é uma função* $\delta : S \to D$.

Definição 2.3. *Seja* D *um espaço arbitrário de decisões. A função de valor real não negativa* $L(\theta_0, \delta) : \Theta \times D \to R$ *é chamada de função perda.*

Então para cada par $(\theta_0, d) \in \Theta \times D$ consideramos uma sequência em termos da perda $L(\theta_0, d)$ que mede quanto perdemos ao decidir por d quando $\theta = \theta_0$. Em particular, se a decisão for correta deveríamos ter $L(\theta, d) = 0$.

Definição 2.4. *O valor esperado em $P_\theta(X)$ da função perda $L(\theta, d(X))$ é chamado de função risco para qualquer $d(X) = \delta \in D$ e é dado por*

$$r(\theta, \delta) \quad = \quad \int_\chi L(\theta, \delta) dP_\theta(X)$$

Especificamente no contexto Bayesiano temos:

Definição 2.5. *O risco de Bayes é definido para todas as $\delta \in D$ por*

$$R(\theta, \delta) = \int_\Theta r(\theta, \delta)\pi(\theta)d\theta \tag{2.5}$$

Onde $\pi(\theta)$ é a distribuição a priori de θ.

Esta função mede o desempenho da regra δ quando podemos repetir o experimento que gera a observação de X. A regra δ, que tem o menor risco de Bayes, é chamada de regra ou estimador de Bayes com respeito à priori e é denotada por δ^π.

Agora, consideramos a Regra de Bayes dadas três das funções perda mais comuns.

Função perda quadrática

Seja $L(\theta, \delta) = (\delta - \theta)^2$ a função perda quadrática. Para calcular o risco de Bayes para esta perda, seja

$$b \quad = \quad E_{\pi(\theta|\mathbf{x})}(\theta) = \int \theta\pi(\theta|\mathbf{x})d\theta,$$

a média da distribuição a posteriori $\pi(\theta|x)$, então:

$$\begin{aligned} E[L(\theta, \delta)] &= \int L(\theta, \delta)\pi(\theta|\mathbf{x})d\theta \\ &= \int (\delta - b + b - \theta)^2\pi(\theta|\mathbf{x})d\theta \\ &= (\delta - b)^2 \int (b - \theta)^2\pi(\theta|\mathbf{x})d\theta \\ &\geqslant \int (b - \theta)^2\pi(\theta|\mathbf{x})d\theta, \end{aligned}$$

para qualquer valor de δ. A desigualdade vira a igualdade quando $\delta = b$, pelo que a regra de Bayes sob uma função perda quadrática é a média da distribuição a posteriori.

Função perda erro absoluto

Seja $L(\theta, \delta) = |\delta - \theta|$ a função de perda absoluta. O risco é minimizado tomando δ como a mediana da distribuição a posteriori, digamos δ^*. Isto é, a mediana é a regra

(estimador) de Bayes quando a função perda é o modulo. Para mostrar isto suponhamos outra decisão, digamos d, tal que $d > \delta^*$. Então, dado que $(d + \delta^* - 2\theta) > (\delta^* - d)$ quando $\delta^* < \theta < d$, então

$$|\theta - d| - |\theta - \delta^*| = \begin{cases} \delta^* - d & \theta \geqslant d \\ d + \delta^* - 2\theta & \delta^* < \theta < d \\ d - \delta^* & \theta \leqslant \delta^* \end{cases}$$

Dado que $(d + \delta^* - 2\theta) > (\delta^* - d)$ quando $\delta^* < \theta < d$, então

$$\begin{aligned} E(|\theta - d| - |\theta - \delta^*|) &\geqslant (\delta^* - d)P(\theta \geqslant d) + (\delta^* - d)P(\delta^* < \theta < d) \\ &\quad + (d - \delta^*)P(\theta \leqslant \delta^*) \\ &= (d - \delta^*)\left[P(\theta \leqslant \delta^*) - P(\theta > \delta^*)\right] \geqslant 0 \end{aligned}$$

A última desigualdade se dá porque δ^* é mediana da distribuição de θ. Portanto $E(|\theta - d| \geqslant E(|\theta - \delta^*|)$ e a igualdade se cumpre se, e somente se, d é também mediana. De forma análoga, pode-se provar para $d < \delta^*$.

Função perda escalonada

Seja

$$L(\theta, \delta) = \begin{cases} 0 & |\delta - \theta| \leqslant \epsilon \\ 1 & |\delta - \theta| > \epsilon \end{cases}$$

onde ϵ é um número positivo predeterminado, usualmente pequeno. Para minimizar o risco de Bayes, é necessário maximizar $\pi(\delta|x)$ com relação a δ e o estimador de Bayes será então:

$$\begin{aligned} E[L(\theta, \delta)] &= \int I(|\delta - \theta| > \epsilon)\pi(\theta|\mathbf{x})d\theta \\ &= \int I(1 - |\delta - \theta| \leqslant \epsilon)\pi(\theta|\mathbf{x})d\theta \\ &= 1 - \int_{\delta-\epsilon}^{\delta+\epsilon} \pi(\theta|\mathbf{x})d\theta \\ &\geqslant 1 - 2\epsilon\pi(\delta|\mathbf{x}) \end{aligned}$$

Isto implica que neste caso, o estimador é a moda da distribuição a posteriori, devido que a igualdade vale para esse valor.

2.2 Teste de hipótese

De forma geral, podemos dizer que uma hipótese é uma formulação provisória, com intenções de ser posteriormente demostrada ou verificada (testada). Nosso contexto

estatístico de inferência, as hipóteses vão ser afirmações ou conjeturas sobre o parâmetro ou a distribuição. Antes de definir o teste de hipótese como um tipo de problema de decisão, relembremos a notação.

- θ: Parâmetro
- Θ: Espaço paramétrico ou conjunto de possíveis valores para θ.
- D: Espaço de possíveis decisões (rejeitar (d_1) ou não a hipótese nula (d_0)).
- d: Um elemento do espaço D.
- $L(a, d)$: Função perda.

Suponha que temos uma partição de Θ em $\Theta = \Theta_H \cup \Theta_A$, onde $\Theta_H \cap \Theta_A = \varnothing$. A afirmação $\theta \in \Theta_H$ é uma *hipótese* que vamos chamar de *hipótese nula* e vai ser denotada por H. A correspondente *alternativa* (hipótese alternativa), denotada por A, é a afirmação $\theta \in \Theta_A$. O problema de decisão $H : \theta \in \Theta_H$ e $A : \theta \in \Theta_A$ é chamado *teste de hipótese* se $D = \{d_0, d_1\}$ e $L(t, d)$ satisfaz que $L(t, 1) > L(t, 0)$ para $\theta \in \Theta_H$ e $L(t, 1) < L(t, 0)$ para $\theta \in \Theta_A$. Se rejeitamos H, mas H é verdadeira, cometemos um *erro de tipo I*. Se aceitamos H quando esta é falsa, então cometemos um *erro de tipo II*.

Uma forma simples de função perda para testar hipótese é

$$L(\theta, d_0) = \begin{cases} 0 & \theta \in \Theta_H \\ 1 & \theta \in \Theta_A \end{cases} \quad \text{e} \quad L(\theta, d_1) = \begin{cases} 0 & \theta \in \Theta_A \\ 1 & \theta \in \Theta_H \end{cases}$$

Esta função é chamada de função perda 0-1 que pode ser generalizada

$$L(\theta, d_0) = \begin{cases} 0 & \theta \in \Theta_H \\ c_2 & \theta \in \Theta_A \end{cases} \quad \text{e} \quad L(\theta, d_1) = \begin{cases} 0 & \theta \in \Theta_A \\ c_1 & \theta \in \Theta_H \end{cases}$$

onde c_1 é o custo do erro tipo I e c_2 o custo do erro tipo II; $D = d_0, d_1$ é o espaço de decisões onde d_0: Não Rejeitar H e d_1:Rejeitar H. Então, para a regra de decisão $\delta : S \to D$, temos

$$\begin{aligned} &\{\mathbf{x} : \delta(\mathbf{x}) = d_0\} \quad \text{Região de Aceitação (Não Rejeição)} \\ &\{\mathbf{x} : \delta(\mathbf{x}) = d_1\} \quad \text{Região de Rejeição} \end{aligned}$$

A função risco é da forma

$$R(\theta, \delta) \quad = \quad \begin{cases} 0 \times P(\delta(x) = d_0) + c_1 P(\delta(x) = d_1) & \theta \in \Theta_H \\ c_2 P(\delta(x) = d_0) + 0 \times P(\delta(x) = d_1) & \theta \in \Theta_A \end{cases}$$

A solução Bayesiana para um problema de teste de hipótese, com a função perda 0-1 generalizada, é estritamente teórica ([Schervish, 2012]). O risco posterior para a escolha da decisão $d = d_1$ é $c_2 P(\theta \in \Theta_H | X = x)$ e o risco posterior de escolher a decisão $d = d_0$ é $c_1 P(\theta \in \Theta_A | X = x)$. A decisão ótima é escolher $d = d_1$ se

$$c_2 P(\theta \in \Theta_H | X = x) \quad < \quad c_1 P(\theta \in \Theta_A | X = x)$$

que é equivalente a

$$P(\theta \in \Theta_H | X = x) \quad < \quad \frac{c_1}{c_1 + c_2}$$

Portanto, a solução Bayesiana é rejeitar a hipótese H se sua probabilidade a posteriori é menor do que $c_1/(c_1 + c_2)$. Teoricamente, isso é tudo o que precisamos para testar uma hipótese Bayesiana com função perda 0-1 generalizada, mas na prática, calcular essa probabilidade muitas vezes não é tão trivial, pelo que resulta interessante achar alternativas.

2.2.1 Full Bayesian significance test

O *Teste de Significância Genuinamente Bayesiano* (FBST - Full Bayesian Significance Test) é baseado no cálculo de uma quantidade denominada evidência a favor da hipótese (e-valor). De fato, é uma alternativa genuinamente Bayesiana ao p-valor, que é a probabilidade de achar valores mais extremos da estatística do teste $T(x)$, assumindo que a hipótese nula é verdadeira. Esta seção esta baseada principalmente no artigo de [de Bragança Pereira and Stern, 1999].

Considere uma variável aleatória X cuja observação é representada por x. O espaço estatístico é representado pela tripla (S, X, Θ) definida como na Seção 2.1. Definamos também o modelo a priori (Θ, B, π), que é um espaço de probabilidade definido sobre Θ onde B é o conjunto de valores observaveis para θ e π é um modelo probabilistico associado. Claramente, depois de observar x, podemos obter o modelo de probabilidade a posteriori (Θ, B, π_x), onde π_x é a medida de probabilidade condicional sobre B dado o ponto amostral observado x.

Para definições posteriores, vamos nos concentrar apenas no espaço de probabilidade a posteriori (Θ, B, π_x), e seja $f(\theta|x)$ a função de densidade de probabilidade associada a π_x. Definamos inicialmente T_φ como o subconjunto do espaço paramétrico, onde a densidade a posteriori é maior que uma quantidade φ

$$T_\varphi \quad = \quad \{\theta \in \Theta | f(\theta) > \varphi\}.$$

A credibilidade de T_φ é a sua probabilidade a posteriori dada por:

$$\kappa \quad = \quad \int_{T_\varphi} f(\theta|x) d\theta = \int_\Theta f_\varphi(\theta|x) d\theta,$$

onde $f_\varphi(x) = f(x)$ se $f(x) > \varphi$ e zero em outro caso. Agora definimos f^* como o máximo da função de densidade a posteriori sob a hipótese nula, atingido no argumento θ^*

$$\theta^* \in \arg\max_{\theta \in \Theta_H} f(\theta), \qquad f^* = f(\theta^*),$$

e definimos $T^* = T_{f^*}$ como o conjunto tangente à hipótese nula, H, cuja credibilidade é κ^*. Então, a medida de evidência proposta por ([de Bragança Pereira and Stern, 1999]) é o complemento da probabilidade do conjunto T^*, isto é

$$Ev(H) = 1 - \kappa^* \quad \text{ou} \quad 1 - \pi_x(T^*)$$

Se a probabilidade do conjunto $T*$ é grande, significa que o conjunto nulo se acha na região de baixa probabilidade e a evidência nos dados é contraria à hipótese nula. Por outro lado, se a probabilidade de $T*$ é pequena, então o conjunto nulo está na região de maior probabilidade e a evidência nos dados está a favor da hipótese nula.

Este procedimento foi criado com o objetivo de testar hipóteses precisas, isto é, uma hipótese nula cuja dimensão é menor do que o espaço paramétrico $dim(\Theta_H) < dim(\Theta_A)$.

2.3 Inferência preditiva

Na prática, muitas vezes estamos interessados em fazer inferências sobre observações futuras de uma variável aleatória, cuja distribuição depende de um número finito de parâmetros (desconhecidos). Esta distribuição é chamada *distribuição preditiva* e segundo [Smith, 1998], às vezes fazer afirmações preditivas sobre variáveis aleatórias não observadas, faz mais sentido do que a estimação tradicional de parâmetros.

Supondo que $\pi(\theta)$ é a distribuição a priori e que $\pi(\theta|x)$ é a distribuição a posteriori de θ, seja z qualquer observação futura da variável aleatória X. Então definimos a distribuição preditiva Bayesiana para z como sendo:

$$\begin{aligned}
p(z|\mathbf{x}) &= \frac{p(z,\mathbf{x})}{p(\mathbf{x})} \\
&= \frac{\int_\Theta p(z,\mathbf{x},\theta)d\theta}{\int_\Theta p(\mathbf{x},\theta)d\theta} \\
&= \frac{\int_\Theta p(z,\mathbf{x}|\theta)\pi(\theta)d\theta}{\int_\Theta p(\mathbf{x}|\theta)\pi(\theta)d\theta} \\
&= \frac{\int_\Theta p(z|\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int_\Theta p(\mathbf{x}|\theta)\pi(\theta)d\theta} \\
&= \int_\Theta p(z|\theta)\left\{\frac{p(\mathbf{x}|\theta)\pi(\theta)}{\int_\Theta p(\mathbf{x}|\theta)\pi(\theta)d\theta}\right\}d\theta \\
&= \int_\Theta p(z|\theta)\pi(\theta|\mathbf{x})d\theta
\end{aligned}$$

que podemos escrever como

$$p(z|\mathbf{x}) = \int_{\Theta} p(z|\theta)\pi(\theta|\mathbf{x})d\theta = E_{\theta|\mathbf{x}}\left[p(z|\theta)\right] \tag{2.6}$$

onde a função $p(z|\theta)$ é a função densidade (ou massa no caso de uma v.a discreta) sendo que $p(x|\theta)$ avaliada em z

Exemplo 2.3. *Suponha que $x_1, ..., x_n$ é uma amostra aleatória de uma Bernoulli(θ) e suponha também que a distribuição a priori para θ é uma Beta(α, β) como no Exemplo 2.2. Encontremos a distribuição preditiva para uma observação futura z que representa uma pessoa que responde à pergunta (considera votar candidato A?). Temos*

$$p(z|\mathbf{x}) = \int_{\Theta} p(z|\theta)\pi(\theta|\mathbf{x})d\theta$$

Agora

$$p(z|\theta) = \theta^z(1-\theta)^{1-z}, \quad z = 0, 1$$

e

$$\pi(\theta|\mathbf{x}) \propto \theta^{\sum x_i+\alpha-1}(1-\theta)^{n-\sum x_i+\beta-1}$$

Então, se denotamos $\alpha* = \sum x_i + \alpha$ e $\beta* = n - \sum x_i + \beta$ temos que

$$\begin{aligned} p(z|\mathbf{x}) &= \int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha*)\Gamma(\beta*)}\theta^{z+\alpha*-1}(1-\theta)^{\beta*+1-z-1}d\theta \\ &= \frac{\Gamma(n+\alpha+\beta)\Gamma(z+\alpha*)\Gamma(1-z+\beta*)}{\Gamma(\alpha*)\Gamma(\beta*)\Gamma(n+\alpha+\beta+1)} \end{aligned}$$

Então a probabilidade de que uma pessoa responda não é

$$\begin{aligned} P(z=0|\mathbf{x}) &= \frac{\Gamma(n+\alpha+\beta)\Gamma(\alpha*)\Gamma(1+\beta*)}{\Gamma(\alpha*)\Gamma(\beta*)\Gamma(n+\alpha+\beta+1)} \\ &= \frac{\beta*}{\alpha+\beta+1} \\ &= \frac{\beta*}{\alpha*+\beta*} \end{aligned}$$

E a probabilidade de que sua resposta seja sim é

$$\begin{aligned} P(z=1|\mathbf{x}) &= \frac{\Gamma(n+\alpha+\beta)\Gamma(1+\alpha*)\Gamma(\beta*)}{\Gamma(\alpha*)\Gamma(\beta*)\Gamma(n+\alpha+\beta+1)} \\ &= \frac{\alpha*}{\alpha*+\beta*} \end{aligned}$$

Note que esta última representa a média posterior de θ.

3 Metodologia de estimação de cópulas

No Capítulo 1 descrevemos tipos de cópulas e demos como exemplo algumas das famílias mais conhecidas. Estas famílias (parametrizadas) vêm indexadas por um parâmetro que geralmente caracteriza a dependência das variáveis representadas por uma cópula particular, pelo que o interesse do investigador se reduz a procurar o valor do parâmetro que melhor descreve a relação entre as variáveis estudadas. Esse processo é conhecido como estimação, e em estatística temos diferentes métodos para resolver esta situação. Em particular, neste trabalho vamos nos focar na estimação Bayesiana que foi introduzida no Capítulo 2.

Neste Capítulo vamos apresentar uma metodologia que permita encontrar a cópula que melhor se ajusta a um conjunto de dados. Como primeiro passo, o pesquisador deve determinar uma ou mais famílias de cópulas, que segundo seu critério, seriam apropriadas para refletir a relação entre as variáveis de interesse. Dentro de cada uma delas, deve selecionar um representante (a partir da estimação de um ou vários parâmetros da família) que vai ser aquele que melhor (dentro da família) reflete a relação observada nos dados. Finalmente deve decidir qual desses representantes é quem melhor descreve aos dados, isto baseado em algum ou alguns critérios de seleção (qualidade de ajuste).

3.1 Como selecionar cópulas?

Conforme já destacado, as cópulas são funções que cumprem com as caraterísticas da Definição 1.4, pelo que contamos com praticamente infinitas famílias de funções para escolher. Isto implicaria um extenso trabalho de ajuste devido a que teríamos que estimar o parâmetro θ para cada família (e se consideramos famílias multiparametricas este processo seria muito mais trabalhoso), pelo que se torna importante saber como selecionar um número reduzido de famílias entre as quais esperamos achar a cópula que melhor descreve nossos dados.

Para solucionar este problema, vai ser muito importante relembrar a Seção 1.7 e os tipos de cópulas segundo o tipo de dependência que refletem. Na verdade, se espera que quem trabalha com cópulas tenha conhecimento das propriedades que caracterizam as diferentes famílias existentes de tal forma que possa selecionar aquelas que refletem a relação, a priori, que ele supõe entre as variáveis. Também podemos construir cópulas usando os métodos que descrevemos na Seção 1.9, se as famílias existentes não satisfazem nossos interesses ou se queremos ampliar nossas opções.

Em geral, para este passo o pesquisador precisa de um conhecimento razoável

sobre as variáveis que quer analisar e estabelecer o tipo de relação de dependência que precisa estudar.

3.2 Como determinar o representante dentro de uma família de cópulas?

Este passo consiste em estimar o parâmetro de cada família de cópulas selecionada no passo anterior. Para isso utilizamos a inferência Bayesiana, pelo que precisamos definir a função de verossimilhança dos dados, uma distribuição a priori e uma função de perda.

Para ilustrar melhor os procedimentos, vamos usar uma cópula da família Sarmanov Generalizada. Seja $C(u, v)$ a cópula da familia Sarmanov da forma:

$$C(u, v) = uv[1 + \theta(u^{p-1} + v^{p-1} - u^{p-1}v^{p-1} - 1)] \tag{3.1}$$

com $(u, v) \in I^2$ e $\theta \in [-1/(p-1), 1/(p-1)^2]$ para $p > 2$ (veja Exemplo 2.2 de [Bairamov et al., 2011]). Esta familia na verdade pertence à família FGM generalizada descrita por [Úbeda-Flores et al., 2004].

Função de verossimilhança.

Sejam $X_1, ..., X_n$ e $Y_1, ..., Y_n$ amostras de duas variáveis aleatórias com funções de distribuição $F(x)$ e $G(y)$ respetivamente e seja $H(x, y)$ a função de distribuição conjunta de X e Y. Então, considerando a cópula da expressão (3.1) com $p = 3$ temos, usando o Teorema de Sklar, temos que a distribuição conjunta $H(u, v)$ pode ser escrita,

$$H(u, v) \quad = \quad uv[1 + \theta(u^2 + v^2 - u^2v^2 - 1)],$$

com $u = F(x)$ e $v = G(y)$. Assim, a densidade conjunta de X e Y seria a segunda derivada parcial da cópula da forma,

$$h(u, v) = \frac{\partial}{\partial u \partial v} C(u, v) = 1 + 3\theta[u^2 + v^2 - 3u^2v^2 - 1]. \tag{3.2}$$

Agora sejam $x_1, ..., x_n$ e $y_1, ..., y_n$ observações das variáveis X e Y, então escrevemos a verossimilhança $h(u_1, ...u_n, v_1, ..., v_n|\theta)$ como

$$h(\mathbf{u}, \mathbf{v}) = \prod_{i=1}^{n} 1 + 3\theta[u_i^2 + v_j^2 - 3u_i^2v_j^2 - 1], \tag{3.3}$$

onde $u_i = F(x_i)$ e $v_i = G(y_i)$.

Distribuição a priori.

Uma das grandes dificuldades da análise Bayesiana é a identificação, seleção e justificação da distribuição a priori para os parâmetros. Muitas vezes, a informação que temos sobre nosso parâmetro de interesse é mínima ou quase nula, pelo que devemos escolher uma distribuição que reflete nosso desconhecimento. Neste caso, conhecemos o domínio de θ pela condição $\theta \in [-0.5, 0.25]$ (substituindo $p = 3$ nas condições do Exemplo 2.2 de [Bairamov et al., 2011]), assim, podemos sugerir uma distribuição uniforme nesse intervalo que daria mesmo peso a todos os valores.

Então a densidade a priori para θ é:

$$\pi(\theta) = \begin{cases} \frac{4}{3}, & \text{se } \theta \in [-0.5, 0.25], \\ 0 & \text{caso contrario.} \end{cases}$$

Distribuição a posteriori.

Pelo Teorema de bayes, sabemos que $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$. Nesse caso:

$$\begin{aligned} \pi(\theta|\mathbf{u}, \mathbf{v}) &\propto h(u, v)\pi(\theta) \\ &\propto \prod_{i=1}^{n} [1 + 3\theta(u_i^2 + v_j^2 - 3u_i^2 v_j^2 - 1)\{4/3\}] \\ &\propto \prod_{i=1}^{n} [1 + 3\theta(u_i^2 + v_j^2 - 3u_i^2 v_j^2 - 1)] \end{aligned}$$

Em geral, para obter $\pi(\theta_j|u_i, v_i)$ neste contexto de cópulas, vamos considerar amostras aleatórias de tamanho n das variáveis X e Y, e uma cópula $C_\theta(u, v)$.

- Definimos uma distribuição a priori $\pi(\theta)$ para θ.
- Definimos a função densidade da cópula considerada $c_\theta(u, v)$.
- Calculamos $u_i = F(x_i)$ e $v_i = G(y_i)$.
- Consideramos m valores de θ_j.
- Calculamos $\pi(\theta_j|u_i, v_i) \propto \prod_{i=1}^{n} c_{\theta_j}(u_i, v_i)\pi(\theta_j)$.

Uma vez que temos a expressão da $\pi(\theta|U, V)$ podemos, dada uma função perda, calcular a regra de Bayes (estimador Bayesiano) para θ. Se, por exemplo, a nossa $L(\theta, d)$ fosse a perda escalonada, então o estimador $\hat{\theta}$ de θ vem dado por:

$$\hat{\theta} = \theta \in \overset{\arg\max}{[-0.5, 0.25]} \left\{ \frac{\prod_{i=1}^{n} 1 + 3\theta[u_i^2 + v_j^2 - 3u_i^2 v_j^2 - 1]\{4/3\}}{\int_{-0.5}^{0.25} \prod_{i=1}^{n} 1 + 3\theta[u_i^2 + v_j^2 - 3u_i^2 v_j^2 - 1]\{4/3\} d\theta} \right\}. \tag{3.4}$$

Neste caso, e provavelmente em muitos outros no contexto das cópulas, não é possível maximizar a densidade a posteriori de forma analítica mas, de novo, é possível fazer uso de ferramentas computacionais.

Temos uma expressão explícita da densidade a posteriori pois, dado que $x_1, ..., x_n$ e $y_1, ..., y_n$ são observações das variáveis X e Y com funções de distribuição F e G respectivamente, os valores u_i e v_j são conhecido, porem nada em $\pi(\theta|u, v)$ é desconhecido. Uma forma de achar esse máximo é avaliar essa $\pi(\theta|X, Y)$ para diferentes valores de θ no intervalo $[-0.5, 0.25]$ e observar seu gráfico.

Exemplo 3.1. *Por exemplo, suponha que temos $x_1, ..., x_{1000}$ e $y_1, ..., y_{1000}$ observações de duas variáveis aleatórias X e Y. Seja também $S = (F(x), G(y)) = (u, v)$ um vetor bivariado simulado de uma cópula de Sarmanov da forma:*

$$C(u, v) \quad = \quad uv[1 + \theta(u^{p-1} + v^{p-1} - u^{p-1}v^{p-1} - 1)],$$

com $\theta = -0.3$. Se supomos que θ é desconhecido e tentamos estimá-lo a partir da equação 3.4, podemos calcular de forma explícita a expressão $\pi(\theta|X, Y) = \pi(\theta|S)$ para, por exemplo, 100 valores de θ dentro do intervalo $[-0.5, 0.25]$, para ter uma ideia de onde está o máximo. A Figura 8 exibe o gráfico de θ e $\pi(\theta|S)$, onde podemos observar que o máximo é atingido praticamente em $\theta = -0.3$, que é o verdadeiro valor de θ.

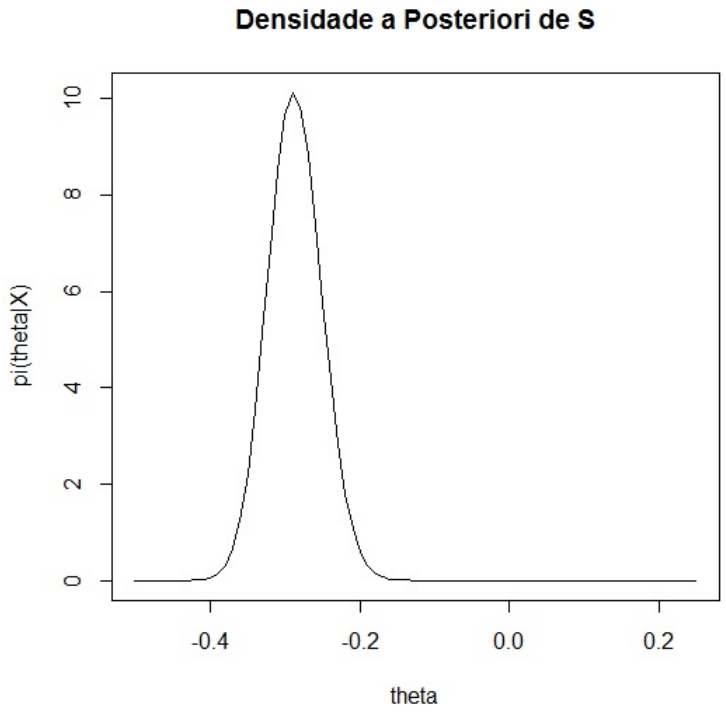

Figura 8 – Densidade a posteriori de θ.

Obviamente, esta aproximação apenas pode nos dar uma ideia da localizaçã do máximo (nos casos onde θ é univariado ou bivariado), que vai melhorar conforme aumentamos o número de pontos onde avaliamos, mas é um bom começo. Para resultados mais exatos é preciso implementar algum algoritmo de otimização.

Desta forma, para cada família de cópulas selecionada, o pesquisador acha o melhor representante de cada uma delas, segundo determinada função perda.

3.3 Seleção de modelos

Uma vez que contamos com um representante de cada família de cópulas que estamos considerando, desejamos estabelecer critérios a partir dos quais selecionar dentre eles a cópula que melhor descreve nossas variáveis objeto de estudo. Nesta seção apresentaremos alguns métodos de seleção que serão de utilidade para a aplicação do Capítulo 4.

3.3.1 Uso da cópula empírica.

Na seção 1.7, descrevemos a definição e as expressões a partir das quais é construída a cópula empírica. Uma forma de determinar qual das nossas cópulas possíveis é a mais adequada é minimizar uma distância como faz [Romano, 2002]. Neste trabalho, vamos usar duas distâncias: a distância de Hellinger e a distância de Kolmogorov. Segue a definição de cada uma delas.

Definição 3.1 (Distância de Hellinger)**.** *Para distribuições (discretas) de probabilidade* $P = p_{i\{i\in[n]\}}$ *e* $Q = q_{i\{i\in\mathbb{N}\}}$ *com suporte em* $\mathbb{N}$*, a distância de Hellinger entre elas se define como sendo*

$$H(P,Q) \quad = \quad \sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2 \tag{3.5}$$

Por definição, a distância de Hellinger é uma medida que satisfaz a desigualdade triangular e tal que $0 \leqslant H(P,Q) \leqslant 2^{1/2}$*. [Bar-Yossef et al., 2002].*

Definição 3.2 (Distância de Kolmogorov)**.** *Esta distância foi proposta por [Kolmogorov, 1950] e o teste baseado nela é um dos mais populares na hora de avaliar bondade de ajuste. Quando definimos como hipótese nula que os dados tem uma função distribuição acumulada dada, digamos* $H(x,y)$*, podemos testar a hipótese calculando a seguinte distância*

$$D = \sup_{(x,y)} |H_D - H(x,y)|, \tag{3.6}$$

onde H_D *representa a função de distribuição acumulada empírica dos dados observados [Arnold and Emerson, 2011].*

Para o problema de seleção de cópulas, vamos escolher a cópula cujo ajuste minimize as distância.

3.3.2 Métodos gráficos

Essas distancias vão fornecer um critério numérico, porem muitas vezes é interessante contar com métodos gráficos. Se esse é o nosso desejo, então poderíamos usar as curvas de nível da cópula ajustada e graficá-la conjuntamente com os pontos que representam as observações das variáveis analisadas.

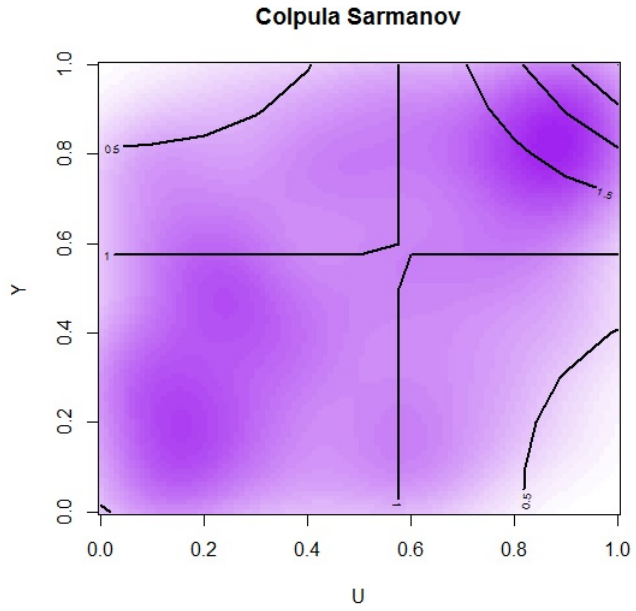

Figura 9 – Gráfico de contorno de densidade Sarmanov vs. a imagem da densidade empírica de uma amostra Sarmanov com $\theta = -0.3$.

O gráfico da Figura 9 ajuda-nos a verificar a coerência dos resultados numéricos, isto é, algumas vezes a cópula que minimiza a distância que consideramos não é coerente com a densidade observada nos dados, e por isso é importante ter uma ferramenta gráfica que suporte nossos resultados analíticos.

Também poderíamos pensar que o objetivo de ajustar uma cópula a nossos dados é, por exemplo, estimar uma probabilidade condicional, pelo que a cópula mais adequada para nosso problema é aquela que estima melhor tal probabilidade. Usando de novo a cópula empírica, poderíamos calcular essa probabilidade nos dados observados e em cada uma das cópulas que estamos considerando. Vamos escolher aquela cuja distribuição seja mais coerente com o valor observado nos dados. Para comparar estas distribuições podemos construir um gráfico como na Figura 10 onde temos um boxplot para cada cópula.

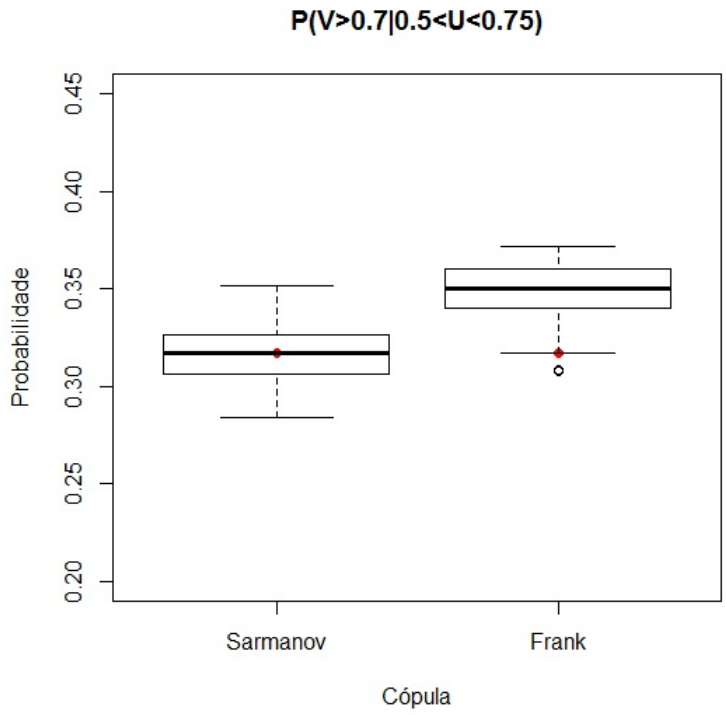

Figura 10 – Gráfico de boxplot da probabilidade condicional $P(V \geqslant 0.7|0.5 \leqslant U < 0.75)$ calculada em 100 amostras de tamanho 1000 de cada cópula, Sarmanov e Frank.

Nesse caso, por exemplo, a cópula melhor comportada com respeito à probabilidade condicional que estamos considerando é a cópula de Sarmanov, isto porque o valor observado nos dados se encontram dentro da caixa e muito perto da mediana da mesma, pelo que é coerente pensar que dito valor pertence a essa distribuição.

3.3.3 Seleção a partir da distribuição preditiva

Outra forma de selecionar a melhor cópula é fazer as análises propostas anteriormente, usando uma porcentagem dos dados (por exemplo o 80%) e usar o restante 20% e cada cópula em estudo para construir uma distribuição preditiva para a densidade desses dados, como foi mostrado na Seção 2.3, ou para uma função dos mesmos. Assim podemos verificar com que probabilidade cada modelo considerado prediz algum comportamento de interesse.

A partir da preditiva também, podemos avaliar as distâncias propostas na Seção 3.3.1 e em geral, fazer a mesma análise descrita anteriormente, mas dessa vez aplicada a aquele 20% que não foi considerado para a estimação.

4 Aplicação

Usando a teoria descrita nos Capítulos 1 e 2, e a metodologia proposta no Capítulo 3, vamos analisar a relação entre os resultados dos aspirantes no Vestibular da UNICAMP, e o seu desempenho acadêmico no primeiro semestre após serem admitidos nos cursos pertencentes às áreas de Exatas e Engenharia. Para medir esse desempenho, vamos usar as variáveis coeficiente de rendimento (CR) e a nota final numa das disciplinas mais básicas dos cursos pertencentes às áreas de interesse, Cálculo I (MA111). Inicilamente, descrevemos os dados e fazemos uma análise exploratória com o propósito de descrever mais um pouco cada uma das variáveis que vamos utilizar na nossa análise. Seguiremos com a seleção de um subconjunto de cópulas a serem ajustadas de onde escolheremos, segundo certos critérios, a melhor. Finalmente analisaremos os resultados obtidos no contexto do problema.

4.1 O Vestibular da UNICAMP

O processo de seletivo para o ingresso na UNICAMP, consiste de uma série de provas conhecidas como o Vestibular, que consta de duas fases. A partir do desenvolvimento do aspirante nestas duas fases, lhe é atribuída uma pontuação que vai determinar se é admitido ou não no curso ao qual pretende ingressar. A UNICAMP permite a cada estudante escolher dois cursos para inscrição, indicando uma preferência.

Até 2014, a primeira fase do Vestibular consistia de uma única prova dividida em duas partes,

- Redação: composta de duas propostas de textos a serem desenvolvidos pelos candidatos.
- Conhecimentos Gerais: composta de várias questões de múltipla escolha sobre as áreas do conhecimento desenvolvidas no ensino médio, isto é, matemática, biologia, química, física, história, geografia, língua portuguesa, literatura portuguesa e inglês.

O candidato precisa ser aprovado nesta primeira fase para continuar no processo seletivo. A segunda fase consiste de cinco exames específicos relativos às áreas de conhecimento, ciências humanas, ciências da natureza e matemáticas, junto com exames das línguas portuguesa e inglesa.

Quando as provas são corrigidas, elas recebem uma nota bruta. Essas notas variam entre 0 e 96 em cada prova. Porém, para calcular a classificação das duas fases e a classificação final dos candidatos não são utilizadas as notas brutas, mas sim as notas padronizadas. A padronização consiste em uma mudança de escala baseada na média e no desvio padrão de cada prova. A padronização evita que uma prova muito difícil num determinado ano faça diferença no desempenho dos candidatos daquele ano. Este processo de padronização, ocorre tanto na primeira quanto na segunda, fase e atribui 500 pontos à nota média de cada prova e 100 pontos para cada desvio padrão. A formula geral da nota padronizada para cada questão é a seguinte,

$$NP = \frac{(N - M) \times 100}{DP} + 500$$

onde N representa a nota bruta do estudante na prova, M representa a média geral de todos os candidatos na questão e DP o desvio padrão correspondente.

A classificação da primeira para a segunda fase se faz por cursos. São convocados os candidatos que optaram pelo curso em questão em primeira opção e que obtiveram nota igual ou superior à nota mínima exigida pelo curso na prova da primeira fase (esta nota pode variar de 450 a 600 pontos). O número de convocados será no mínimo três vezes e no máximo de seis vezes o número de vagas de cada curso.

Cada curso tem até duas provas consideradas prioritárias. Para cada prova prioritária é atribuído um peso, que é utilizado no cálculo da Nota Padronizada por Opção (NPO) e a Nota Mínima de Opção (NMO), que são utilizadas entre os critérios de classificação e convocação dos candidatos em cada opção.

Finalmente, é calculada uma pontuação geral NPO, que é a média ponderada das notas padronizadas dos candidatos nas provas, para cada um dos cursos aos que o estudante está aspirando. A nota final do estudante para cada uma das suas opções, vai depender da ponderação das provas consideradas prioritárias (que é diferente em cada curso). Aquelas provas que não são prioritárias em cada curso têm a seguinte ponderação,

- Peso 0,5 (meio): para a prova de língua inglesa;
- Peso 1 (um): para as provas língua portuguesa e matemática;
- Peso 2 (dois): para as provas da primeira fase, ciências humanas e artes, ciências da natureza e a prova de habilidades específicas (se houver).

Em cada curso, serão convocados por ordem decrescente de NPO os candidatos que optaram pelo curso em primeira opção, e que obtiveram nota padronizada nas provas

prioritárias maior ou igual às NMO estabelecidas.

Desta descrição, temos que é muito importante para a Comissão Permanente para os Vestibulares (CONVEST), contar com uma ferramenta efetiva de classificação destinada a selecionar o pessoal melhor preparado. Com esse propósito, uma das questões de interesse que abordaremos, é determinar como se relaciona o resultado de um indivíduo em cada prova com o seu desempenho como estudante ativo da UNICAMP. Mais especificamente, analisaremos o rendimento geral e na disciplina de Cálculo I dos alunos das áreas de Exatas e Engenharia. O rendimento geral vai se medir a partir do Coeficiente de Rendimento (CR), que é o índice que mede o desempenho acadêmico do aluno ao longo de seu curso e é assim calculado,

$$CR = \frac{\sum_{i=1}^{n} N_i C_i}{10 \sum_{i=1}^{n} C_i} \tag{4.1}$$

onde N_i é a nota obtida na i-ésima disciplina dentre as n disciplinas cursadas e C_i é número de créditos correspondentes a i-ésima disciplina.

4.1.1 Análise exploratória de dados

Um primeiro passo na análise que pretendemos implementar é conhecer mais um pouco a natureza, e as caraterísticas das variáveis envolvidas no problema. Contamos com uma base de dados com 6073 registros que correspondem a todos os alunos que ingressaram nos cursos oferecidos nas áreas de Exatas e Engenharia entre os anos 2011 e 2014. Desses 6073, contamos com 4560 registos de notas em MA111-Cálculo I. Na Tabela 1 encontramos a descrição de todas as variáveis incluídas na base de dados mencionada.

Tabela 1 – Descrição das variáveis

Variável	**Descrição**
ANO_ING	Ano de ingresso
ENS	Tipo de ensino medio recebido (publico ou não)
AREA	Área do curso matriculado (Exatas ou Engenharia)
NPT	Nota padronizada para o curso matriculado
VF1	Nota Vestibular UNICAMP - 1a Fase
CH	Nota Vestibular UNICAMP - Ciências Humanas
CN	Nota Vestibular UNICAMP - Ciências da Natureza
PT	Nota Vestibular UNICAMP - Português
ING	Nota Vestibular UNICAMP - Inglês
MA	Nota Vestibular UNICAMP - Matemática
CR1	CR do ingressante no semestre 1
MA111	Nota obtida na disciplina Cálculo I

Comecemos descrevendo as variáveis do Vestibular e do rendimento no primeiro semestre para cada uma das variáveis categóricas, Ano e tipo de ensino medio (ENS).

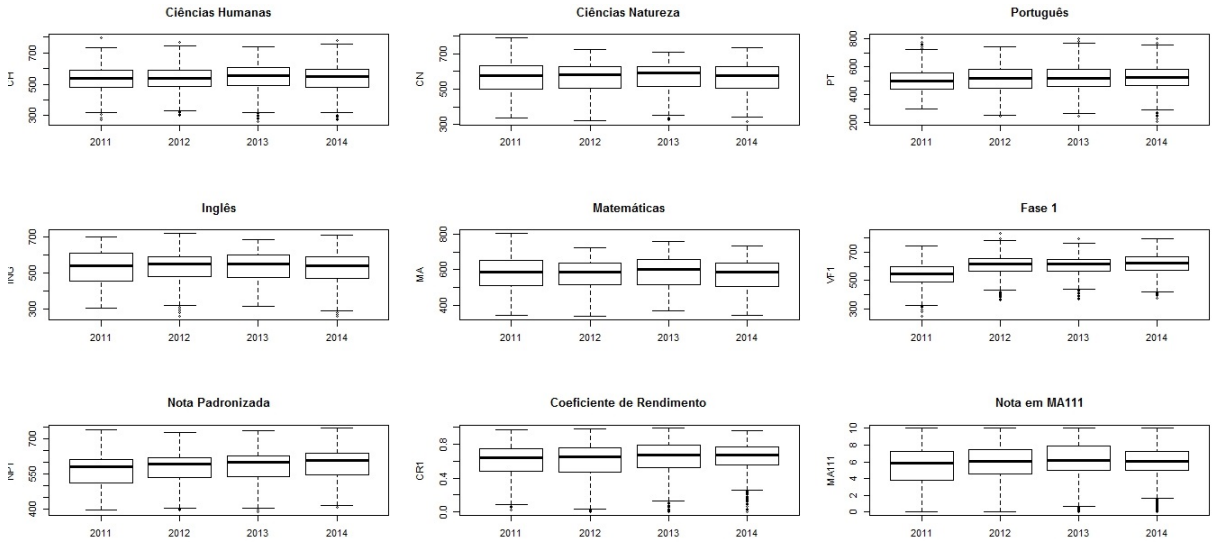

Figura 11 – Boxplots das notas do Vestibular, o coeficiente de rendimento e a nota em MA111 para cada um dos anos(2011, 2012, 2013, 2014).

Na Figura 11, encontramos os boxplot de todas as variáveis para cada ano da análise. Em geral, não se observam diferencias claramente significativas, embora se percebem algumas diferencias sobretudo na variabilidade. Então, inicialmente poderíamos considerar os dados sem discriminar pelo ano de ingresso do estudante. Isto sugere, que apesar das mudanças que existem no conteúdo das provas, estas conservam uma estrutura básica que é constante neste período de anos.

Uma história diferente pode-se observar na Figura 12, onde vemos que em geral para

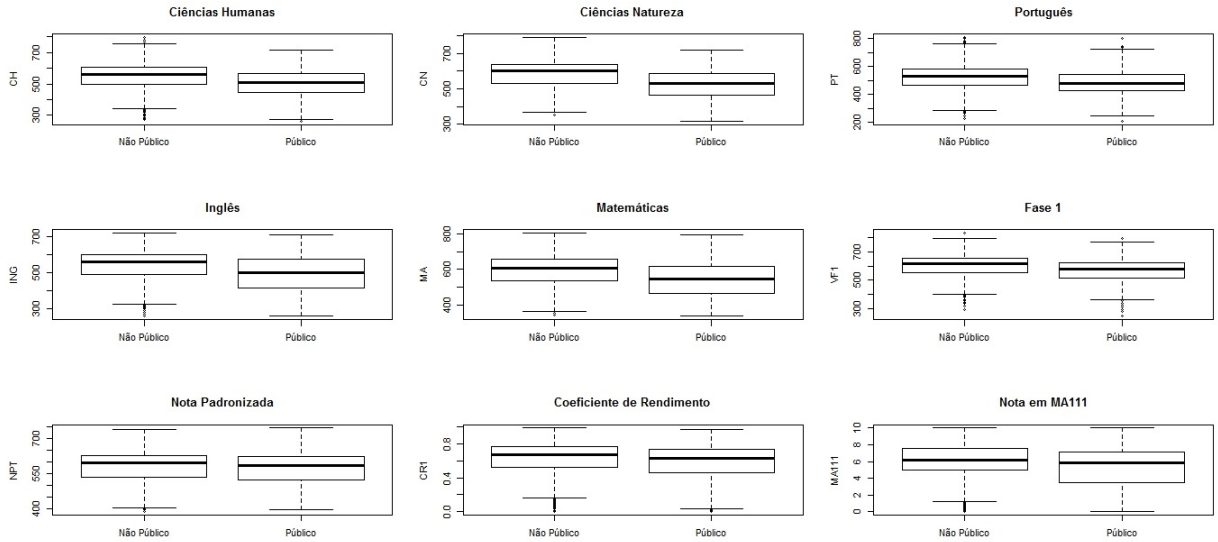

Figura 12 – Boxplots das notas do Vestibular, o coeficiente de rendimento e a nota em MA111 por tipo de ensino.

todas as variáveis, o pessoal que recebeu educação numa instituição de ensino não público

obteve melhores resultados que o pessoal proveniente do ensino público. Para nossa análise, no entanto, vamos desconsiderar esta diferença, mesmo cientes de que ela existe, dado que o Vestibular como ferramenta de medição não faz esta distinção.

Dado que nosso objetivo é analisar a relação entre pares de variáveis, formados pelas combinações de cada uma das provas do Vestibular e as nossas variáveis de interesse, isto é, o CR e a nota em MA111, uma forma de descrever esta relação é descrevendo a dependência probabilística entre essas variáveis, razão pela qual estamos considerando às cópulas como ferramenta de análise. A forma mais simples de dependência que conhecemos é a dependência linear, que pode ser medida a partir do coeficiente de correlação de Pearson, como visto no Capítulo 1.

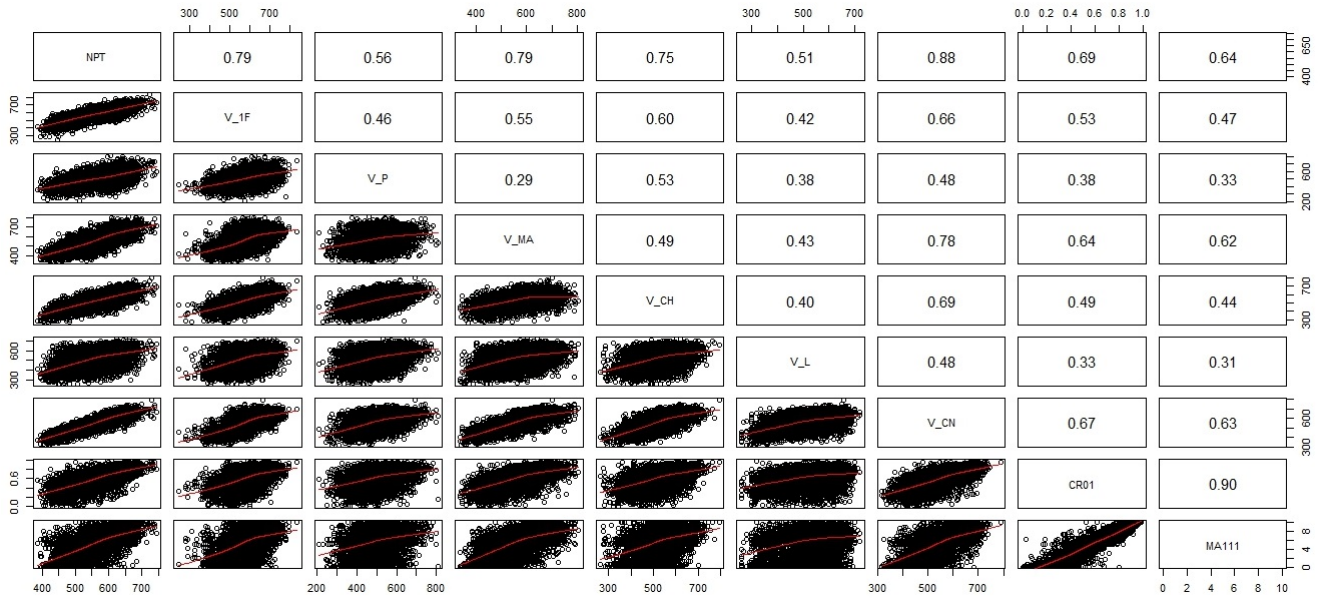

Figura 13 – Matriz de correlação das variáveis do Vestibular, o coeficiente de rendimento e a nota em Cálculo I.

Da Figura 13 podemos destacar que as variáveis que tem maior relação de dependência (linear) com o CR são CN, MA e NPT. Este resultado é esperado, dado que para a maioria dos cursos considerados, CN e MA são as áreas de conhecimento que se relacionam diretamente com as disciplinas oferecidas por estes. Também, podemos observar que CN é a variável que tem mais peso na NPT. Um cenário parecido, encontramos com respeito à variável MA111, onde de novo se destacam as variáveis CN, MA e NPT. Além disso, vemos que a correlação linear entre CR e MA111 é muito alta, refletindo a importância que esta disciplina tem como medida do desempenho acadêmico dos estudantes. Infelizmente, pela forma em que são suminitrados os dados, não podemos excluir a nota em MA111 do CR e dada a alta correlação entre eles, temos uma alta probabilidade de

que as análises sejam muito parecidas.

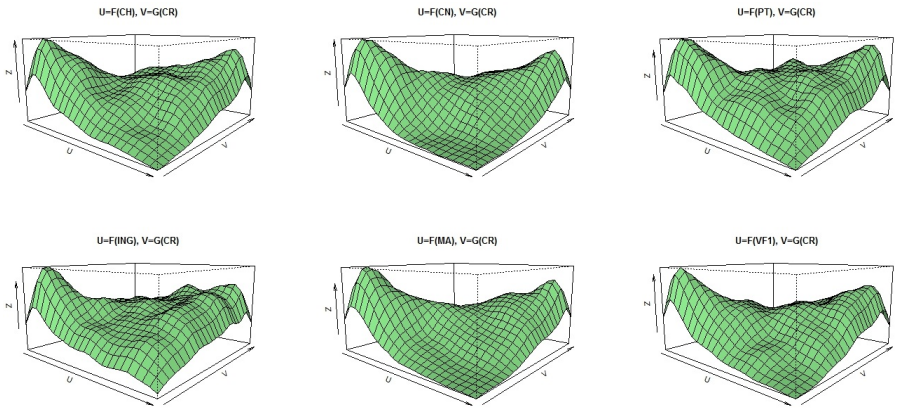

Figura 14 – Densidades bivariadas associadas aos pares (U, V) formados pelas provas do Vestibular e o CR.

Para a nossa metodologia de cópulas, precisamos dar uma ideia da distribuição bivariada associada a cada par de variáveis, fato que vai nos ajudar na hora de selecionar um conjunto de cópulas possíveis. Mais especificamente, gostariamos observar a densidade correspondente aos pares da forma (U, V), onde $U = F(X)$, $V = G(Y)$, X é alguma das variáveis do Vestibular, Y é o CR ou a nota de Cálculo I-MA111, e F e G são as funções de distribuição univariadas de X e Y respectivamente.

Das Figuras 14 e 15, podemos dizer que estas densidades apresentam dois máximos, um para valores altos de ambas variáveis e outro para valores pequenos de ambas variáveis. Isto, tanto para os casos com CR quanto para os casos com MA111, pelo que deveremos procurar que as cópulas selecionadas tenham esta caraterística.

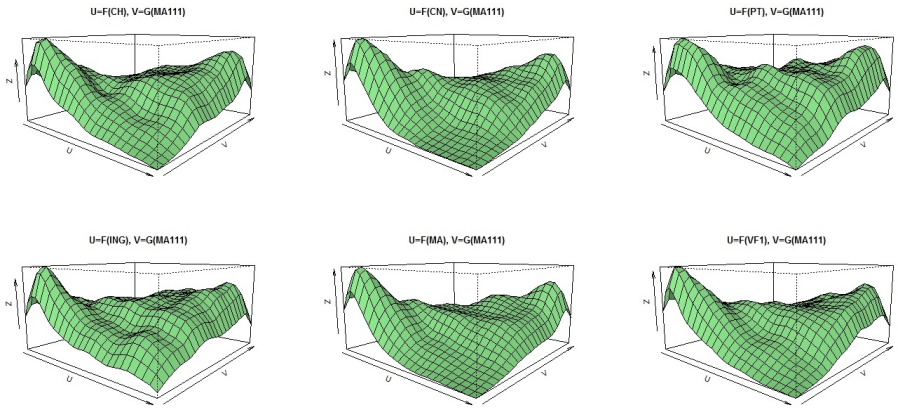

Figura 15 – Densidades bivariadas associadas aos pares (U, V) formados pelas provas do Vestibular e a nota de MA111.

4.2 Escolha de um conjunto de cópulas

Nesta seção, vamos discutir um pouco mais os resultados obtidos na seção anterior e procuraremos dentre as opções propostas no Capítulo 1, algumas funções que satisfaçam as nossas necessidades.

Da nossa análise exploratória, concluímos que uma boa cópula para os dados que estamos considerando, deveria ser capaz de modelar a presença de dois máximos junto com uma tendencia a que um deles seja maior (assimetria). Este tipo de comportamento, pode-se encontrar em funções bivariadas com seções polinomiais. Em particular, poderíamos considerar um polinômio de grau 3 e usar a teoria descrita na Seção 1.9.2.3, especificamente as cópulas descritas a partir das equações (1.22) e (1.23) correspondentes a duas cópulas de seção cúbica. Relembrando estas duas famílias de cópulas, consideramos a cópula C_{SC1} (cópula de seção cúbica em U e quadrática em V) como sendo,

$$C_{SC1}(u,v) \quad = \quad uv + uv(1-u)(1-v)\left[\alpha(1-u) + \beta u\right],$$

onde $-1 \leqslant \alpha, \beta \leqslant 1$. Esta cópula se caracteriza por ser assimétrica. Com respeito à dependência, temos que quando $\alpha \geqslant \beta$ a relação entre as variáveis é positiva, caso contrário dita relação é negativa. De forma mais geral, podemos considerar uma cópula com seções cúbicas em ambas variáveis, C_{SC2} (cópula seção cúbica tanto em U quanto em V) como sendo,

$$C_{SC2}(u,v) \quad = \quad uv + uv(1-u)(1-v)\left[(a-b)v(1-u) + b\right]$$

onde $|b| \leqslant 1$ e $\left[b - 3 - (9 + 6b - 3b^2)^{1/2}\right]/2 \leqslant a \leqslant 1$.

É fácil ver, que a família de cópulas de seção quadrática FGM é caso especial quando $\alpha = \beta$, para a C_{SC1}, pelo que também vamos considerá-la. A expressão matemática para a cópula C_{FGM} vem dada por,

$$C_{FGM}(u,v) \quad = \quad uv + \gamma uv(1-u)(1-v)$$

com $\gamma \in [-1, 1]$.

Em capítulos anteriores descrevemos uma família de cópulas chamada de família Sarmanov. No artigo de **??** se exibem vários exemplos de generalizações desta família, entre eles encontramos uma que é também uma cópula de seção cúbica. Seja C_s uma cópula dada pela seguinte expressão,

$$C_s(u,v) \quad = \quad uv[1 + \lambda(u^2 + v^2 - u^2v^2 - 1)]$$

onde $-0.5 \leqslant \lambda \leqslant 0.25$.

Pensando no tipo de dependência, podemos considerar alguma cópula da classe arquimediana. Na Figura 13 observamos que, pelo menos no caso linear, a relação de dependência entre as variáveis é sempre positiva. Este fato faz sentido dado que em geral esperamos que uma pessoa com boas notas nas provas do Vestibular tenha um bom rendimento como estudante universitário. Nesse sentido, a melhor opção é a família Frank, que é capaz de modelar dependência positiva de forma flexível, ou seja, tanto se é muito forte quanto se é muito fraca. Seja C_f uma cópula definida por,

$$C_f(u,v) \quad = \quad -\frac{1}{\theta} ln\left(1 + \frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)}\right)$$

com $\theta \in \mathbb{R}\backslash\{0\}$. Embora, a Frank seja muito flexível no sentido da dependência, ela exige que esta tenha simetria radial, isto é, mesma magnitude de dependência entre valores grandes das duas variáveis e pequenos das mesmas. Nos dados, o comportamento tende a ser de dependência mais forte no extremo superior da diagonal principal do que no extremo inferior, pelo que gostaríamos de incluir algum fator de assimetria. Desta forma, vamos considerar uma soma convexa entre a C_f e a C_s, dado que esta última é a mais assimétrica de todas as consideradas anteriormente. Então, seja C_{fs} a cópula,

$$C_{fs}(u,v) = \mu C_f(u,v) + (1-\mu)C_s(u,v)$$

com $\mu \in [0,1]$.

4.3 Ajuste de cópulas

Para o ajuste, usamos estimação baseada na teoria de análise Bayesiana exibida no Capítulo 3. Nesta seção, vamos a apresentar os cálculos para cada uma das seis cópulas selecionadas na seção anterior, detalhando o processo e os resultados obtidos. Os dados que vamos a utilizar correspondem ao 80% dos dados originais, ou seja 4858 observações no caso do CR, e 3648 no caso da nota em MA111. Isto, com o propósito de usar o 20% restante em cada caso para, a partir de inferência preditiva, selecionar o melhor modelo em cada caso.

Para nossas estimações, vamos considerar uma função perda escalonada, pelo que o nosso processo de estimação se reduz a um problema de maximização. Para todos os casos, consideramos a log posterior (o logaritmo natural da densidade a posteriori), como a função a ser maximizada. A expressão da log posterior é a seguinte,

$$log[\pi(\theta|u,v)] \quad \propto \quad \log(\pi(\theta)) \sum_{i=1}^{4858} \log[f(u_i,v_i|\theta)],$$

onde a $f(u_i, v_i|\theta)$, para o nosso contexto de cópulas, é a densidade da cópula que pretendemos estimar. Os gráficos destas log posteriores, para os casos das cópulas $C_{SC1}, C_{SC2}, C_s, C_{FGM}$ e C_f, podemo ser encontrados no Anexo B.

4.3.1 Cópula C_{SC1}

Para começar com nossos cálculos, vamos definir uma distribuição a priori para os parâmetros. Neste caso, como $-1 \leqslant \alpha \leqslant 1$ e $-1 \leqslant \beta \leqslant 1$, poderíamos pensar em distribuições uniformes $U(-1, 1)$ para cada um deles. Além disso, vamos supor por simplicidade que α e β são independentes. Definimos então,

$$\pi(\alpha, \beta) \quad = \quad \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4},$$

se $\alpha, \beta \in [-1, 1]$. A expressão análitica desta cópula vem dada pela equação **??**. Para a verossimilhança temos a seguinte expressão,

$$L(\alpha, \beta|u, v) \quad = \quad \prod_{i=1}^{4858} \left(1 - (\beta(1 - 3u_i)u_i + \alpha(u_i - 1)(3u_i - 1))(2v_i - 1)\right),$$

pelo que a densidade a posteriori é,

$$\pi(\alpha, \beta|u, v) \quad \propto \quad \frac{1}{4} \prod_{i=1}^{4858} \left(1 - (\beta(1 - 3u_i)u_i + \alpha(u_i - 1)(3u_i - 1))(2v_i - 1)\right),$$

e finalmente a log posterior,

$$log[\pi(\alpha, \beta|u, v)] \quad \propto \quad -log(4) \sum_{i=1}^{4858} log\left[1 - (\beta(1 - 3u_i)u_i + \alpha(u_i - 1)(3u_i - 1))(2v_i - 1)\right].$$

Usando métodos computacionais para otimizar a função correspondente a cada par de variáveis analisadas, e obtivemos os resultados apresentados na Tabela 2.

Tabela 2 – Valores estimados de α e β para a cópula $C_{SC1}(u, v)$

		CN	CH	PT	ING	MA	VF1	NPT
CR	α	0,9999	1	1	0,999	1	0,997	0,9999
	$\hat{\beta}$	0,9999	0,9995	0,9995	0,6628	0,9999	1	1
MA111	$\hat{\alpha}$	0,9995	1	0,9411	0,999	0,9995	1	0,9995
	$\hat{\beta}$	1	0,9999	0,8553	0,5113	1	0,9999	1

Em geral, podemos considerar que as estimações da Tabela 2 são coerentes com a ideia de que a relação entre as variáveis do Vestibular e, o coeficiente de rendimento CR e a nota em Cálculo I-MA111 é positiva, isto dado que em quase todos os casos $\alpha \geqslant \beta$ aproximadamente.

4.3.2 Cópula C_{SC2}

Para definir uma distribuição a priori para os parâmetros a e b, devemos considerar que existe uma relação entre eles, pois $\left[b-3-(9+6b-3b^2)^{1/2}\right]/2 \leqslant a \leqslant 1$. Dado que $|b| \leqslant 1$, poderíamos assinar uma uniforme $U(-1,1)$ como na cópula anterior, e definir a distribuição de a como sendo $U(\left[b-3-(9+6b-3b^2)^{1/2}\right]/2, 1)$, pelo que sua densidade conjunta poderia se escrever,

$$\pi(a,b) \;=\; \left(\frac{1}{2}\right)\frac{1}{[b-3-(9+6b-3b^2)^{1/2}]/2} = \frac{1}{b-3-(9+6b-3b^2)^{1/2}},$$

para $a \in [\left[b-3-(9+6b-3b^2)^{1/2}\right]/2, 1]$ e $b \in [-1,1]$. A expressÃo desta cópula foi apresentada no Cápitulo 2 na equação **??**. Para a verossimilhança temos,

$$L(a,b|u,v) \;=\; \prod_{i=1}^{4858}\left(1-a(u_i-1)(3u_i-1)v_i(3v_i-2)+b(1+3(u_i-1)v_i)(1-v_i+u_i(3v_i-2))\right),$$

pelo que a densidade a posteriori se escreve,

$$\begin{aligned}\pi(a,b|u,v) \;\propto\; & \frac{1}{b-3-(9+6b-3b^2)^{1/2}}\prod_{i=1}^{4858}[1-a(u_i-1)(3u_i-1)v_i(3v_i-2)+ \\ & b(1+3(u_i-1)v_i)(1-v_i+u_i(3v_i-2))],\end{aligned}$$

e finalmente a log posterior,

$$\begin{aligned}log[\pi(a,b|u,v)] \;\propto\; & -log(\left[b-3-(9+6b-3b^2)^{1/2}\right])\sum_{i=1}^{4858} log[1-a(u_i-1)(3u_i-1)v_i(3v_i-2)+ \\ & b(1+3(u_i-1)v_i)(1-v_i+u_i(3v_i-2))].\end{aligned}$$

De novo, a partir de métodos computacionais, otimizamos a função correspondente a cada par de variáveis analisadas, e obtivemos os resultados apresentados na Tabela 3.

Tabela 3 – Valores estimados de a e b para a cópula $C_{SC2}(u,v)$

		CN	CH	PT	ING	MA	VF1	NPT
CR	$\hat{a}$	1	1	1	1	1	0,9847	1
	$\hat{b}$	1	1	1	0,8	1	1	1
MA111	$\hat{a}$	1	0.84	0,6435	0,9389	1	1	1
	$\hat{b}$	1	1	1	0,7358	1	1	1

Note que neste caso, vemos que na Tabela 3 as estimações para quase todos os pares de variáveis considerados são iguais a 1, pelo que esta família de cópulas está sugerindo que uma mesma cópula consegue modelar as relações entre as diferentes variáveis, mas isto não faz sentido porque pelo menos na relação linear existe uma clara diferença como visto na Figura 13. Também, podemos dizer que o fato de $a = b = 1$, onde 1 é o máximo valor possível quando a relação entre as variáveis é positiva, sugere que essa relação é igual o maior do que a máxima relação modelável a partir desta família de cópulas.

4.3.3 Cópula Sarmanov

Neste caso, a definição da distribuição a priori é muito mais simples do que no caso anterior, dada a condição $-0.5 \leqslant \lambda \leqslant 0.25$, podemos considerar uma $U(-0.5, 0.25)$ cuja densidade é,

$$\pi(\lambda) = \frac{4}{3}$$

se $\lambda \in [-0,5; 0,25]$. Para a verossimilhança, consideramos a análise feita na Seção 3.2, especificamente a equação **??**, que para nosso caso fica

$$L(\lambda|u,v) \quad = \quad \prod_{i=1}^{4858} \left[1 + \lambda(3u_i^2 + 3v_i^2 - 9u_i^2v_i^2 - 1)\right],$$

pelo que a densidade a posteriori se escreve,

$$\pi(\lambda|u,v) \quad \propto \quad \left(\frac{4}{3}\right) \prod_{i=1}^{4858} \left[1 + \lambda(3u_i^2 + 3v_i^2 - 9u_i^2v_i^2 - 1)\right],$$

e finalmente a log posterior,

$$log[\pi(\lambda|u,v)] \quad \propto \quad -log(0.75) \sum_{i=1}^{4858} log\left[1 + \lambda(3u_i^2 + 3v_i^2 - 9u_i^2v_i^2 - 1)\right].$$

Neste caso, todas as estimações são negativas, ou seja, $\hat{\lambda} < 0$ e segundo o estabelecido

Tabela 4 – Valores estimados de λ para a cópula $C_s(u,v)$

	CN	CH	PT	ING	MA	VF1	NPT
$\hat{\lambda}$ CR	-0,4999	-0,4999	-0,4665	-0,3568	-0,4999	-0,4999	-0,4999
$\hat{\lambda}$ MA111	-0,4999	-0,4517	-0,3516	-0,2875	-0,4999	-0,4775	-0,4999

por [Bairamov et al., 2011] isto implica que todos estes pares têm dependência possitiva ($0 < \rho_C \leqslant 0,375$).

4.3.4 Cópula FGM

Da mesma forma como feito com a cópula Sarmanov, a distribuição a priori neste caso seria uma $U(-1,1)$ com densidade,

$$\pi(\gamma) \quad = \quad \frac{1}{2}$$

se $\gamma \in [-1,1]$. Para a verossimilhança temos,

$$L(\gamma|u,v) \quad = \quad \prod_{i=1}^{4858} \left(1 + \gamma(1 - 2v_i - 2u_i + 4u_iv_i)\right),$$

pelo que a densidade a posteriori se escreve,

$$\pi(\gamma|u,v) \quad \propto \quad \left(\frac{1}{2}\right) \prod_{i=1}^{4858} \left[1 + \gamma(1 - 2v_i - 2u_i + 4u_iv_i)\right],$$

e finalmente a log posterior,

$$log[\pi(\gamma|u,v)] \quad \propto \quad -log(2)\sum_{i=1}^{4858} log\left[1+\gamma(1-2v_i-2u_i+4u_iv_i)\right].$$

Vemos nesse caso, que a maioria das estimações são iguais a $0,9999$ isto é porque esta

Tabela 5 – Valores estimados de γ para a cópula $C_{FGM}(u,v)$

	CN	CH	PT	ING	MA	VF1	NPT
$\hat{\gamma}$ CR	0,9999	0,9999	0,9999	0,8977	0,9999	0,9999	0,9999
$\hat{\gamma}$ MA111	0,9999	0,9999	0,9020	0,8196	0,9999	0,9999	0,9999

cópula serve para modelar dependências fracas, e para alguns dos nossos pares de variáveis sabemos que pelo menos a sua correlação linear é maior do que o máximo que consegue modelar a cópula FGM (0,33). No entanto, consideramos esta cópula porque temos algumas variáveis cuja dependência pode ser desse tipo.

4.3.5 Cópula Frank

Para as cópulas desta família, a escolha da distribuição a priori não é tão imediata como para as outras cópulas que estamos considerando. Sabemos que o parâmetro θ pode ser qualquer número real menos o zero, o que nos faz considerar alguma distribuição contínua com domínio em $\mathbb{R}$. Então, vamos considerar uma distribuição normal $N(0,\sigma^2 = 5)$, restringindo-a para que θ não possa ser igual a zero. Então consideramos a densidade,

$$\pi(\theta) \quad = \quad \frac{1}{\sqrt{2\pi(5)}}\exp\{-\theta^2/10\},$$

se $\theta \neq 0$. Para a verossimilhança temos,

$$L(\theta|u,v) \quad = \quad \prod_{i=1}^{4858}\left[\frac{\theta e^{\theta(1+u_i+v_i)}\left(e^{\theta}-1\right)}{\left(e^{\theta(u_i+v_i)}-e^{\theta}\left(e^{\theta u_i}+e^{\theta v_i}-1\right)\right)^2}\right],$$

pelo que a densidade a posteriori se escreve,

$$\pi(\theta|u,v) \quad \propto \quad \left(\frac{1}{\sqrt{10\pi}}\right)\exp\{-\theta^2/10\}\prod_{i=1}^{4858}\left[\frac{\theta e^{\theta(1+u_i+v_i)}\left(e^{\theta}-1\right)}{\left(e^{\theta(u_i+v_i)}-e^{\theta}\left(e^{\theta u_i}+e^{\theta v_i}-1\right)\right)^2}\right],$$

e finalmente a log posterior,

$$log[\pi(\theta|u,v)] \quad \propto \quad \frac{-\theta^2}{10}\sum_{i=1}^{4858} log\left[\frac{\theta e^{\theta(1+u_i+v_i)}\left(e^{\theta}-1\right)}{\left(e^{\theta(u_i+v_i)}-e^{\theta}\left(e^{\theta u_i}+e^{\theta v_i}-1\right)\right)^2}\right].$$

Desta cópula sabemos que quanto maior é θ, mais relacionadas em sentido positivo estão as variáveis analisadas. Assim, podemos dizer a partir da Tabela 6 que as variáveis mais relacionadas, tanto com o CR quanto com a nota em MA111, são: NP, CN

Tabela 6 – Valores estimados de θ para a cópula $C_f(u, v)$

	CN	CH	PT	ING	MA	VF1	NPT
$\hat{\theta}$ CR	4,3626	3,1	2,5827	1,9267	4,1049	3,2292	4,5266
$\hat{\theta}$ MA111	4,7089	2,8468	1,9984	1,7755	4,5361	3,0529	4,7288

e MA, e a menos relacionada é ING. Também, poderiamos comparar estes resultados com os obtidos no caso da FGM devido a que esta ultima é uma aproximação de primeira ordem da Frank. Nesse sentido, temos que para ambos casos as estimações dos respectivos parâmetros são possitivas pelo que a relação entre todos os pares de variavéis é possitiva, mas no caso da Frank as estimativas são mais variadas devido a que é capaz de perceber correlações mais fortes do que a FGM.

4.3.6 Cópula Frank Sarmanov

Dado que esta cópula é uma soma convexa de duas cópulas analisadas anteriormente, vamos usar os resultados que já foram obtidos para simplificar a análise. Se consideramos que os três parâmetros deste modelo são independentes, e que λ e θ têm as mesmas distribuições a priori definidas nos casos individuais da Sarmanov e a Frank, então

$$\pi(\theta, \mu, \lambda) = \left(\frac{4}{3}\right) \frac{1}{\sqrt{(10)}} \exp\{-\theta^2/10\},$$

quando $\theta \in \mathbb{R}\backslash\{0\}$, $\lambda \in [-0,5; 0,25]$ e $\mu \in [0, 1]$, supondo que a distribuição a priori para μ é uniforme U(0,1). Para a verossimilhança temos,

$$L(\theta, \mu, \lambda|u, v) = \prod_{i=1}^{4858} \mu \frac{\theta e^{\theta(1+u_i+v_i)} \left(e^{\theta} - 1\right)}{\left(e^{\theta(u_i+v_i)} - e^{\theta} \left(e^{\theta u_i} + e^{\theta v_i} - 1\right)\right)^2} + (1 - \mu) \left(1 + \lambda(3u_i^2 + 3v_i^2 - 9u_i^2 v_i^2 - 1)\right),$$

pelo que a densidade a posteriori se escreve,

$$\begin{aligned}\pi(\theta, \mu, \lambda|u, v) \propto\ & \frac{4}{3\sqrt{(10)}} \exp\{-\theta^2/10\} \prod_{i=1}^{4858} \mu \frac{\theta e^{\theta(1+u_i+v_i)} \left(e^{\theta} - 1\right)}{\left(e^{\theta(u_i+v_i)} - e^{\theta} \left(e^{\theta u_i} + e^{\theta v_i} - 1\right)\right)^2} + \\ & (1 - \mu) \left(1 + \lambda(3u_i^2 + 3v_i^2 - 9u_i^2 v_i^2 - 1)\right),\end{aligned}$$

e finalmente a log posterior,

$$\begin{aligned}log[\pi(\theta, \mu, \lambda|u, v)] \propto\ & \frac{-\theta^2}{10} \sum_{i=1}^{4858} log[\mu \frac{\theta e^{\theta(1+u_i+v_i)} \left(e^{\theta} - 1\right)}{\left(e^{\theta(u_i+v_i)} - e^{\theta} \left(e^{\theta u_i} + e^{\theta v_i} - 1\right)\right)^2} + \\ & (1 - \mu) \left(1 + \lambda(3u_i^2 + 3v_i^2 - 9u_i^2 v_i^2 - 1)\right)].\end{aligned}$$

Neste caso, temos que os parâmetros θ e λ vão ter a mesma interpretação que nos casos individuais das cópulas C_f e C_s respectivamente. Na Tabela 7, observamos que no caso de θ a maioria das estimações são maiores às obtidas para a C_f, mas a ordem de

Tabela 7 – Valores estimados de θ, μ e λ para a cópula $C_{fs}(u, v)$

		CN	CH	PT	ING	MA	VF1	NPT
CR	$\hat{\theta}$	5,2148	3,4196	2,8555	1,9210	5,0265	3,9330	6,2322
	$\hat{\mu}$	0,7370	0,6821	0,6740	0,9953	0,7022	0,5503	0,6388
	λ	-0,4979	-0,4802	-0,4432	-0,4532	-0,4727	-0,4993	-0,4999
MA111	$\hat{\theta}$	4,9342	3,0318	2,1101	2,5981	4,5640	3,0875	4,8359
	$\hat{\mu}$	0,9772	0,9659	0,6740	0,7911	0,9997	0,9975	0,9853
	λ	0,2443	0,2443	0,248	0,2272	0,1386	0,2189	0,2254

maior a menor se mantem, tanto para o CR quanto para MA111. No caso de λ, vemos que estas estimações apresentam mais variabilidade que as obtidas para a cópula C_s, mas o sinal segue sendo negativo e a magnitude muito próxima de -0.5, pelo que seguimos considerando uma relação positiva entre as variáveis do Vestibular e, tanto o CR quanto a nota em MA111.

Pelo lado de μ, temos que é o parâmetro que nos reflete o peso de cada cópula considerada da seguinte forma: se μ é muito próximo de 1 então diremos que não faz muito sentido incluir à C_s na análise, enquanto que se μ é muito próximo de 0 então poderíamos tdesconsiderar a C_f. Na Tabela 7, vemos que o único valor que se acha em algum dos casos extremos mencionados anteriormente é o $\mu = 0,9953$, no caso do CR (correspondente ao par (ING, CR)), pelo que poderíamos considerar nesse caso um ajuste apenas com a C_f. No caso da nota em MA111, observamos que a maioria de valores de μ estão muito proximos de 1 (exceto para os pares $(PT, MA111)$ e $(ING, MA111)$).

4.4 Seleção da melhor cópula: CR

Na seção anterior, ajustamos cada uma das nossas opções de famílias de cópulas a todos os pares formados entre as variáveis do Vestibular e, o CR e a nota em MA111. Este ajuste não é mais que a eleição de uma cópula representante de cada uma das seis famílias que escolhemos na Seção 4.2, ou seja, temos seis modelos possíveis para cada um dos sete pares de variáveis (no caso do CR) e devemos selecionar aquele que seja melhor em cada caso.

No Cápitulo 4, definimos diferentes critérios a partir dos quais identificar quando um modelo é melhor do que outro. Nesta seção vamos fazer uso desses critérios, tanto gráficos quanto numéricos, para selecionar uma cópula para cada um dos sete pares de variáveis.

4.4.1 Ciências da Natureza

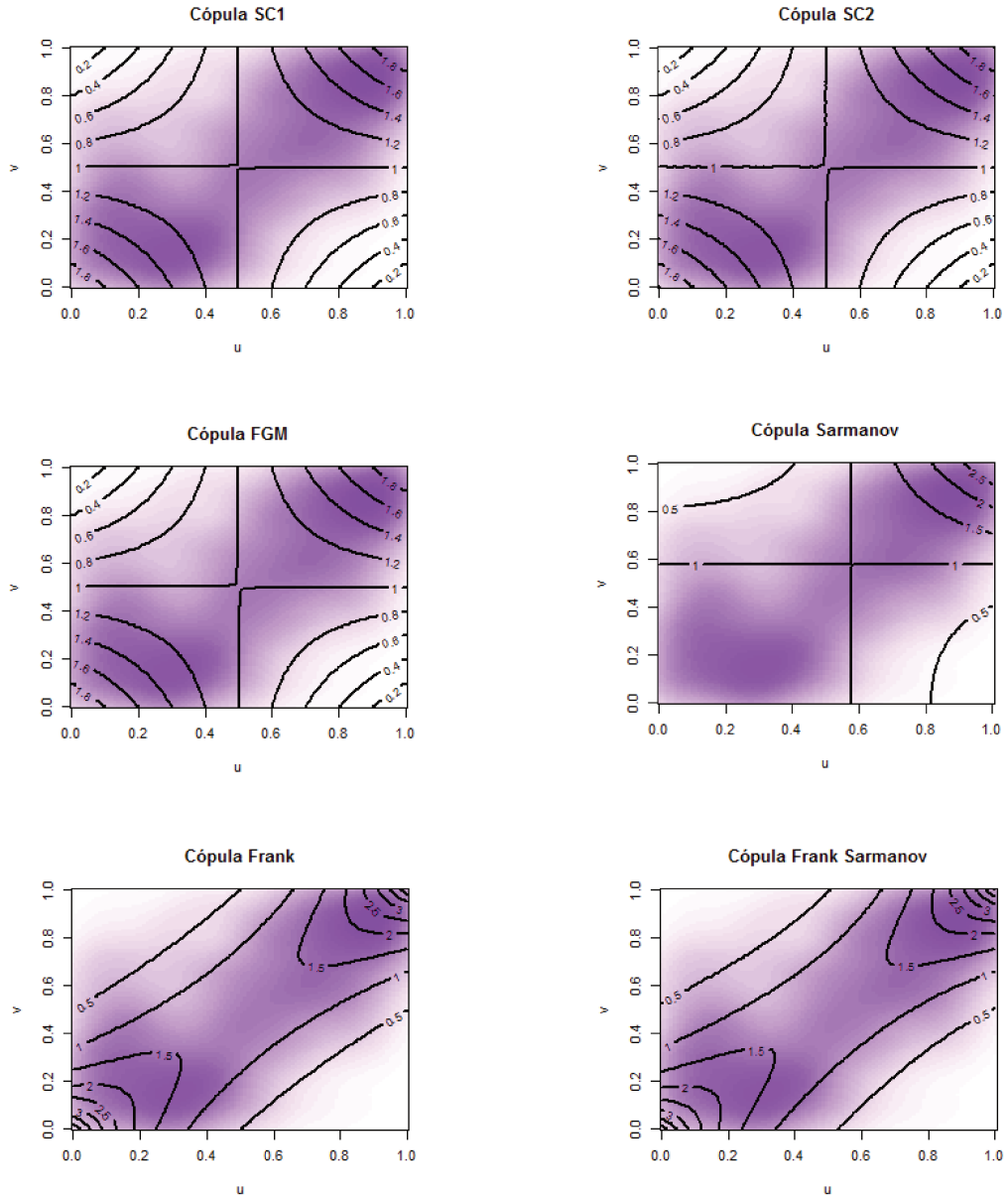

Figura 16 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de (CN, CR) para cada um dos modelos possíveis.

O nosso objetivo é encontrar o melhor ajuste para este par de variáveis dentre aqueles feitos na Seção 4.3. Um melhor ajuste será aquele cujo comportamento seja mais semelhante com o comportamento amostral dos dados, pelo que vamos começar analisando a Figura 16, onde as linhas pretas representam os contornos da cópula ajustada em cada caso, e a área sombreada representa a cópula empírica associada a (CN, CR).

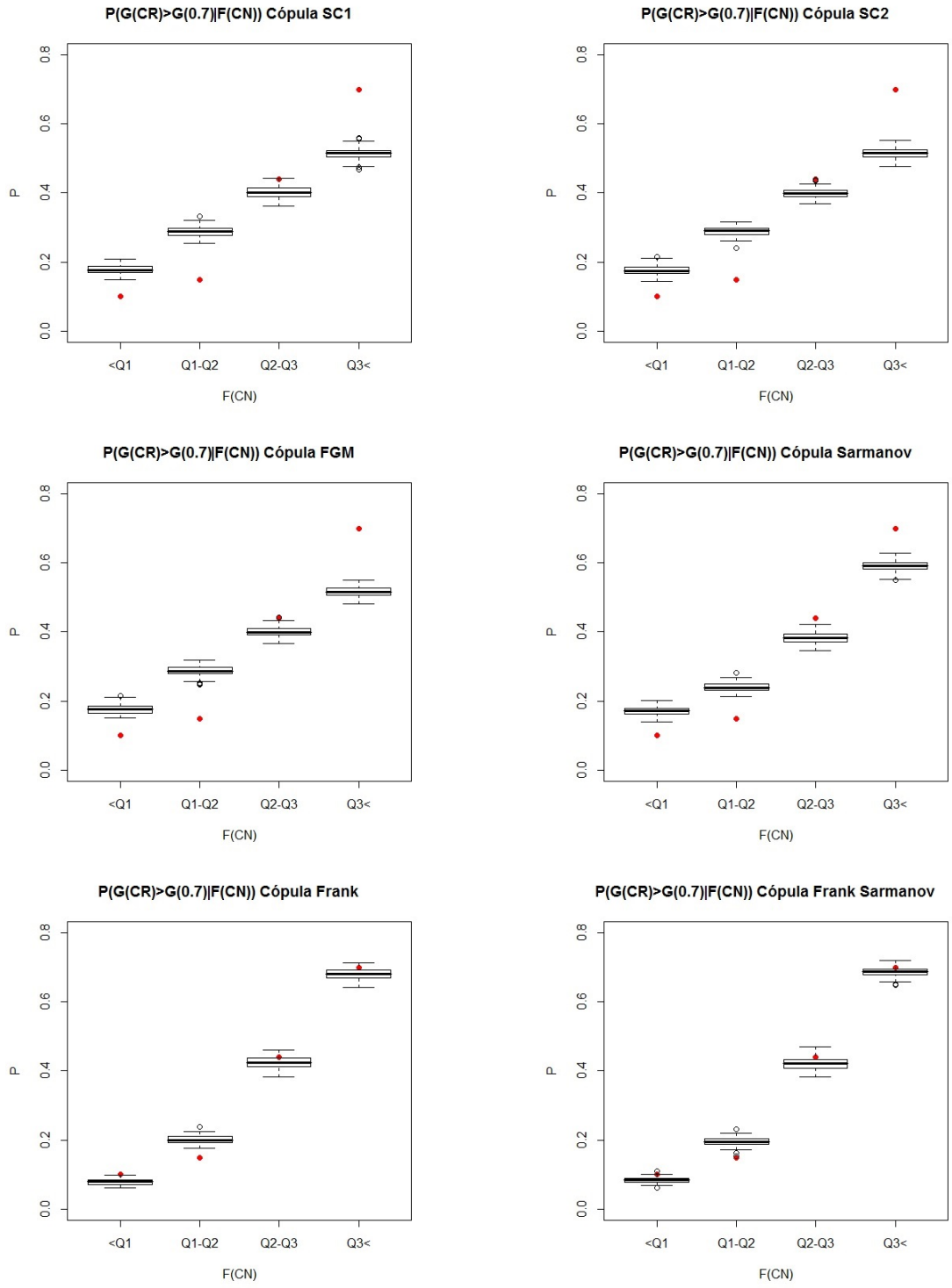

Figura 17 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(CN))$ calculadas a partir do ajuste das cópulas.

Desta figura, podemos dizer que os contornos das cópulas C_{SC1}, C_{SC2} e C_{FGM} são praticamente os mesmos, pelo que analisar um é equivalente a analisar os três, e encontramos que o gráfico correspondente não é apropriado, dado que nos extremos inferior direito e superior esquerdo, a cópula empírica decresce muito mais rápido do que o ajuste

sugere. É claro, que o pior modelo neste sentido é a cópula Sarmanov, dado que não consegue modelar o que acontece quando ambas das variáveis (CN, CR) tomam valores pequenos. Finalmente ficamos com as cópulas, C_f e C_{fs}, que pelo menos graficamente não são claramente diferenciáveis. Agora vamos analisar a capacidade que tem as cópulas para modelar o valor de uma probabilidade.

A Figura 17 mostra os boxplot da probabilidade $P(G(CR) \geqslant G(0.7)|u_1 \leqslant F(CN) \leqslant u_2)$, onde (u_1, u_2) são os intervalos formados pelos quartís $(0 - Q_1, Q_1 - Q_2, Q_2 - Q_3, Q_3 - 1)$ e o ponto vermelho é o valor amostral da probabilidade. Dado que na análise da Figura 16 encontramos que as cópulas $C_{SC1}, C_{SC2}, C_{FGM}$ e C_s não são adequadas, unicamente analisaremos os boxplot correspondentes aos modelos Frank e Frank Sarmanov. Vemos que, embora sendo muito parecidos, os boxplot do modelo Frank Sarmanov tem mediana mais perto do ponto vermelho para a maioria dos intervalos considerados.

Tabela 8 – Distâncias entre cópula empírica e a preditiva C caso (CN, CR)

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D, Hellinger	1,057	1,0567	1,0565	1,0587	0,4875	0,4422
D. Kolmogorov	0,0519	0,0519	0,0585	0,0519	0,0173	0,0148

A Tabela 8, mostra as distâncias de Hellinger e Kolmogorov entre a cópula empírica e sua correspondente preditiva. De novo, para ambos os casos, as distâncias mínimas são obtidas para os modelos Frank e Frank Sarmanov.

Finalmente, para selecionar entre estes dois modelos (dado que um é caso especial do outro), vamos utilizar o FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$. Para isto, vamos simular por MCMC uma amostra aleatória de 1000 valores da $\pi(\theta, \mu, \lambda)$ e calcular a log posterior para todos os (θ, μ, λ) tais que $\mu \geqslant 0,95$ e selecionamos φ como sendo o máximo valor da log posterior nesse subconjunto. Assim, o nosso κ vai ser a probabilidade de encontrar na log posterior da nossa amostra, valores maiores ou iguais a φ. Dessa forma achamos que $\kappa = 0,976$ e então e-valor$= 1 - 0,976 = 0,033$ pelo que temos evidência suficiente para rejeitar H_0 e concluir que o melhor modelo para o par (CN, CR) é C_{fs}.

4.4.2 Ciências Humanas

Seguindo a mesma sequência de análise feita para o caso anterior, apresentamos as Figuras 18 e 19, e a Tabela 9.

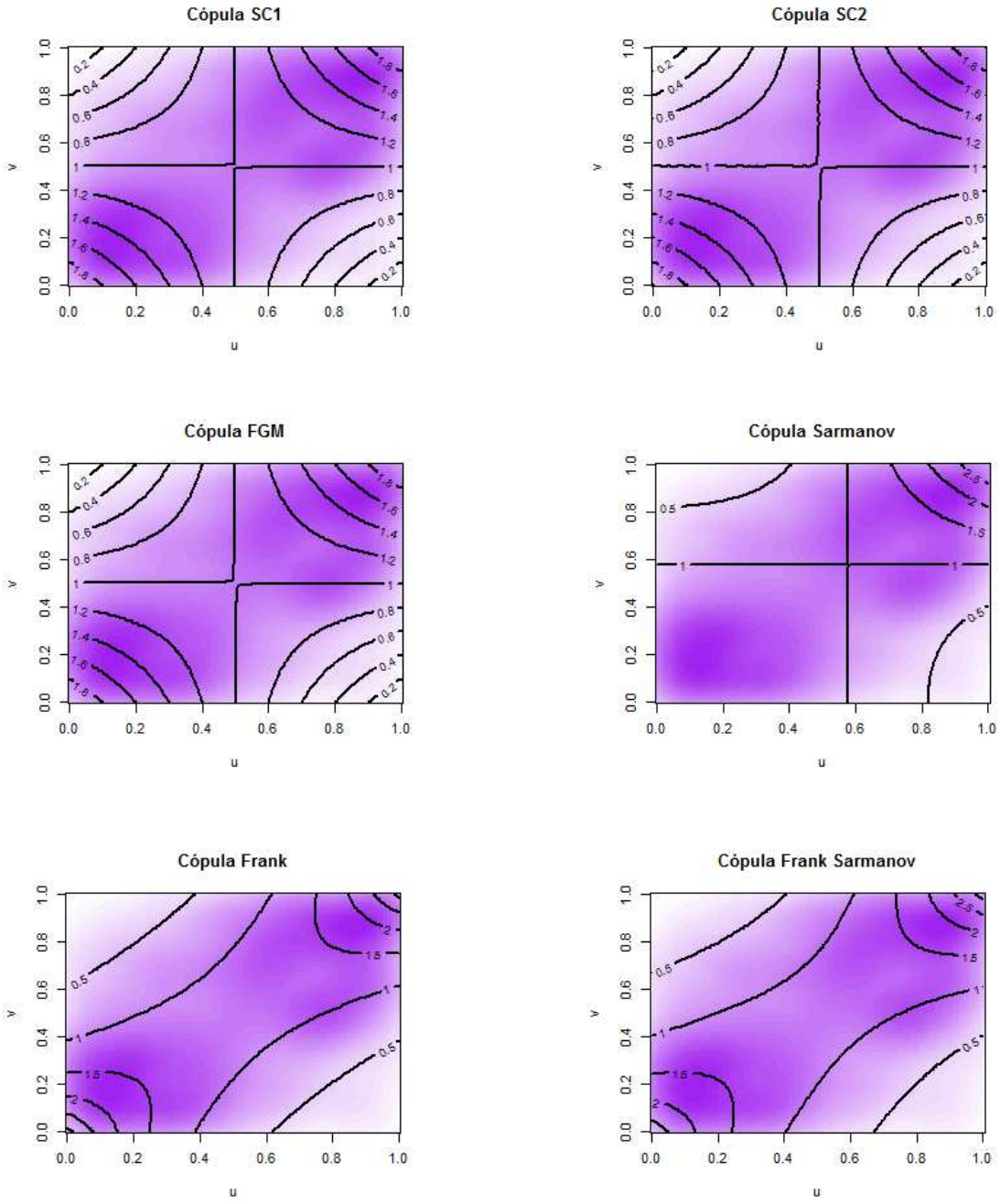

Figura 18 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de (CH, CR) para cada um dos modelos possíveis.

Para este caso (Figura 18), encontramos um cenário parecido com o achado para (CN, CR), no sentido das cópulas C_{SC1}, C_{SC2} e C_{FGM}, pois o gráfico é igual para as três. De novo a cópula C_s é a pior de todas, mas dessa vez dá para perceber uma diferença entre os gráficos dos modelos Frank e Frank Sarmanov, sendo este último um pouco mais apropriado dado que reflete a assimetria entre os extremos da diagonal principal.

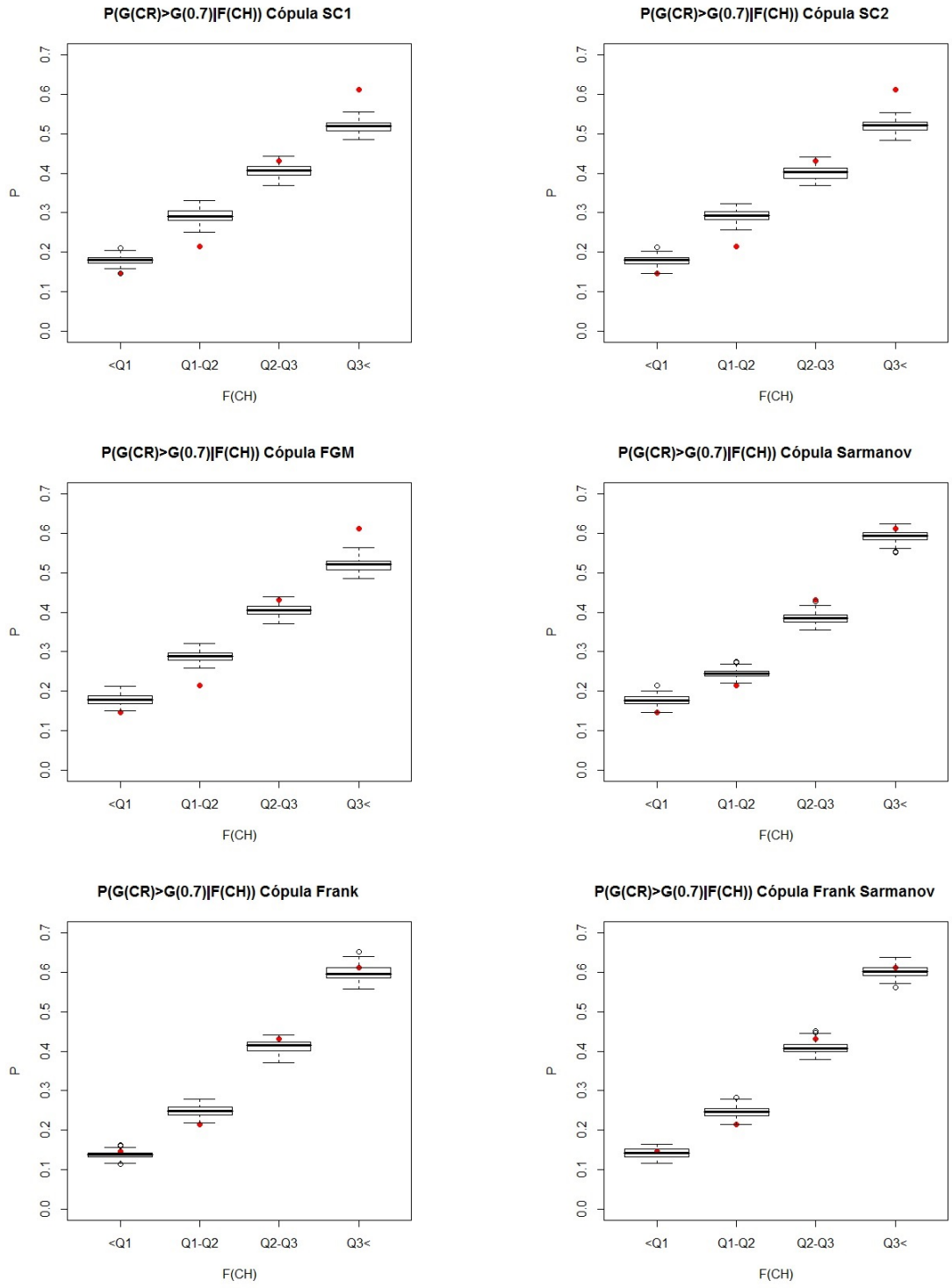

Figura 19 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(CH))$ calculadas a partir do ajuste das cópulas.

Com respeito à Figura 19, vemos claramente que os modelos Frank e Frank Sarmanov são os melhores, observando-se que o ultimo melhora o ajuste da probabilidade no intervalo $Q_1 \leqslant F(CH) \leqslant Q_2$.

Tabela 9 – Distâncias entre cópula empírica e a preditiva C caso (CH, CR)

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,5583	0,5618	0,6431	0,5623	0,2584	0,2596
D. Kolmogorov	0,0362	0,0362	0,0395	0,0362	0,0181	0,0181

A partir da Tabela 9, podemos afirmar que em termos da distância de Hellinger que o melhor modelo é o modelo Frank. No caso da distância de Kolmogorov, vemos que os valores para as cópulas C_f e C_{fs} são iguais, pelo que não podemos dizer qual deles é melhor usando esse critério.

Até o momento temos que, tanto C_f quanto C_{fs} são boas opções. Em termos da probabilidade da Figura 19 e das curvas de nível da Figura 18, escolheríamos o modelo C_{fs}, cuja expressão é mais complexa que para C_f, que segundo a distância de Hellinger é melhor do que C_{fs}. Para finalmente decidir entre estas duas opções, usemos a metodologia do FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$, seguindo o mesmo procedimento feito no caso (CN, CR). Como obtemos $\kappa = 1$, o e-valor é nulo, e concluímos que a melhor cópula para este caso é C_{fs}.

4.4.3 Inglês

No caso da Figura 20, que dessa vez todos os gráficos são diferentes, e de novo o correspondente à C_s tem o pior ajuste de todos. Também, podemos observar que os gráficos das C_{SC2} e C_{FGM} são os que menos refletem a imagem dos dados, por causa de assimetria destes. Portanto, vamos ficar por enquanto com as cópulas C_{SC1}, C_f e C_{fs}.

No caso da Figura 21 a análise é mais complexa, dado que não é possível visualizar qual modelo é o melhor, pois todos parecem estar razoavelmente perto dos valores das probabilidades calculadas nas observações. No entanto, o gráfico correspondente à C_{SC1} mostra que os pontos vermelhos estão quase todos acima de mediana dos boxplot, pelo que diremos que segundo este critério a cópulas C_{SC1} é melhor modelo.

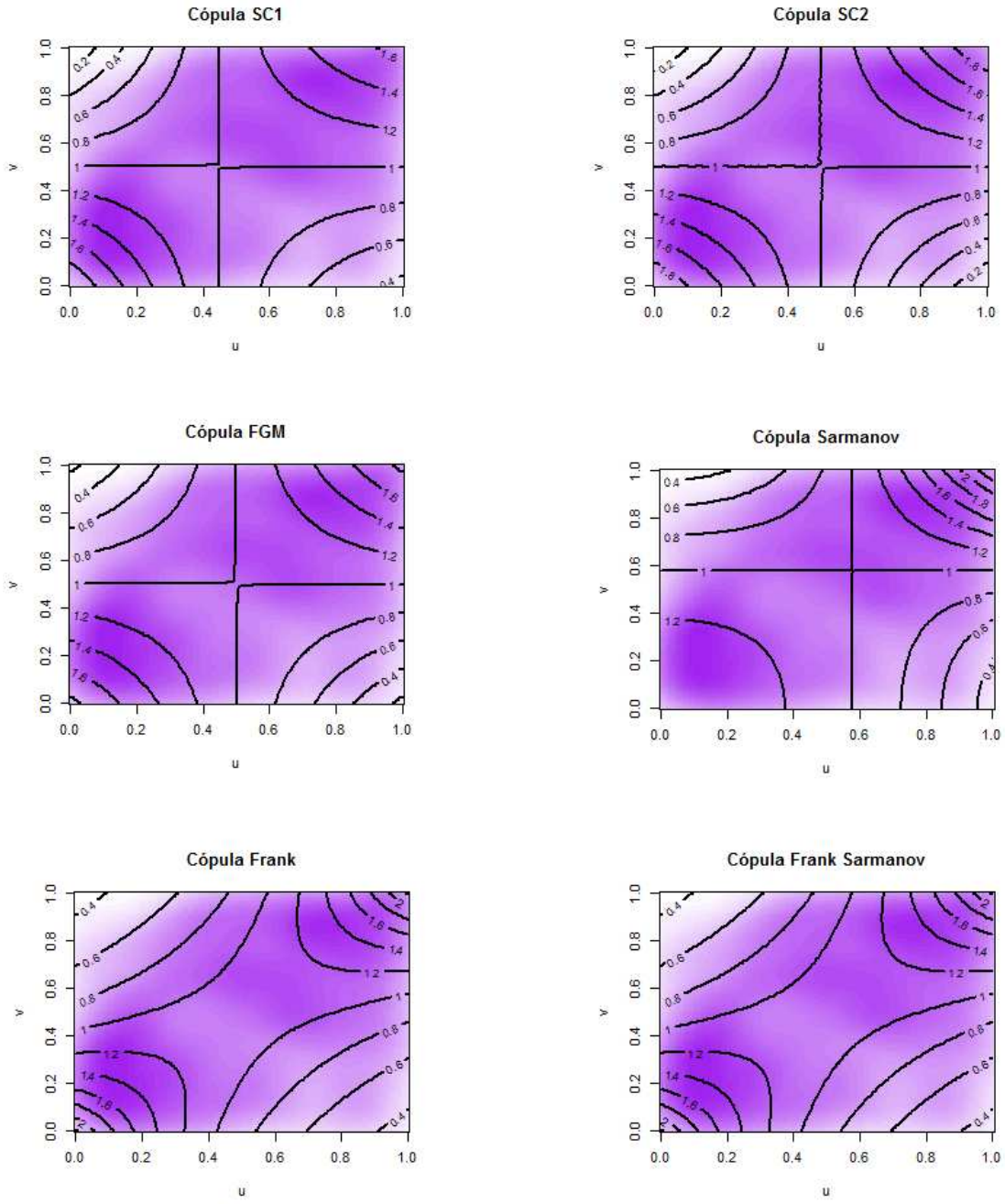

Figura 20 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de (ING, CR) para cada um dos modelos possíveis.

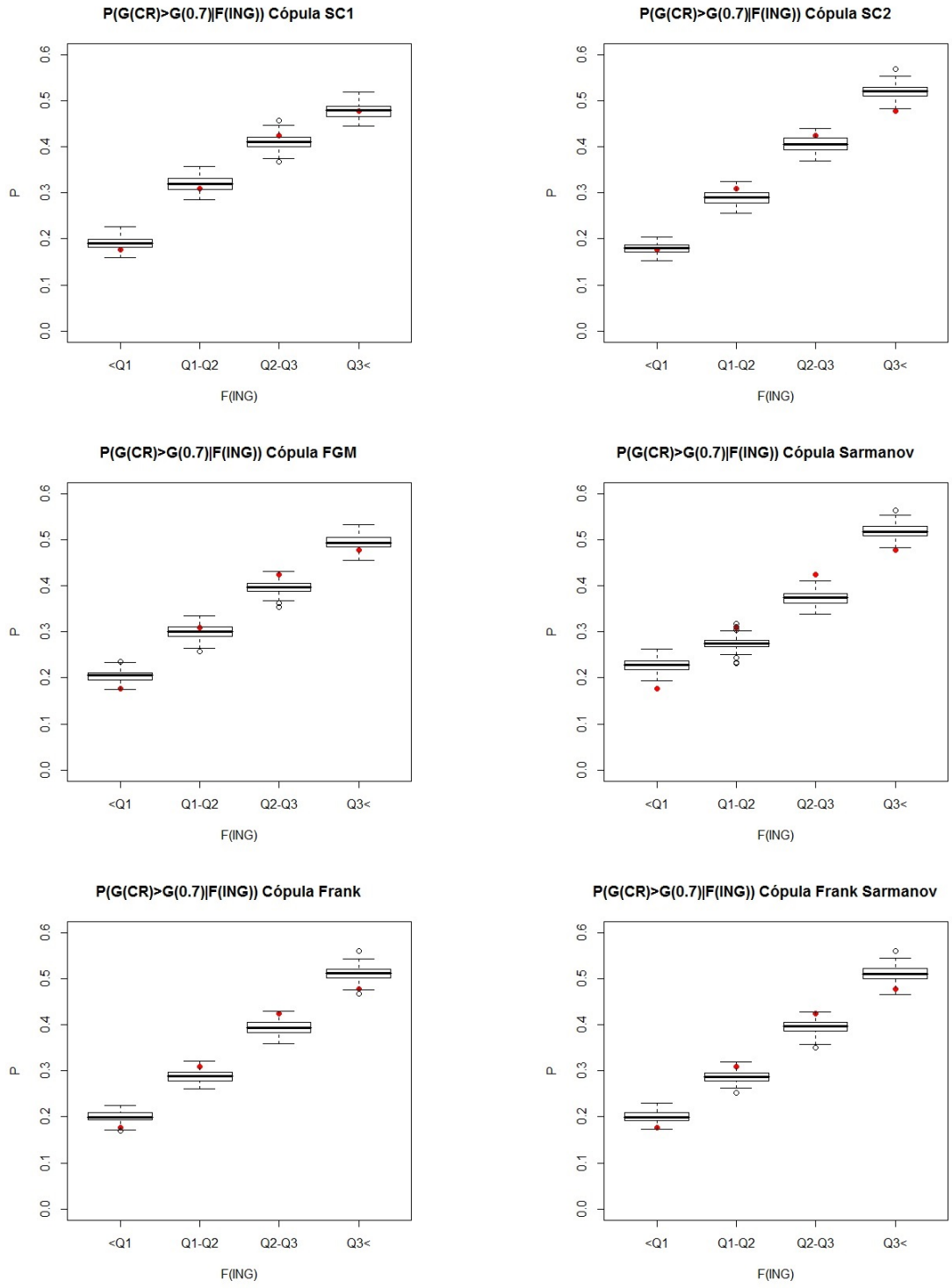

Figura 21 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(ING))$ calculadas a partir do ajuste das cópulas.

Tabela 10 – Distâncias entre cópula empírica e a preditiva C caso (ING, CR)

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,2671	0,2914	0,4843	0,288	0,2952	0,2868
D. Kolmogorov	0,0208	0,0230	0,0419	0,0223	0,0177	0,0188

Com respeito à Tabela 10, temos uma decisão por cada distâncias, segundo a distância de Hellinger o melhor modelo corresponde à cópula C_{SC1}, enquanto que para a distância de Kolmogorov o melhor modelo é a cópula de Frank. Finalmente, concluímos que a cópula C_{SC1} é o modelo escolhido pois, na maioria dos casos teve os melhores resultados.

4.4.4 Matemática

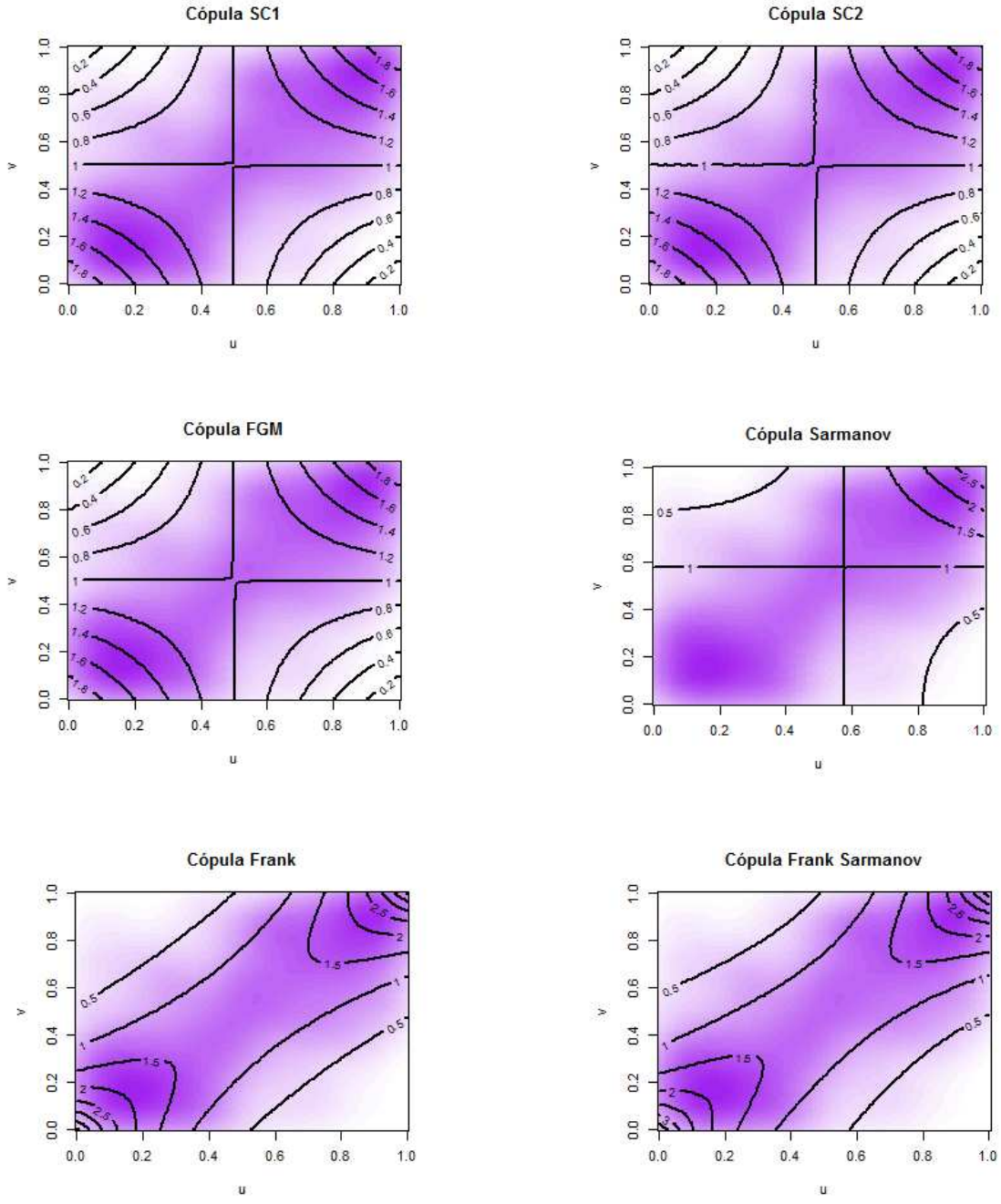

Figura 22 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de (MA, CR) para cada um dos modelos possíveis.

Começando com a análise dos gráficos de contorno da Figura 22, encontramos um cenário similar ao encontrado no caso de (CN, CR), isto é, as cópulas C_{SC1}, C_{SC2}, C_{FGM} e C_s podem ser desconsideradas, dado que nenhuma parece se ajustar bem aos dados e por isso continuamos a nossa análise para as cópulas C_f e C_{fs}.

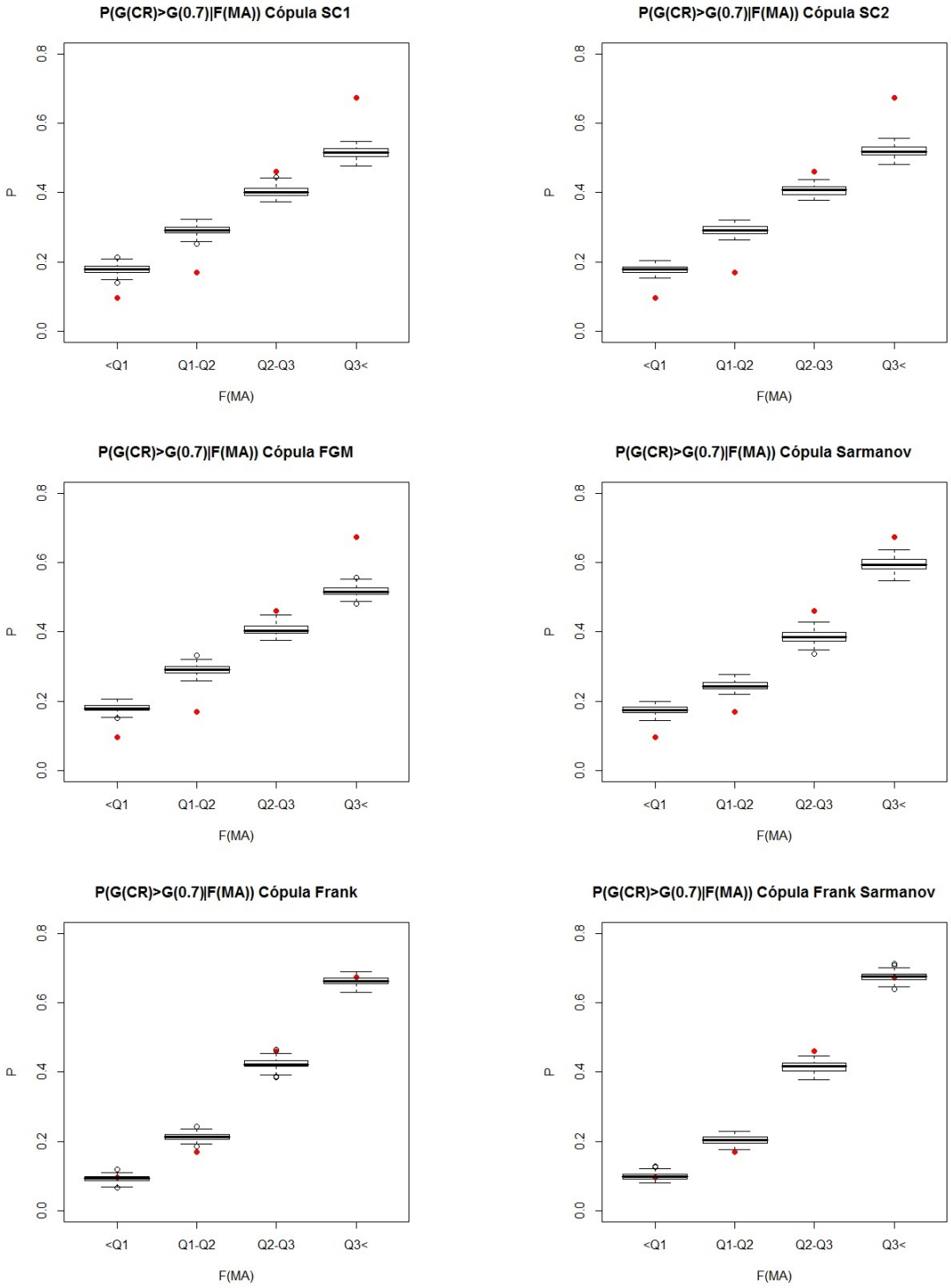

Figura 23 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(MA))$ calculadas a partir do ajuste das cópulas.

Com respeito à Figura 23, diremos que não da para perceber uma diferença entre o modelo Frank e Frank Sarmanov, mas é claro que com respeito aos outros, esses são os melhores modelos.

Pelo lado da Tabela 11, diremos que no caso da distância de Hellinguer o melhor modelo seria a cópula Frank Sarmanov, enquanto que no caso da distância de Kolmogorov diremos que o melhor modelo seria a cópula de Frank.

Tabela 11 – Distâncias entre cópula empírica e a preditiva C para (MA, CR)

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	1,0108	1,0112	1,0292	1,0132	0.4295	0,3995
D. Kolmogorov	0,0362	0,0362	0,0453	0,0362	0,0189	0,0198

Dado o resultado achado no cálculo das distâncias, vamos testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$, seguindo o mesmo procedimento do FBST. Assim, obtemos $\kappa = 1$ pelo que nosso e-valor é nulo, e finalmente decidimos que o melhor modelo é a cópula C_{fs}.

4.4.5 Português

Para este caso, temos que de novo as cópulas C_{SC1}, C_{SC2} e C_{FGM} apresentam o mesmo gráfico e mais uma vez não é o adequado por causa da simetria. Como em todos os casos anteriores, a cópula de Sarmanov é a que apresenta pior ajuste. Dado que a cópula de Frank decresce nos extremos superior esquerdo e inferior direito de forma mais parecida aos dados, diremos que segundo este critério a cópula C_f é o melhor modelo.

Dos gráficos da Figura 25, podemos dizer que em todos os casos os pontos vermelhos estão bem perto das medianas dos boxplot. No entanto, destacamos que nos casos das cópulas C_s,C_f e C_{fs} os pontos ficam dentro das caixas em todos os intervalos, sendo a C_{fs} a que parece dar um melhor ajuste.

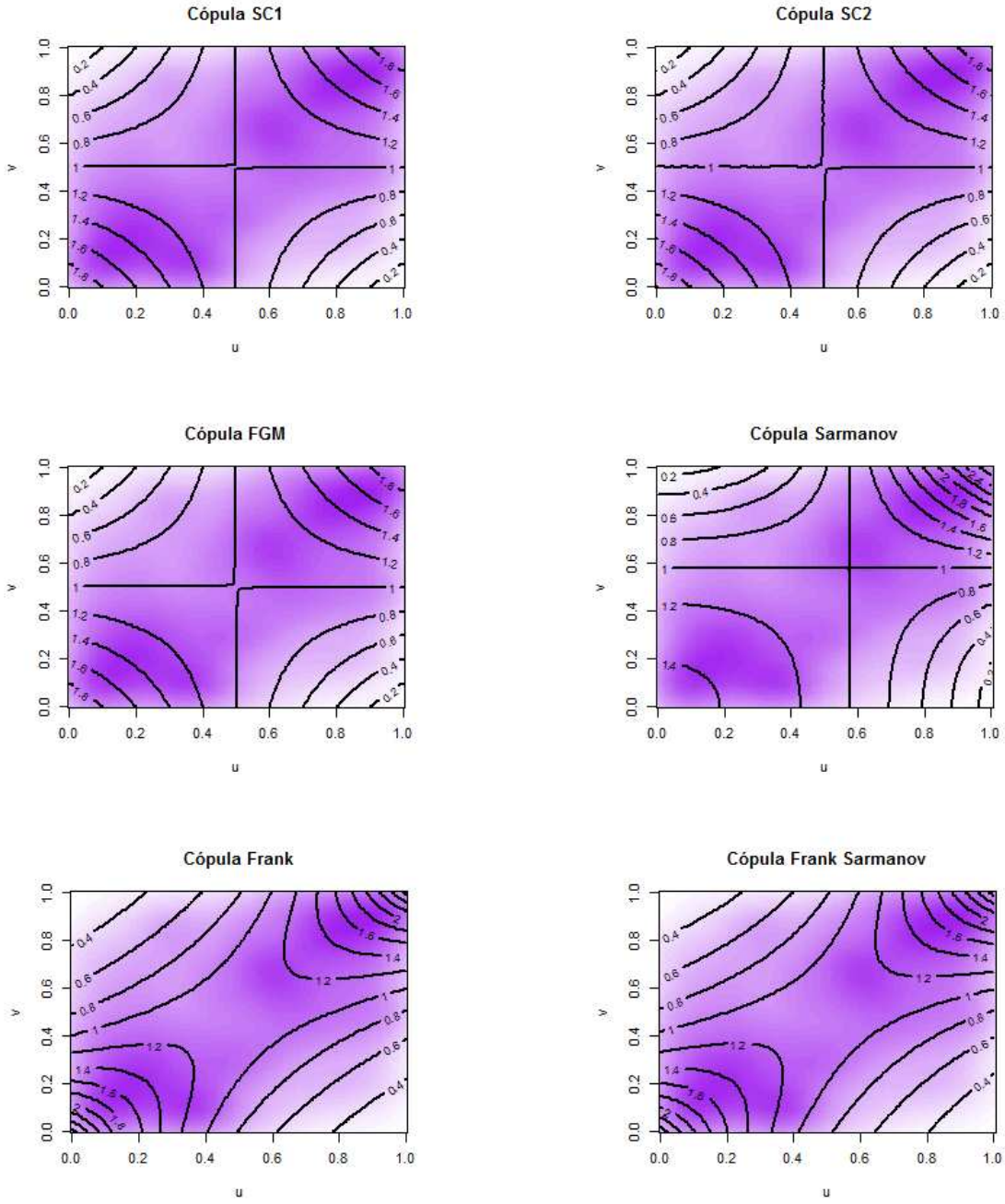

Figura 24 – Gráfico de contorno das densidades conjuntas ajustadas e empÍrica de (PT, CR) para cada um dos modelos possíveis.

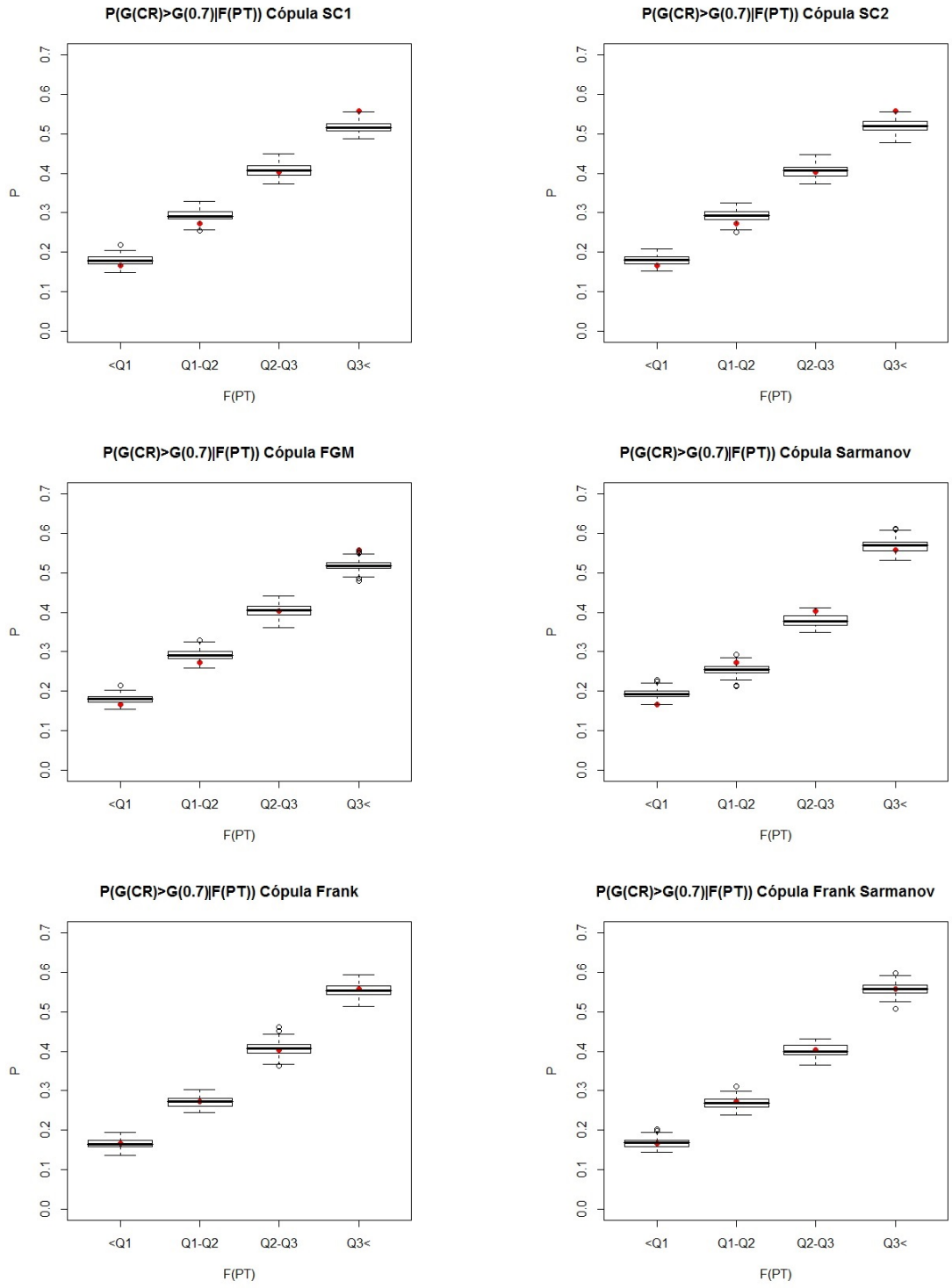

Figura 25 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(PT))$ calculadas a partir do ajuste das cópulas.

Respeito às distâncias, temos uma única decisão, dado que para ambos cálculos a melhor cópula é a cópula de Frank.

Então, mais uma vez devemos selecionar entre a cópula de Frank e a cópula

Tabela 12 – Distâncias entre cópula empírica e a preditiva C para (PT, CR)

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,3694	0,3714	0,5319	0,3724	0,2224	0,222
D. Kolmogorov	0,0283	0,0425	0,0585	0,0283	0,0161	0,0203

Frank Sarmanov. Usando o FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$ e segundo o procedimento descrito para o caso (CN, CR), obtivemos um $\kappa = 0,821$. Então, o e-valor$= 1 - 0,821 = 0,179$ e dado que nossa evidência a favor de H_0 é pequena, decidimos que o melhor modelo é a cópula C_{fs}.

4.4.6 Vestibular Fase 1

Os gráficos de contorno (Figura 26) para este caso, sugerem desconsiderar mais uma vez as cópulas C_{SC1}, C_{SC2} e C_{FGM}, por causa da simetria, e a C_s por que é o pior ajuste. Desta vez, os gráficos das cópulas C_f e C_{fs} são claramente diferentes, sendo o correspondente à C_f aquele que parece se ajustar melhor. Os gráficos das probabilidades, Figura 27, sugerem que tanto C_f quanto C_{fs} dão bons resultados, mas não é possível selecionar apenas um modelo segundo este critério, porque não dá para perceber diferenças entre eles.

Segundo o critério das distâncias, temos dois modelos possíveis, a cópula C_f e a cópula C_{fs}.

Tabela 13 – Distâncias entre cópula empírica e a preditiva C para $(VF1, CR)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,6544	0,6614	0,681	0,6553	0,3994	0,3632
D. Kolmogorov	0,0208	0,0230	0,0419	0,0223	0,0177	0,0188

Agora, vamos selecionar entre a cópula de Frank e a cópula Frank Sarmanov. De novo, usamos o FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$ e obtivemos um $\kappa = 1$ pelo que nosso e-valor é nulo e dado que nossa evidência está contra H_0, decidimos que o melhor modelo é a cópula C_{fs}.

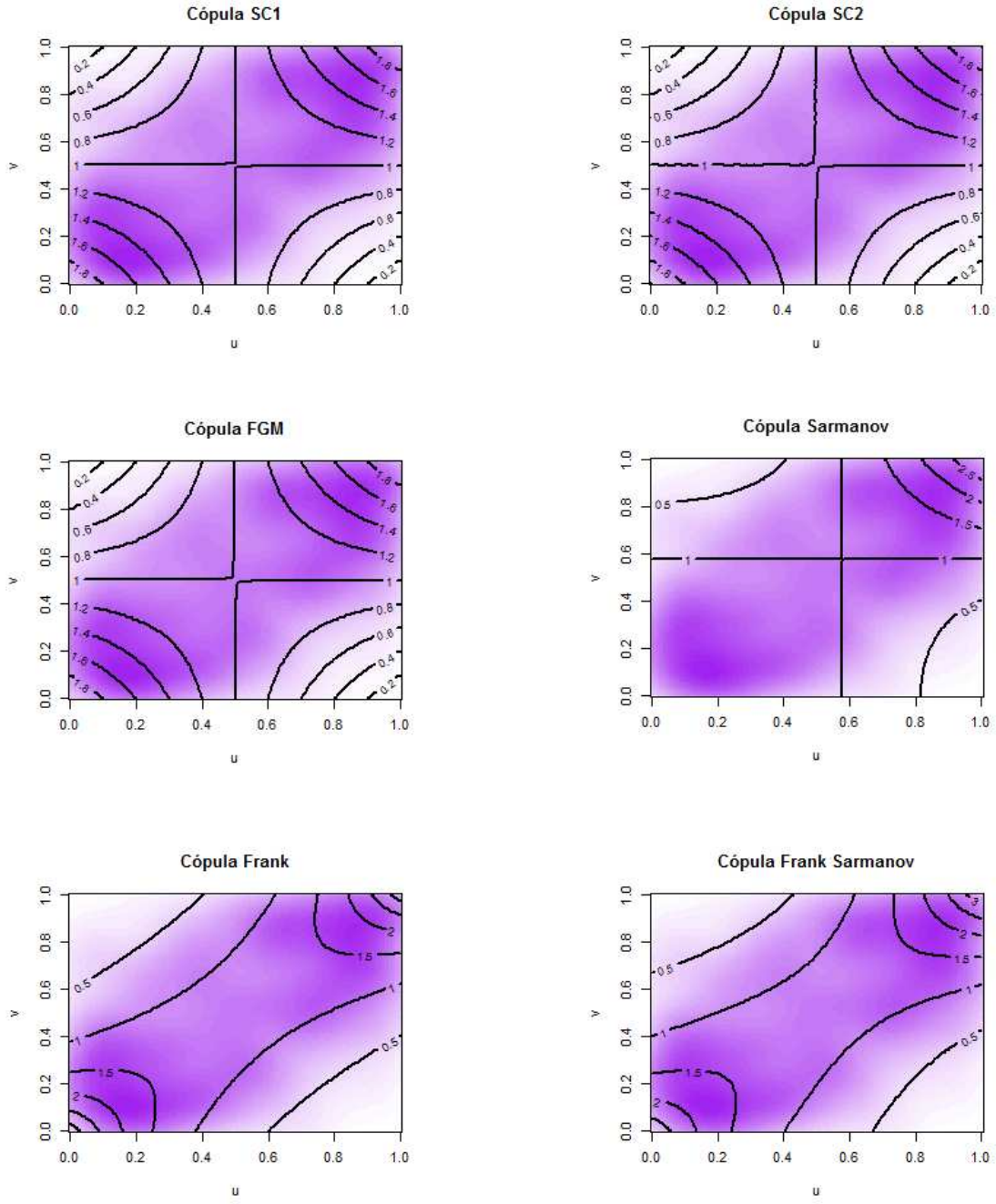

Figura 26 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(VF1, CR)$ para cada um dos modelos possíveis.

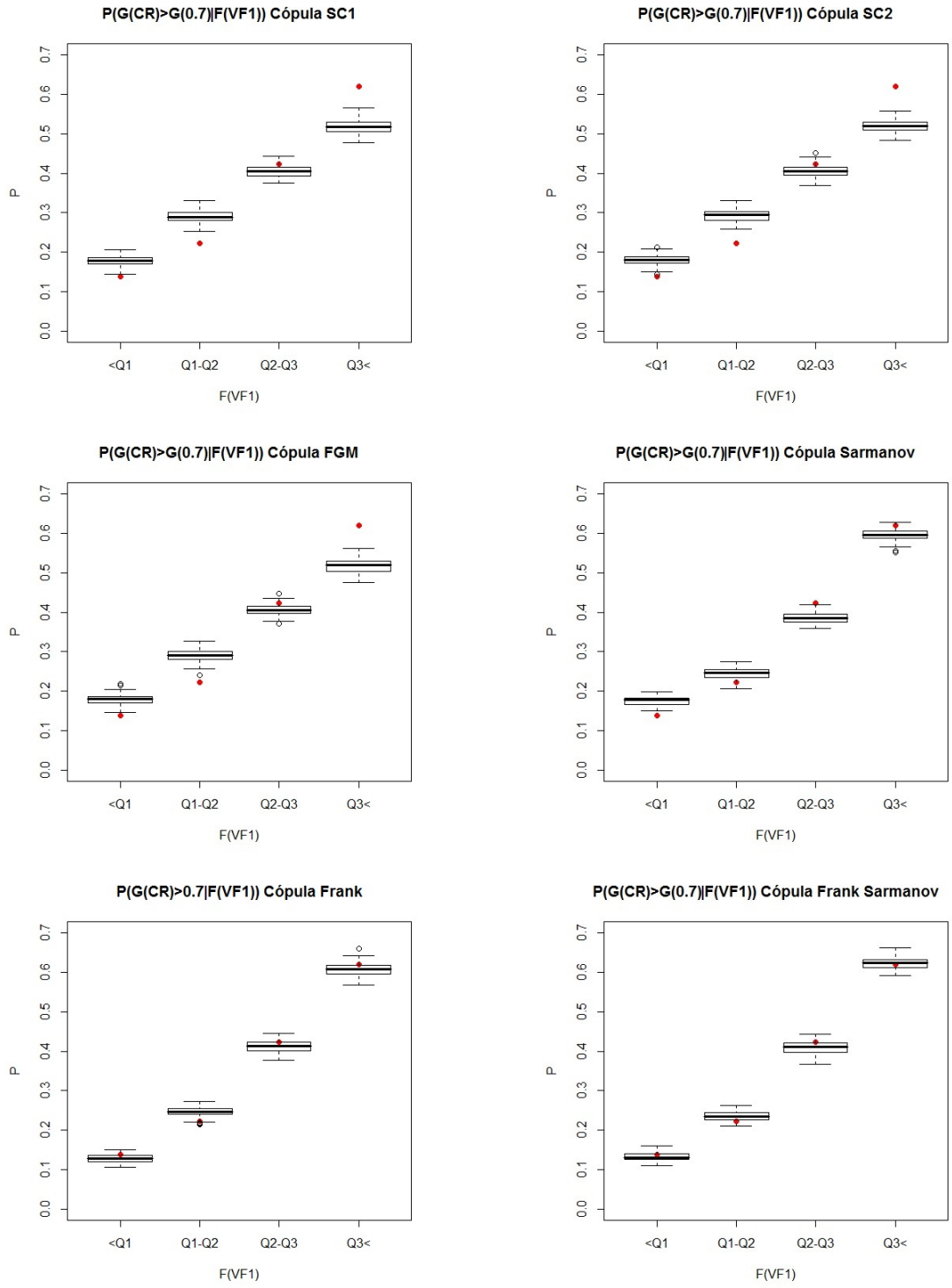

Figura 27 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(VF1))$ calculadas a partir do ajuste das cópulas.

4.4.7 Nota Padronizada

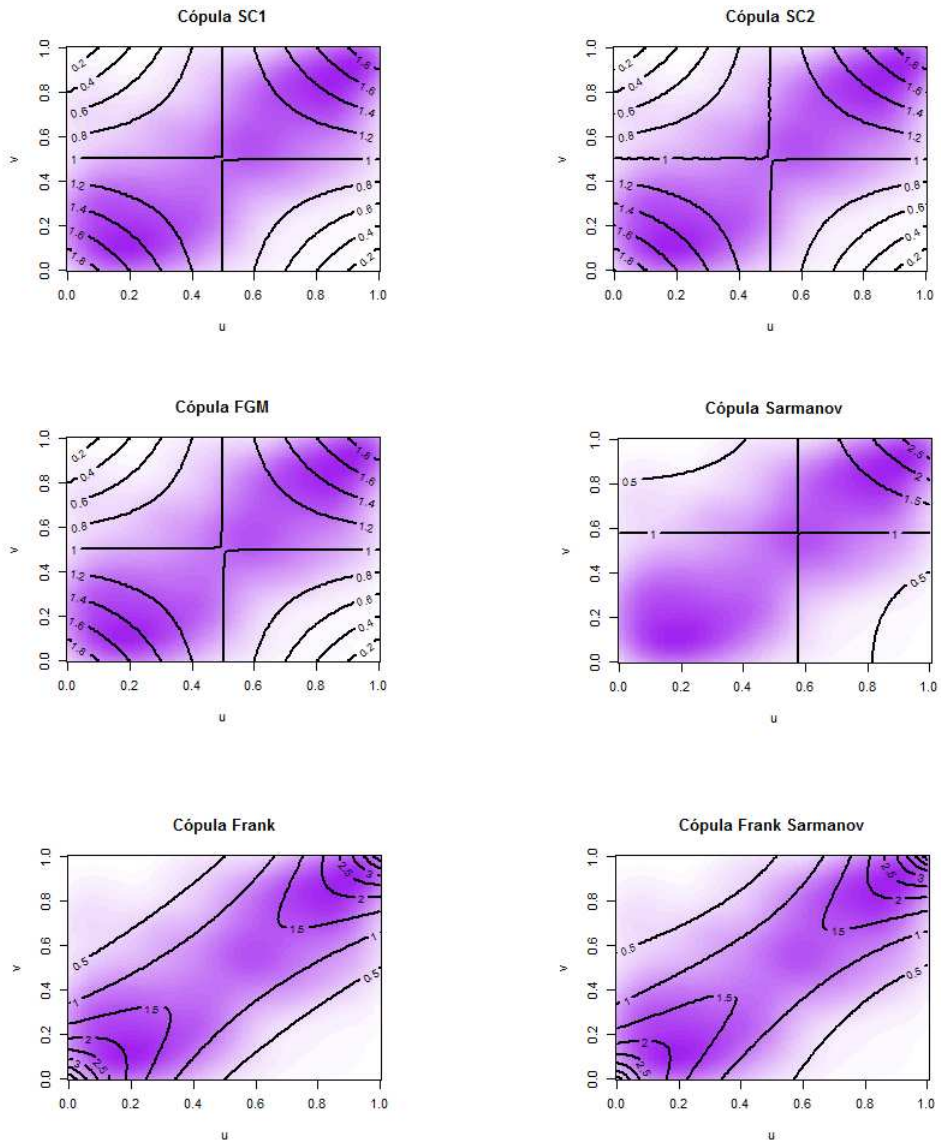

Figura 28 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de (NPT, CR) para cada um dos modelos possíveis.

Finalmente selecionemos o melhor modelo para a nota padronizada (NPT). Com respeito à Figura 28, encontramos um cenário parecido com o achado para os casos (CN, CR) e (MA, CR), onde é claro que os melhores ajustes se dão para as cópulas C_f e C_{fs}.

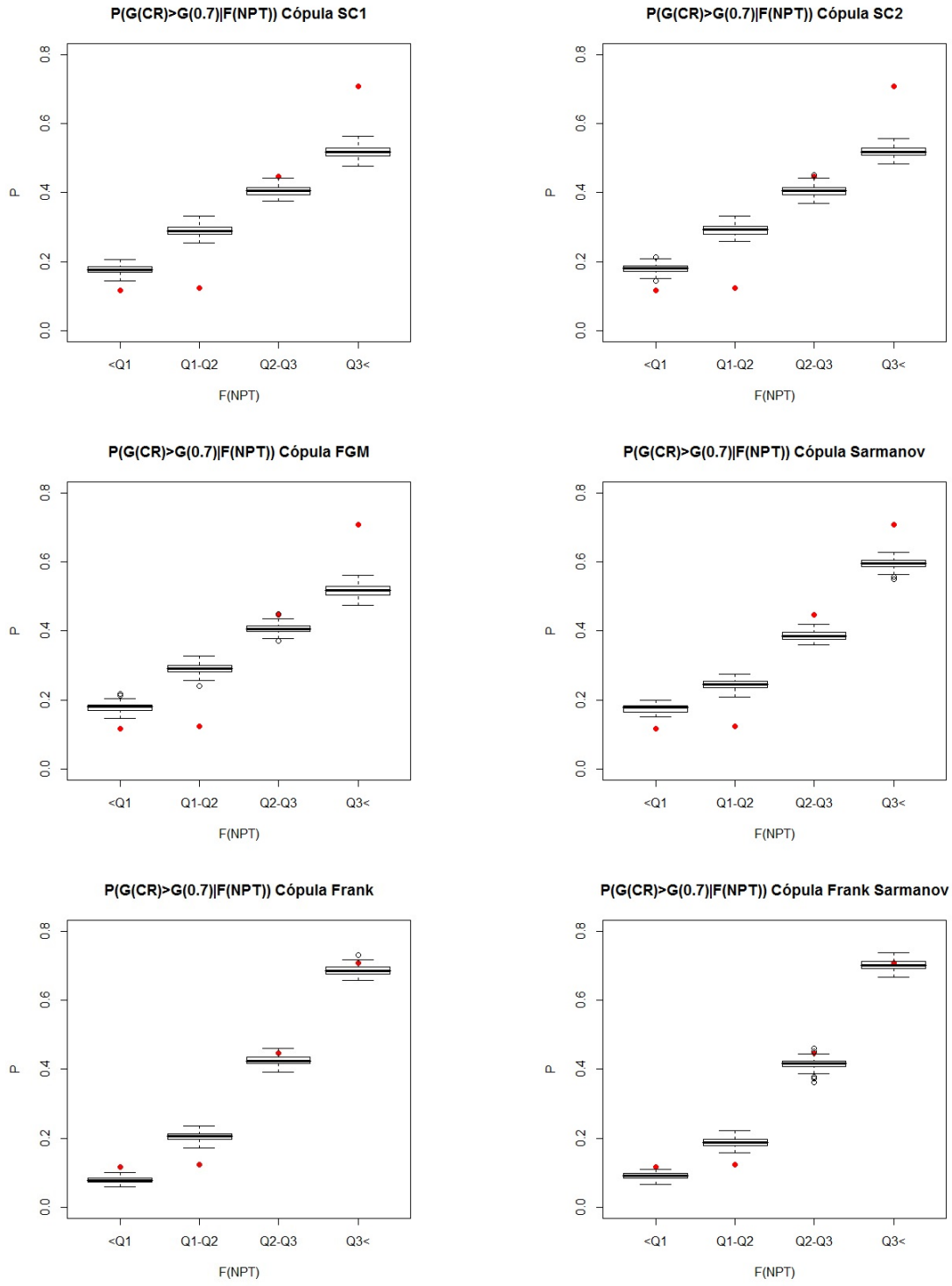

Figura 29 – Boxplots probabilidades $P(G(CR) \geqslant G(0.7)|F(NPT))$ calculadas a partir do ajuste das cópulas.

Analisando as distâncias, tanto a distância de Hellinger quanto a distância de Kolmogorov sugerem que o melhor modelo seria a cópula C_{fs}.

Nos gráficos da Figura 29 é claro que os melhores modelos são as cópulas C_f e

Tabela 14 – Distâncias entre cópula empírica e a preditiva C para (NPT, CR)

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	1,1485	1,1508	1,156	1,1513	0,5246	0,4829
D. Kolmogorov	0,0605	0,0605	0,0571	0,0605	0,0337	0,0308

C_{fs}, devido a que nesses casos os pontos vermelhos estão dentro das caixas em pelo menos três dos quatro intervalos de $F(NPT)$. Embora as diferenças entre C_f e C_{fs} não sejam tão facilmente identificáveis a partir do gráfico, podemos dizer que para este critério a C_{fs} é o melhor modelo.

Verifiquemos que nossa escolha é a melhor. Dado que a cópula C_f é um caso especial e mais simples da C_{fs}, consideramos as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$ e usamo a metodologia proposta para o uso do FBST para testá-las. Após os cálculos correspondentes, encontramos que $\kappa = 1$, pelo que o nosso e-valor é nulo e concluímos que não temos evidência a favor da H_0. Finalmente, selecionamos à C_{fs} como o melhor modelo neste caso.

4.5 Seleção da melhor cópula: MA111

Seguindo com a metodologia proposta no Capítulo 3, e da mesma forma como fizemos na Seção 4.4, vamos selecionar dentre as seis cópulas possíveis aquela que apresenta melhor ajuste para cada par de variáveis formado entre as provas do Vestibular e a nota em MA111-Cálculo I.

4.5.1 Ciências da Natureza

Nesta oportunidade, vamos a apresentar gráficos análogos aos apresentados para cada prova do Vestibular da Seção 4.4. Inicialmente temos a Figura 30.

O pior modelo neste sentido é a cópula Sarmanov, dado que não consegue modelar o comportamento da densidade conjunta das variáveis $(CN, MA111)$ quando tomam valores pequenos. Finalmente ficamos com as cópulas, C_f e C_{fs}, que pelo menos graficamente não são claramente diferenciáveis. Agora, vamos analisar a capacidade que tem as cópulas para modelar o valor de uma probabilidade.

A Figura 31 mostra os boxplot da probabilidade $P(G(MA111) \geqslant G(5)|u_1 \leqslant F(CN) \leqslant u_2)$, onde lembramos que o ponto vermelho é o valor real de dita probabilidade. Temos que, em geral, nenhum dos modelos conseguiu modelar inteiramente a probabilidade

de interesse, mas, os boxplot dos modelos Frank e Frank Sarmanov, têm mediana que coincide com o ponto vermelho no intervalo $0 - Q_1$, e para os outros intervalos, a mediana está mais perto do ponto vermelho do que para os outros modelos.

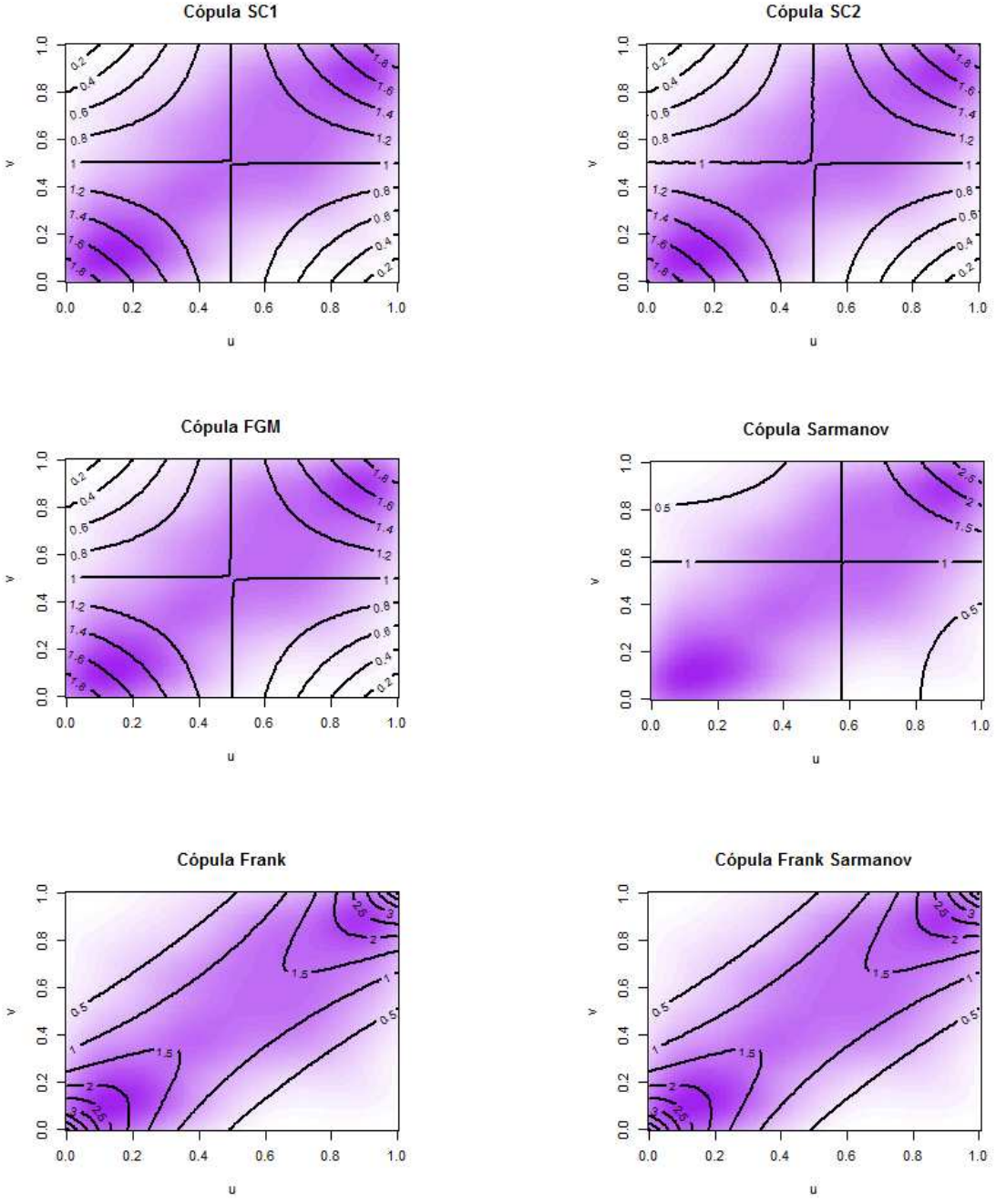

Figura 30 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(CN, MA111)$ para cada um dos modelos possíveis.

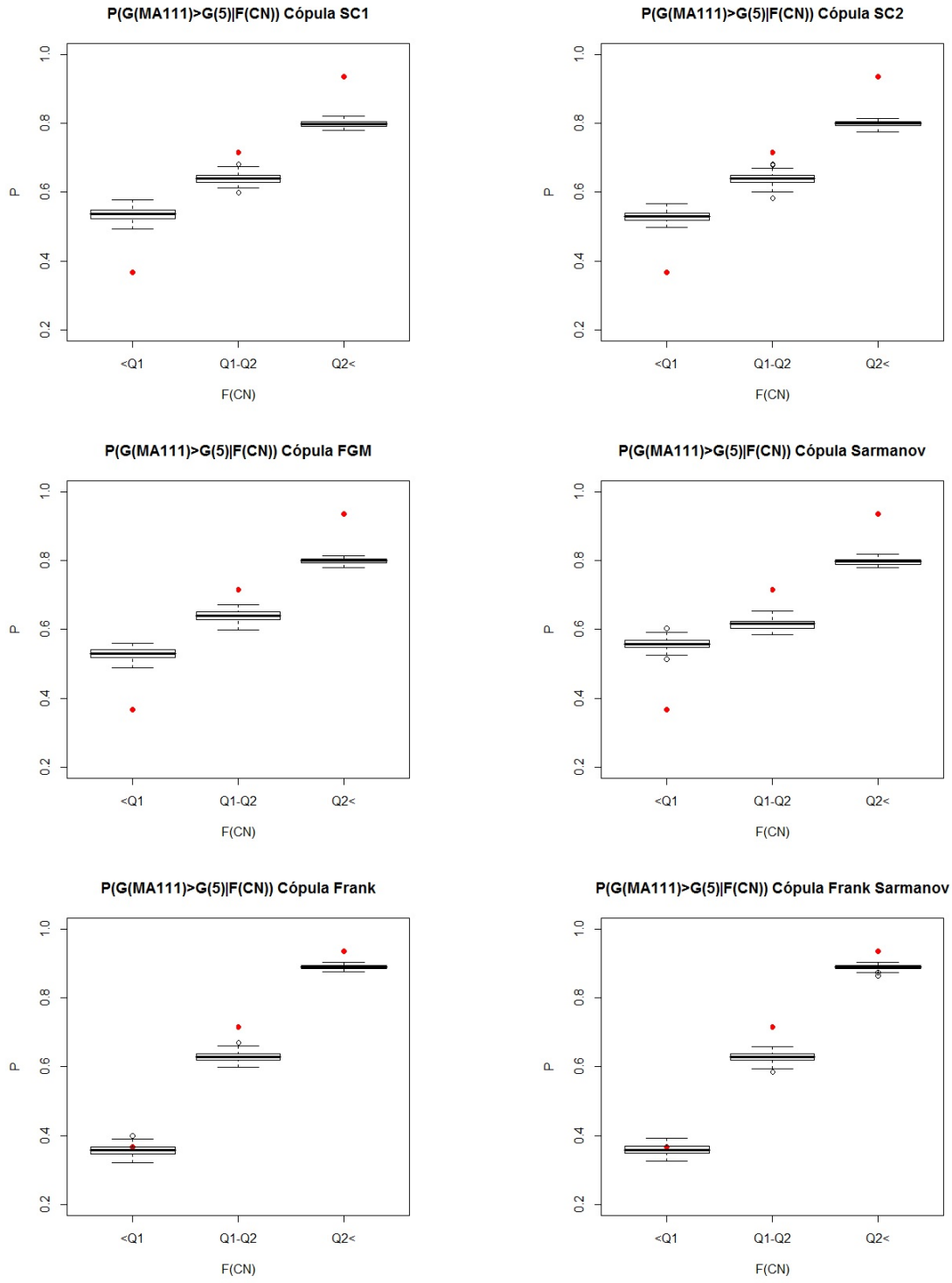

Figura 31 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(CN))$ calculadas a partir do ajuste das cópulas.

Tabela 15 – Distâncias entre cópula empírica e a preditiva C caso $(CN, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	1,11	1,111	1,2227	1.113	0,3397	0,3357
D. Kolmogorov	0,073	0,073	0,083	0,073	0,0303	0,0303

A Tabela 15, mostra as distâncias de Hellinger e Kolmogorov entre a cópula empírica e sua correspondente preditiva. Usando o critério da distância mínima, os modelos escolhidos são Frank e Frank Sarmanov.

Finalmente, para selecionar entre estes dois modelos, vamos utilizar um FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$, como explicado na Seção 4.4. Dessa forma encontramos que $\kappa = 0,001$ e o e-valor$= 1 - 0,001 = 0,999$ pelo que não temos evidência suficiente para rejeitar a hipótese nula e concluímos que o melhor modelo para o par $(CN, MA111)$ é C_f.

4.5.2 Ciências Humanas

Para esta variável podemos afirmar, a partir da Tabela 16 que, em termos da distância de Hellinger, o melhor modelo é o Frank Sarmanov. No caso da distância de Kolmogorov, vemos que o valor para as cópulas C_f e C_{fs} é igual, pelo que não podemos dizer qual deles é melhor usando esse critério.

Tabela 16 – Distâncias entre cópula empírica e a preditiva C caso $(CH, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,5742	0,5928	0,7877	0,5768	0,3449	0,3387
D. Kolmogorov	0,0434	0,0434	0,0558	0,0434	0,0293	0,0293

Temos da Figura 32, que os gráficos correspondentes às cópulas C_{SC1}, C_{SC2} e C_{FGM} são equivalentes e mais uma vez a cópula C_s é a pior de todas. Dos modelos Frank e Frank Sarmanov podemos dizer que, embora sendo diferentes entre eles são coerentes com a densidade empírica.

Na Figura 33, vemos que unicamente os modelos Frank e Frank Sarmanov conseguem que pelo menos uma das caixas contenha o ponto vermelho, mas como para a variável anterior o ajuste para os intervalos $Q_1 \leqslant F(CH) \leqslant Q_2$ e $Q_2 \leqslant F(CH) \leqslant 1$ segue sem ser satisfatório.

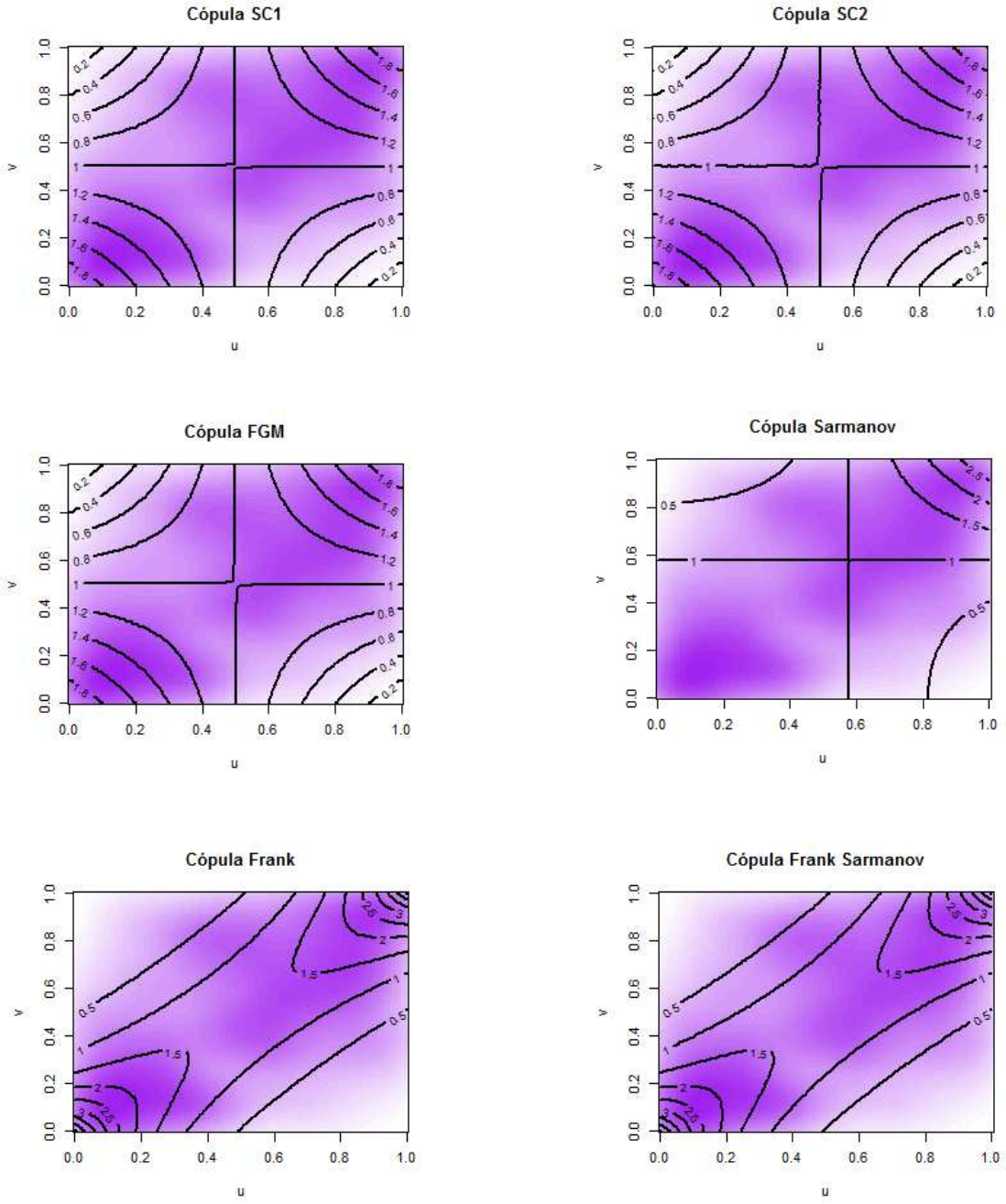

Figura 32 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(CH, MA111)$ para cada um dos modelos possíveis.

Seguindo todos os critérios, temos que qualquer um dos modelos Frank ou Frank Sarmanov é adequado, mas até o momento não conseguimos estabelecer uma diferença clara entre eles. Para finalmente escolher uma destas opções, vamos usar a metodologia do FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$, seguindo o mesmo procedimento feito para os casos da Seção 4.4. É assim, como obtemos $\kappa = 0,002$ pelo que nosso e-valor$= 0,998$ e concluímos que a melhor cópula para este caso é C_f.

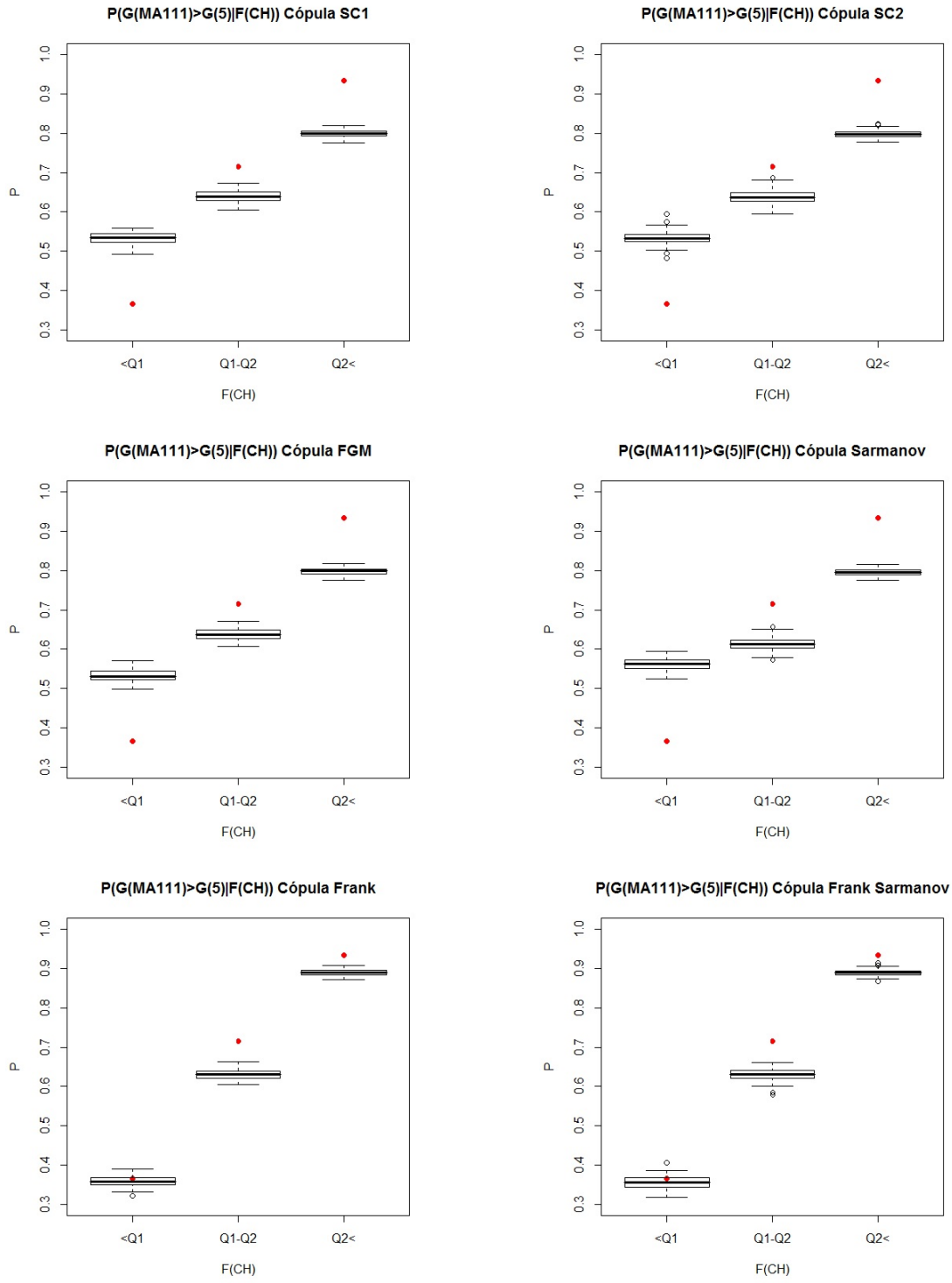

Figura 33 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(CH))$ calculadas a partir do ajuste das cópulas.

4.5.3 Inglês

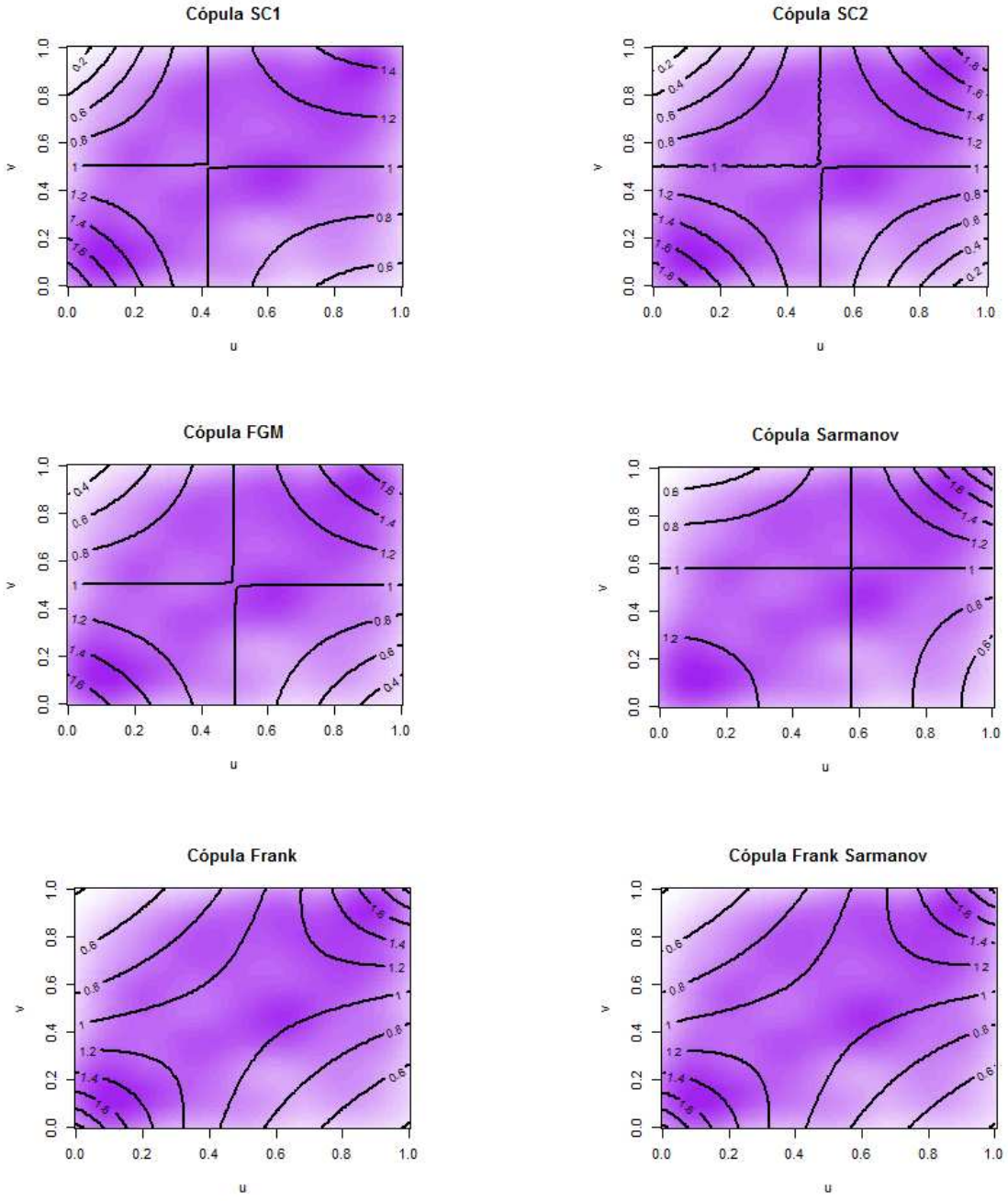

Figura 34 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(ING, MA111)$ para cada um dos modelos possíveis.

Neste caso, vemos na Figura 34, que é claro que o pior ajuste é aquele correspondente ao modelo Sarmanov, isto porque as curvas de nível desse gráfico mostram uma assimetria extrema que não mostram os dados. Podemos dizer que os dados presentam uma assimetria sim, mas é leve pelo que escolheríamos a cópula C_{SC1} por enquanto, embora os modelos Frank e Frank Sarmanov também parecem coerentes.

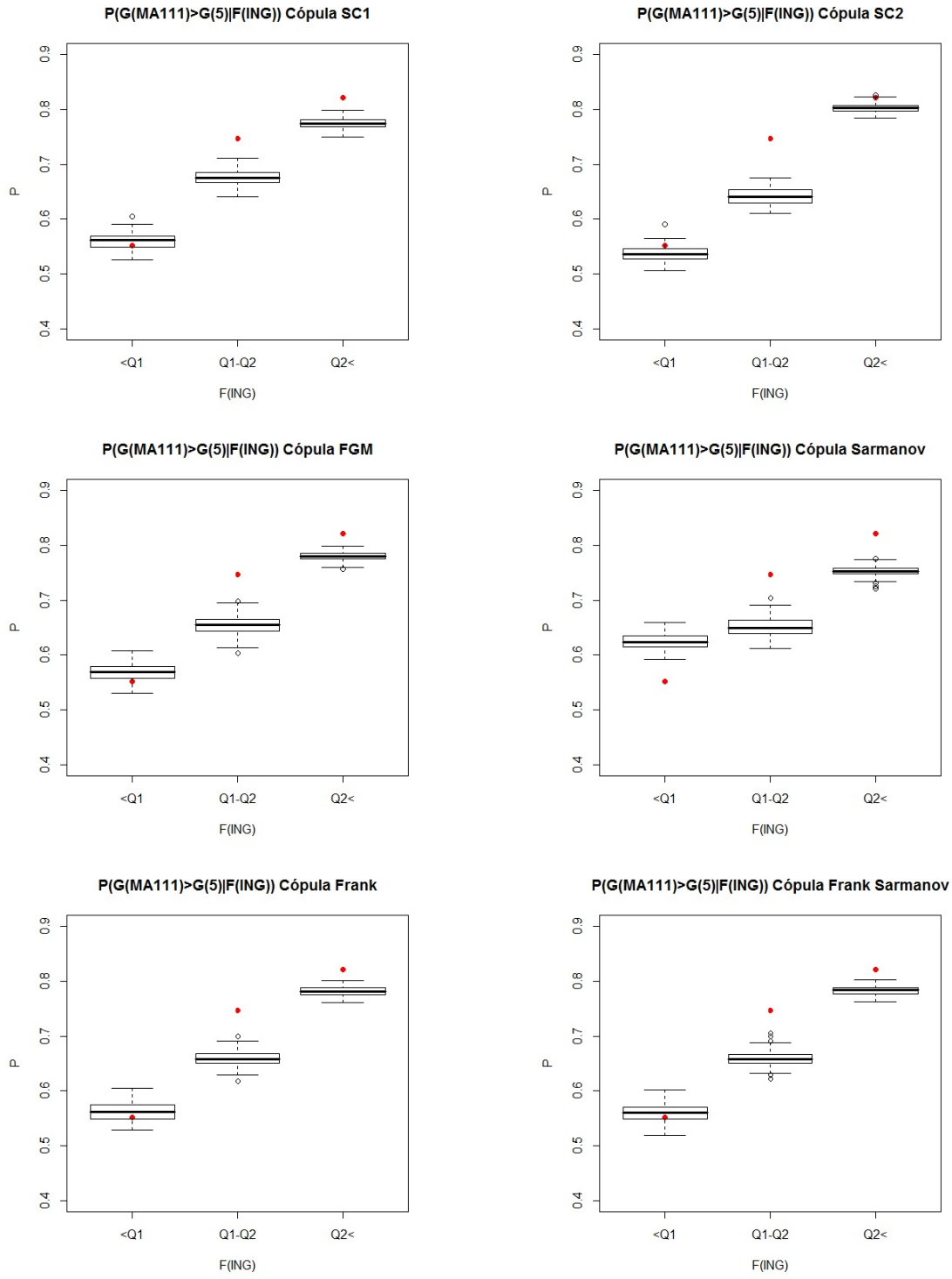

Figura 35 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(ING))$ calculadas a partir do ajuste das cópulas.

Na Figura 35, encontramos que a única cópula cujo ajuste da probabilidade condicional é claramente insatisfatório é a cópula de Sarmanov, isto porque o seu gráfico é o único onde os pontos vermelhos não coincidem com nenhum dos boxplot. Os outros são todos parecidos, pelo que é difícil identificar unicamente uma cópula como a melhor de

todas. No entanto, poderíamos dizer que o melhor é o modelo C_{SC2}, porque dois dos três pontos vermelhos caem dentro dos boxplot correspondentes.

Dados os resultados obtidos das distância da Tabela 17, podemos dizer que tanto para a Hellinger quanto a Kolmogorov as cópulas com menores distâncias são a C_{SC1}, Frank e Frank Sarmanov, as mesmas escolhidas a partir dos gráficos da Figura 34. Nesse sentido, diremos que o melhor modelo será aquele mais simples de implementar.

Tabela 17 – Distâncias entre cópula empírica e a preditiva C caso $(ING, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,2875	0,3433	0,5294	0,3437	0,3184	0.2887
D. Kolmogorov	0,031	0,0336	0,0453	0,0342	0,0293	0,0293

Para finalmente escolher uma destas opções, vamos primeiro usar a metodologia do FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$, para saber se podemos ficar apenas com o modelo Frank. Assim, como obtemos $\kappa = 0,878$ o e-valor$= 1 - 0,878 = 0,122$, pelo que é melhor ficar com o modelo Frank Sarmanov e portanto, como este tem mais parâmetros do que o modelo C_{SC1}, concluímos que a melhor cópula para este caso é C_{SC1}.

4.5.4 Matemática

Para matemática temos, em termos das distâncias da Tabela 18, que no caso da distância de Hellinguer o melhor modelo seria a cópula Frank Sarmanov, enquanto que no caso da distância de Kolmogorov diremos que o melhor modelo seria a cópula de Frank.

Tabela 18 – Distâncias entre cópula empírica e a preditiva C para $(MA, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	1,17	1,171	1,2936	1,1723	0,4212	0,417
D. Kolmogorov	0,066	0,066	0,075	0,066	0,0238	0,0227

Olbservando os gráficos de contorno da Figura 36, encontramos um cenário similar ao encontrado no caso de $(CN, MA111)$, ou seja, as cópulas C_{SC1}, C_{SC2}, C_{FGM} e C_s podem ser desconsideradas, dado que nenhuma parece se ajustar bem aos dados, pelo que continuamos a nossa analise para as cópulas C_f e C_{fs}.

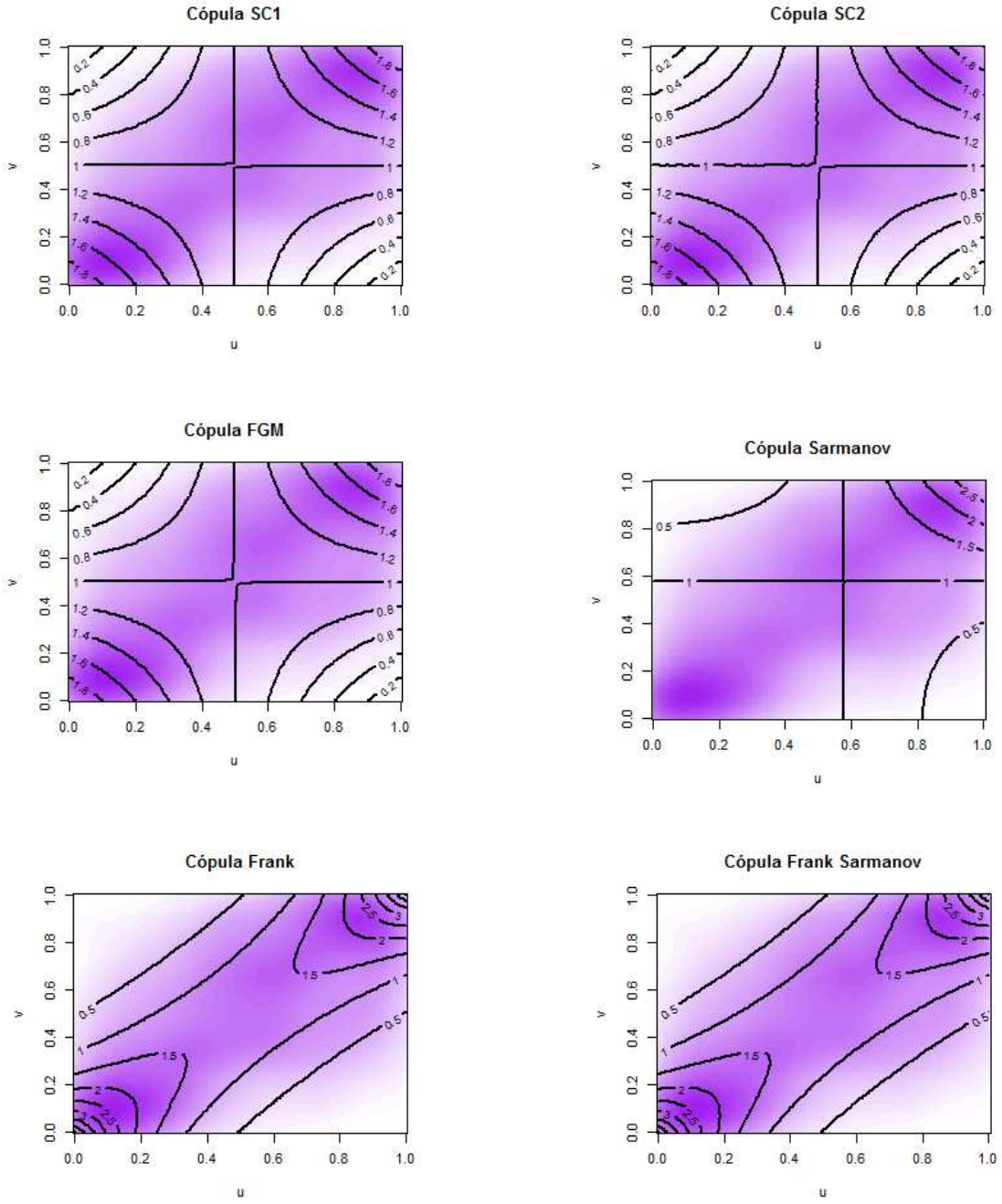

Figura 36 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(MA, MA111)$ para cada um dos modelos possíveis.

Com respeito à Figura 37, temos que o melhor ajuste, no sentido dessa probabilidade de interesse, é dado por os modelos Frank e Frank Sarmanov, devido a que correspondem aos únicos gráficos onde pelo menos um dos pontos vermelhos está dentro das caixas dos boxplot.

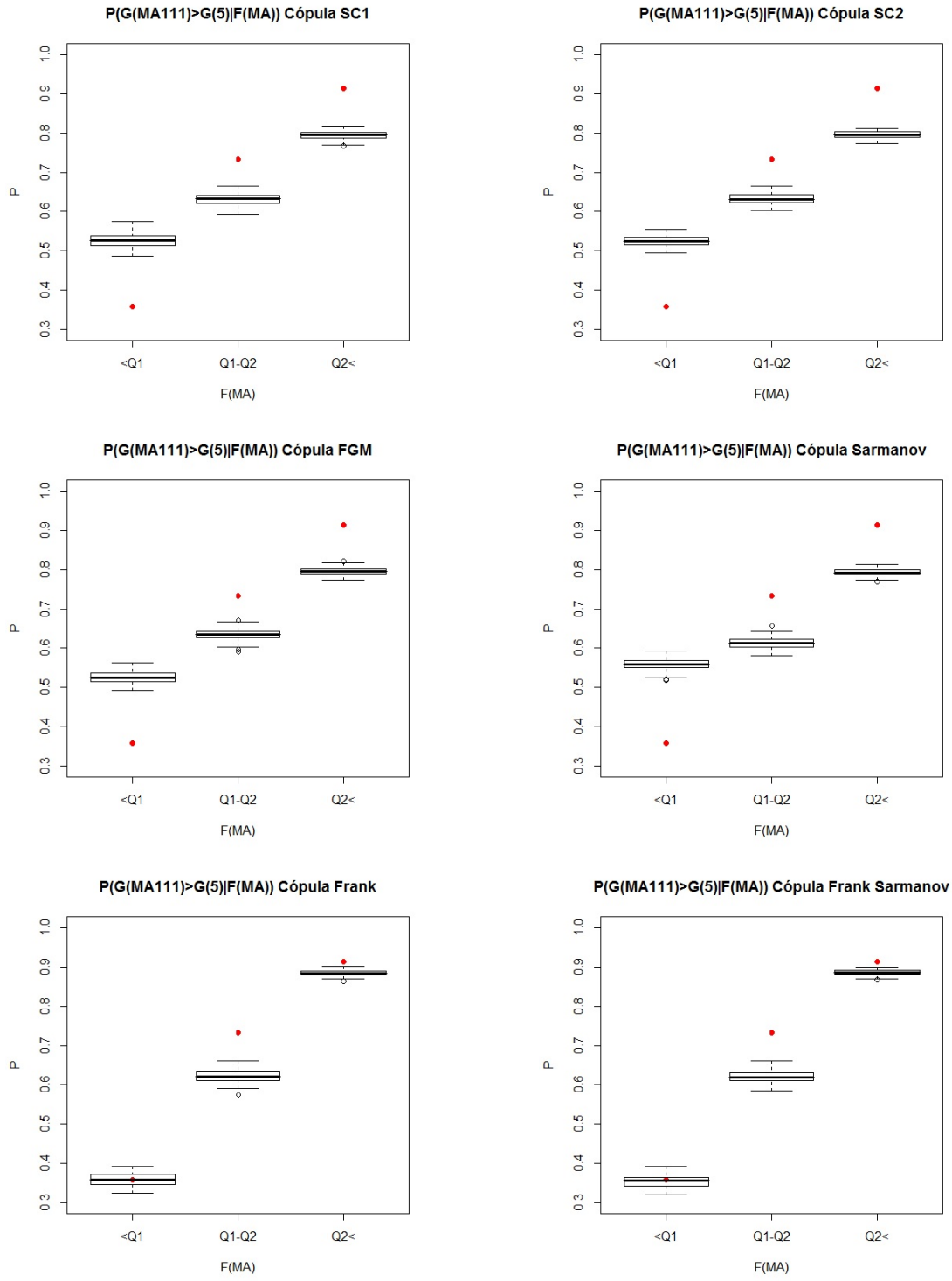

Figura 37 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(MA))$ calculadas a partir do ajuste das cópulas.

Dado o resultado achado no cálculo das distâncias, vamos testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$, seguindo o mesmo procedimento do FBST. É assim, como obtemos $\kappa = 0,002$ pelo que nosso e-valor$= 0,998$, e finalmente decidimos que o melhor modelo é a cópula C_f.

4.5.5 Português

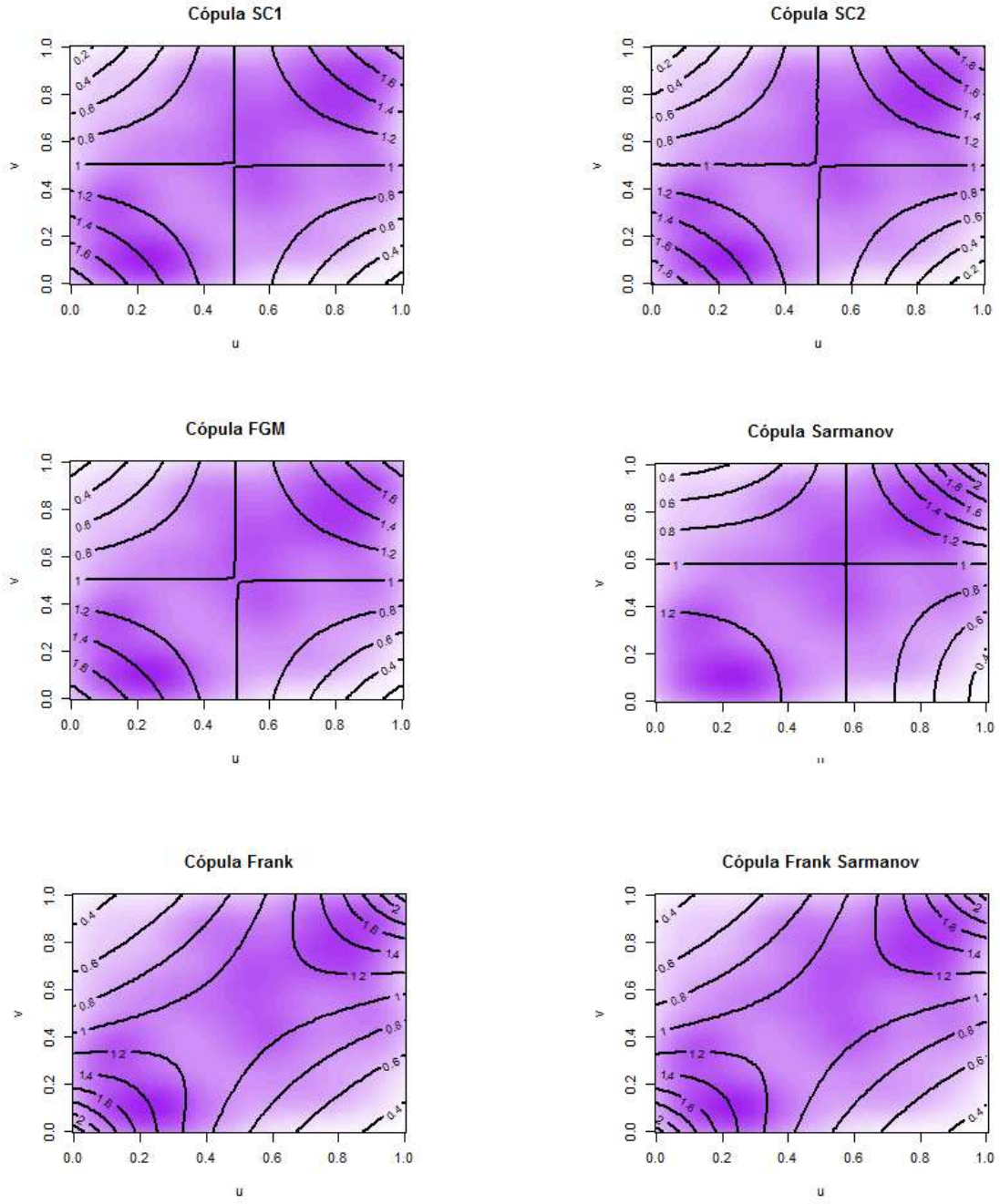

Figura 38 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(PT, MA111)$ para cada um dos modelos possíveis.

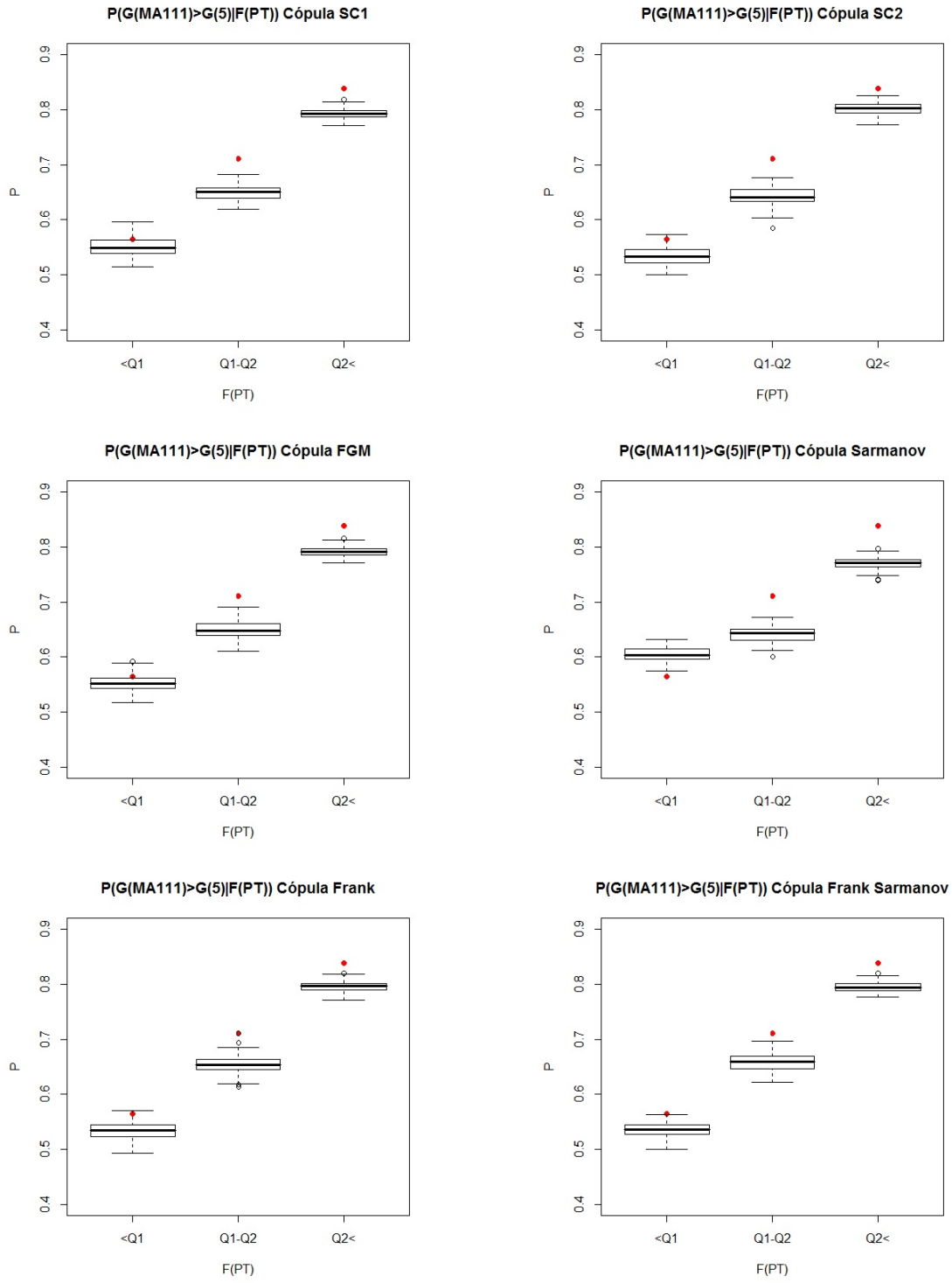

Figura 39 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(PT))$ calculadas a partir do ajuste das cópulas.

A partir da Figura 38, temos que mais uma vez o pior ajuste é dado pela C_s. Dessa vez podemos dizer que, os gráficos correspondentes às cópulas C_{SC1} e C_{SC2} são equivalentes, mas o ajuste não parece o melhor. Sob este critério, o ajuste mais coerente graficamente, é aquele dado pelas cópulas C_f e C_{fs}. Dos gráficos da Figura 39, podemos

dizer que em geral nenhum dos ajustes é satisfatorio.

Com respeito às distâncias, temos uma única decisão, dado que para ambos os cálculos a melhor cópula é a cópula de Frank Sarmanov.

Tabela 19 – Distâncias entre cópula empírica e a preditiva C para $(PT, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D, Hellinger	0,3237	0,306	0,568	0,3228	0,2682	0,2683
D. Kolmogorov	0,0185	0,0185	0,0327	0,0214	0,024	0,0217

Então, mais uma vez estamos na dúvida entre a cópula de Frank e a cópula Frank Sarmanov. Usando o FBST para testar as hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$ e segundo o procedimento descrito para o caso (CN, CR) na Seção 4.4 obtivemos um $\kappa = 0,001$, então o e-valor$= 1 - 0,001 = 0,999$, indicando forte evidência a favor da H_0 e por isso decidimos que o melhor modelo é a cópula C_f.

4.5.6 Vestibular Fase 1

Os gráficos de contorno (Figura 40) para este caso, sugerem desconsiderar às cópulas C_{SC1}, C_{SC2} e C_{FGM}, pois neles o decrescimento nos extremos superior esquerdo e inferior direito é lento, contrário ao mostrado pelos dados. Também desconsideramos à C_s porque é o pior ajuste. Desta vez, os gráficos das cópulas C_f e C_{fs} são praticamente equivalentes, pelo que podemos dizer que o melhor ajuste é um deles mas não podemos especificar qual. Os gráficos da Figura 41 sugerem que os modelos C_f e C_{fs} dão os melhores resultados, mas não é possível selecionar apenas um modelo segundo este critério, porque não dá para perceber diferenças entre eles.

Segundo o critério das distâncias, o melhor modelo em ambos os casos é a cópula C_f.

Tabela 20 – Distâncias entre cópula empírica e a preditiva C para $(VF1, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	0,6969	0,7014	0,8729	0,7004	0,3791	0,3862
D. Kolmogorov	0,049	0,049	0,0601	0,0488	0,025	0,0252

Dado que segundo todos os critérios a cópula de Frank é a vencedora, selecionamos este modelo como o melhor para este caso.

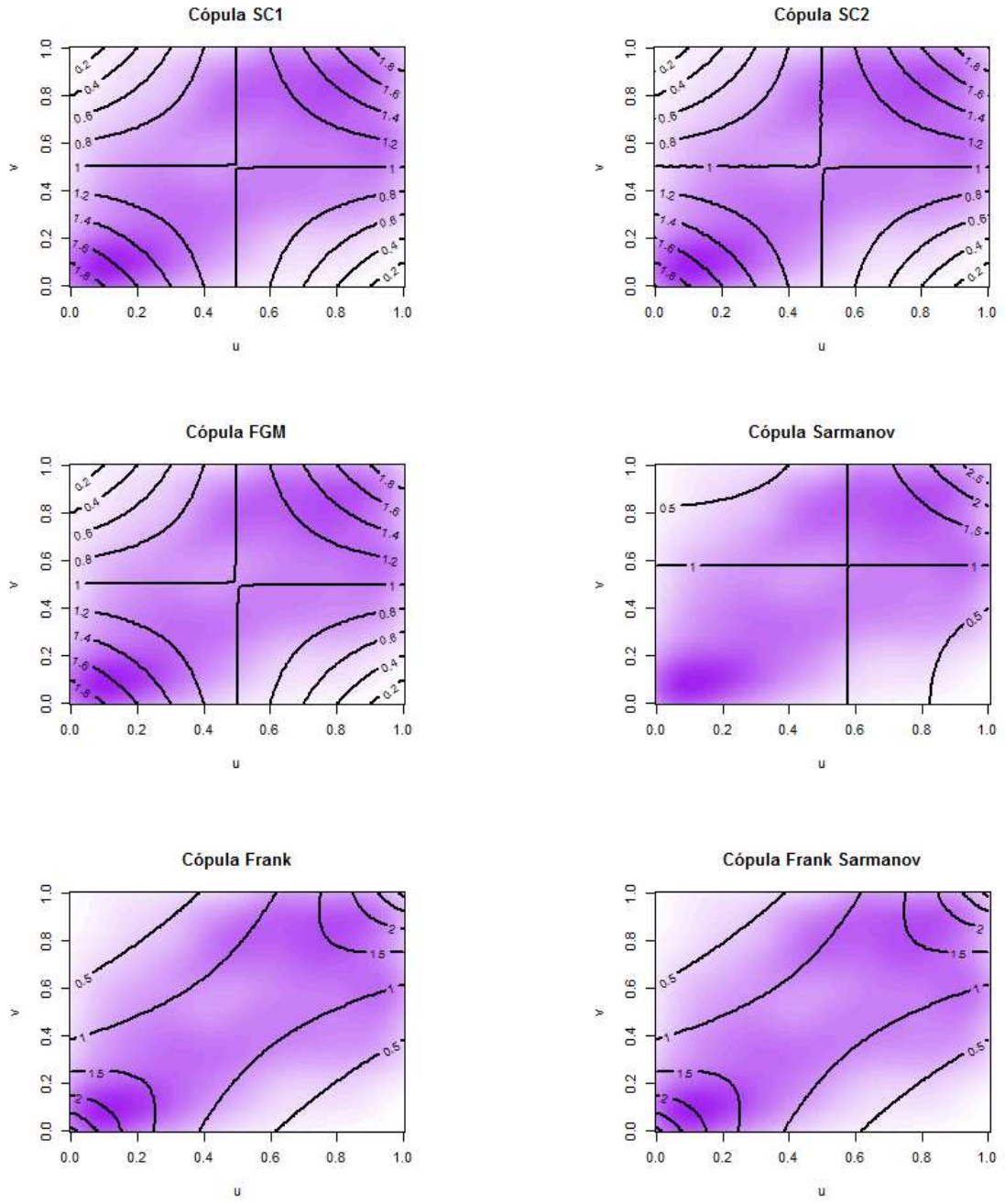

Figura 40 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(VF1, MA111)$ para cada um dos modelos possíveis.

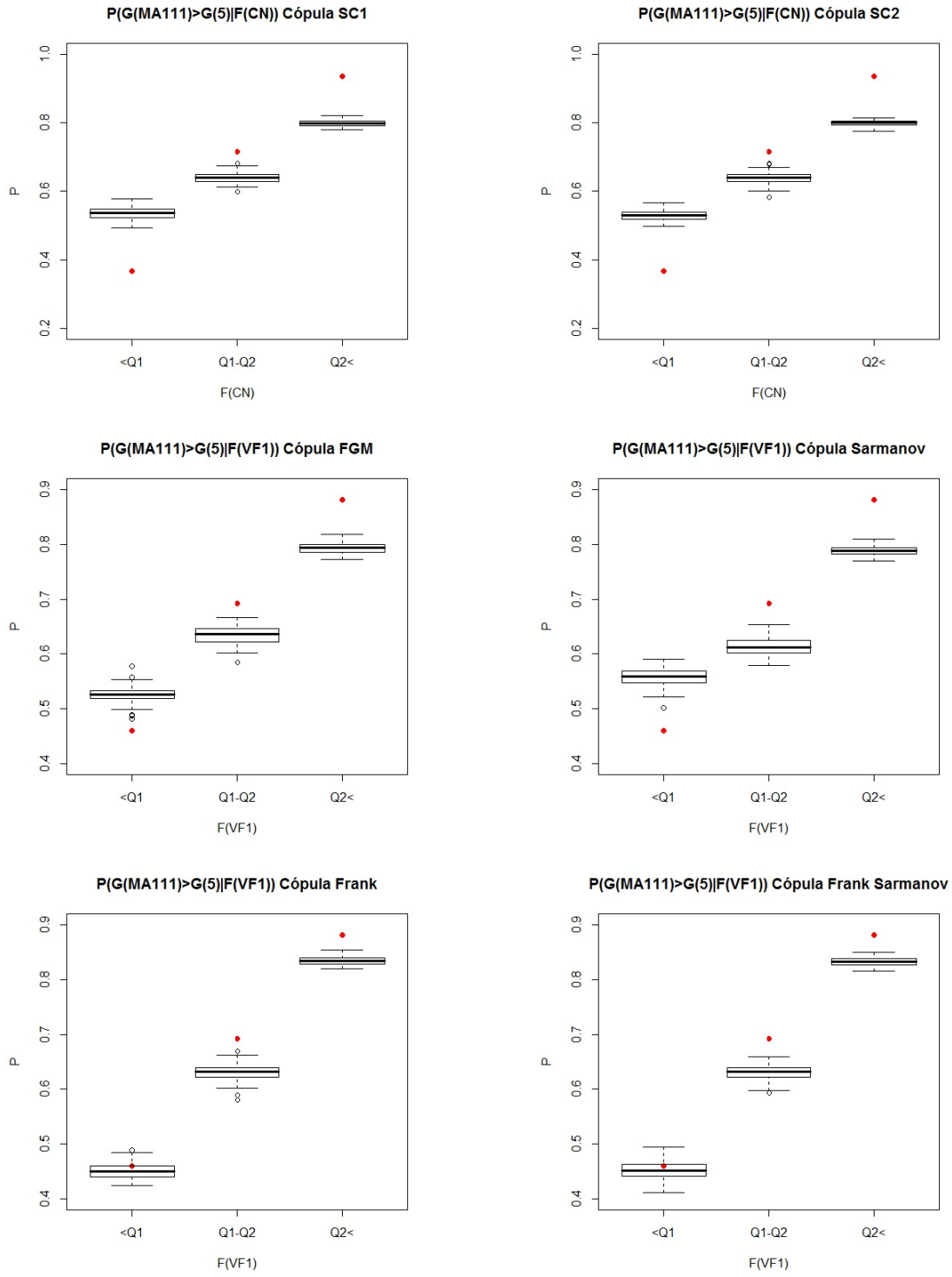

Figura 41 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(VF1))$ calculadas a partir do ajuste das cópulas.

4.5.7 Nota Padronizada

Finalmente fazemos a análise para a nota padronizada (NPT). Olhando a Figura 42, encontramos um cenário parecido a dos $(CN, MA111)$ e $(MA, MA111)$, onde

é claro que os melhores ajustes se dão para as cópulas C_f e C_{fs}.

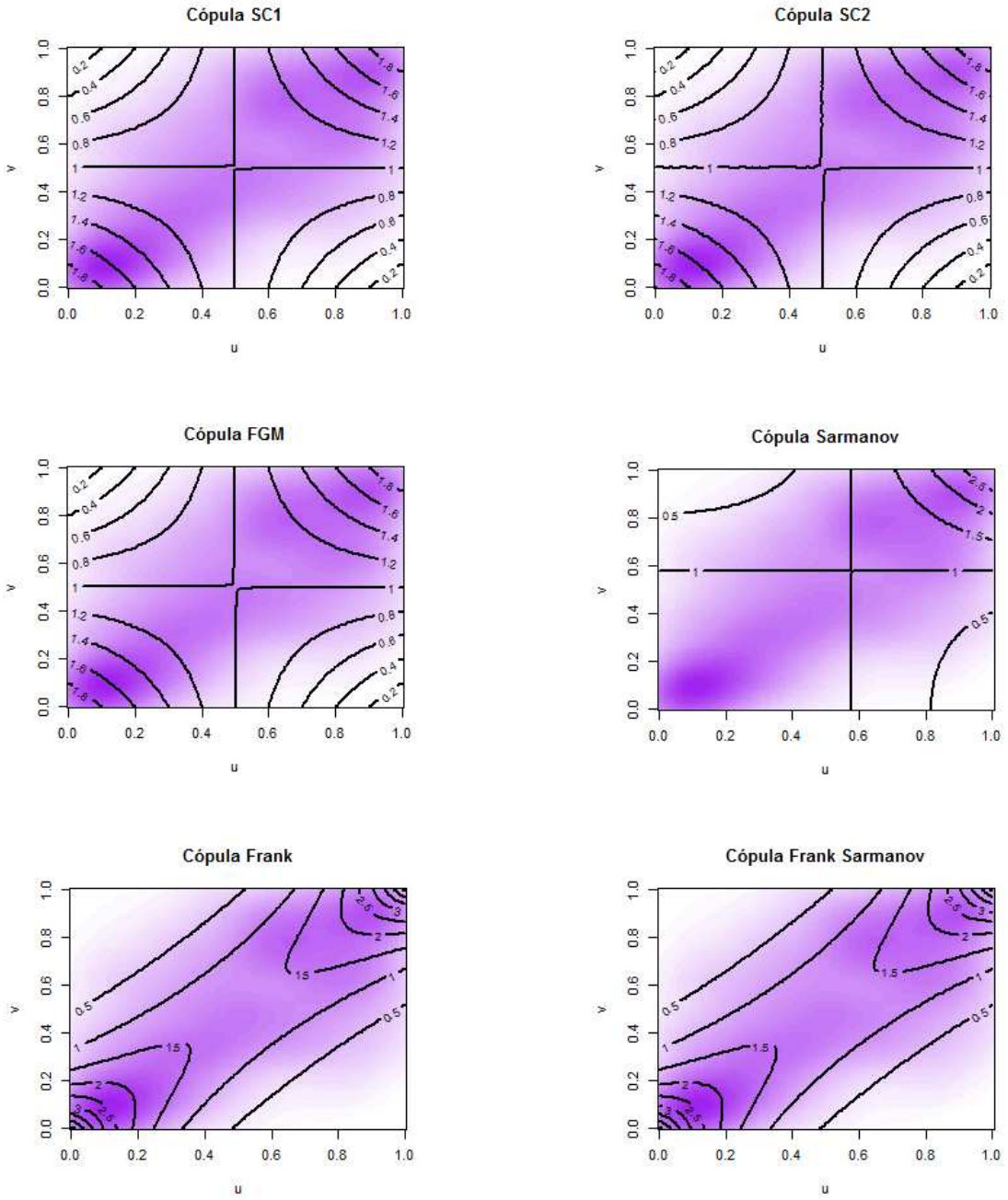

Figura 42 – Gráfico de contorno das densidades conjuntas ajustadas e empírica de $(NPT, MA111)$ para cada um dos modelos possíveis.

Nos gráficos da Figura 43, é claro que os melhores modelos são as cópulas C_f e C_{fs}, porque unicamente nesses casos pelo menos um ponto vermelho está dentro do boxplot correspondente. Embora as diferenças entre C_f e C_{fs} não sejam tão facilmente identificáveis, podemos dizer que a C_f é o melhor modelo, pensando em que pelo menos a partir deste critério o incremento no número de parâmetros não fez muito diferença.

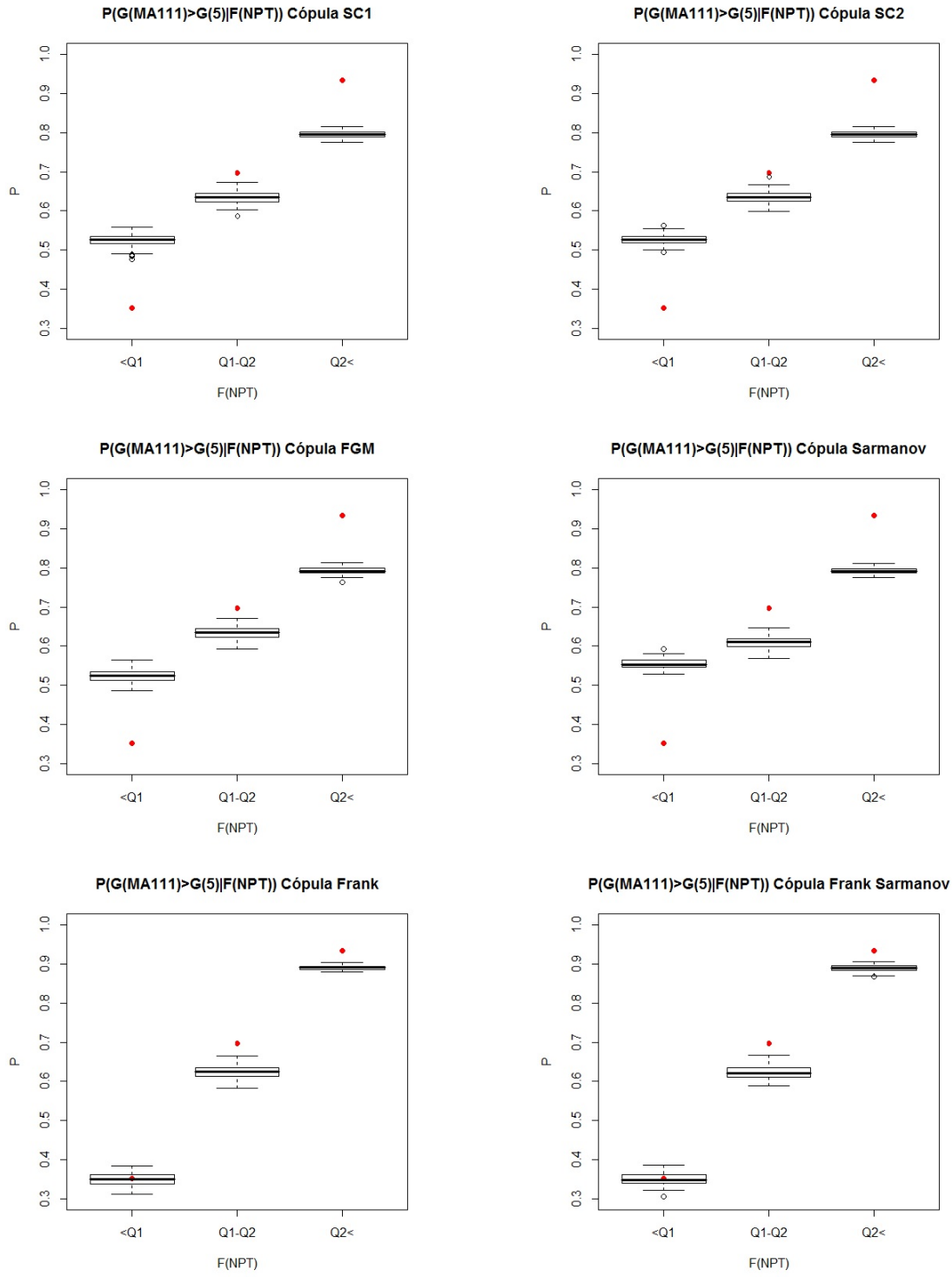

Figura 43 – Boxplots probabilidades $P(G(MA111) \geqslant G(5)|F(NPT))$ calculadas a partir do ajuste das cópulas.

Com respeito às distâncias, tanto a distância de Hellinger quanto a distância de Kolmogorov sugerem que o melhor modelo seria a cópula C_{fs}.

A partir dos critérios, a melhor escolha é C_{fs}, mas dado que a cópula C_f é

Tabela 21 – Distâncias entre cópula empírica e a preditiva C para $(NPT, MA111)$

Cópula C	C_{SC1}	C_{SC2}	C_s	C_{FGM}	C_f	C_{fs}
D. Hellinger	1,21	1,2109	1,3245	1,2124	0,377	0,3702
D. Kolmogorov	0,0763	0,0763	0,0836	0,0763	0,0289	0,0277

um caso especial e mais simples da C_{fs} e que a estimativa do parâmetro $\hat{\mu}$ é $0,9853$, consideremos as seguintes hipóteses $H_0 : \mu = 1$ vs $H_a : \mu \neq 1$ e as testamos usando a metodologia proposta para o uso do FBST. Após os cálculos correspondentes, encontramos que $\kappa = 0,001$ com o e-valor$= 0,999$ e concluímos que temos evidência a favor da H_0, e nos levando a selecionar a C_f como o melhor modelo neste caso.

4.6 Interpretação de resultados

Resumindo os resultado encontrados nas seções anteriores, apresentamos as Tabelas 22 e 23 contendo também as medidas de concordância, tanto amostrais quanto as calculadas a partir da cópula ajustada para cada caso, tanto para o CR quanto para a nota em MA111-Cálculo I.

Tabela 22 – Resultados CR

Variavel	**Cópula (parâmetros)**	ρ_{cop}	ρ_N	τ_{cop}	τ_N
Ciências da Natureza	Frank Sarmanov (5,21 ; 0,74 ;-0,5)	0,58	0,59	0,41	0,41
Ciências Humanas	Frank Samanov (3,42 ; 0,68 ;-0,48)	0,45	0,45	0,31	0,31
Inglês	C_{SC1} (0,99 ; 0,66)	0,28	0,3	0,18	0,21
Matemáticas	Frank Sarmanov (5,03 ; 0,7 ;-0,55)	0,55	0,56	0,39	0,39
Português	Frank Sarmanov (2,86 ; 0,67;-0,48)	0,41	0,38	0,28	0,26
Vestibular Fase 1	Frank Sarmanov (3,93 ; 0,55 ; -0,47)	0,48	0,48	0,33	0,33
Nota Padronizada	Frank Sarmanov (6,23 ; 0,6388 ; -0,5)	0,61	0,6	0,43	0,42

Tabela 23 – Resultados MA111

Variavel	**Cópula (parâmetros)**	ρ_{cop}	ρ_N	τ_{cop}	τ_N
Ciências da Natureza	Frank (4,71)	0,62	0,61	0,44	0,43
Ciências Humanas	Frank (2,85)	0,43	0,42	0,29	0,29
Inglês	C_{SC1} (0,99 ; 0,51)	0,28	0,27	0,18	0,19
Matemáticas	Frank (4,54)	0,61	0,6	0,43	0,42
Português	Frank (1,99)	0,32	0,32	0,21	0,22
Vestibular Fase 1	Frank (3,05)	0,45	0,45	0,31	0,31
Nota Padronizada	Frank(4,73)	0,62	0,61	0,44	0,43

Das Tabela 22 e 23, podemos concluir que apenas as variáveis Ciências da Natureza (CN), Matemáticas (MA) e Nota Padronizada (NPT), apresentam medidas de concordância superior ao $0,5$ e, portanto, são aquelas que têm uma relação mais forte com o CR e a nota em MA111-Calculo I.

Estamos considerando conjuntamente as áreas de Exatas e Engenharia, mas dado que encontramos que em ambos casos as variáveis mais relacionadas são CN e MA, poderíamos perguntar se existe diferença entre a nota nestas provas de alunos de cada uma de estas áreas.

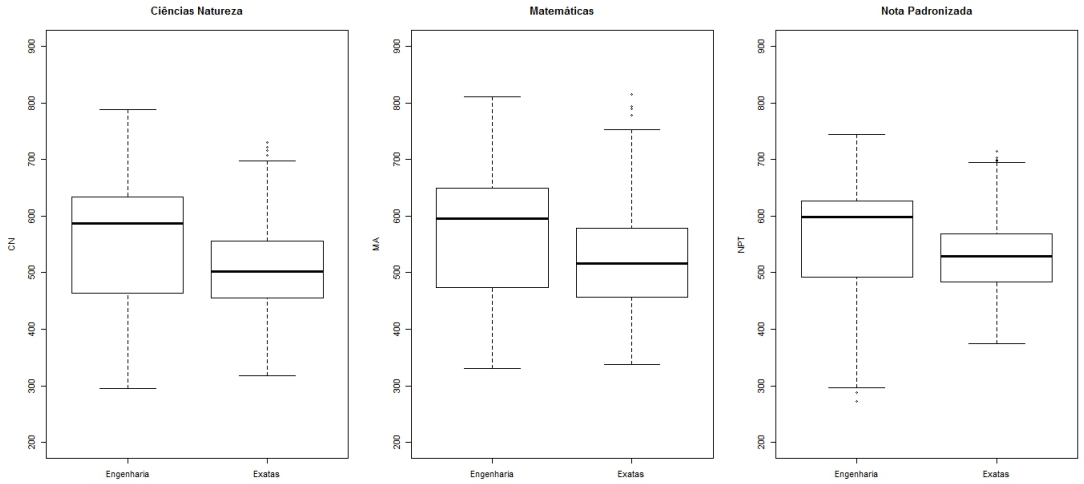

Figura 44 – Boxplot das variáves CN, MA e NPT por àrea.

Nesse sentido, vemos na Figura 44 que os alunos pertencentes a cursos da área de Engenharias presentam uma maior variabilidade, ou seja, a amplitude de valores das notas obtidas por este grupo é maior que para o caso da área de exatas. Embora os alunos de exatas presentam uma amplitude de notas menor, a maioria dos estudantes dessa área obtiveram notas menores que os alunos de engenharias, pelo que poderíamos sugerir fazer a mesma análise, mas dessa vez para cada área em separado.

Além disso, se estamos interessados em alguns cursos especificamente, e dado que já identificamos as variáveis CN, MA e NPT como as mais importantes, poderíamos analisar e comparar estas variáveis a partir de um gráfico como os apresentados pela Figura 45. Nessa Figura, temos os boxplot dos estudantes da licenciatura em matemática, do cursão (o curso que inclui todas ciências) e todos os demais cursos incluindo aqueles que são da área de engenharia.

Dessa comparação, podemos concluir que os alunos que entraram no cursão foram melhores nas duas provas analisadas e na nota padronizada final, do que os alunos que entram na licenciatura. Dado que a ementa da disciplina de Cálculo I é a mesma para todos cursos, poderiamos dizer que os alunos do curso de licenciatura em matemática estão em desvantagem con respeito aos outros cursos.

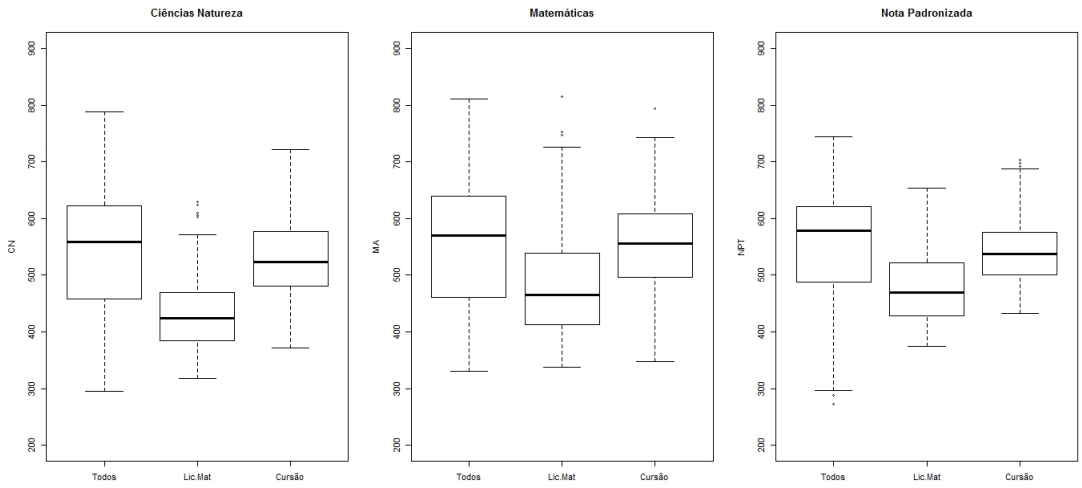

Figura 45 – Boxplot das variáveis CN, MA e NPT por cursos.

5 Conclusões

O objetivo desta dissertação foi estabelecer uma metodologia de estimação Bayesiana para cópulas bivariadas e ilustrar a partir de uma aplicação como implementá-la. Nesta seção, vamos apresentar as conclusões que surgem a partir dos resultados obtidos no Capítulo 4 e algumas recomendações para trabalhos futuros.

5.1 Conclusões Gerais

Sobre a metodologia, encontramos que pelo menos para as cópulas consideradas, a escolha de uma distribuição a priori é simples, quando é baseada apenas no conhecimento do domínio dos parâmetros, no entanto, não foi possível achar distribuições conjugadas de forma direta, devido à complexidade das expressões das verossimilhanças. Nesse sentido, dada a forma das densidades das cópulas, a definição de distribuições conjugadas é um problema de alta dificuldade nesse contexto.

Com respeito à estimação, encontramos que os cálculos são de pouco custo computacional, mesmo com um volume de dados acima da 3000 observações, e portanto, tentar trabalhar com mais do que 6 famílias de cópulas é bem possível. No entanto, na hora de comparar modelos a partir dos cálculos das probabilidades condicionais, implicou um custo computacional maior do que realizar conjuntamente o ajuste e os demais cálculos comparativos como as distâncias, as distribuições preditivas e o e-valor.

Falando especificamente das distîcias, encontramos que a distância de Kolmogorv geralmente não percebe diferenças entre modelos, quando um é caso particular de outro, mesmo como no caso das cópulas Frank e Frank Sarmanov, onde o e-valor estabelece que existe diferencia entre eles. Portanto concluímos que, em termos de comparação a partir de distâncias, a distância de Hellinger é o melhor critério.

Com respeito à análise dos dados, encontramos que as variáveis do Vestibular mais relacionadas com o coeficiente de rendimento (CR) de um aluno são Ciências da Natureza (CN), Matemáticas (MA) e a Nota Padronizada (NPT). Embora não seja uma surpresa que esta última seja a mais relacionada, encontramos que a diferença entre as medidas de concordância obtidas para NPT e para CN é bem pequena. Assim, concluímos que o desempenho de um estudante no seu primeiro semestre está mais fortemente relacionado com seu desempenho na prova de CN no Vestibular.

No caso da análise para a variável nota em MA111-Cálculo I, encontramos que em geral o ajuste não é tão bom como para o caso do CR, devido a que, pelo menos em termos da probabilidade condicional, nenhum dos modelos conseguiu reproduzir

exatamente o comportamento da probabilidade encontrada na amostra. No entanto, com base nos modelos escolhidos para cada variável, chegamos a uma conclusão semelhante à obtida para o CR, isto é, de novo as variáveis mais relacionadas com o rendimento no primeiro semestre do aluno, medido a través da nota em Cálculo I, foram CN, MA e NPT.

5.2 Trabalhos futuros

A partir da análise feita neste trabalho surgem algumas perguntas que podem servir de base para trabalhos futuros. Uma primeira recomendação, é tentar estabelecer um melhor representante do que a nota padronizada NPT, já que o propósito de estabelecer uma medida como esta, é resumir o conhecimento integral do aluno. Porém na prática vemos que trabalhar a partir desta variável não faz muita diferença com trabalhar apenas com a nota na prova de Ciências da Natureza.

Outro aspecto interessante, é o que acontece com o comportamento da probabilidade condicional $P(V > 5|u1 < U < u2)$ nos ajustes para a variável MA111-Cálculo I, onde encontramos que o melhor modelo escolhido em cada caso, consegue reproduzir a probabilidade amostral com sucesso apenas para o primeiro intervalo (até o primeiro quartil). Embora, para alguns casos, no último intervalo dos modelos escolhidos os boxplot estão perto da probabilidade condicional amostral (ponto vermelho). Esse senánrio desejável nunca acontece quando a nota na prova do Vestibular analisada está entre o primeiro e segundo quartil, pelo que uma sugestão é considerar uma soma ordinal de cópulas e tentar ajustar uma cópula com parâmetros diferentes para cada um dos três intervalos considerados.

Referências

[Albert, 1997] Albert, J. (1997). Matlab as an environment for bayesian computation. In *Proceedings of the Section on Bayesian Statistical Science*, page 60. The Association.

[Arnold and Emerson, 2011] Arnold, T. B. and Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3(2):34–39.

[Bairamov et al., 2011] Bairamov, I., Altinsoy, B., and Kerns, G. J. (2011). On generalized sarmanov bivariate distributions. *Journal of Applied and Engineering Mathematics*, 1:86–97.

[Bar-Yossef et al., 2002] Bar-Yossef, Z., Jayram, T., Kumar, R., and Sivakumar, D. (2002). An information statistics approach to data stream and communication complexity. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 209–218. IEEE.

[Clayton, 1978] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

[Conway, 1983] Conway, D. (1983). The farlie-gumbel-morgenstern distributions. *Encyclopedia of Statistical Sciences*, 3:28–31.

[Danaher and Smith, 2011] Danaher, P. J. and Smith, M. S. (2011). Modeling multivariate distributions using copulas: applications in marketing. *Marketing Science*, 30(1):4–21.

[de Bragança Pereira and Stern, 1999] de Bragança Pereira, C. A. and Stern, J. M. (1999). Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy*, 1(4):99–110.

[DeGroot, 2005] DeGroot, M. H. (2005). *Optimal statistical decisions*, volume 82. John Wiley & Sons.

[Deheulves, 1981] Deheulves, P. (1981). A kolmogorov-smirnov test of independence. Technical report, Rev. Roumaine Math. Pures. Appl.

[Deheuvels, 1979] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.(5)*, 65(6):274–292.

[Deheuvels, 1980] Deheuvels, P. (1980). Non parametric tests of independence. In *Statistique non Paramétrique Asymptotique*, pages 95–107. Springer.

[dos Santos Silva and Lopes, 2008] dos Santos Silva, R. and Lopes, H. F. (2008). Copula, marginal distributions and model selection: a bayesian note. *Statistics and Computing*, 18(3):313–320.

[Fernández et al., 2014] Fernández, M., González-López, V. A., and Rifo, L. R. (2014). A note on conjugate distributions for copulas. *Mathematical Methods in the Applied Sciences*.

[Frechét, 1951] Frechét, M. (1951). Sur les tableaux de corrélation dont les marges sont données. Technical report, Ann Univ. Lyon Seet A 9.

[Frees and Valdez, 1998] Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American actuarial journal*, 2(1):1–25.

[Genest and Rémillard, 2004] Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369.

[Hoeffding, 1940] Hoeffding, W. (1940). Scale invariant correlation theory. In *The collected works of Wassily Hoeffding*, pages 57–107. Springer.

[Hoeffding, 2012] Hoeffding, W. (2012). *The collected works of Wassily Hoeffding*. Springer.

[Huard et al., 2006] Huard, D., Evin, G., and Favre, A.-C. (2006). Bayesian copula selection. *Computational Statistics & Data Analysis*, 51(2):809–822.

[Hutchinson and Lai, 1990] Hutchinson, T. and Lai, C. (1990). Continuos bivariate distributions, emphasising applications. Technical report, Rumsby Scientific Publishing.

[Joe, 1997] Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

[Kendall, 1938] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, pages 81–93.

[Kolmogorov, 1950] Kolmogorov, A. N. (1950). Grundbegriffe der wahrscheinlichkeitsrechnung, berlin, 1933. *English translation, Chelsea, New York*.

[Kruskal, 1958] Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861.

[Lambert and Vandenhende, 2002] Lambert, P. and Vandenhende, F. (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in medicine*, 21(21):3197–3217.

[Lehmann, 1966] Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, pages 1137–1153.

[Mesiar and Sempi, 2010] Mesiar, R. and Sempi, C. (2010). Ordinal sums and idempotents of copulas. *Aequationes mathematicae*, 79(1-2):39–52.

[Nelsen, 2013] Nelsen, R. B. (2013). *An introduction to copulas*, volume 139. Springer Science.

[Nikoloulopoulos and Karlis, 2008] Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27(30):6393–6406.

[Oakes, 1989] Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493.

[Patton, 2006] Patton, A. J. (2006). Modelling asymmetric exchange rate dependence*. *International economic review*, 47(2):527–556.

[Romano, 2002] Romano, C. (2002). Calibrating and simulating copula functions: an application to the italian stock market. *Centro Interdipartimentale sul Diritto e lÉconomia dei Mercati, Working paper*.

[Schervish, 2012] Schervish, M. (2012). *Theory of Statistics*. Springer Series in Statistics. Springer New York.

[Schweizer, 1991] Schweizer, B. (1991). Thirty years of copulas. In *Advances in probability distributions with given marginals*, pages 13–50. Springer.

[Schweizer and Sklar, 1974] Schweizer, B. and Sklar, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica*, 52(1):43–52.

[Segers, 2004] Segers, J. (2004). Non-parametric inference for bivariate extreme-value copulas. Technical report, Center Discussion Paper.

[Sklar, 1959] Sklar, A. (1959). Fonctions de repartition à n-dimensions et leur marges. Technical report, Publ.Ins.Statist.Univ. París 8.

[Smith, 2000] Smith, M. (2000). Modeling and short-term forecasting of new south wales electricity system load. *Journal of Business & Economic Statistics*, 18(4):465–478.

[Smith, 2011] Smith, M. S. (2011). Bayesian approaches to copula modelling. *Available at SSRN 1974297*.

[Smith, 1998] Smith, R. L. (1998). Bayesian and frequentist approaches to parametric predictive inference. In *BAYESIAN STATISTICS, JM BERNARDO, JO BERGER, AP DAWID, AFM SMITH (EDS.)*. Citeseer.

[Spearman, 1904] Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101.

[Ting Lee, 1996] Ting Lee, M.-L. (1996). Properties and applications of the sarmanov family of bivariate distributions. *Communications in Statistics-Theory and Methods*, 25(6):1207–1222.

[Úbeda-Flores et al., 2004] Úbeda-Flores, M. et al. (2004). A new class of bivariate copulas. *Statistics & probability letters*, 66(3):315–325.

Anexos

ANEXO A – Gráficos de algumas cópulas comuns

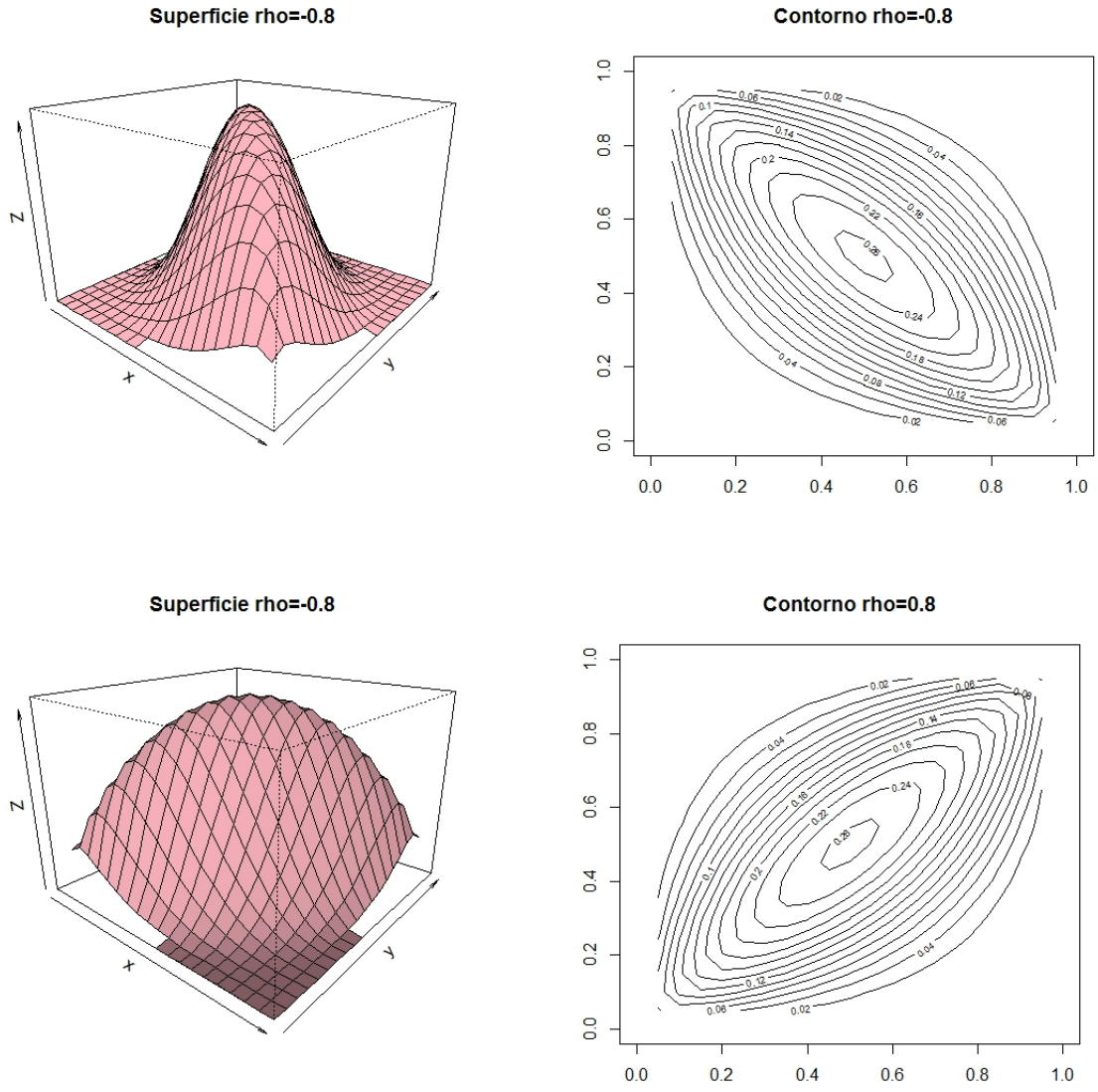

Figura 46 – Gráficos de contorno e superfície da densidade de duas cópulas gaussianas para diferentes valores de ρ

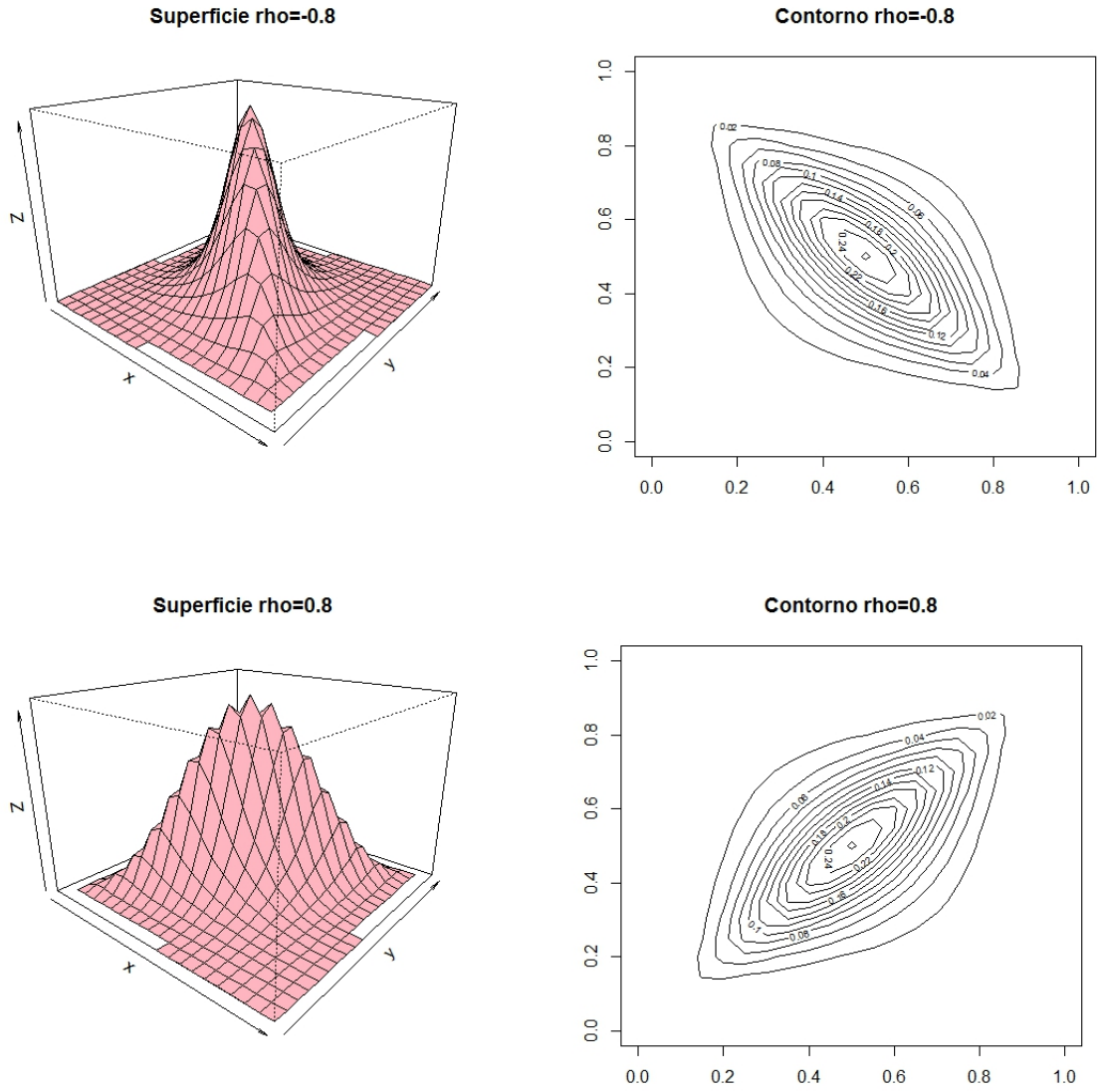

Figura 47 – Gráficos de contorno e superfície da densidade de duas cópulas t-Student com $\nu = 1$ gl para diferentes valores de ρ

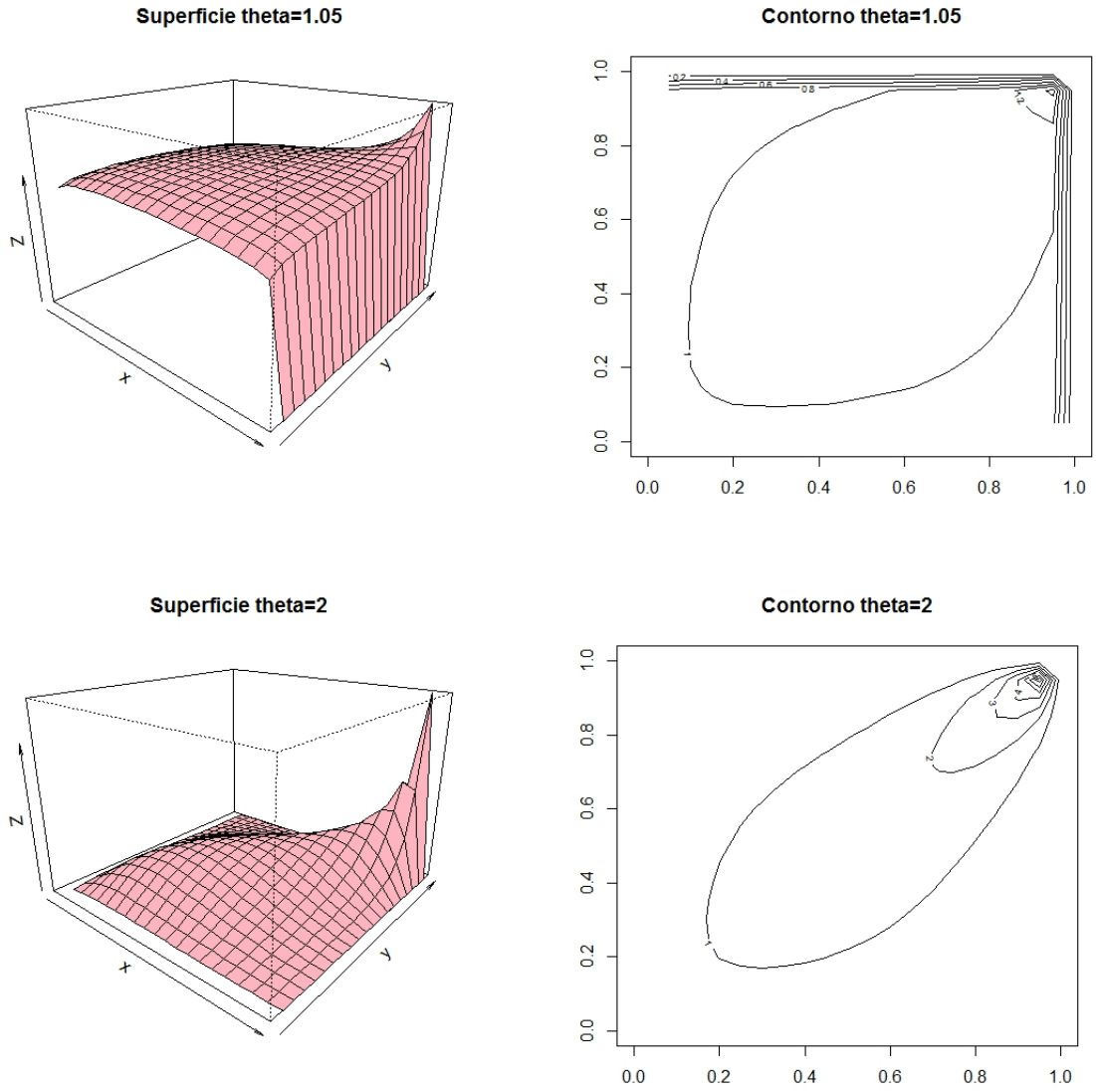

Figura 48 – Gráficos de contorno e superfície da densidade de duas cópulas Gumbel para diferentes valores de θ.

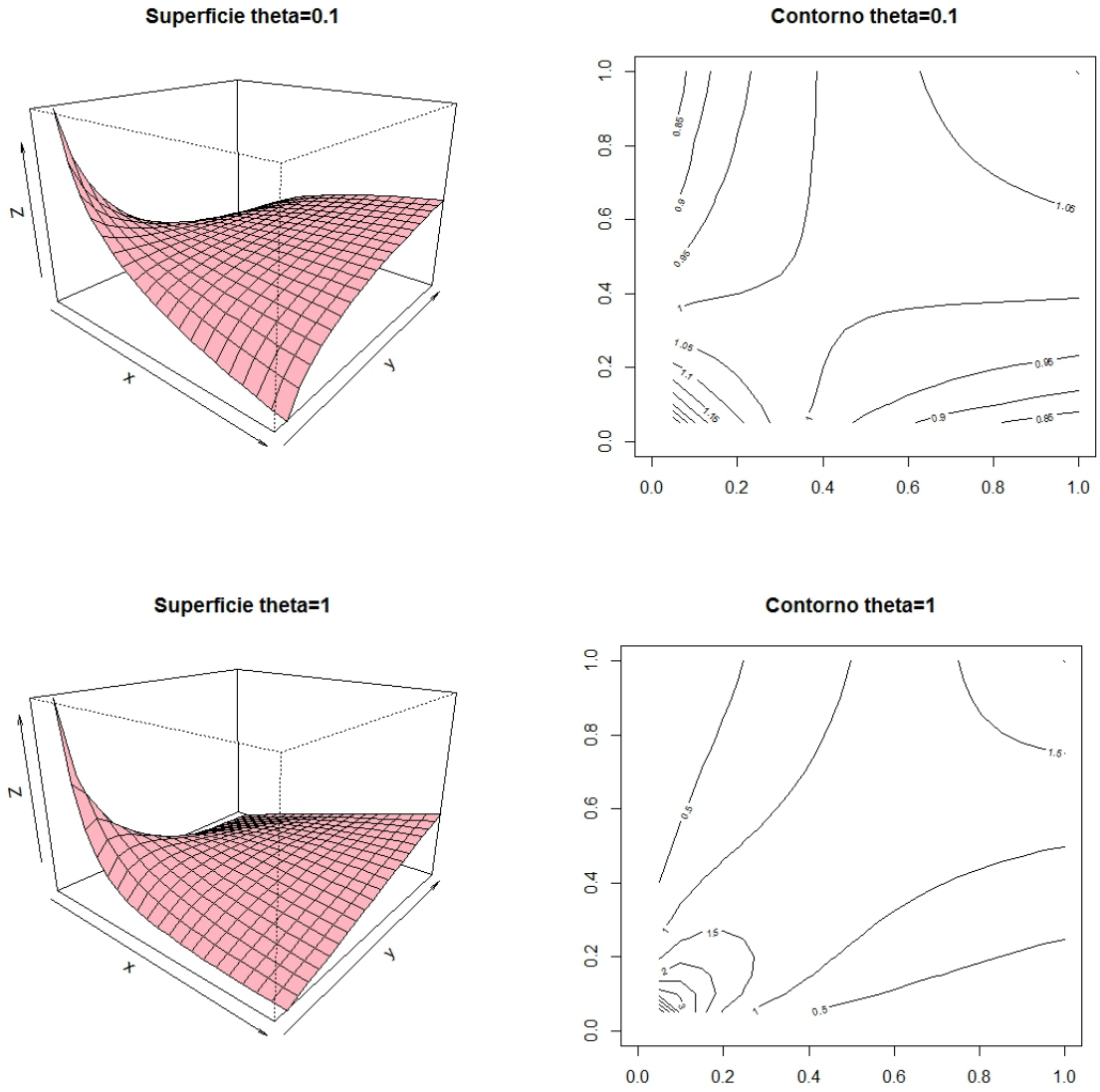

Figura 49 – Gráficos de contorno e superfície da densidade de duas cópulas Clayton para diferentes valores de θ.

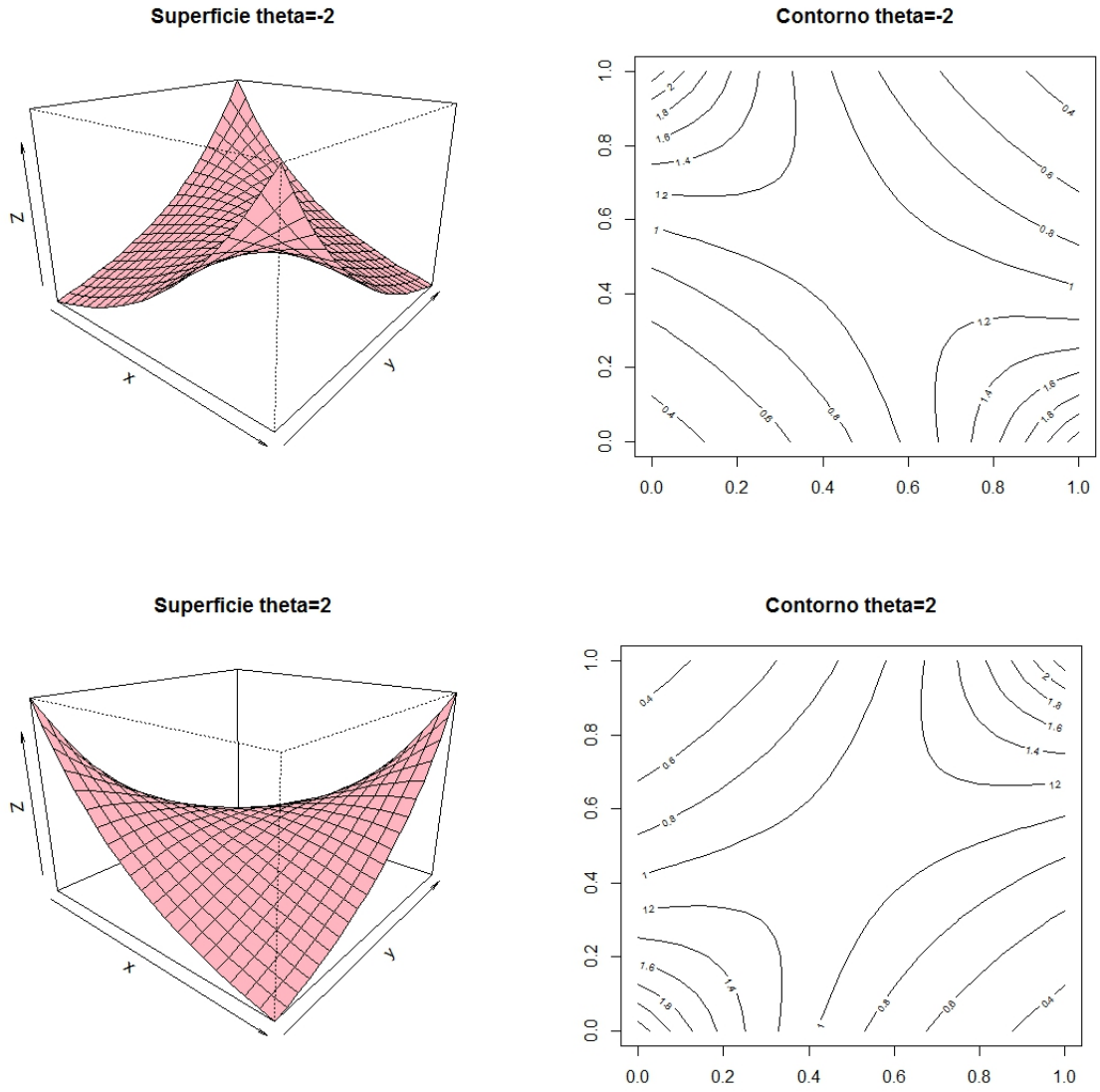

Figura 50 – Gráficos de contorno e superfície da densidade de duas cópulas Frank para diferentes valores de θ.

ANEXO B – Gráficos densidade log posterior

B.1 Ajustes com o CR

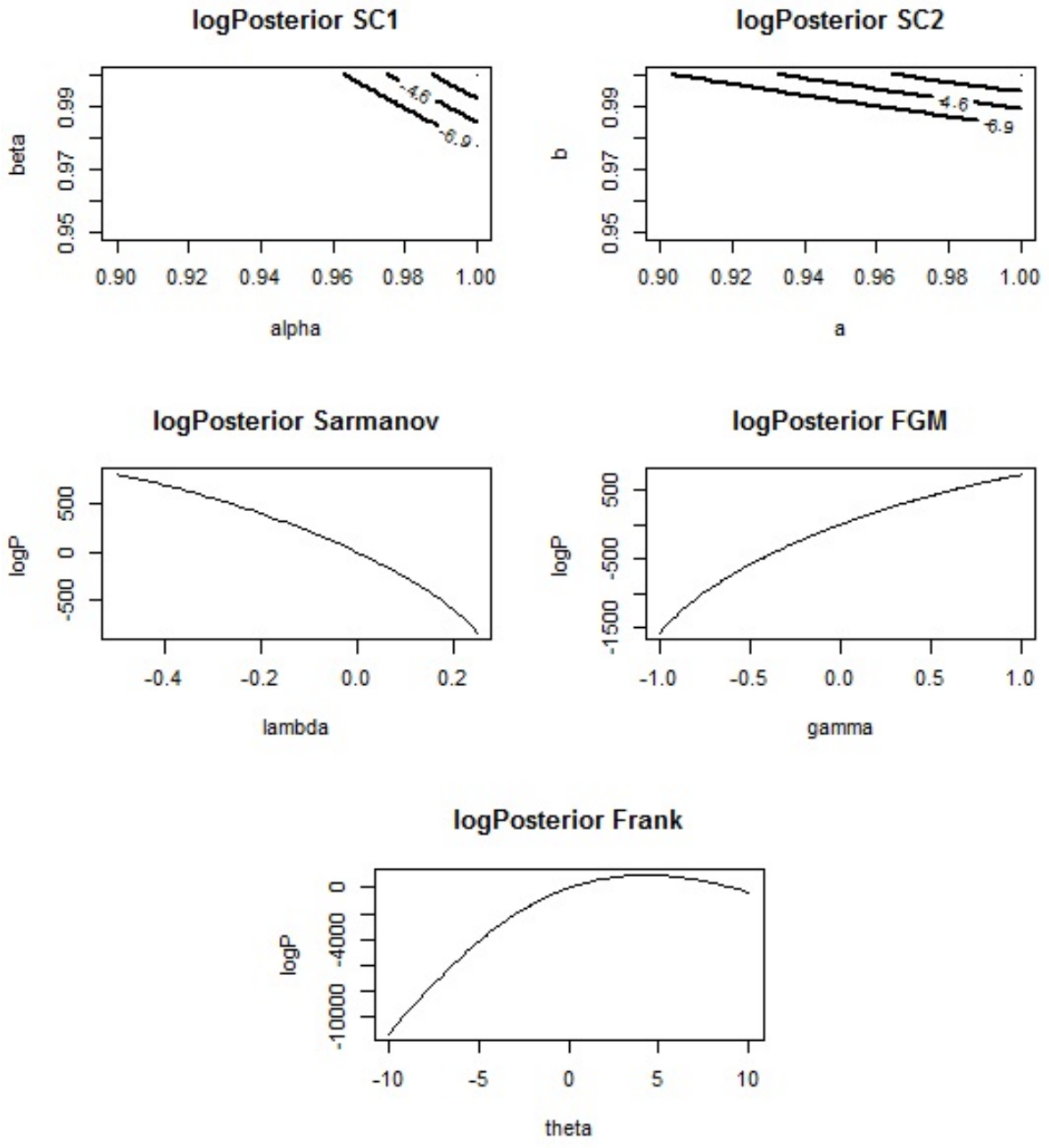

Figura 51 – Densidades log posterior dos parâmetros de cada cópula no caso (CN, CR)

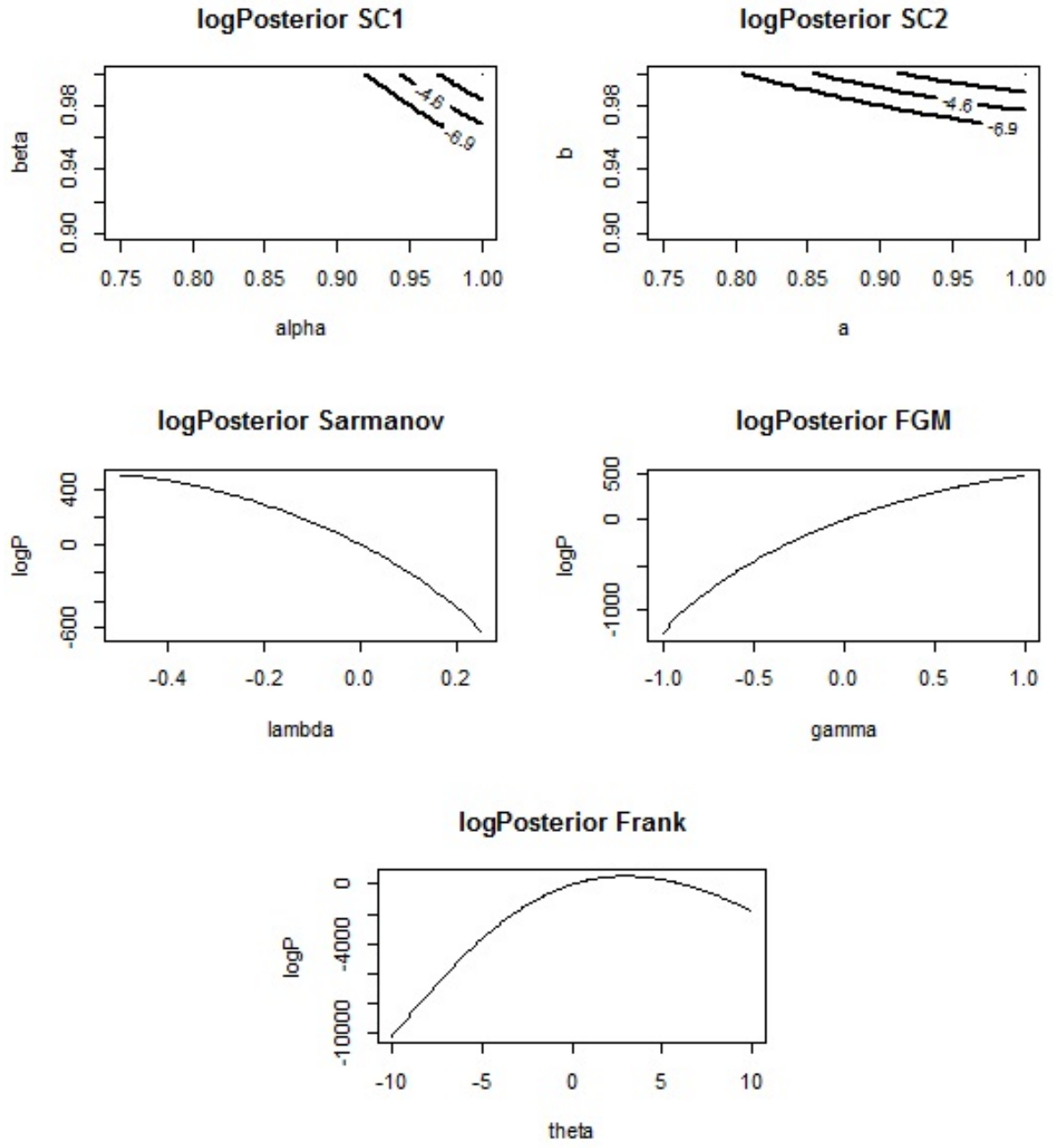

Figura 52 – Densidades log posterior dos parâmetros de cada cópula no caso (CH, CR)

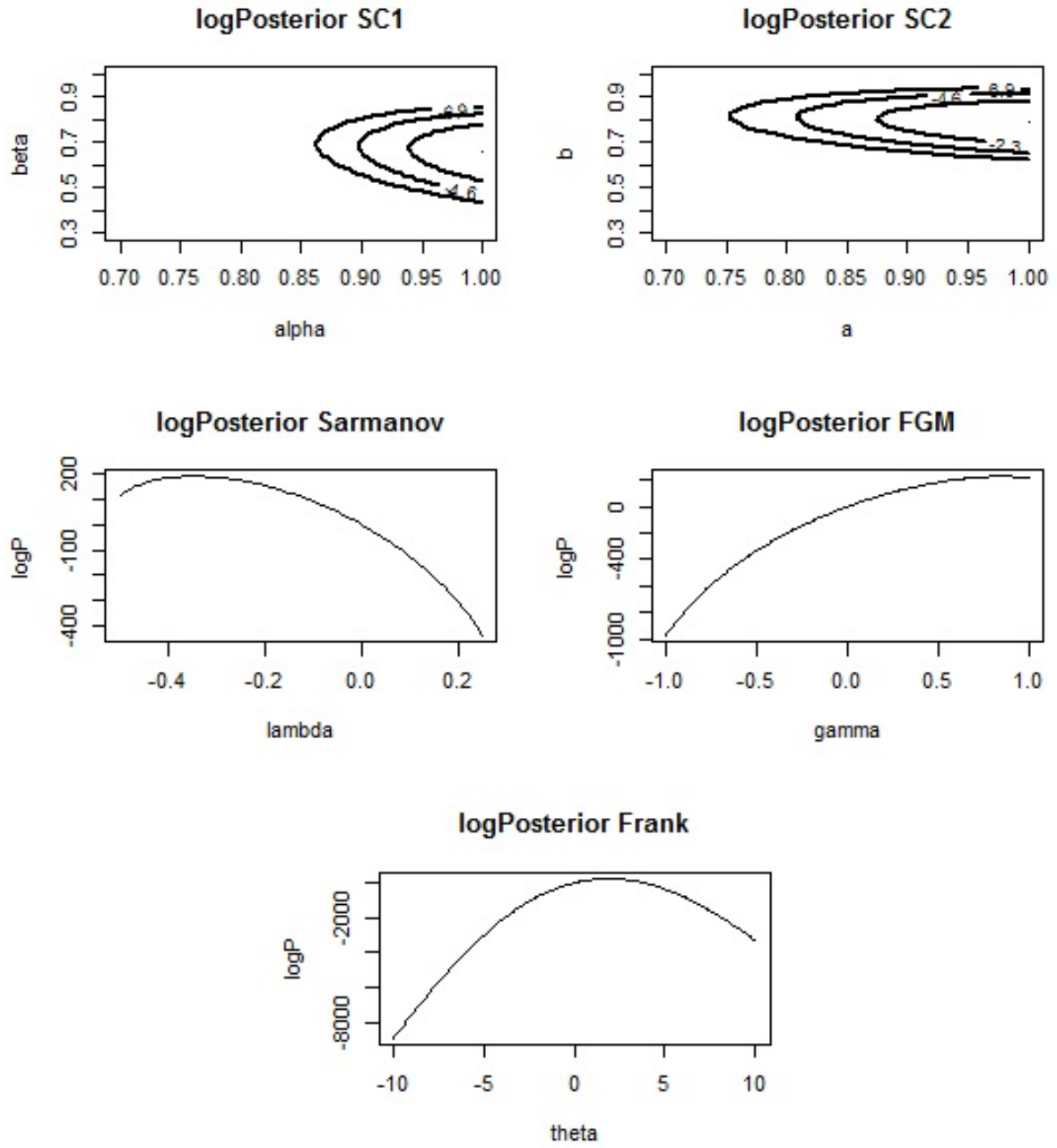

Figura 53 – Densidades log posterior dos parâmetros de cada cópula no caso (ING, CR)

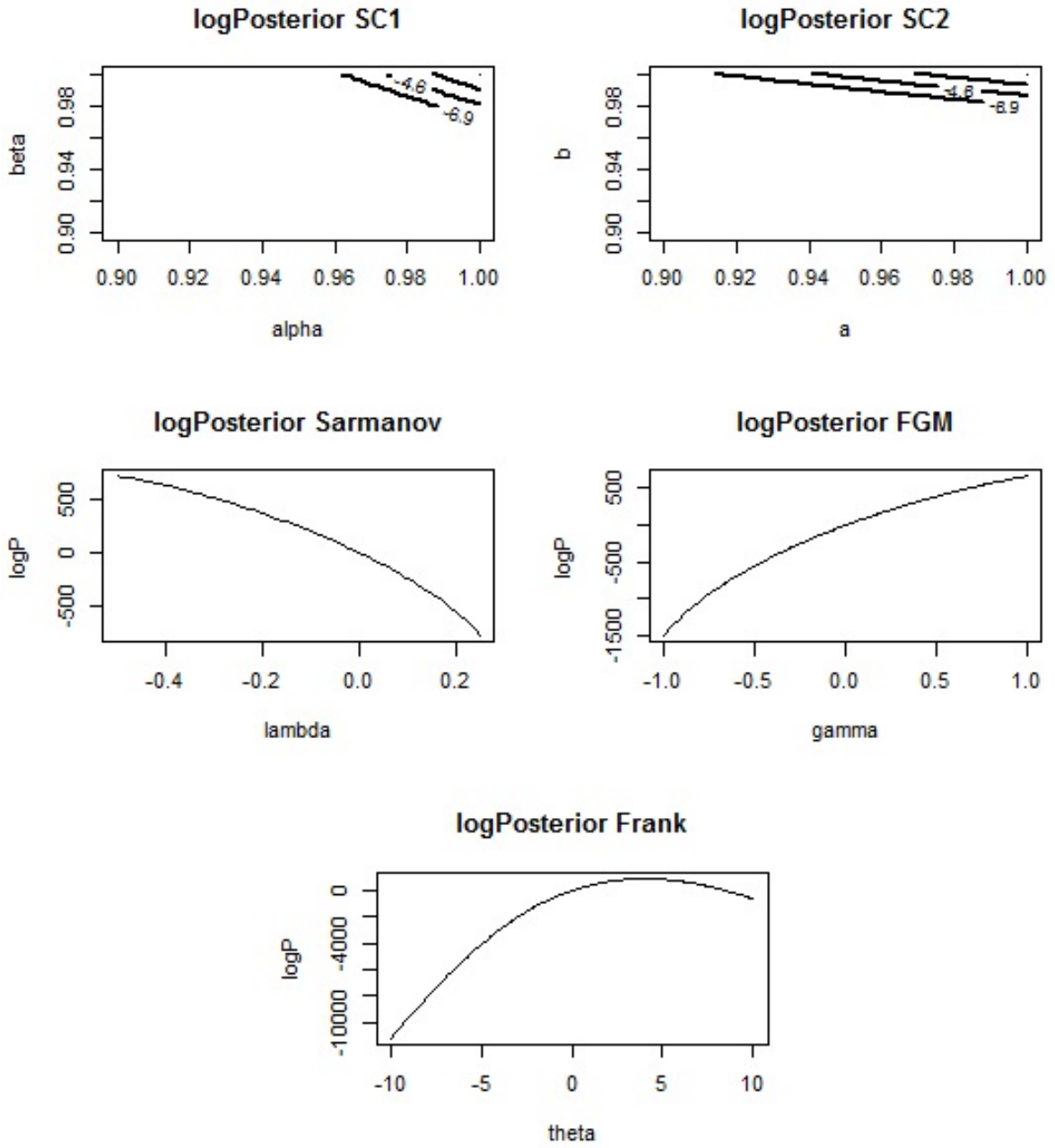

Figura 54 – Densidades log posterior dos parâmetros de cada cópula no caso (MA, CR)

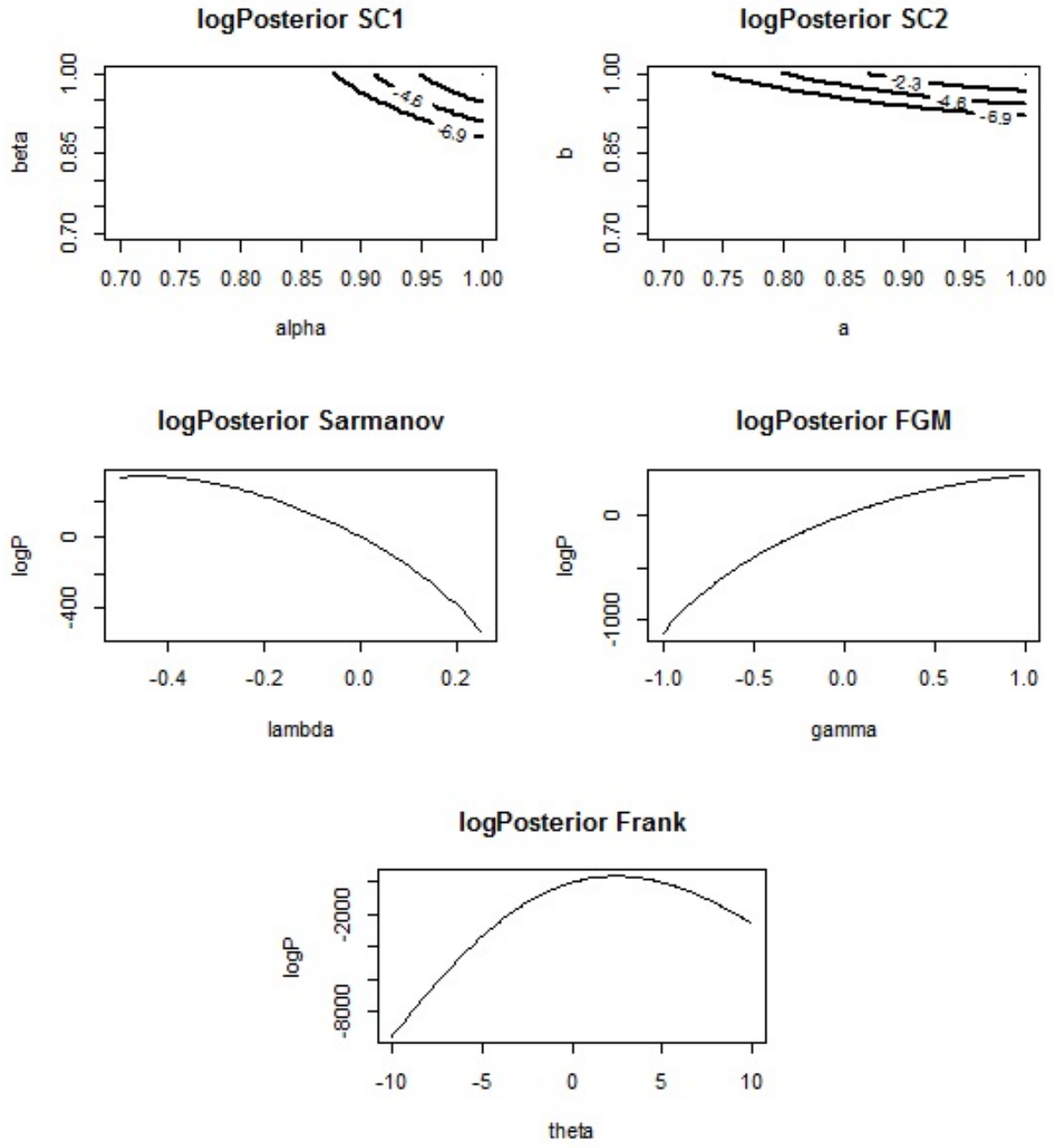

Figura 55 – Densidades log posterior dos parâmetros de cada cópula no caso (PT, CR)

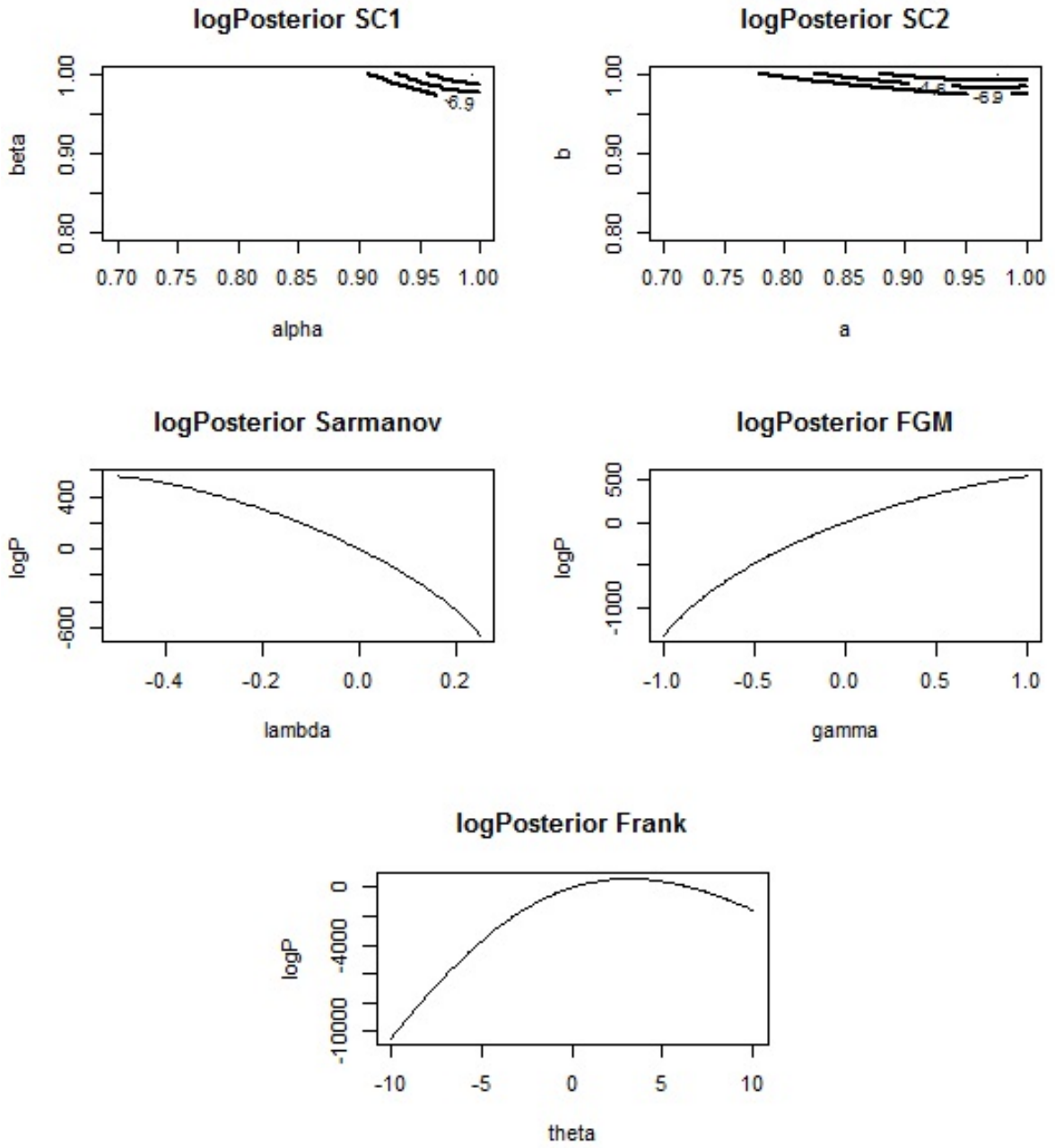

Figura 56 – Densidades log posterior dos parâmetros de cada cópula no caso $(VF!, CR)$

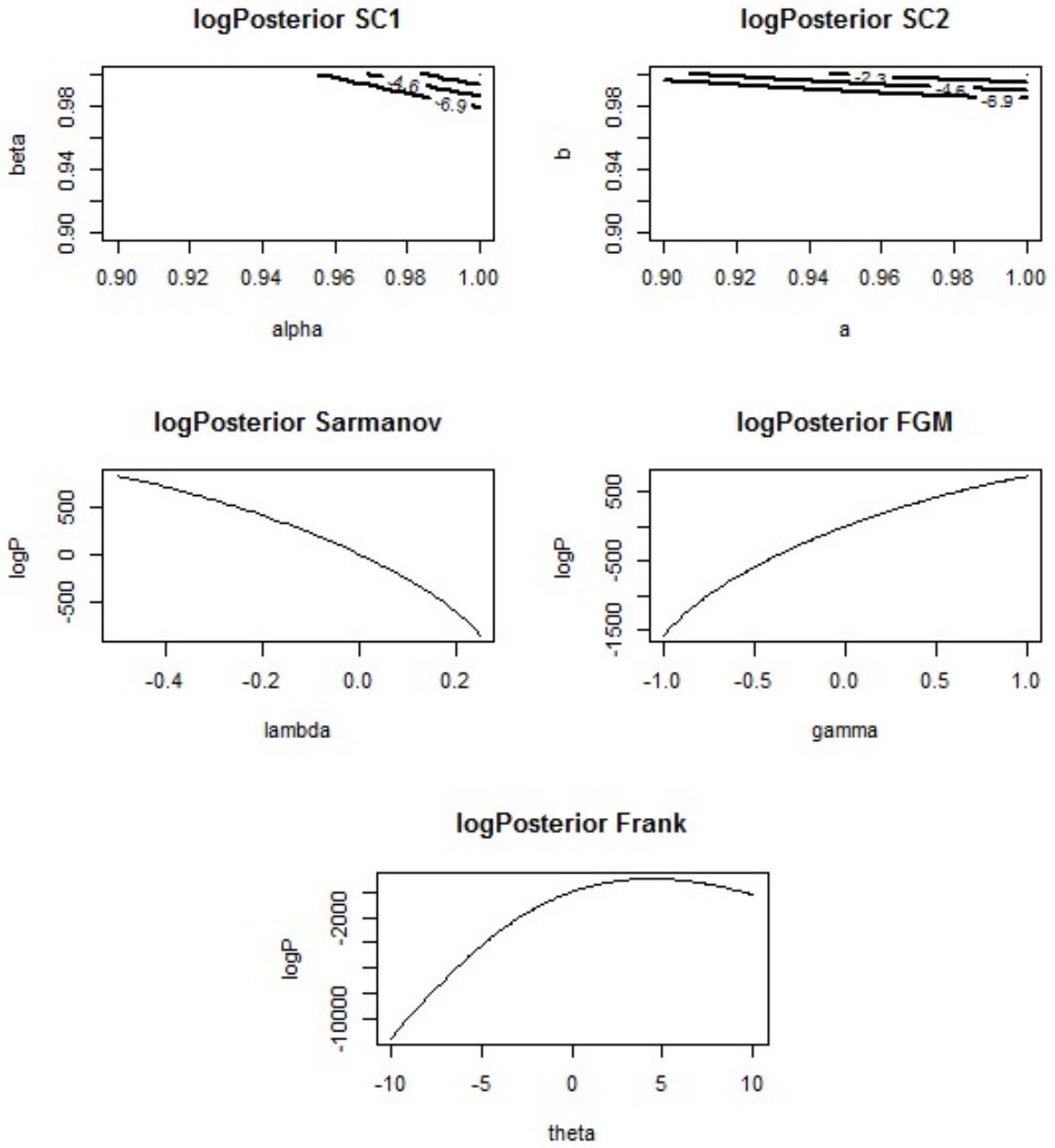

Figura 57 – Densidades log posterior dos parâmetros de cada cópula no caso (NPT, CR)

B.2 Ajustes com a nota em MA111

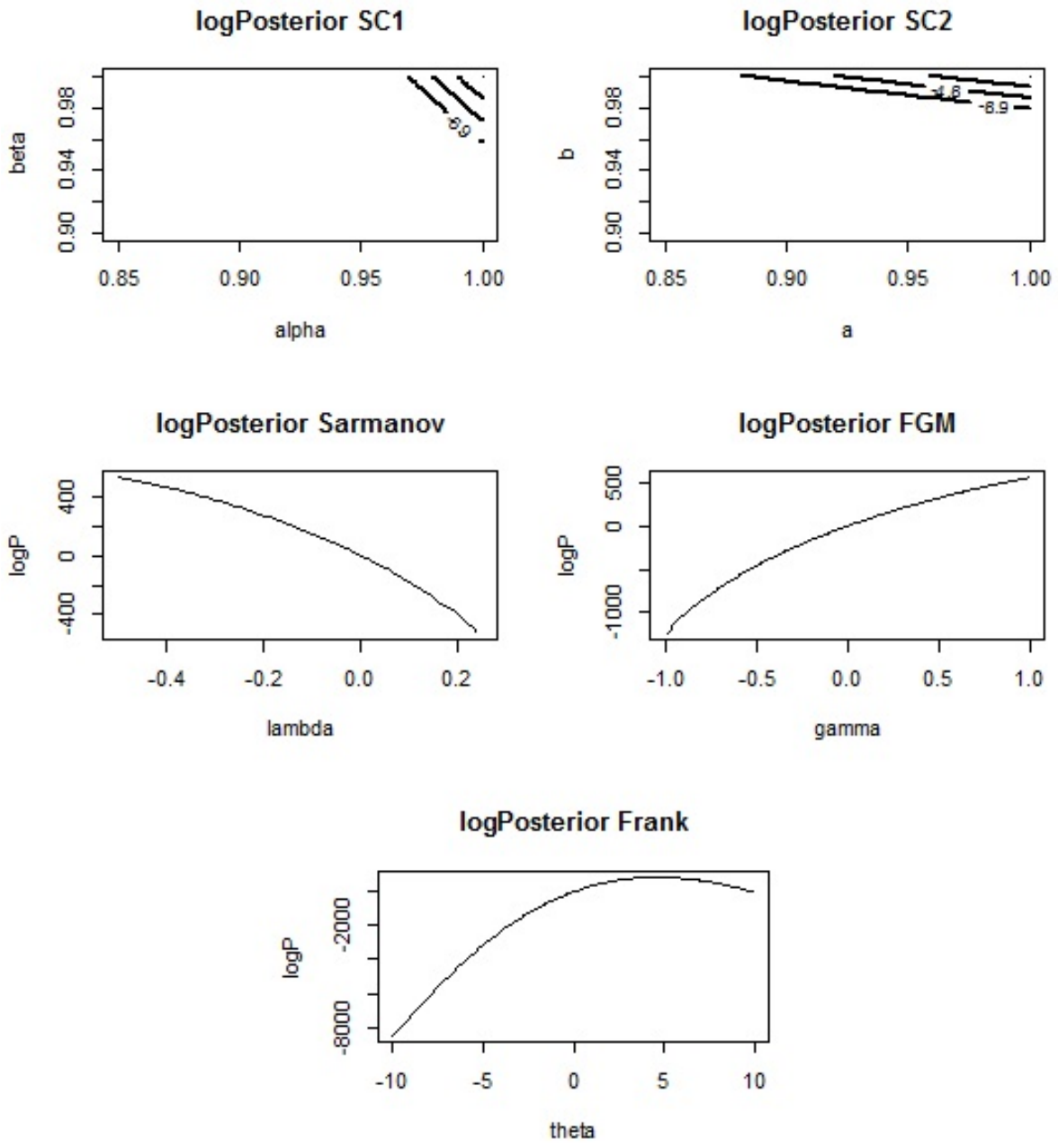

Figura 58 – Densidades log posterior dos parâmetros de cada cópula no caso $(CN, MA111)$

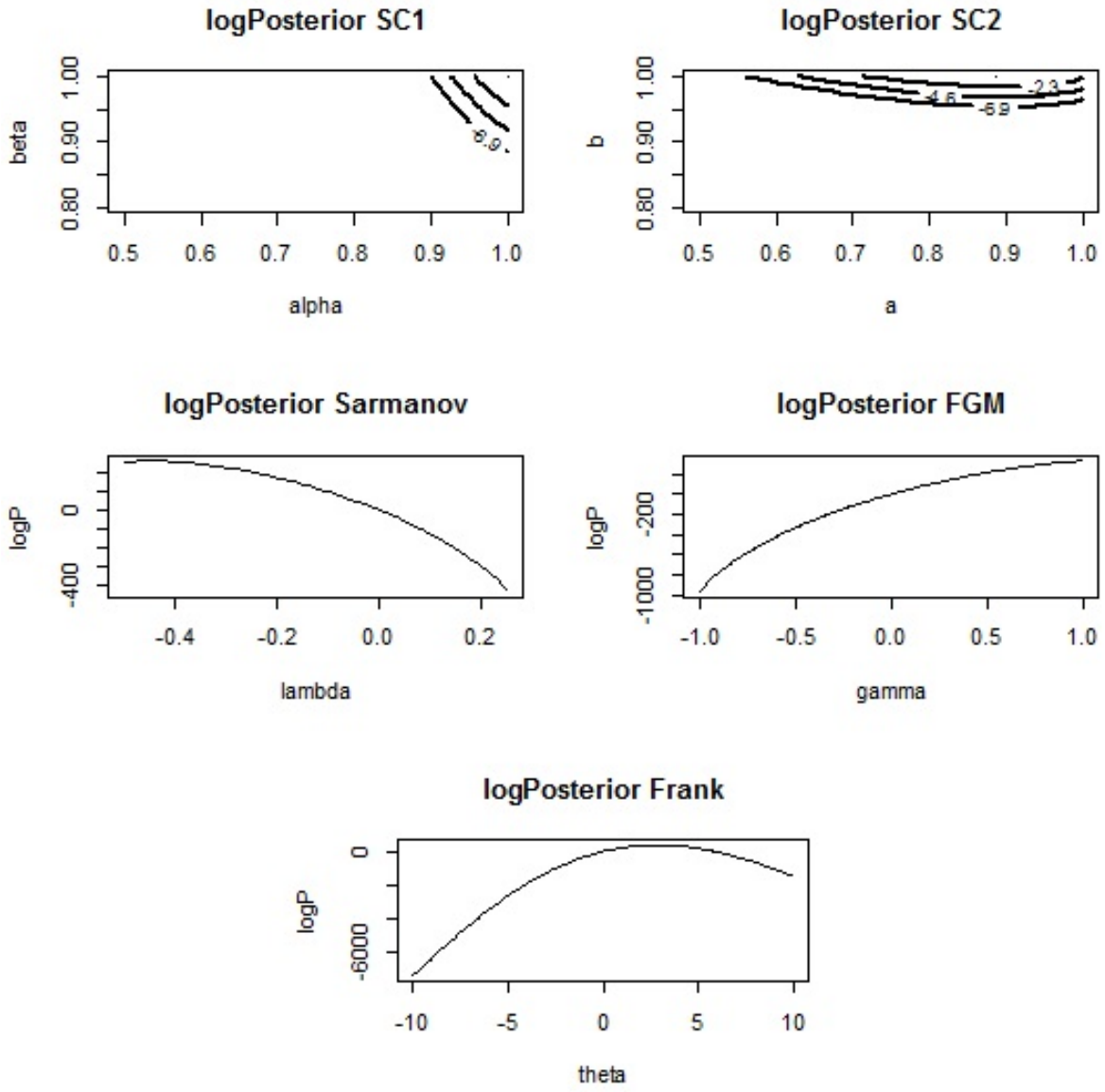

Figura 59 – Densidades log posterior dos parâmetros de cada cópula no caso $(CH, MA111)$

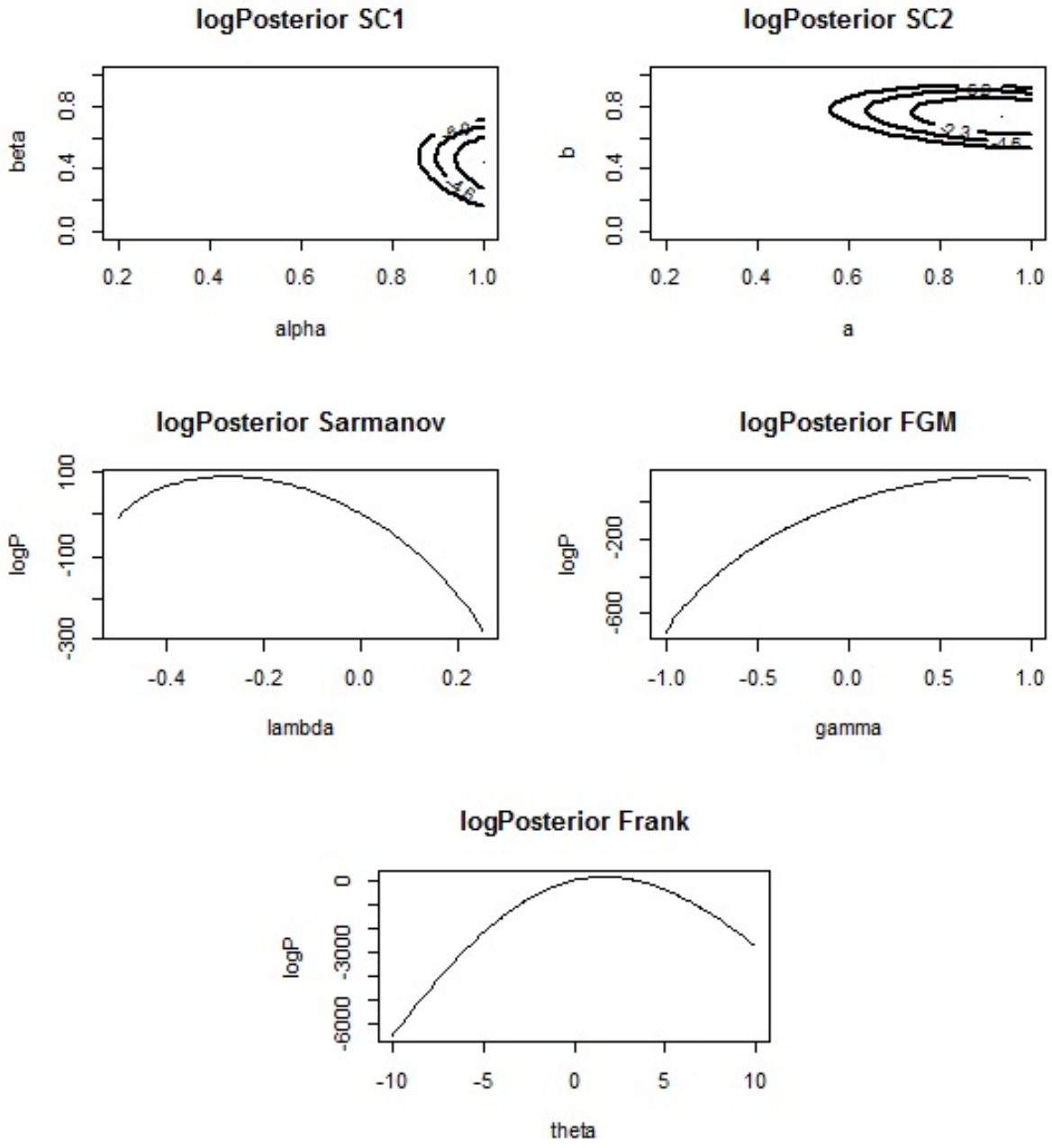

Figura 60 – Densidades log posterior dos parâmetros de cada cópula no caso $(ING, MA111)$

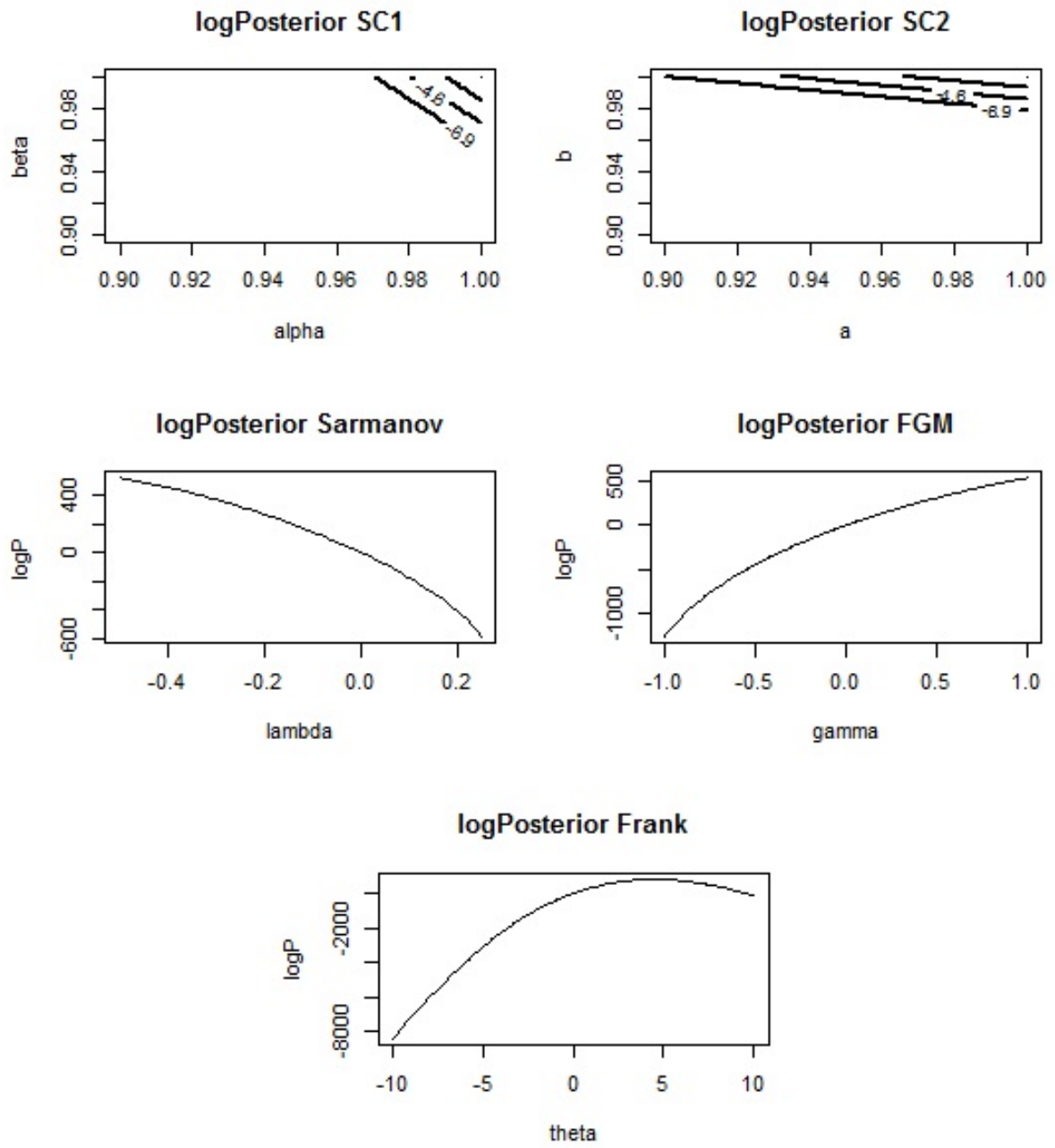

Figura 61 – Densidades log posterior dos parâmetros de cada cópula no caso $(MA, MA111)$

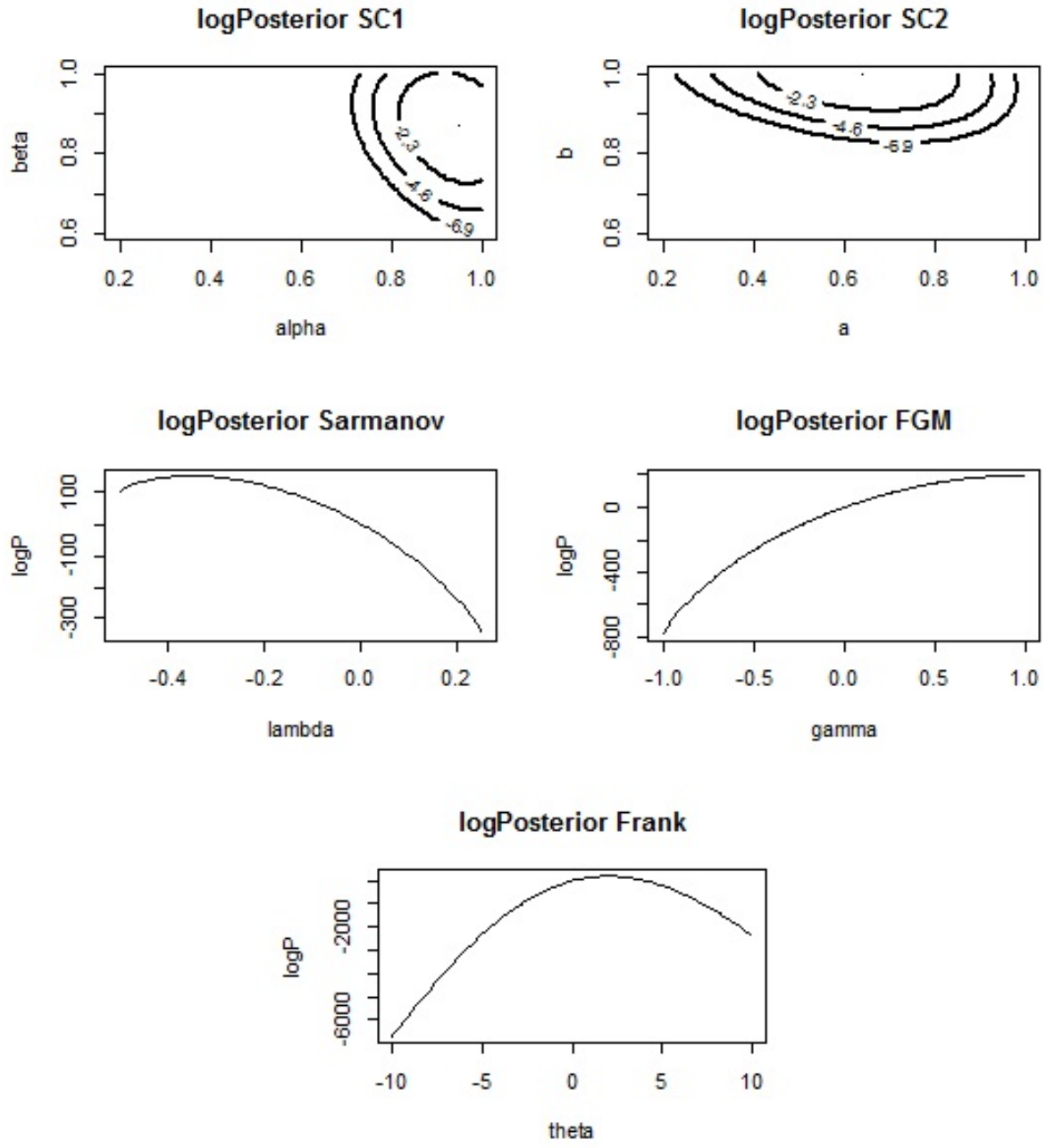

Figura 62 – Densidades log posterior dos parâmetros de cada cópula no caso $(PT, MA111)$

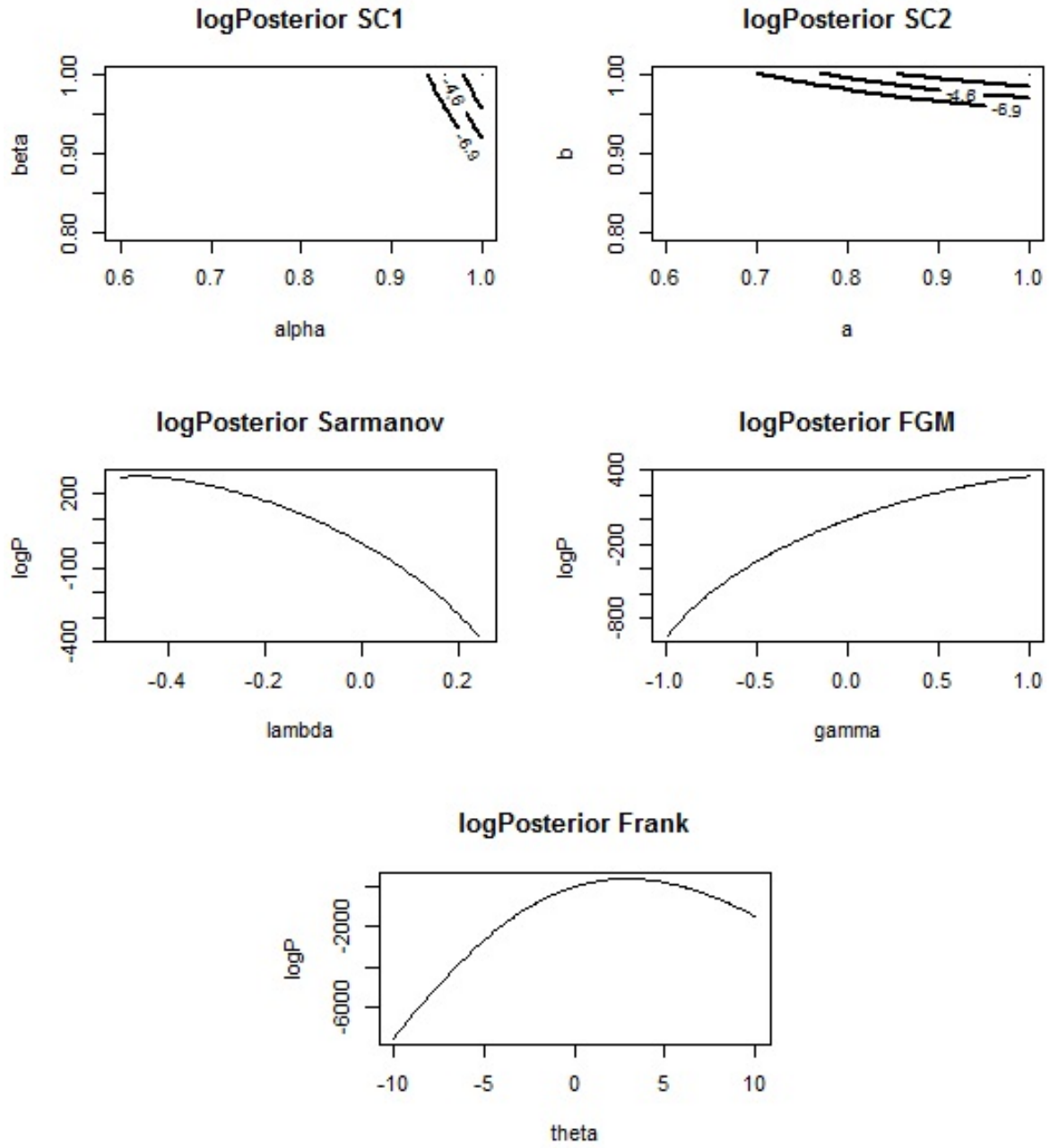

Figura 63 – Densidades log posterior dos parâmetros de cada cópula no caso $(VF1, MA111)$

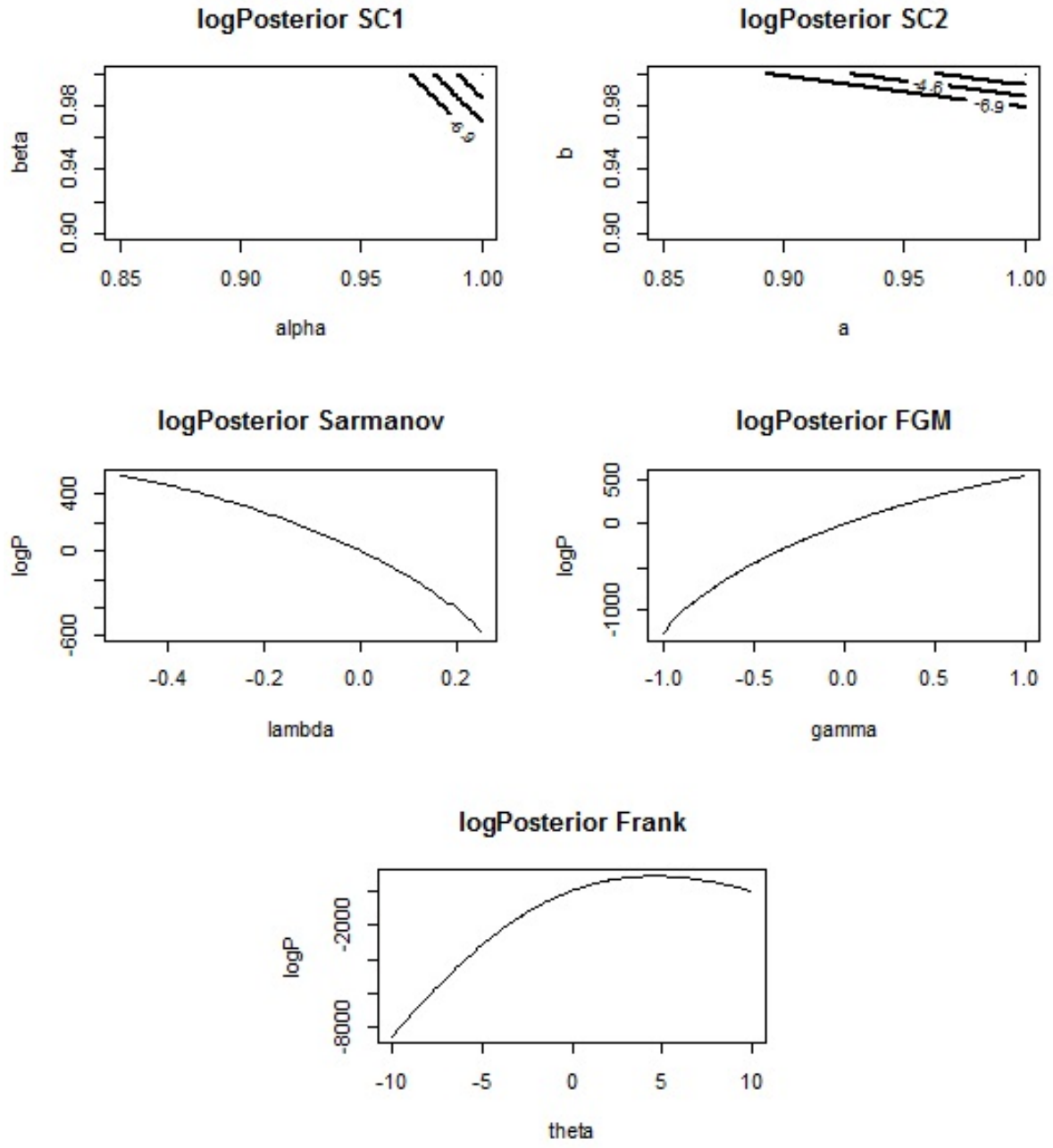

Figura 64 – Densidades log posterior dos parâmetros de cada cópula no caso $(NPT, MA111)$