

Modelo Alfa Normal Assimétrico Multivariado para Redes de Classificação

Anderson Ara

Orientador: Prof. Dr. Francisco Louzada

São Carlos

Março de 2016

Modelo Alfa Normal Assimétrico Multivariado para Redes de Classificação

Anderson Ara

Orientador: Prof. Dr. Francisco Louzada

Tese apresentada ao Programa de Pós-Graduação em Estatística, curso de Doutorado do Departamento de Estatística - CCET UFSCar, como requisito para obtenção do título de Doutor.

São Carlos

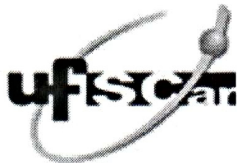
Março de 2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

A658m Ara, Anderson
Modelo alfa normal assimétrico multivariado para
redes de classificação / Anderson Ara. -- São Carlos
: UFSCar, 2016.
63 p.

Tese (Doutorado) -- Universidade Federal de São
Carlos, 2016.

1. Distribuição Multivariada. 2. Inferência
Estatística. 3. Assimetria. 4. Bimodalidade. 5. Data
mining. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Estatística

Folha de Aprovação

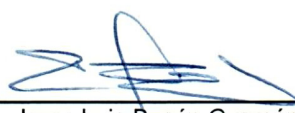
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Anderson Luiz Ara Souza, realizada em 16/03/2016:



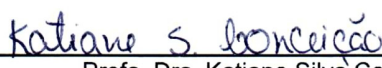
Prof. Dr. Francisco Louzada Neto
USP



Prof. Dr. Estevam Rafael Hruschka Junior
UFSCar



Prof. Dr. Jorge Luis Bazán Guzmán
USP



Profa. Dra. Katiane Silva Conceição
USP



Prof. Dr. Rafael Izbicki
UFSCar

Agradecimentos

Aos meus pais, Carmen Aparecida Ara de Souza e Valdeci Felisberto de Souza, por todo o suporte de vida que me deram.

A minha irmã Crystiane Fernanda de Souza, por toda a ajuda.

Aos meus avós, Aparecida Baptista Ara e Benedicto José Ara (*in memoriam*), por toda experiência e criação.

A Caio César Marani Silva por toda compreensão, paciência, carinho, apoio e dedicação.

Ao professor Benedito Galvão Benze, o qual me despertou para as Ciências Estatísticas.

A meu orientador, Francisco Louzada Neto, pela parceria que já nos rendeu diversos frutos, pela amizade e o vasto conhecimento me transferido.

Aos grandes e pequenos desafios cotidianos, os quais nos inspiram a buscar estratégias contínuas de resolução.

'É a vida desse meu lugar'

(M. Nascimento, F. Brant)

Resumo

Esta Tese expõe a proposição de uma nova classe de distribuições de probabilidade, denominada alfa normal assimétrica multivariada, uma extensão da distribuição alfa normal assimétrica univariada, introduzida por Elal-Olivero (2010). A distribuição proposta é muito flexível, capaz de assumir até duas modas e generaliza a distribuição proposta por Elal-Olivero em suas componentes marginais. Além disso, aplicamos esta nova distribuição na construção de dois novos métodos de *data mining* para classificação. Os procedimentos aqui desenvolvidos incrementam a capacidade preditiva da classificação na presença de dados assimétricos e/ou bimodais. Os resultados indicam que a nova proposição é significativamente mais apropriada que a modelagem usual por meio da distribuição normal clássica, além de ser igualmente adequada para conjuntos de dados sem a presença de assimetria. Nesta Tese são apresentados, utilizando dados reais e artificiais, os procedimentos de construção, estimação e validação tanto da nova distribuição de probabilidade quanto para as redes para classificações binárias, particularmente para redes probabilísticas de k-dependência.

Abstract

In this Thesis we expose the proposition of a new class of probability distributions, the so called alpha skew normal multivariate, an extension of the univariate Normal Alpha distribution, introduced by Elal-Olivero (2010). It can accommodate up to two modes and generalizes the distribution proposed by Elal-Olivero in its marginal components. In addition, we apply this new distribution in the construction of two new data mining methods for classification. The procedures developed here increment the predictive ability of the classification in the presence of asymmetric and / or bimodal data. The results indicate that the new proposal is significantly more appropriate than the usual modeling by classical normal distribution, and is also suitable for datasets without the presence of asymmetry. In this thesis it is shown, using real and synthetic data, the procedures of construction, estimation and validation for the new probability distribution and for probabilistic networks for binary classifications, particularly for the k-dependence probabilistic networks.

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 1.1 | Assimetria e distribuições de probabilidade | 3 |
| 2 | Distribuição alfa normal assimétrica bivariada | 6 |
| 2.1 | Gênese | 6 |
| 2.1.1 | Algumas propriedades | 7 |
| 2.1.2 | Modelo geral | 10 |
| 2.1.3 | Distribuições marginais | 10 |
| 2.2 | Inferência | 10 |
| 2.3 | Estudo de Simulação | 12 |
| 2.4 | Uma aplicação real ilustrativa | 13 |
| 2.5 | Comentários Finais | 16 |
| 3 | Distribuição alfa normal assimétrica multivariada | 17 |
| 3.1 | Gênese | 17 |
| 3.1.1 | Algumas propriedades | 18 |
| 3.1.2 | Distribuições marginais e condicionais | 23 |
| 3.1.3 | Modelo geral | 25 |
| 3.2 | Inferência | 27 |
| 3.3 | Estudo de simulação | 28 |
| 3.4 | Uma aplicação real ilustrativa | 29 |
| 3.5 | Comentários Finais | 33 |
| 4 | Redes de Classificação de K-dependência | 34 |
| 4.1 | Entropia e Informações Mútuas | 37 |
| 4.2 | Rede KDB e Alfa Normal Assimétrica | 38 |
| 4.3 | Comentários Finais | 40 |

| | | |
|----------|--|-----------|
| 5 | Validação do modelo proposto para classificação | 41 |
| 5.1 | Dados artificiais | 43 |
| 5.2 | Dados reais | 48 |
| 5.3 | Comentários Finais | 55 |
| 6 | Considerações Finais | 56 |
| | Referências Bibliográficas | 58 |

Lista de Figuras

| | | |
|-----|--|----|
| 1.1 | Principais técnicas de classificação na área de escore de crédito: HYBRID = métodos híbridos, COMBINED = métodos de reamostragem, SVM = máquina de vetores de suporte, NN = redes neurais, REG.LIN = regressão linear, TREES= árvores de decisão, LR = regressão logística, GENETIC = algoritmos genéticos, BN = redes probabilísticas, FUZZY = métodos <i>fuzzy</i> , DA = análise discriminante. | 2 |
| 1.2 | Noções de assimetria. | 3 |
| 2.1 | Representação da distribuição $ANAB(\alpha_1, \alpha_2, \rho)$ para diferentes valores de parâmetros. | 8 |
| 2.2 | Conjunto de dados reais: gráfico de dispersão e comportamento da distribuição ajustada (a) normal assimétrica bivariada e (b) alfa normal assimétrica bivariada. | 15 |
| 2.3 | Comportamento de G_n e $y_1 \times y_2$ para o teste de Kolmogorov-Smirnov bivariado. | 15 |
| 3.1 | Comportamento da distribuição alfa normal assimétrica trivariada sendo $\alpha = (0, 0, 0)$, $\rho_{12} = 0.5$, $\rho_{13} = 0$ e $\rho_{23} = -0.5$ | 20 |
| 3.2 | Comportamento da distribuição alfa normal assimétrica trivariada sendo $\alpha = (1, 0, 2)$, $\rho_{12} = 0$, $\rho_{13} = 0$ e $\rho_{23} = 0.7$ | 20 |
| 3.3 | Comportamento da distribuição alfa normal assimétrica trivariada sendo $\alpha = (-2, 1, 1)$, $\rho_{12} = 0$, $\rho_{13} = 0.5$ e $\rho_{23} = 0$ | 21 |
| 3.4 | Comportamento de Z_1 quando (a) $\alpha_3 = 0$, (b) $\alpha_3 = 1$ e (c) $\alpha_3 = 2$ | 25 |
| 3.5 | Gráfico de dispersão e distribuição normal multivariada ajustada. | 32 |
| 3.6 | Gráfico de dispersão e distribuição normal assimétrica multivariada ajustada. | 32 |
| 3.7 | Gráfico de dispersão e distribuição alfa normal assimétrica multivariada ajustada. | 33 |

| | | |
|-----|---|----|
| 4.1 | Em (a) apresentamos a estrutura tradicional da rede <i>naïve</i> Bayes e em (b) apresentamos uma rede probabilística com seis variáveis aleatórias. | 34 |
| 4.2 | Exemplificação da estrutura das redes probabilísticas com: (a) 0-dependência, (b) 1-dependência e (c) 2-dependência. | 36 |
| 5.1 | Distribuição das métrica F1 para os métodos em comparação quando $a = -1$. | 45 |
| 5.2 | Distribuição das métrica F1 para os métodos em comparação quando $a = 0$. | 45 |
| 5.3 | Distribuição das métrica F1 para os métodos em comparação quando $a = 1$. | 46 |
| 5.4 | Distribuição das métrica <i>MCC</i> para os métodos em comparação quando $a = -1$ | 46 |
| 5.5 | Distribuição das métrica <i>MCC</i> para os métodos em comparação quando $a = 0$ | 47 |
| 5.6 | Distribuição das métrica <i>MCC</i> para os métodos em comparação quando $a = 1$ | 47 |
| 5.7 | Distribuição das métricas (a) <i>ACC</i> , (b) <i>MCC</i> e (c) <i>F1</i> para o conjunto de dados HABERMAN. | 52 |
| 5.8 | Distribuição das métricas (a) <i>ACC</i> , (b) <i>MCC</i> e (c) <i>F1</i> para o conjunto de dados HEART. | 53 |
| 5.9 | Distribuição das métricas (a) <i>ACC</i> , (b) <i>MCC</i> e (c) <i>F1</i> para o conjunto de dados NORMTEMP. | 54 |

Lista de Tabelas

| | | |
|-----|---|----|
| 2.1 | Distribuição padrão: resultados da simulação para as 1000 estimativas de máxima verossimilhança e as duas diferentes configurações. | 13 |
| 2.2 | Distribuição não padrão: resultados da simulação para as 1,000 estimativas de máxima verossimilhança e as duas diferentes configurações. | 13 |
| 2.3 | Análise descritiva do tempos de erupção e espera para as 272 observações no gêiser de Yellowstone, EUA. | 14 |
| 2.4 | Estimativas dos parâmetros para os modelos Normal Bivariado (NB), Normal Assimétrico Bivariado (NAB) e alfa normal assimétrico bivariado ($ANAB$) para o conjunto de dados gêiser. | 14 |
| 3.1 | Distribuição padrão: média das 1 000 EMVs para quatro diferentes tamanhos amostrais, erros padrões e probabilidade de cobertura para intervalos de 95% de confiança. | 29 |
| 3.2 | Distribuição não padrão: média das 1 000 EMVs para quatro diferentes tamanhos amostrais, erros padrões e probabilidade de cobertura para intervalos de 95% de confiança. | 30 |
| 3.3 | Estatísticas básicas do logaritmo da massa, logaritmo do período e raiz quadrada da excentricidade dos 101 Exoplanetas descobertos até Outubro de 2002. | 30 |
| 3.4 | Estimativas dos parâmetros, valor máximo do logaritmo da função de verossimilhança e teste de aderência para a distribuição normal trivariada (NT), normal assimétrica trivariada (NAT) e alfa normal assimétrica trivariada ($ANAT$) para o conjunto de dados dos Exoplanetas. | 31 |
| 5.1 | Matriz de confusão. | 41 |

| | | |
|-----|---|----|
| 5.2 | Médias e desvios padrões das medidas <i>ACC</i> , <i>MCC</i> , <i>P</i> , <i>R</i> e <i>F1</i> para as diferentes técnicas de classificação | 44 |
| 5.3 | Estatísticas descritivas para o conjunto HABERMAN. | 49 |
| 5.4 | Estatísticas descritivas para o conjunto NORMTEMP. | 49 |
| 5.5 | Estatísticas descritivas para o conjunto HEART. | 50 |
| 5.6 | Médias e desvios padrões das medidas <i>ACC</i> , <i>MCC</i> , <i>R</i> , <i>P</i> e <i>F1</i> para os três conjuntos de dados e as diferentes técnicas de classificação. | 51 |

Capítulo 1

Introdução

As redes probabilísticas, também conhecidas como redes Bayesianas, redes de crença, redes causais ou gráficos de dependência probabilística, surgiram na década de 80 (Pearl, 1988) e têm sido aplicadas em uma grande variedade de atividades do mundo real. Algumas aplicações atuais se estendem a áreas como finanças (Xueling, 2010; Peng *et al.*, 2011), saúde (Bodén *et al.*, 2010; Burkom *et al.*, 2013), meio ambiente (Aguilera *et al.*, 2011; Chan *et al.*, 2012), desenvolvimento de jogos (Cheng *et al.*, 2010; Orun & Seker, 2012), descoberta de padrões em textos (de Campos & Romero, 2009; Xu *et al.*, 2013), entre outras.

Além disso, para Nagarajan *et al.* (2013) as redes probabilísticas nascem da interseção de várias áreas do conhecimento, como a Estatística, Teoria dos grafos e a Teoria das probabilidades, bem como fornecem um quadro abrangente e intuitivo de modelar dependências de variáveis aleatórias. Para Witten *et al.* (2011) a técnica de redes probabilísticas é um caso particular de modelo gráfico, sendo sua estrutura um grafo direcionado acíclico, assunto que vem se popularizando no contexto de aprendizagem de máquina e mineração de dados.

Ao mesmo tempo, a área de aprendizagem de máquina é o processo computacional de descobrir padrões em grandes conjuntos de dados e, de acordo com Mitchell (1997), um programa de computador é dito aprender com a experiência E com relação a alguma classe de tarefas T e medida de desempenho P se, seu desempenho em tarefas T , medida por P , melhora com a experiência E . Neste contexto, Mitchell (2006) considera o aprendizado de máquina uma consequência natural da intersecção entre a Ciência da Computação e a Estatística. Como caso particular de tarefas no processo de aprendizagem de máquina, a

classificação consiste na descoberta de modelos de previsão para auxílio no planejamento e tomada de decisões, sendo uma ferramenta indispensável e um tema bastante discutido na literatura (Tenev *et al.*, 2013).

No contexto de classificação, as redes probabilísticas possuem estruturas específicas e também são conhecidas como classificadores bayesianos. Alguns exemplos de estruturas são a rede *naïve* Bayes, *Tree Augmented Network* (TAN) (Friedman *et al.*, 1997) e as redes de k-dependência (KDB) (Sahami, 1996). Alguns exemplos de aplicações destas estruturas podem ser vistos em Madden (2009); Zhang *et al.* (2009); Murakami & Mizuguchi (2010); Rubio & Gámez (2011); Louzada & Ara (2012).

Diversas técnicas podem ser utilizadas em tarefas de classificação, o que torna as redes probabilísticas ainda um cenário incipiente e pouco explorado. Para evidenciar tal fato, realizamos uma revisão sistemática de literatura que envolveu 187 artigos relativos a classificação no contexto de escore de crédito. A Figura 1.1 exhibe as técnicas de classificação utilizadas dentre 1992-2015. Os demais resultados desta revisão estão condensados em Louzada *et al.* (2016).

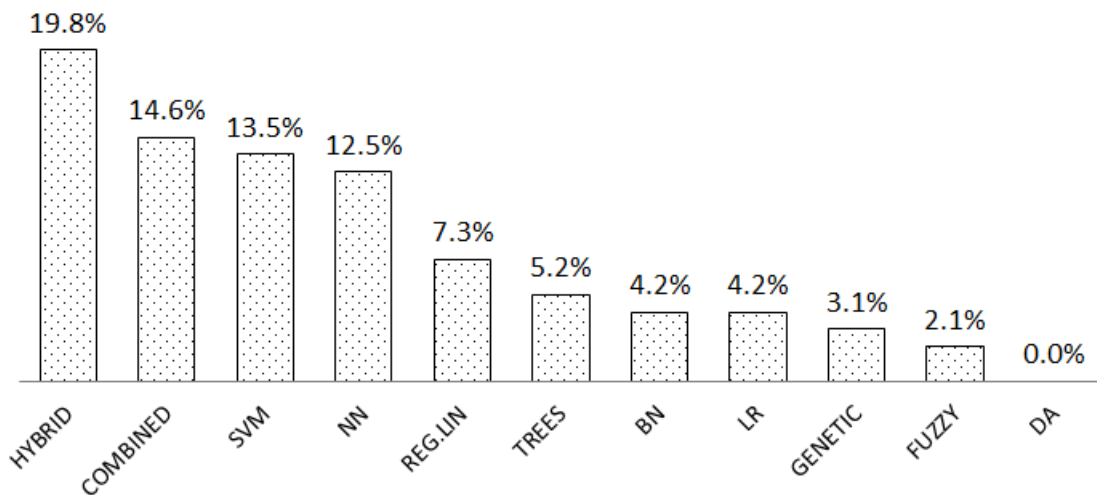


Figura 1.1: Principais técnicas de classificação na área de escore de crédito: HYBRID = métodos híbridos, COMBINED = métodos de reamostragem, SVM = máquina de vetores de suporte, NN = redes neurais, REG.LIN = regressão linear, TREES = árvores de decisão, LR = regressão logística, GENETIC = algoritmos genéticos, BN = redes probabilísticas, FUZZY = métodos *fuzzy*, DA = análise discriminante.

Este trabalho tem como foco as técnicas de redes probabilísticas para tarefas de classificação, sumariamente denominamos redes para classificação. Em especial, consideramos

as redes de k-dependência ajustadas para conjuntos de dados assimétricos.

1.1 Assimetria e distribuições de probabilidade

A assimetria ocorre quando os valores de uma variável ocorrem com frequências irregulares, levando a média, a mediana e a moda a ocorrerem em diferentes pontos. Em outras palavras, denomina-se assimetria o grau de afastamento de uma distribuição com relação ao seu eixo de simetria. Em suma, quando a média, a mediana e a moda coincidem, a distribuição é simétrica, caso contrário, a distribuição é dita assimétrica. Caso a média seja menor que a mediana a assimetria é dita a esquerda, caso a média seja maior que a mediana a assimetria é dita à direita. Tais comportamentos são evidenciados na Figura 1.2.

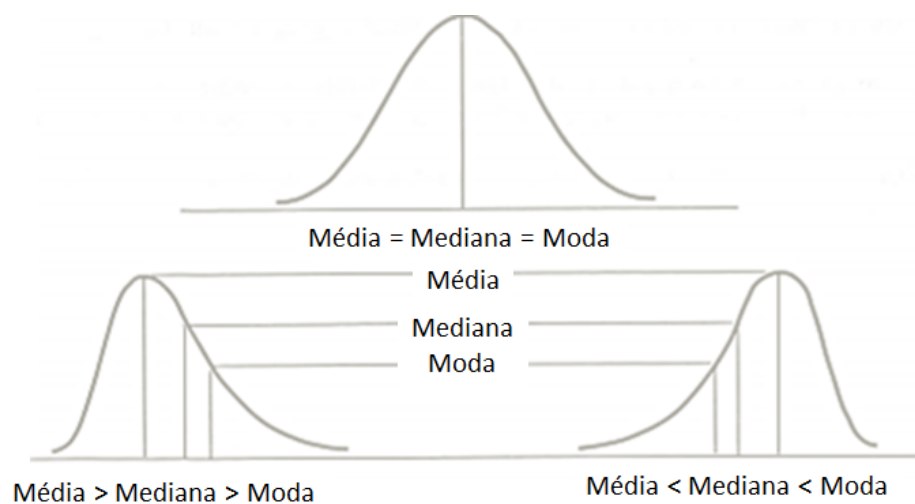


Figura 1.2: Noções de assimetria.

A distribuição normal é uma distribuição de probabilidade simétrica e amplamente utilizada em diversas aplicações em muitas áreas, como engenharia, física, finanças, biologia, entre outras. Entretanto, frequentemente encontramos conjuntos de dados reais em que as suposições de simetria desta distribuição são violadas. Este fato motivou o surgimento de distribuições mais gerais, as quais consideram uma perturbação da distribuição normal para captura possíveis assimetrias. Azzalini (1985) introduziu uma distribuição denominada normal assimétrica, a qual controla o formato de assimetria da distribuição por meio de um parâmetro, tendo a distribuição normal clássica como um caso particular. Desde então, a possibilidade de construção de classes de probabilidade mais flexíveis, no

que se diz respeito a assimetria e normalidade, tem motivado o surgimento de inúmeros outros trabalhos. Arellano-Valle *et al.* (2004) introduziram um modelo ainda mais geral que o proposto por Azzalini, denominado normal assimétrica generalizada. Gómez *et al.* (2006) introduziram uma classe de probabilidade que possui dois parâmetros de assimetria e possui a distribuição normal assimétrica como caso particular. Outros exemplos da extensão da distribuição proposta por Azzalini podem ser encontrados em Jamalizadeh *et al.* (2008), Bahrami *et al.* (2009), Arellano-Valle *et al.* (2010).

Para casos relativos a duas ou mais variáveis, Azzalini & Dalla Valle (1996) propuseram um modelo multivariado em que cada variável marginal possui uma distribuição normal assimétrica. Azzalini & Capitanio (1999) discutem alguns problemas de estimação para o caso da distribuição normal assimétrica multivariada. Branco & Dey (2001) introduziram uma classe de distribuições elípticas assimétricas. Gupta *et al.* (2004) definiram uma nova classe de distribuições relativas a normal assimétrica multivariada.

Paralelamente, Elal-Olivero (2010) propôs uma forma alternativa e flexível para capturar a assimetria em uma distribuição normal univariada, a qual considera unimodalidade ou, até mesmo, bimodalidade da distribuição. A classe de probabilidade introduzida por Elal-Olivero, conhecida como alfa normal assimétrica, tem levado a recentes extensões deste modelo por outros autores. Handam (2012) modificou esta distribuição para introduzir uma distribuição univariada chamada alfa normal assimétrica generalizada. Gui *et al.* (2012) produziram uma mistura de variáveis aleatórias simétricas de componente alfa normal e variáveis aleatórias uniformes para generalizar a distribuição *slash*-normal. Harandi & Alamatsaz (2013) propuseram uma classe alfa assimétrica para a distribuição Laplace univariada.

Neste sentido, introduzimos um novo tipo de rede de classificação baseada em uma nova classe de distribuições de probabilidade, designada alfa normal assimétrica multivariada, uma extensão da distribuição alfa normal assimétrica univariada, introduzida por Elal-Olivero (2010). A distribuição proposta é muito flexível, capaz de assumir até duas modas e generaliza a distribuição proposta por Elal-Olivero em suas componentes marginais. Tais fatos incrementam a capacidade preditiva da classificação para dados assimétricos ou bimodais.

Esta Tese está organizada em 6 capítulos. No Capítulo 2 exibimos a proposição da dis-

tribuição alfa normal assimétrica bivariada, bem como algumas de suas propriedades, procedimentos de inferência e aplicabilidade. No Capítulo 3 exibimos a proposição da distribuição alfa normal assimétrica multivariada, bem como algumas de suas propriedades e procedimentos de inferência. No Capítulo 4 é exibida a metodologia de construção da rede probabilística de k-dependência para a distribuição proposta. O Capítulo 5 exhibe a aplicação do método em conjuntos de dados reais e artificiais. Por fim, o Capítulo 6 apresenta os comentários finais e perspectivas futuras.

Capítulo 2

Distribuição alfa normal assimétrica bivariada

Neste capítulo introduzimos uma extensão bivariada do modelo alfa normal assimétrico, denominada distribuição alfa normal assimétrica bivariada. Sendo esta uma classe muito flexível e capaz de generalizar a distribuição alfa normal assimétrica univariada em suas distribuições marginais. Além disso, possui uma função de densidade de probabilidade com no máximo duas modas bem como possui, como caso particular, a distribuição normal bivariada clássica.

2.1 Gênese

A extensão para o caso bivariado resulta diretamente do caso univariado da distribuição alfa normal assimétrica (Elal-Olivero, 2010) e é definida a seguir.

Seja Z uma variável aleatória alfa normal assimétrica com função densidade de probabilidade dada por

$$f(z) = \frac{((1 - \alpha z)^2 + 1)}{(2 + \alpha^2)} \phi(z), \quad (2.1)$$

em que $\phi(\cdot)$ é a função densidade de probabilidade da distribuição normal padrão, α o parâmetro regulador de assimetria, com $-\infty < \alpha < \infty$ e $-\infty < z < \infty$. Assim, denotamos $Z \sim ANA(\alpha)$. É importante notar que se $\alpha = 0$ então $Z \sim N(0, 1)$; se $\alpha > 0$, Z possui uma assimetria positiva e torna-se bimodal a direita quando $\alpha \rightarrow \infty$; se $\alpha < 0$, Z possui uma assimetria negativa e torna-se bimodal à esquerda quando $\alpha \rightarrow -\infty$. O quociente dado em (2.1) pode ser interpretado como um termo de perturbação e gerador

de assimetria. Da mesma forma, a extensão bivariada é naturalmente obtida através da generalização da função densidade de probabilidade normal univariada para o caso bivariado e pela extensão do termo de perturbação de modo que possa levar em conta as contribuições aleatórias das duas variáveis, como exibido a seguir.

Seja $\mathbf{Z}' = (Z_1, Z_2)$ um vetor aleatório bivariado. Então, \mathbf{Z} possui uma distribuição alfa normal assimétrica bivariada (ANAB) se sua função densidade de probabilidade é dada por

$$f(z_1, z_2|\theta) = \frac{1 + (1 - \alpha_1 z_1 - \alpha_2 z_2)^2}{K} \phi_2(z), \quad (2.2)$$

em que $K = 2 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2\rho$, $\theta = (\alpha_1, \alpha_2, \rho)$ e $\phi_2(\cdot)$ é a função densidade de probabilidade da distribuição normal padrão bivariada $N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$, sendo $\mathbf{Z} \sim ANAB(\alpha_1, \alpha_2, \rho)$, com α_1 e $\alpha_2 \in \mathbb{R}$ e $\rho \in [-1; 1]$.

Proposição 1. *A função $f(z_1, z_2|\theta)$ é uma função densidade de probabilidade.*

Prova. Esta demonstração é simplesmente obtida através da integração,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z_1, z_2|\theta) dz_1 dz_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{-1} (1 + (1 - \alpha_1 z_1 - \alpha_2 z_2)^2) \phi_2(z) dz_1 dz_2 =$$

$$K^{-1} (2 - 2\alpha_1 E_\phi(Z_1) - 2\alpha_2 E_\phi(Z_2) + \alpha_1^2 E_\phi(Z_1^2) + \alpha_2^2 E_\phi(Z_2^2) + 2\alpha_1\alpha_2 E_{\phi_2}(Z_1 Z_2)) = 1$$

sendo E_ϕ e E_{ϕ_2} as esperanças univariada e bivariada da distribuição normal clássica, respectivamente. \square

2.1.1 Algumas propriedades

As propriedades a seguir são imediatamente deduzidas da função densidade de probabilidade (2.2):

- Se $\alpha_1 = 0$ e $\alpha_2 = 0$, então $\mathbf{Z} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$;
- Se $\alpha_1 = \alpha_2 \rightarrow \pm\infty$, então $f(z_1, z_2|\theta) = \frac{(z_1+z_2)^2}{2(1+\rho)} \phi_2(z)$;
- $-\mathbf{Z} \sim ANAB(-\alpha_1, -\alpha_2, \rho)$.

De maneira similar a densidade univariada (2.1), é possível mostrar que a função dada em (2.2) possui pelo menos duas modas, uma vez que o sistema de equações das derivadas de primeira ordem são dadas por

$$\frac{\partial f(z)}{\partial z_1} = \frac{\phi_2(z)}{K} \{2\alpha_1^2 z_1 + 2\alpha_1 \alpha_2 z_2 - 2\alpha_1 + [(-z_1 + \rho z_2)(1 + (1 - \alpha_1 z_1 - \alpha_2 z_2)^2)] / (1 - \rho^2)\} = 0$$

e

$$\frac{\partial f(z)}{\partial z_2} = \frac{\phi_2(z)}{K} \{2\alpha_2^2 z_2 + 2\alpha_1 \alpha_2 z_1 - 2\alpha_2 + [(\rho z_1 - z_2)(1 + (1 - \alpha_1 z_1 - \alpha_2 z_2)^2)] / (1 - \rho^2)\} = 0,$$

possui no máximo três raízes.

A Figura 2.1 exibe o comportamento da distribuição alfa normal assimétrica para diferentes valores de parâmetros.

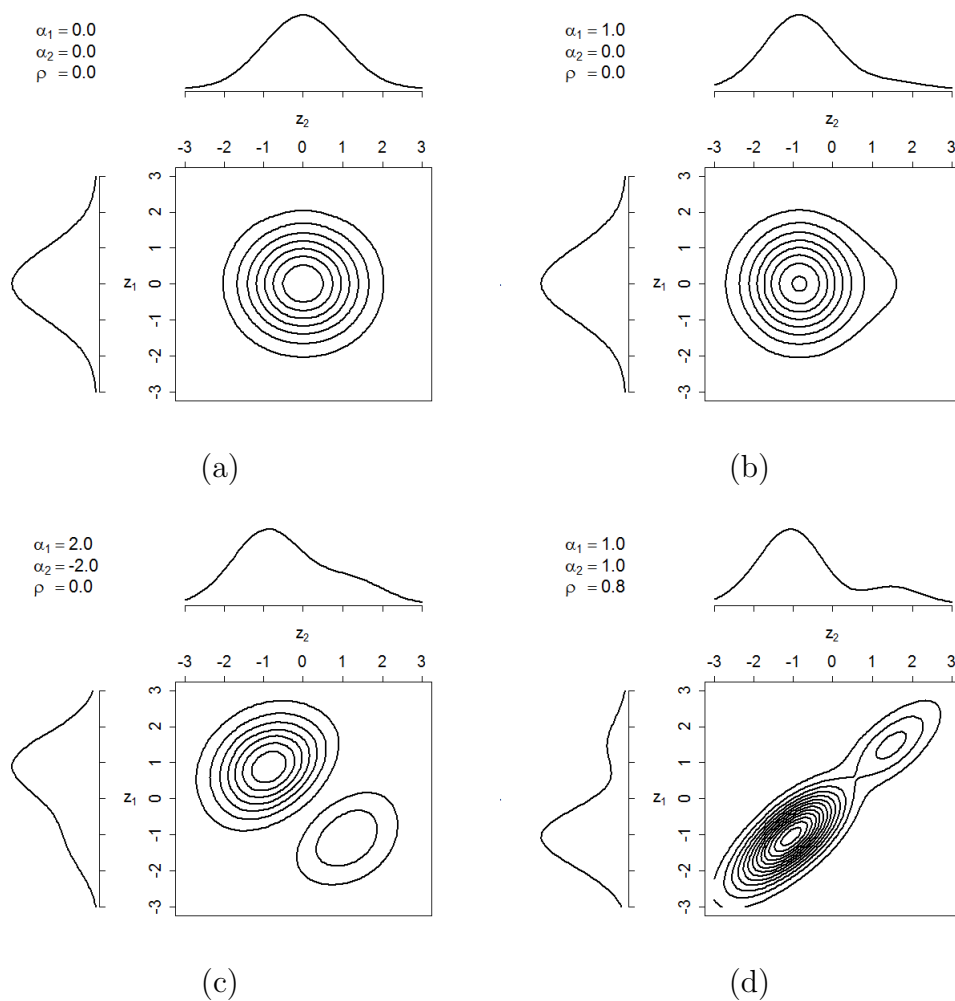


Figura 2.1: Representação da distribuição $ANAB(\alpha_1, \alpha_2, \rho)$ para diferentes valores de parâmetros.

Proposição 2. A função geradora de momentos de $\mathbf{Z}' = (Z_1, Z_2)$ é representada por $M_{Z_1, Z_2}(t_1, t_2) = E(\exp\{Z_1 t_1 + Z_2 t_2\})$ e dada por,

$$M_{Z_1, Z_2}(t_1, t_2) = K^{-1} \exp\left\{\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)\right\} \left[2 + a_1^2(1 + (t_1 + \rho t_2)^2) + a_2^2(1 + (\rho t_1 + t_2)^2) - 2a_1(t_1 + \rho t_2) - 2a_2(\rho t_1 + t_2) + 2a_1 a_2(t_1 t_2 + \rho(1 + t_1^2 + t_2^2) + t_1 t_2 \rho^2)\right]. \quad (2.3)$$

Prova. A prova de (2.3) é diretamente obtida por meio da solução

$$\begin{aligned} M_{Z_1, Z_2}(t_1, t_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{t_1 z_1 + t_2 z_2\} K^{-1} \left(1 + (1 - \alpha_1 z_1 - \alpha_2 z_2)^2\right) \phi(z_1, z_2) dz_1 dz_2 \\ &= M_{Z_1, Z_2}^*(t_1, t_2) \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(1 + (1 - \alpha_1 z_1 - \alpha_2 z_2)^2\right) \phi^*(z_1, z_2) dz_1 dz_2 \right) \\ &= M_{Z_1, Z_2}^*(t_1, t_2) (2 + \alpha_1^2 E(Z_1^2) + \alpha_2^2 E(Z_2^2) - 2\alpha_1 E(Z_1) - 2\alpha_2 E(Z_2) + 2\alpha_1 \alpha_2 E(Z_1 Z_2)), \end{aligned}$$

em que $M_{Z_1, Z_2}^*(t_1, t_2) = K^{-1} \exp\left\{\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)\right\}$, $\phi^*(z_1, z_2)$ é a função densidade de probabilidade de $N_2\left(\begin{bmatrix} -(t_1 + \rho t_2) \\ -(t_2 + \rho t_1) \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$, $E(Z_1^2) = 1 + (t_1 + \rho t_2)^2$, $E(Z_2^2) = 1 + (t_2 + \rho t_1)^2$ e $E(Z_1 Z_2) = \rho + (t_1 - \rho t_2)(t_2 - \rho t_1) = t_1 t_2 + \rho(1 + t_1^2 + t_2^2) + t_1 t_2 \rho^2$. \square

De uma forma geral, os momentos de \mathbf{Z} são obtidos por meio de (2.3) e (2.4),

$$E(Z_1^m Z_2^n) = \frac{\partial^{m+n}}{\partial t_1^m \partial t_2^n} M_{Z_1, Z_2}(t_1, t_2) \Big|_{t_1 = t_2 = 0}; \quad (2.4)$$

A esperança, e a matriz de variância e covariância de \mathbf{Z} são exibidas abaixo,

$$E \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = -2K^{-1} \begin{pmatrix} a_1 + a_2 \rho \\ a_1 \rho + a_2 \end{pmatrix}$$

e

$$COV \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} + K^{-2} (2K - 4) \begin{pmatrix} (a_1 + a_2 \rho)^2 & (a_1 \rho + a_2)(a_1 + a_2 \rho) \\ (a_1 \rho + a_2)(a_1 + a_2 \rho) & (a_1 \rho + a_2)^2 \end{pmatrix},$$

com K definido em (2.2).

2.1.2 Modelo geral

Seja a transformação: $X_1 = \sigma_1 Z_1 + \mu_1$ e $X_2 = \sigma_2 Z_2 + \mu_2$, em que μ_i é a esperança e σ_i o desvio padrão de X_i para a distribuição normal clássica, com $i = \{1, 2\}$. Por meio de (2.2), o vetor aleatório bivariado $X' = (X_1, X_2)$ possui função densidade de probabilidade dada por (2.5)

$$f(x_1, x_2 | \theta) = \frac{1 + \left(1 - \alpha_1 \frac{(x_1 - \mu_1)}{\sigma_1} - \alpha_2 \frac{(x_2 - \mu_2)}{\sigma_2}\right)^2}{K} \phi_2(x), \quad (2.5)$$

em que $K = 2 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2\rho$, $\theta = (\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \sigma_{12})$ e $\phi_2(\cdot)$ é a função densidade de probabilidade da distribuição normal bivariada, $N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$, uma vez que $\sigma_{12} = \sigma_1\sigma_2\rho$, sendo $X \sim ANAB(\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \sigma_{12})$.

2.1.3 Distribuições marginais

Por meio da integração marginal da função densidade de probabilidade conjunta (2.5), obtemos as seguintes funções densidade de probabilidade marginais

$$f(x_1 | \theta) = \frac{1 + \alpha_2^2(1 - \rho^2) + \left(1 - (\alpha_1 + \alpha_2\rho) \left(\frac{x_1 - \mu_1}{\sigma_1}\right)\right)^2}{K} \phi(x_1) \quad (2.6)$$

e

$$f(x_2 | \theta) = \frac{1 + \alpha_1^2(1 - \rho^2) + \left(1 - (\alpha_1\rho + \alpha_2) \left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right)^2}{K} \phi(x_2). \quad (2.7)$$

Um ponto importante a ser notado é que (2.6) e (2.7) são generalizações da distribuição alfa normal assimétrica univariada proposta por Elal-Olivero (2010), uma vez que $\alpha_2 = 0$, então $X_1 \sim ANA(\alpha_1, \mu_1, \sigma_1)$ e se $\alpha_1 = 0$ então $X_2 \sim ANA(\alpha_2, \mu_2, \sigma_2)$.

2.2 Inferência

Uma possível abordagem para realizar a inferência deste modelo é baseada no método usual de estimação de máxima verossimilhança. Considere X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da distribuição $ANAB(\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Desse modo, o logaritmo natural da função de verossimilhança é dado por,

$$\ell(\theta) = \sum_{i=1}^n \ln \left[\left(1 - \alpha_1 \frac{(x_{i,1} - \mu_1)}{\sigma_1} - \alpha_2 \frac{(x_{i,2} - \mu_2)}{\sigma_2} \right)^2 + 1 \right] - n \ln \left[\sigma_1 \sigma_2 \sqrt{1 - \rho^2} (2 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \rho) \right] \\ - (1 - \rho^2)^{-1} \sum_{i=1}^n \left[\left(\frac{x_{i,1} - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_{i,1} - \mu_1}{\sigma_1} \right) \left(\frac{x_{i,2} - \mu_2}{\sigma_2} \right) + \left(\frac{x_{i,2} - \mu_2}{\sigma_2} \right)^2 \right], \quad (2.8)$$

em que $\theta = (\mu_1, \sigma_1, \alpha_1, \mu_2, \sigma_2, \alpha_2, \rho)$. Os estimadores de máxima verossimilhança são obtidos diretamente através da maximização da função (2.8). Em outras palavras, $\hat{\theta} = \arg \max_{\theta} \ell(\theta)$.

A estimação intervalar é baseada distribuição normal assintótica dos estimadores de máxima verossimilhança, bem como seus erros padrões obtidos da matriz de Informação de Fisher.

Entretanto, a singularidade da matriz de informação de Fisher para modelos assimétricos que generalizam a distribuição normal é um problema recorrente, sendo necessária uma transformação do espaço paramétrico. Este assunto é tratado por Azzalini (1985), Chiogna (2005), Elal-Olivero (2010), Ley & Paindaveine (2010) e Hallin & Ley (2012).

Neste trabalho, não encontramos problemas de convergência considerando uma transformação similar a utilizada por Elal-Olivero (2010) para o caso univariado, sendo que μ_1 e μ_2 são substituídos por $\mu_1^* = \mu_1 - (\alpha_1 + \alpha_2 \rho) \sigma_1$ e $\mu_2^* = \mu_2 - (\alpha_2 + \alpha_1 \rho) \sigma_2$. Além disso, a matriz de Informação de Fisher obtida por meio de (2.8) não possui expressão analítica simples e pode ser obtida através de métodos numéricos como, por exemplo, o pacote `maxLik` do Software R (R Core Team, 2013; Henningsen & Toomet, 2011).

Entretanto, a matriz de informação de Fisher para o modelo padrão, em que $\mu_1 = \mu_2 = 0$ e $\sigma_1 = \sigma_2 = 1$, pode ser expressa por,

$$J(\alpha_1, \alpha_2, \rho) = K^{-2} \begin{bmatrix} G_{\alpha_1 \alpha_1} & G_{\alpha_1 \alpha_2} & G_{\alpha_1 \rho} \\ & G_{\alpha_2 \alpha_2} & G_{\alpha_2 \rho} \\ & & G_{\rho \rho} \end{bmatrix},$$

onde $G_{\alpha_1 \alpha_1} = 4 - 2(\alpha_1^2 + 2\alpha_1 \alpha_2 \rho + \alpha_2^2(2\rho^2 - 1)) + 2K^2 b(2, 0)$, $G_{\alpha_1 \alpha_2} = -4\alpha_1 \alpha_2 - 2(\alpha_1^2 + \alpha_2^2 - 2)\rho + 2K^2 b(1, 1)$, $G_{\alpha_2 \alpha_2} = 4 - 2(\alpha_2^2 + 2\alpha_1 \alpha_2 \rho + \alpha_1^2(2\rho^2 - 1)) + 2K^2 b(0, 2)$, $G_{\alpha_1 \rho} = 2\alpha_2(2 - \alpha_1^2 + \alpha_2^2)$, $G_{\alpha_2 \rho} = 2\alpha_1(2 + \alpha_1^2 - \alpha_2^2)$, e $G_{\rho \rho} = (\rho^2 - 1)^{-2} [4\alpha_1 \alpha_2 K \rho (\rho^2 - 1) - K(K(\rho^2 - 1) + 2\alpha_2^2(\rho^2 - 1)) - 2\alpha_1^2(\rho^2 - 1)(K + 2\alpha_2^2(\rho^2 - 1))]$, com

$b(m, n) = E \left(Z_1^m Z_2^n \frac{(1 - \alpha_1 Z_1 - \alpha_2 Z_2)^2 - 1}{((1 - \alpha_1 Z_1 - \alpha_2 Z_2)^2 - 1)^2} \right)$, m e n em $\{0, 1, 2\}$ termo que deve ser obtido numericamente. Fato que demonstra que nem mesmo a matriz de informação de Fisher para o modelo bivariado padrão possui forma analítica explícita.

2.3 Estudo de Simulação

Este estudo de simulação está baseado em 1000 conjuntos de dados artificiais gerados de uma distribuição ANAB em duas diferentes configurações paramétricas, o modelo padrão e não padrão desta distribuição. As replicações foram realizadas para os tamanhos amostrais 100, 250, 500 e 1000. Para a geração destas amostras, consideramos o procedimento MCMC de Gibbs *sampling*, também utilizado por Kotecha & Djuric (1999) para a geração aleatório de valores de uma distribuição normal multivariada truncada. Para a geração das condicionais no método de Gibbs, consideramos como o método da transformação inversa numericamente adaptado ao método de Brent (Brent, 1971; Toral & Chakrabarti, 1993).

Os resultados são condensados nas Tabela 2.1 e 3.1, que exhibe a média das estimativas de máxima verossimilhança (EMV), a média dos erros padrões associados (EP) e a probabilidade de cobertura (PC) dos intervalos de confiança assintóticos (95%). Visualmente, com o aumento do tamanho amostral as estimativas de máxima verossimilhança estão mais próximas do valor real dos parâmetros, há a diminuição do erro padrão e um aumento da probabilidade de cobertura para ambos os casos. Para a situação padrão, temos que os procedimentos de inferência são também satisfatórios para o caso da normal clássica bivariada.

Tabela 2.1: Distribuição padrão: resultados da simulação para as 1000 estimativas de máxima verossimilhança e as duas diferentes configurações.

| Tamanho Amostral | Parâmetro | μ_1 | σ_1 | α_1 | μ_2 | σ_2 | α_2 | ρ |
|------------------|-----------|---------|------------|------------|---------|------------|------------|--------|
| | | 0 | 1 | 0 | 0 | 1 | 0 | -0.3 |
| 100 | EMV | 0.04 | 0.97 | 0.08 | 0.03 | 0.98 | 0.08 | -0.29 |
| | EP | 1.53 | 0.07 | 1.66 | 1.46 | 0.07 | 1.56 | 0.09 |
| | PC | 0.93 | 0.91 | 0.93 | 0.92 | 0.92 | 0.93 | 0.93 |
| 250 | EMV | 0.03 | 0.98 | 0.04 | 0.02 | 0.99 | 0.05 | -0.30 |
| | EP | 1.01 | 0.04 | 1.06 | 0.99 | 0.04 | 1.03 | 0.06 |
| | PC | 0.95 | 0.92 | 0.96 | 0.96 | 0.94 | 0.95 | 0.95 |
| 500 | EMV | 0.03 | 0.99 | 0.03 | 0.02 | 1.00 | 0.04 | -0.30 |
| | EP | 1.04 | 0.03 | 1.06 | 0.96 | 0.03 | 0.98 | 0.04 |
| | PC | 0.96 | 0.93 | 0.96 | 0.96 | 0.94 | 0.96 | 0.95 |
| 1000 | EMV | 0.02 | 0.99 | 0.03 | 0.01 | 1.00 | 0.04 | -0.30 |
| | EP | 0.92 | 0.02 | 0.92 | 0.89 | 0.02 | 0.88 | 0.03 |
| | PC | 0.96 | 0.94 | 0.96 | 0.95 | 0.96 | 0.95 | 0.95 |

Tabela 2.2: Distribuição não padrão: resultados da simulação para as 1,000 estimativas de máxima verossimilhança e as duas diferentes configurações.

| Tamanho Amostral | Parâmetro | μ_1 | σ_1 | α_1 | μ_2 | σ_2 | α_2 | ρ |
|------------------|-----------|---------|------------|------------|---------|------------|------------|--------|
| | | -2 | 3 | 1 | 4 | 9 | -2 | 0.6 |
| 100 | EMV | -1.96 | 2.97 | 1.04 | 4.12 | 8.93 | -2.03 | 0.59 |
| | EP | 0.40 | 0.21 | 0.35 | 1.12 | 0.60 | 0.42 | 0.06 |
| | PC | 0.95 | 0.94 | 0.95 | 0.95 | 0.93 | 0.95 | 0.95 |
| 250 | EMV | -1.96 | 2.99 | 1.01 | 4.04 | 8.95 | -1.99 | 0.60 |
| | EP | 0.25 | 0.13 | 0.22 | 0.70 | 0.38 | 0.25 | 0.04 |
| | PC | 0.96 | 0.96 | 0.94 | 0.96 | 0.96 | 0.94 | 0.96 |
| 500 | EMV | -1.96 | 3.00 | 1.01 | 4.02 | 8.96 | -1.99 | 0.60 |
| | EP | 0.18 | 0.10 | 0.15 | 0.49 | 0.27 | 0.18 | 0.03 |
| | PC | 0.96 | 0.95 | 0.95 | 0.97 | 0.97 | 0.95 | 0.96 |
| 1000 | EMV | -1.96 | 3.00 | 1.01 | 4.02 | 8.96 | -1.99 | 0.60 |
| | EP | 0.12 | 0.07 | 0.11 | 0.34 | 0.19 | 0.12 | 0.02 |
| | PC | 0.96 | 0.97 | 0.95 | 0.95 | 0.96 | 0.95 | 0.96 |

2.4 Uma aplicação real ilustrativa

Nesta seção, avaliamos o desempenho do modelo alfa normal assimétrico bivariado proposto em um conjunto de dados conhecido na literatura e extraído de Härdle (1991) e explorado com mais detalhes por Azzalini & Bowman (1990). Este conjunto de dados é referente as erupções de um gêiser, cilindro natural de gás, localizado no Parque Nacional

de Yellowstone, EUA. A amostra corresponde a 272 observações das duas variáveis considerando o tempo de erupção (em minutos) e o tempo de espera até a próxima erupção (em minutos). A Tabela 2.3 exhibe uma análise descritiva dos dados.

Tabela 2.3: Análise descritiva do tempos de erupção e espera para as 272 observações no géiser de Yellowstone, EUA.

| Variável | Mínimo | Q1 | Mediana | Média | Q3 | Máximo | Desvio Padrão |
|----------------------------|--------|-------|---------|-------|-------|--------|---------------|
| Tempo de Erupção (em min.) | 1.60 | 2.20 | 4.00 | 3.50 | 4.50 | 5.10 | 1.10 |
| Tempo de Espera (em min.) | 43.00 | 58.00 | 76.00 | 70.90 | 82.00 | 96.00 | 13.60 |

Os resultados da estimação são exibidos na Tabela 2.4, sendo os erros padrões estimados numericamente através da matriz de informação de Fisher. A Figura 2.2 exhibe o comportamento da distribuição normal assimétrica ajustada e da distribuição alfa normal assimétrica ajustada, bem como suas respectivas distribuições marginais. Todos os resultados indicam que a distribuição alfa normal assimétrica é, para este conjunto de dados, mais adequada que a distribuição normal bivariada e a distribuição normal assimétrica bivariada.

Tabela 2.4: Estimativas dos parâmetros para os modelos Normal Bivariado (NB), Normal Assimétrico Bivariado (NAB) e alfa normal assimétrico bivariado ($ANAB$) para o conjunto de dados géiser.

| Estimativas dos Parâmetros | NB | NAB | $ANAB$ |
|----------------------------|-------------|---------------|---------------|
| α_1 | - | -11.42 (3.78) | -4.80 (-1.63) |
| α_2 | - | -1.88 (1.00) | -1.63 (1.02) |
| μ_1 | 3.49 (0.07) | 4.91 (0.05) | 3.23 (0.03) |
| μ_2 | 70.9 (0.86) | 86.58 (0.67) | 68.03 (0.48) |
| σ_1 | 1.14 (0.05) | 1.82 (0.44) | 0.69 (0.02) |
| σ_2 | 0.9 (0.57) | 20.74 (4.31) | 8.95 (0.27) |
| ρ | 0.9 (0.01) | 0.96 (0.02) | 0.76 (0.02) |
| $l(\hat{\theta})$ | -789.8942 | -1243.464 | -675.05 |
| AIC | 1589.7884 | 2496.928 | 1360.1 |

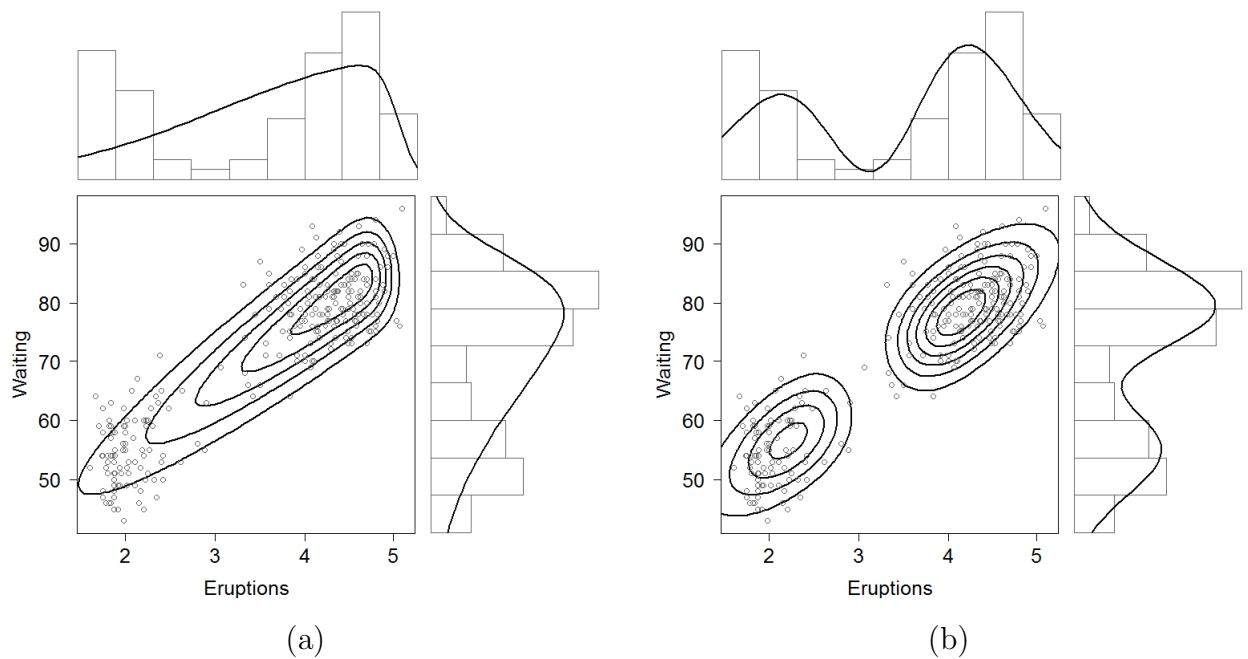


Figura 2.2: Conjunto de dados reais: gráfico de dispersão e comportamento da distribuição ajustada (a) normal assimétrica bivariada e (b) alfa normal assimétrica bivariada.

Para uma comparação mais formal do modelo ajustado e os dados reais, verificamos a qualidade de ajuste por meio da aplicação do teste de Kolmogorov-Smirnov multivariado proposto por Justel *et al.* (1997), no qual a estatística K-S bivariada é definida por $D_n = \max\{D_n^1, D_n^2\}$, onde $D_n^1 = \sup[G_n(y_1, y_2) - y_1 \times y_2]$, através das transformações $y_1 = F(x_1)$ e $y_2 = F(x_2|x_1)$, com $D_n^2 = \sup[G_n(y_2, y_1) - y_2 \times y_1]$, $y_2 = F(x_2)$ e $y_1 = F(x_1|x_2)$, em que G_n é a função de distribuição empírica da amostra transformada. A Figura 2.3 exibe o relacionamento de G_n e $y_1 \times y_2$ para as duas permutações consideradas no teste.

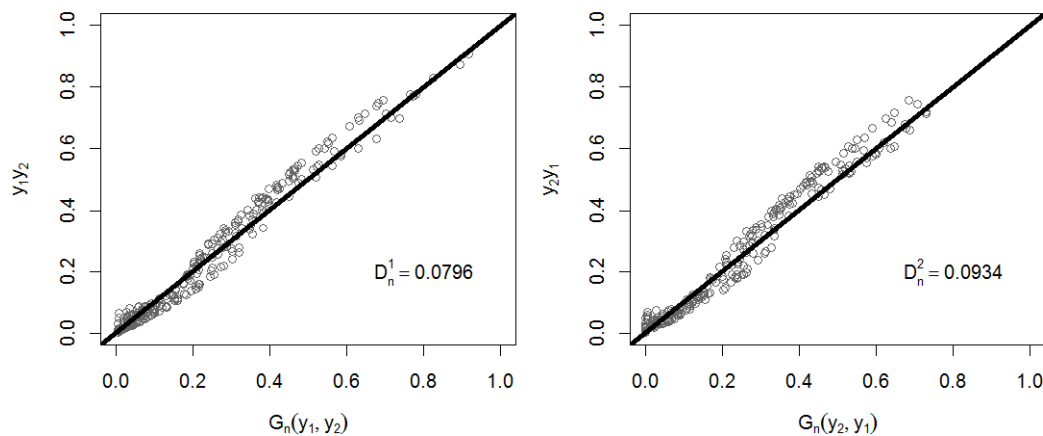


Figura 2.3: Comportamento de G_n e $y_1 \times y_2$ para o teste de Kolmogorov-Smirnov bivariado.

Dado que $D_n = 0.0934$ é menor que o respectivo valor crítico em uma significância global de 5% (0.1041), como indicado pelos autores (Justel *et al.* (1997), Tabela 1, página 254), concluimos que o tempo de erupção e espera seguem uma distribuição alfa normal assimétrica bivariada, fato também sustentado visualmente pela Figura 2.2. Além disso, temos que para a distribuição normal bivariada $D_n = 0,8361$ e a distribuição normal assimétrica bivariada $D_n = 0.9852$, então podemos concluir que o tempo de espera entre erupções e a duração da erupção não segue uma distribuição normal bivariada nem uma distribuição normal assimétrica bivariada.

A critério de análise, vale ressaltar que ambas as variáveis possuem sua média amostral menor que sua mediana amostral, o que indica uma assimetria à esquerda em cada uma das dimensões, este fato é corroborado pelo modelo alfa normal bivariado nas estimativas $\hat{\alpha}_1 < 0$ e $\hat{\alpha}_2 < 0$.

2.5 Comentários Finais

Neste capítulo apresentamos a distribuição alfa normal assimétrica bivariada, bem como sua função densidade de probabilidade, suas propriedades, comportamento gráfico, função geradora de momentos, os momentos principais - esperança e variância, seu modelo geral, as funções densidade de probabilidade marginal, sua respectiva função de verossimilhança e matriz de informação de Fisher. Além disso, apresentamos sua aplicação em dados artificiais e reais. Os procedimentos aqui apresentados apresentam-se como necessários para a modelagem de dados bivariados na presença de assimetria e/ou bimodalidade, bem como se sobrepõem a distribuição normal clássica bivariada e a distribuição normal assimétrica bivariada para o caso analisado.

Capítulo 3

Distribuição alfa normal assimétrica multivariada

Neste capítulo apresentamos a distribuição alfa normal assimétrica multivariada, bem como suas propriedades estatísticas e sua aplicação a dados artificiais e reais.

3.1 Gênese

Seja $\mathbf{Z}' = (Z_1, Z_2, \dots, Z_p)$ um vetor aleatório composto por p variáveis aleatórias e $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \dots, \alpha_p)$ um vetor composto por p parâmetros controladores de assimetria. Através do mesmo critério adotado por Seber (1977) para estender a distribuição normal univariada para o caso multivariado, por analogia a função densidade de probabilidade da distribuição univariada alfa normal assimétrica,

$$\begin{aligned} f(z) &= \frac{((1-\alpha z)^2+1)}{2+\alpha^2} \phi(z) \\ &= ((1-\alpha z)^2+1) K^{-1} \phi(z), \end{aligned}$$

com $\phi(z)$ a função densidade de probabilidade da distribuição normal padrão com $-\infty < z < \infty$. Podemos definir a função densidade de probabilidade da distribuição alfa normal assimétrica multivariada (ANAM) padrão como,

$$f(\mathbf{z}) = \left((1 - \boldsymbol{\alpha}' \mathbf{z})^2 + 1 \right) K^{-1} \phi_p(\mathbf{z}), \quad (3.1)$$

sendo $-\infty < z_i < \infty$, $i = 1, \dots, p$, $\boldsymbol{\alpha} \in \mathbb{R}^p$ e $\phi_p(\mathbf{z})$ a função densidade de probabilidade da distribuição normal padrão p -variada.

Proposição 3. Se $\mathbf{Z}' = (Z_1, Z_2, \dots, Z_p)$ é um vetor aleatório com função densidade de probabilidade expressa por 3.1, então $K = 2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}$, com $\boldsymbol{\Sigma}$ a matriz de correlação de \mathbf{Z} .

Prova.

$$\begin{aligned}
& \int_{\mathbf{z}} \left[(1 - \boldsymbol{\alpha}'\mathbf{z})^2 + 1 \right] \phi_p(\mathbf{z}) d\mathbf{z} \\
&= \int_{\mathbf{z}} (1 - \boldsymbol{\alpha}'\mathbf{z})^2 \phi_p(\mathbf{z}) d\mathbf{z} + 1 \\
&= \int_{\mathbf{z}} (1^2 - 2\boldsymbol{\alpha}'\mathbf{z} + \boldsymbol{\alpha}'\mathbf{z}\mathbf{z}'\boldsymbol{\alpha}) \phi_p(\mathbf{z}) d\mathbf{z} + 1 \\
&= -2\boldsymbol{\alpha}' \int_{\mathbf{z}} \mathbf{z} \phi_p(\mathbf{z}) d\mathbf{z} + \int_{\mathbf{z}} \boldsymbol{\alpha}'\mathbf{z}\mathbf{z}'\boldsymbol{\alpha} \phi_p(\mathbf{z}) d\mathbf{z} + 2 \\
&= -2\boldsymbol{\alpha}'E(\mathbf{Z}) + \boldsymbol{\alpha}'E(\mathbf{Z}\mathbf{Z}')\boldsymbol{\alpha} + 2 \\
&= 2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}.
\end{aligned}$$

Desta forma, $(2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha})^{-1} \int_{\mathbf{z}} \left((1 - \boldsymbol{\alpha}'\mathbf{z})^2 + 1 \right) \phi_p(\mathbf{z}) d\mathbf{z} = 1$, sendo (3.1) uma função densidade de probabilidade. \square

Vale ressaltar que o caso bivariado, caso particular da demonstração acima, já foi apresentado anteriormente na Seção 2.1.

3.1.1 Algumas propriedades

As propriedades a seguir são deduzidas da função densidade de probabilidade 3.1:

- a. Se $\boldsymbol{\alpha} = \underset{\sim}{0}$, então $\mathbf{Z} \sim N_p(\underset{\sim}{0}, \boldsymbol{\Sigma})$;
- b. Se $\boldsymbol{\alpha} = \mathbf{1}c$ e $c \rightarrow \pm\infty$, então $f(\mathbf{z}) = \frac{(\mathbf{1}'\mathbf{z})^2}{2(1+S)} \phi_p(\mathbf{z})$, sendo $S = \frac{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}-p}{p-1}$ a soma dos coeficientes da matriz de correlação.
- c. $-\mathbf{Z} \sim ANAM(-\boldsymbol{\alpha}, \boldsymbol{\Sigma})$

Prova.

- a. Quando $\boldsymbol{\alpha} = \underset{\sim}{0}$, o primeiro termo do produto deduzido é reduzido para 1.

b. Considere o caso $p = 3$, com $\boldsymbol{\alpha} = (c, c, c)'$, $c \in \mathbb{R}$ e $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$. Assim,

$$\frac{(1 - \boldsymbol{\alpha}'\mathbf{Z})^2 + 1}{2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}} = \frac{z_1^2 + z_2^2 + z_3^2 + 2(z_1z_2 + z_1z_3 + z_2z_3) - 2(z_1 + z_2 + z_3)c^{-1} + 2c^{-2}}{2(c^{-2} + 1 + \rho_{12} + \rho_{13} + \rho_{23})}$$

quando $c \rightarrow \pm\infty$ $\frac{(1 - \boldsymbol{\alpha}'\mathbf{Z})^2 + 1}{2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}} = \frac{(z_1 + z_2 + z_3)^2}{2(1+S)}$ sendo $S = \rho_{12} + \rho_{13} + \rho_{23}$.

c. Quando $((1 - \boldsymbol{\alpha}'(-\mathbf{z}))^2) = (1 - (-\boldsymbol{\alpha})'\mathbf{z})^2$. \square

Da mesma forma que realizado para o caso bivariado, a função densidade de probabilidade dada por 3.1 possui pelo menos duas modas para o caso multivariado, uma vez que o sistema de equações das derivadas de primeira ordem, de ordem p , é dado por

$$\frac{\partial f(\mathbf{z})}{\partial \mathbf{z}} = K^{-1}\phi_p(\mathbf{z}) \left\{ -2\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\alpha}(\mathbf{H} + \mathbf{I})\mathbf{z} - \left[(1 - \boldsymbol{\alpha}'\mathbf{z})^2 + 1 \right] \boldsymbol{\Sigma}^{-1}\mathbf{z} \right\} = \underset{\sim}{0}, \quad (3.2)$$

o qual possui no máximo três raízes, sendo \mathbf{H} uma matriz de 1's de ordem p e \mathbf{I} a matriz identidade de ordem p . A prova é bastante simples, uma vez que $\phi_p(\mathbf{z})$ é uma função densidade de probabilidade, sendo estritamente positiva e K não depende de \mathbf{z} . Portanto, 3.2 é um sistema de equações que envolve um polinômio de grau três.

As Figuras 3.1-3.3 exibem o comportamento da distribuição alfa normal assimétrica trivariada padrão para diferentes valores de parâmetros, sendo o comportamento apresentado pela Figura 3.1 o mesmo comportamento simétrico apresentado pela distribuição normal trivariada padrão.

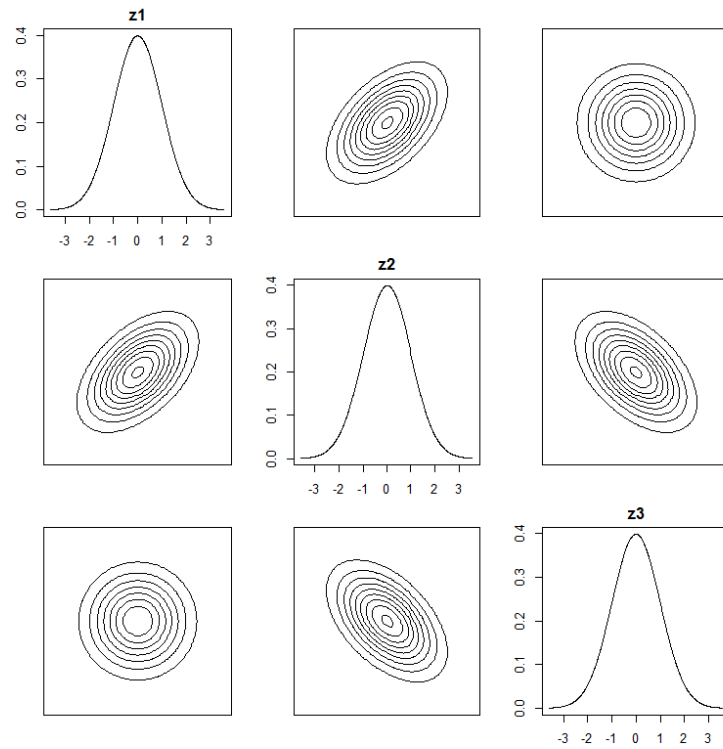


Figura 3.1: Comportamento da distribuição alfa normal assimétrica trivariada sendo $\alpha = (0, 0, 0)$, $\rho_{12} = 0.5$, $\rho_{13} = 0$ e $\rho_{23} = -0.5$.

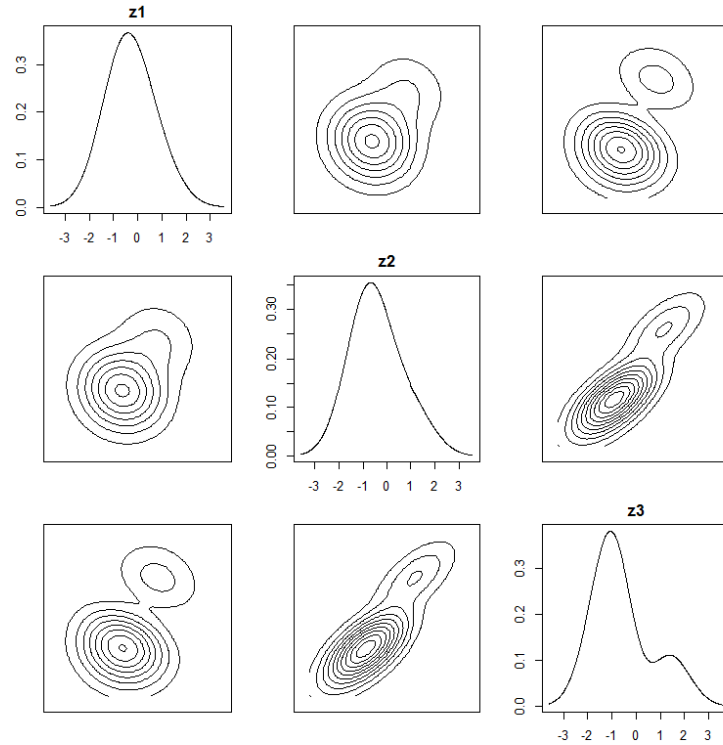


Figura 3.2: Comportamento da distribuição alfa normal assimétrica trivariada sendo $\alpha = (1, 0, 2)$, $\rho_{12} = 0$, $\rho_{13} = 0$ e $\rho_{23} = 0.7$.

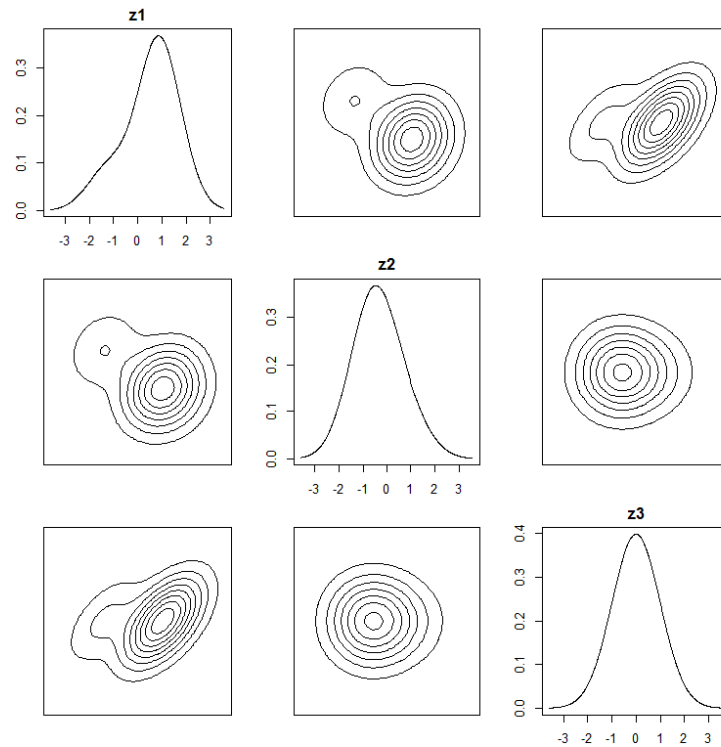


Figura 3.3: Comportamento da distribuição alfa normal assimétrica trivariada sendo $\alpha = (-2, 1, 1)$, $\rho_{12} = 0$, $\rho_{13} = 0.5$ e $\rho_{23} = 0$.

Proposição 4. A função geradora de momentos de $\mathbf{Z}' = (Z_1, Z_2, \dots, Z_p)$ é representada por $M_{\mathbf{Z}}(\mathbf{t}) = E(\exp\{\mathbf{Z}'\mathbf{t}\})$ e dada por,

$$M_{\mathbf{Z}}(\mathbf{t}) = \exp\left\{\frac{\mathbf{t}'\Sigma\mathbf{t}}{2}\right\} [1 - K^{-1}\alpha'\Sigma\mathbf{t}(2 - \mathbf{t}'\Sigma\alpha)]. \quad (3.3)$$

sendo $K = 2 + \alpha'\Sigma\alpha$.

Prova. A prova de (3.3) é diretamente obtida por meio da solução

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) &= E(\exp\{\mathbf{t}'\mathbf{Z}\}) \\ &= \int_{\mathbf{z}} \frac{(1 - \alpha'\mathbf{z})^2 + 1}{2 + \alpha'\Sigma\alpha} \exp\{\mathbf{t}'\mathbf{z}\} \phi_p(\mathbf{z}) d\mathbf{z} \\ &= \int_{\mathbf{z}} \frac{(1 - \alpha'\mathbf{z})^2 + 1}{2 + \alpha'\Sigma\alpha} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}'\Sigma^{-1}\mathbf{z}) + \mathbf{t}'\mathbf{z}\right\} d\mathbf{z} \end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbf{z}} \frac{2-2\boldsymbol{\alpha}'\mathbf{z}+\boldsymbol{\alpha}'\mathbf{z}\mathbf{z}'\boldsymbol{\alpha}}{2+\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}) + \mathbf{t}'\mathbf{z}\right\} d\mathbf{z} \\
&= E_{\mathbf{Z}}\left(\frac{2-2\boldsymbol{\alpha}'\mathbf{Z}+\boldsymbol{\alpha}'\mathbf{Z}\mathbf{Z}'\boldsymbol{\alpha}}{2+\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}\right) \times \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\}, \text{ com } \mathbf{Z} \sim N_p(\boldsymbol{\Sigma}\mathbf{t}, \boldsymbol{\Sigma}) \\
&= \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} [1 - K^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}(2 - \mathbf{t}'\boldsymbol{\Sigma}\boldsymbol{\alpha})].
\end{aligned}$$

□

A esperança, $E(\mathbf{Z})$, e a matriz de variância e covariância, $COV(\mathbf{Z})$, do vetor aleatório \mathbf{Z} são obtidas através de 3.3 e são apresentadas a seguir.

Proposição 5. *A esperança de $\mathbf{Z}' = (Z_1, Z_2, \dots, Z_p)$ com $\mathbf{Z} \sim ANAM(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$ é dada por,*

$$E(\mathbf{Z}) = -K^{-1}2\boldsymbol{\alpha}'\boldsymbol{\Sigma}. \quad (3.4)$$

sendo $K = 2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}$.

Prova. A partir da função geradora de momentos definida em (3.3), a esperança de \mathbf{Z} é dada por $\frac{\partial M_{\mathbf{Z}}^{(1)}(t)}{\partial t}|_{t=\mathbf{0}}$. Desta forma, temos que

$$\begin{aligned}
\frac{\partial M_{\mathbf{Z}}^{(1)}(t)}{\partial t} &= \frac{\partial}{\partial t} \left[\exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} (1 - K^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}(2 - \mathbf{t}'\boldsymbol{\Sigma}\boldsymbol{\alpha})) \right. \\
&\quad \left. + \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} \frac{\partial}{\partial t} \left[\frac{1-2\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}-\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}\mathbf{t}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}{2+\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}} \right] \right] \\
&= \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} \boldsymbol{\Sigma}\mathbf{t} (1 - K^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}(2 - \mathbf{t}'\boldsymbol{\Sigma}\boldsymbol{\alpha})) + \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} \frac{\partial}{\partial t} \left[\frac{-2\boldsymbol{\alpha}'\boldsymbol{\Sigma}-\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\boldsymbol{\alpha}}{2+\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}} \right]. \\
&= \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} \boldsymbol{\Sigma}\mathbf{t} [1 - K^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}\mathbf{t}(2 - \mathbf{t}'\boldsymbol{\Sigma}\boldsymbol{\alpha})] + \exp\left\{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\} \left(\frac{-2\boldsymbol{\alpha}'\boldsymbol{\Sigma}-\boldsymbol{\alpha}'\boldsymbol{\Sigma}(\mathbf{1}'\mathbf{t}+I)\boldsymbol{\Sigma}\boldsymbol{\alpha}}{2+\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}} \right)
\end{aligned}$$

Assim, temos que $E(\mathbf{Z}) = \frac{\partial M_{\mathbf{Z}}^{(1)}(t)}{\partial t}|_{t=\mathbf{0}} = \frac{-2\boldsymbol{\alpha}'\boldsymbol{\Sigma}}{2+\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}$.

□

Proposição 6. A matriz de variância e covariância de $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)'$ com $\mathbf{Z} \sim ANAM(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$ é dada por,

$$COV(\mathbf{Z}) = \boldsymbol{\Sigma} + K^{-2} \boldsymbol{\alpha}' \boldsymbol{\Sigma} [(\mathbf{H} + \mathbf{I})K - 4] \boldsymbol{\Sigma} \boldsymbol{\alpha}. \quad (3.5)$$

sendo $K = 2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}$, \mathbf{H} uma matriz de 1's de ordem p e \mathbf{I} a matriz identidade de mesma ordem.

Prova. A prova de (3.5) é obtida por

$$\begin{aligned} E(\mathbf{Z}\mathbf{Z}') &= \frac{\partial M_{\mathbf{Z}}^{(2)}(\mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}'} \Big|_{\mathbf{t}=\mathbf{0}} = \frac{\partial}{\partial \mathbf{t}} \left[\exp \left\{ \frac{1}{2} \mathbf{t}' \boldsymbol{\Sigma} \mathbf{t} \right\} (1 - K^{-1} \boldsymbol{\alpha}' \boldsymbol{\Sigma} \mathbf{t} (2 - \mathbf{t}' \boldsymbol{\Sigma} \boldsymbol{\alpha})) \right] \Big|_{\mathbf{t}=\mathbf{0}} \\ &= \boldsymbol{\Sigma} + \frac{\boldsymbol{\alpha}' \boldsymbol{\Sigma} (\mathbf{H} + \mathbf{I}) \boldsymbol{\Sigma} \boldsymbol{\alpha}}{2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}}. \end{aligned}$$

Como $COV(\mathbf{Z}) = E(\mathbf{Z}\mathbf{Z}') - E(\mathbf{Z})E(\mathbf{Z})'$, assim

$$COV(\mathbf{Z}) = \boldsymbol{\Sigma} + \frac{\boldsymbol{\alpha}' \boldsymbol{\Sigma} (\mathbf{H} + \mathbf{I}) \boldsymbol{\Sigma} \boldsymbol{\alpha}}{2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}} - \left(\frac{-2\boldsymbol{\alpha}' \boldsymbol{\Sigma}}{2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}} \right) \left(\frac{-2\boldsymbol{\alpha}' \boldsymbol{\Sigma}}{2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}} \right)', \text{ temos que}$$

$$COV(\mathbf{Z}) = \boldsymbol{\Sigma} + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \left(\frac{\mathbf{H} + \mathbf{I}}{2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}} - \frac{4}{(2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha})^2} \right) \boldsymbol{\Sigma} \boldsymbol{\alpha}.$$

□

3.1.2 Distribuições marginais e condicionais

Seja o vetor aleatório \mathbf{Z} particionado como $\mathbf{z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix}$ com tamanho $\begin{bmatrix} q \times 1 \\ (p - q) \times 1 \end{bmatrix}$, da mesma forma $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ com tamanho $\begin{bmatrix} q \times 1 \\ (p - q) \times 1 \end{bmatrix}$ e $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$ com tamanho $\begin{bmatrix} q \times q & q \times (p - q) \\ (p - q) \times q & (p - q) \times (p - q) \end{bmatrix}$.

Por meio da integração multidimensional da função densidade de probabilidade conjunta expressa em (3.1), obtemos a função densidade de probabilidade marginal de \mathbf{Z}_1 e a

função densidade de probabilidade condicional $\mathbf{Z}_1|\mathbf{z}_2$, exibidas nas Proposições 7 e 9, respectivamente.

Proposição 7. *A função densidade de probabilidade de \mathbf{Z}_1 é dada por,*

$$f(\mathbf{z}_1) = \left[(1 - \boldsymbol{\alpha}'_1 \mathbf{z}_1)^2 + 1 + \boldsymbol{\alpha}'_2 \left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \right) \boldsymbol{\alpha}_2 + h_1 (h_1 + 2 (\boldsymbol{\alpha}'_1 \mathbf{z}_1 - 1)) \right] K^{-1} \phi_1(\mathbf{z}_1), \quad (3.6)$$

sendo $h_1 = \boldsymbol{\alpha}'_2 \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{z}_1$.

Prova. Seja $(1 - \boldsymbol{\alpha}' \mathbf{z})^2 + 1 = (1 - \sum_{i=1}^p \alpha_i z_i)^2 + 1 = 2 + \sum_{i=1}^p \alpha_i^2 z_i^2 - 2 \sum_{i=1}^p \alpha_i z_i + 2 \sum_{i=1}^p \sum_{j=i+1}^p \alpha_i \alpha_j z_i z_j$. Podemos escrever,

$$(1 - \boldsymbol{\alpha}' \mathbf{z})^2 + 1 = (1 - \boldsymbol{\alpha}'_1 \mathbf{z}_1)^2 + (1 - \boldsymbol{\alpha}'_2 \mathbf{z}_2)^2 + 2 \boldsymbol{\alpha}'_1 \mathbf{z}_1 \boldsymbol{\alpha}'_2 \mathbf{z}_2.$$

Assim,

$$\begin{aligned} & K^{-1} \int_{\mathbf{z}_2} (1 - \boldsymbol{\alpha}' \mathbf{z})^2 + 1 \phi_p(\mathbf{z}) d\mathbf{z} \\ &= K^{-1} \phi_q(\mathbf{z}_1) \left\{ (1 - \boldsymbol{\alpha}'_1 \mathbf{z}_1)^2 + E \left[(1 - \boldsymbol{\alpha}'_2 \mathbf{z}_2)^2 \mid \mathbf{z}_1 \right] + 2 \boldsymbol{\alpha}'_1 \mathbf{z}_1 E \left[\mathbf{z}'_2 \mid \mathbf{z}_1 \right] \right\} \\ & K^{-1} \phi_q(\mathbf{z}_1) \left\{ (1 - \boldsymbol{\alpha}'_1 \mathbf{z}_1)^2 + 1 + \boldsymbol{\alpha}'_2 \left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \right) \boldsymbol{\alpha}_2 \right. \\ & \left. + \boldsymbol{\alpha}'_2 \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{z}_1 \left(\boldsymbol{\alpha}'_2 \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{z}_1 + 2 (\boldsymbol{\alpha}'_1 \mathbf{z}_1 - 1) \right) \right\}, \end{aligned}$$

uma vez que, utilizando as propriedades da distribuição condicional da normal multivariada $E[\mathbf{Z}_2 | \mathbf{z}_1] = \boldsymbol{\Sigma}_{22} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{z}_1$ e $E[\mathbf{Z}_2 \mathbf{Z}_2' | \mathbf{z}_1] = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{I}_p - \mathbf{z}_1 \mathbf{z}'_1 \boldsymbol{\Sigma}_{11}^{-1}) \boldsymbol{\Sigma}_{12}$.

□

Proposição 8. *A distribuição marginal de \mathbf{Z}_1 depende de, além de seus parâmetros de assimetria $\boldsymbol{\alpha}_1$, dos demais parâmetros de assimetria dados em $\boldsymbol{\alpha}_2$.*

A Proposição 8 é intuitiva e verificada na expressão 3.6. Esta propriedade é verificada graficamente através da Figura 3.4 para a partição $\mathbf{Z}_1 = (Z_1, Z_2)'$ e $\mathbf{Z}_2 = Z_3$.

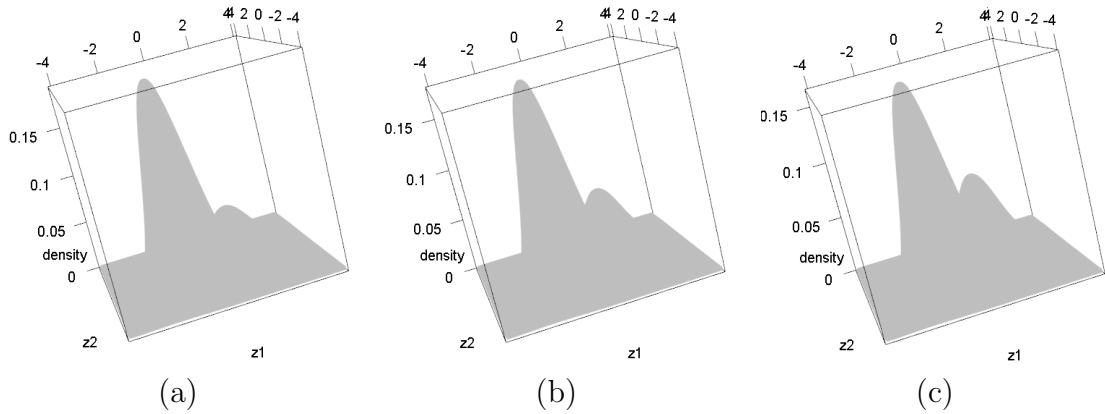


Figura 3.4: Comportamento de \mathbf{Z}_1 quando (a) $\alpha_3 = 0$, (b) $\alpha_3 = 1$ e (c) $\alpha_3 = 2$.

Proposição 9. A função densidade de probabilidade de $\mathbf{Z}_1|\mathbf{z}_2$ é dada por,

$$f(z_1|z_2) = \frac{(1 - \alpha'_1 z_1)^2 + 2\alpha'_1 z_1 z'_2 \alpha_2 + (1 - \alpha'_2 z_2)^2}{(1 - \alpha'_1 z_2)^2 + 1 + \alpha'_1 \left(\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right) \alpha_1 + h_2 (h_2 + 2(\alpha_2' z_2 - 1))} \phi(z_1|z_2), \quad (3.7)$$

sendo $h_2 = \alpha'_1 \Sigma_{12} \Sigma_{22}^{-1} z_2$.

Prova. A prova de (3.7) é obtida simplesmente pela razão de $\frac{f(\mathbf{z})}{f(\mathbf{z}_2)}$, sendo $f(\mathbf{z})$ dada pela expressão (3.1) e $f(\mathbf{z}_2)$ análoga a expressão (3.6).

□

3.1.3 Modelo geral

Através do mesmo critério de extensão já utilizado anteriormente para o caso do modelo padrão univariado, por analogia com a função densidade de probabilidade da distribuição univariada alfa normal assimétrica não padrão, temos

$$\begin{aligned} f(x) &= \frac{(1 - \alpha \left(\frac{x - \mu}{\sigma} \right))^2 + 1}{2 + \alpha^2} \phi(x) \\ &= \left((1 - \alpha v^{-1} (x - \mu))^2 + 1 \right) K^{-1} \phi(x), \end{aligned}$$

com $\phi(x)$ a função densidade de probabilidade da distribuição normal com média μ e desvio padrão σ , $-\infty < x < \infty$ e $v = \sigma$. Podemos definir a função densidade de probabilidade da distribuição normal assimétrica multivariada não padrão como,

$$f(\mathbf{x}) = \left(\left(1 - \boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2 + 1 \right) K^{-1} \phi_p(\mathbf{x}), \quad (3.8)$$

sendo $-\infty < x_i < \infty$, $i = 1, \dots, p$, $\phi_p(\mathbf{x})$ a função densidade de probabilidade da distribuição normal p -variada $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$, com $\boldsymbol{\mu}$ o vetor de médias, $\boldsymbol{\Sigma}^*$ a matriz de variância e covariância e $\mathbf{V} = \text{diag}[(\sigma_1, \dots, \sigma_p)]$ a matriz diagonal dos desvios padrões, sendo que $\boldsymbol{\Sigma}^* = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}$. Deste modo, esta extensão é dada pela transformação linear multivariada $\mathbf{X} = \boldsymbol{\mu} + \mathbf{V}\mathbf{Z}$ sendo a $E(\mathbf{X}) = E(\boldsymbol{\mu} + \mathbf{V}\mathbf{Z}) = \boldsymbol{\mu}$ e $COV(\mathbf{X}) = COV(\boldsymbol{\mu} + \mathbf{V}\mathbf{Z}) = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V} = \boldsymbol{\Sigma}^*$.

Proposição 10. *Se $\mathbf{X}' = (X_1, X_2, \dots, X_p)$ é um vetor aleatório com função densidade de probabilidade expressa por 3.8, então $K = 2 + \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}$, com $\boldsymbol{\Sigma}$ a matriz de correlação de \mathbf{X} .*

Prova.

$$\begin{aligned} & \int_{\mathbf{x}} \left(\left(1 - \boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2 + 1 \right) \phi_p(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{x}} \left(1 - \boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2 \phi_p(\mathbf{x}) d\mathbf{x} + 1 \\ &= \int_{\mathbf{x}} \left(1^2 - 2\boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}^{-1} \boldsymbol{\alpha} \right) \phi_p(\mathbf{x}) d\mathbf{x} + 1 \\ &= -2 \int_{\mathbf{x}} \left(\boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \phi_p(\mathbf{x}) d\mathbf{x} + \int_{\mathbf{x}} \left(\boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}^{-1} \boldsymbol{\alpha} \right) \phi_p(\mathbf{x}) d\mathbf{x} + 2 \\ &= -2\boldsymbol{\alpha}' \mathbf{V}^{-1} \int_{\mathbf{x}} (\mathbf{x} - \boldsymbol{\mu}) \phi_p(\mathbf{x}) d\mathbf{x} + \boldsymbol{\alpha}' \mathbf{V}^{-1} \int_{\mathbf{x}} \left((\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})' \right) \phi_p(\mathbf{x}) d\mathbf{x} \mathbf{V}^{-1} \boldsymbol{\alpha} + 2 \\ &= 2 + \boldsymbol{\alpha}' \mathbf{V}^{-1} \boldsymbol{\Sigma}^* \mathbf{V}^{-1} \boldsymbol{\alpha} \\ &= 2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha}. \end{aligned}$$

Desta forma, $(2 + \boldsymbol{\alpha}' \boldsymbol{\Sigma} \boldsymbol{\alpha})^{-1} \int_{\mathbf{x}} \left(\left(1 - \boldsymbol{\alpha}' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2 + 1 \right) \phi_p(\mathbf{x}) d\mathbf{x} = 1$, sendo 3.8 uma função densidade de probabilidade.

□

As Proposições 11, 12 e 13 exibem, respectivamente, a função geradora de momentos, a esperança e a variância para o caso do modelo de probabilidade geral $\mathbf{X} \sim ANAM(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$. As provas destas Proposições são análogas ao caso padrão.

Proposição 11. *A função geradora de momentos de $\mathbf{X}' = (X_1, X_2, \dots, X_p)$ é representada por $M_{\mathbf{X}}(\mathbf{t}) = E(\exp\{\mathbf{X}'\mathbf{t}\})$ e dada por,*

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathbf{t}'\boldsymbol{\mu} + \frac{\mathbf{t}'\boldsymbol{\Sigma}^*\mathbf{t}}{2}\right\} [1 - K^{-1}\boldsymbol{\alpha}'\mathbf{V}^{-1}\boldsymbol{\Sigma}^*\mathbf{t}(2 - \mathbf{t}'\boldsymbol{\Sigma}^*\mathbf{V}^{-1}\boldsymbol{\alpha})]. \quad (3.9)$$

Proposição 12. *A esperança de $\mathbf{X}' = (X_1, X_2, \dots, X_p)$ com $\mathbf{X} \sim ANAM(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$ é dada por,*

$$E(\mathbf{X}) = \boldsymbol{\mu} - K^{-1}2\boldsymbol{\alpha}'\mathbf{V}^{-1}\boldsymbol{\Sigma}^*. \quad (3.10)$$

Proposição 13. *A matriz de variância e covariância de $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ com $\mathbf{X} \sim ANAM(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$ é dada por,*

$$COV(\mathbf{X}) = \boldsymbol{\Sigma}^* + K^{-2}\boldsymbol{\alpha}'\mathbf{V}^{-1}\boldsymbol{\Sigma}^*[(\mathbf{H} + \mathbf{I})K - 4]\boldsymbol{\Sigma}^*\mathbf{V}^{-1}\boldsymbol{\alpha}. \quad (3.11)$$

3.2 Inferência

Uma possível abordagem para realizar a inferência para amostras independentes e identicamente distribuídas com distribuição alfa normal assimétrica multivariada é baseada no método usual de estimação de máxima verossimilhança. Considere $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ uma amostra aleatória de tamanho n da distribuição $ANAM(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, sendo $\boldsymbol{\Sigma}$ a matriz de variância e covariância. Deste modo, o logaritmo natural da função de verossimilhança para o modelo geral é dado por,

$$\begin{aligned} \ell(\theta) = \sum_{i=1}^n \log \left[(1 - \boldsymbol{\alpha}'\mathbf{V}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}))^2 + 1 \right] - n \left[\log(2 + \boldsymbol{\alpha}'\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{V}^{-1}\boldsymbol{\alpha}) + \frac{\log|\boldsymbol{\Sigma}|}{2} \right] \\ - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})', \quad (3.12) \end{aligned}$$

em que $\theta = (\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ e \mathbf{V} a matriz diagonal composta pelo desvios padrões de cada variável. Os estimadores de máxima verossimilhança são obtidos diretamente através da maximização da função (3.12). Em outras palavras, $\hat{\theta} = \arg \max_{\theta} \ell(\theta)$.

A estimação intervalar é baseada na distribuição normal assintótica dos estimadores de máxima verossimilhança, bem como seus erros padrões obtidos da matriz de Informação de Fisher. A matriz de Informação de Fisher obtida por meio de (3.12), como para o caso bivariado, não possui expressão analítica simples e deve ser obtida através de métodos numéricos como, por exemplo, o pacote `maxLik`.

Em especial, não obtivemos problemas de estimação utilizando o método de decomposição de Cholesky (Stewart, 1998) para decompor $\boldsymbol{\Sigma}$.

3.3 Estudo de simulação

Este estudo de simulação é baseado em 1 000 conjuntos de dados gerados através da distribuição alfa normal assimétrica trivariada em duas diferentes configurações de parâmetros. A primeira considera a distribuição padrão e a segunda a versão não padrão desta distribuição. As replicações foram realizada em tamanhos amostrais iguais a 100, 250, 500 e 1 000. Para gerar estas amostras, consideramos o algoritmo MCMC Gibbs *sampling*, também utilizado por Kotecha & Djuric (1999) para realizar a geração de amostras para a distribuição normal truncada multivariada. Para a geração das condicionais no método de Gibbs, consideramos como o método da transformação inversa numericamente adaptado ao método de Brent (Brent, 1971; Toral & Chakrabarti, 1993).

Os resultados estão condensados nas Tabelas 3.1 e 3.2, as quais exibem a média das estimativas das pelo estimador de máxima verossimilhança (EMV), a média dos erros padrões (EP) e as probabilidades de cobertura (PC) para intervalos assintóticos de 95% de confiança, respectivamente.

De acordo com o aumento do tamanho amostral, observamos que as estimativas estão próximas do verdadeiro valor do parâmetro utilizado na geração, ou seja, os estimadores são assintoticamente não viciados para os parâmetros. Além disso, existe o decaimento do erro padrão para amostras grandes, fato este que exhibe a consistência do estimador. As probabilidades de cobertura se aproximam da confiança nominal conforme o aumento

do tamanho amostral.

Tabela 3.1: Distribuição padrão: média das 1 000 EMVs para quatro diferentes tamanhos amostrais, erros padrões e probabilidade de cobertura para intervalos de 95% de confiança.

| Tamanho Amostrai | Parâmetro | μ_1 | σ_1 | α_1 | μ_2 | σ_2 | α_2 | μ_3 | σ_3 | α_3 | ρ_{12} | ρ_{13} | ρ_{23} |
|------------------|-------------|---------|------------|------------|---------|------------|------------|---------|------------|------------|-------------|-------------|-------------|
| | Varlor real | 0 | 1 | 0.5 | 0 | 1 | -1 | 0 | 1 | 2 | 0.7 | -0.5 | 0 |
| n=100 | EMV | -0.05 | 1.07 | 0.25 | -0.02 | 1.06 | -0.87 | -0.04 | 1.04 | 1.92 | 0.72 | -0.50 | -0.04 |
| | EP | 0.13 | 0.14 | 0.55 | 0.13 | 0.15 | 0.48 | 0.12 | 0.13 | 0.48 | 0.05 | 0.07 | 0.09 |
| | PC | 0.93 | 0.90 | 0.77 | 0.73 | 0.92 | 0.80 | 0.83 | 0.90 | 0.83 | 0.87 | 0.80 | 0.80 |
| n=250 | EMV | -0.06 | 1.04 | 0.43 | -0.04 | 1.04 | -1.08 | -0.01 | 1.00 | 1.95 | 0.71 | -0.50 | -0.01 |
| | EP | 0.08 | 0.09 | 0.33 | 0.08 | 0.09 | 0.29 | 0.07 | 0.08 | 0.28 | 0.03 | 0.04 | 0.06 |
| | PC | 0.93 | 0.90 | 0.87 | 0.80 | 0.94 | 0.93 | 0.90 | 0.93 | 0.90 | 0.87 | 0.90 | 0.90 |
| n=500 | EMV | -0.04 | 1.04 | 0.45 | -0.04 | 1.04 | -1.01 | -0.01 | 0.99 | 1.96 | 0.71 | -0.50 | -0.02 |
| | EP | 0.05 | 0.06 | 0.22 | 0.06 | 0.06 | 0.20 | 0.05 | 0.05 | 0.19 | 0.02 | 0.03 | 0.04 |
| | PC | 0.93 | 0.93 | 0.97 | 0.90 | 0.96 | 0.97 | 0.92 | 0.97 | 0.93 | 0.93 | 0.90 | 0.93 |
| n=1000 | EMV | -0.03 | 1.03 | 0.50 | -0.02 | 1.03 | -0.99 | -0.01 | 1.00 | 2.00 | 0.71 | -0.50 | -0.01 |
| | EP | 0.04 | 0.04 | 0.15 | 0.04 | 0.05 | 0.14 | 0.04 | 0.04 | 0.13 | 0.02 | 0.02 | 0.03 |
| | PC | 0.97 | 0.95 | 0.97 | 0.93 | 0.95 | 0.95 | 0.96 | 0.95 | 0.96 | 0.97 | 0.93 | 0.95 |

3.4 Uma aplicação real ilustrativa

Neste seção apresentamos a aplicação do modelo alfa normal assimétrica multivariada, no caso trivariado, em um conjunto de dados reais extraído de Mayor & Frei (2003). Este conjunto refere-se a algumas propriedades dos Exoplanetas - planetas fora do sistema solar - descobertos até Outubro de 2002 e foi também utilizado por Hothorn & Everitt (2014) como um exemplo didático para análise de agrupamentos e estão disponíveis no pacote 'HSAUR3' do Software R (R Core Team, 2013; Everitt *et al.*, 2014). A amostra corresponde a 101 observações de três variáveis as quais correspondem a massa, período e excentricidade dos planetas. Seja $X1$ = a transformação logarítmica da massa (Jupiter), $X2$ = a transformação logarítmica do período em anos terrestres, $X3$ = a raiz quadrada da excentricidade radial de cada planeta, a Tabela 3.3 exhibe as estatísticas básicas do conjunto de dados.

As estimativas estão apresentadas na Tabela 3.4. Os erros padrões foram obtidos numericamente através da matriz de informação de fisher observada. A distribuição proposta é comparada com a distribuição normal multivariada e com a distribuição normal assimétrica multivariada (Azzalini & Capitanio, 1999). No intuito de realizar uma comparação

Tabela 3.2: Distribuição não padrão: média das 1 000 EMVs para quatro diferentes tamanhos amostrais, erros padrões e probabilidade de cobertura para intervalos de 95% de confiança.

| Tamanho Amostrai | Parâmetro | μ_1 | σ_1 | α_1 | μ_2 | σ_2 | α_2 | μ_3 | σ_3 | α_3 | ρ_{12} | ρ_{13} | ρ_{23} |
|------------------|------------|---------|------------|------------|---------|------------|------------|---------|------------|------------|-------------|-------------|-------------|
| | Valor real | 1 | 4 | -1 | -4 | 9 | 1 | 2 | 1 | 3 | -0.2 | 0.4 | 0.1 |
| n=100 | EMV | 1.04 | 3.77 | -0.66 | -4.04 | 9.35 | 0.68 | 1.96 | 1.37 | 2.12 | -0.17 | 0.36 | 0.22 |
| | EP | 0.27 | 0.54 | 0.37 | 0.42 | 1.32 | 0.35 | 0.16 | 0.18 | 0.59 | 0.09 | 0.08 | 0.09 |
| | PC | 0.88 | 0.74 | 0.71 | 0.80 | 0.91 | 0.80 | 0.91 | 0.63 | 0.62 | 0.91 | 0.92 | 0.70 |
| n=250 | EMV | 1.02 | 3.78 | -0.81 | -3.97 | 9.13 | 0.84 | 1.95 | 1.11 | 2.61 | -0.18 | 0.37 | 0.16 |
| | EP | 0.15 | 0.34 | 0.24 | 0.23 | 0.80 | 0.22 | 0.08 | 0.08 | 0.41 | 0.06 | 0.05 | 0.05 |
| | PC | 0.88 | 0.82 | 0.84 | 0.90 | 0.90 | 0.88 | 0.90 | 0.83 | 0.82 | 0.89 | 0.86 | 0.82 |
| n=500 | EMV | 1.02 | 3.79 | -0.94 | -3.96 | 9.11 | 0.94 | 2.00 | 1.02 | 2.95 | -0.19 | 0.39 | 0.13 |
| | EP | 0.10 | 0.24 | 0.17 | 0.15 | 0.55 | 0.15 | 0.05 | 0.05 | 0.31 | 0.04 | 0.03 | 0.04 |
| | PC | 0.96 | 0.90 | 0.93 | 0.94 | 0.92 | 0.91 | 0.95 | 0.92 | 0.95 | 0.90 | 0.92 | 0.93 |
| n=1000 | EMV | 1.01 | 3.81 | -1.00 | -4.01 | 9.02 | 0.96 | 2.00 | 0.99 | 3.01 | -0.18 | 0.39 | 0.11 |
| | EP | 0.07 | 0.17 | 0.12 | 0.10 | 0.38 | 0.11 | 0.03 | 0.03 | 0.22 | 0.03 | 0.02 | 0.03 |
| | PC | 0.95 | 0.94 | 0.96 | 0.97 | 0.97 | 0.94 | 0.96 | 0.95 | 0.97 | 0.96 | 0.96 | 0.94 |

Tabela 3.3: Estatísticas básicas do logaritmo da massa, logaritmo do período e raiz quadrada da excentricidade dos 101 Exoplanetas descobertos até Outubro de 2002.

| Variável | Mínimo | Mediana | Média | Máximo | Desvio Padrão | Assimetria | Curtose |
|----------|--------|---------|-------|--------|---------------|------------|---------|
| X1 | -2.996 | 0.565 | 0.621 | 2.862 | 1.162 | -0.318 | 2.989 |
| X2 | 1.094 | 5.820 | 5.186 | 8.587 | 2.137 | -0.630 | 2.154 |
| X3 | 0.000 | 0.520 | 0.475 | 0.963 | 0.237 | -0.415 | 2.353 |

ainda mais formal entre as distribuições ajustadas aos dados reais, aplicamos um teste de aderência empírico para distribuições multivariadas proposto por McAssey (2013). De uma forma geral, este teste está baseado na comparação entre a distribuição da distância de Mahalanobis dos dados observados, sob a hipótese nula, com a distribuição da distância de Mahalanobis da distribuição em teste. O teste foi implementado com $N = 10,000$, $T = 25$, $R = 100$, e $B = 100$, sendo N o número de amostras aleatoriamente geradas da distribuição de teste, T o número de quantis avaliados na distribuição das distâncias de Mahalanobis, R o número de repetições realizadas para se obter valores consistentes dos quantis avaliados e B o número de repetições realizadas para se obter a estatística crítica do teste. A estatística observada A_T , o valor crítico, correspondente a 5% de significância, e o p-valor de cada teste são exibidos na Tabela 3.4, bem como o valor máximo do logaritmo da função de verossimilhança $l(\hat{\theta})$ e o critério de informação de Akaike (AIC). Neste caso, considerando um nível de significância de 5%, não rejeitamos a hipótese de

que os dados provêm de uma distribuição alfa normal assimétrica trivariada e rejeitamos a hipótese de que os dados provêm das outras distribuições em teste.

Tabela 3.4: Estimativas dos parâmetros, valor máximo do logaritmo da função de verossimilhança e teste de aderência para a distribuição normal trivariada (*NT*), normal assimétrica trivariada (*NAT*) e alfa normal assimétrica trivariada (*ANAT*) para o conjunto de dados dos Exoplanetas.

| Parâmetros | <i>NT</i> | <i>NAT</i> | <i>ANAT</i> |
|-------------------|-------------|--------------|--------------|
| α_1 | - | -0.08 (0.06) | 0.48 (0.31) |
| α_2 | - | -0.89 (0.08) | -1.98 (0.37) |
| α_3 | - | -0.28 (0.11) | -1.13 (0.32) |
| μ_1 | 0.62 (0.12) | 0.61 (0.11) | 0.42 (0.13) |
| μ_2 | 5.19 (0.21) | 5.13 (0.20) | 4.31 (0.17) |
| μ_3 | 0.48 (0.02) | 0.47 (0.02) | 0.40 (0.02) |
| σ_1 | 1.16 (0.19) | 1.15 (0.18) | 1.11 (0.17) |
| σ_2 | 2.13 (0.63) | 2.06 (0.26) | 1.57 (0.28) |
| σ_3 | 0.24 (0.01) | 0.23 (0.03) | 0.20 (0.01) |
| ρ_{12} | 0.46 (0.27) | 0.45 (0.11) | 0.39 (0.08) |
| ρ_{13} | 0.42 (0.03) | 0.41 (0.06) | 0.33 (0.08) |
| ρ_{23} | 0.52 (0.06) | 0.50 (0.01) | 0.24 (0.08) |
| $l(\hat{\theta})$ | -65.63 | -328.44 | -45.03 |
| AIC | 155.26 | 680.89 | 114.06 |
| A_T | 9.94 | 9.28 | 8.91 |
| $A_{T,b}$ | 9.25 | 9.23 | 9.29 |
| p-valor | 0.00 | 0.01 | 0.71 |

As Figuras 3.5-3.7 exibem a distribuição trivariada ajustada e suas marginais para os modelos normal multivariado, normal assimétrico multivariado e alfa normal assimétrico multivariado, respectivamente. Todos estes resultados indicam que a distribuição alfa normal assimétrica multivariada possui melhor ajuste aos dados observados, em comparação com a normal multivariada e a normal assimétrica multivariada.

A critério de análise, vale ressaltar que X_1 possui sua média amostral maior que sua mediana amostral, o que indica uma assimetria à direita para esta variável. Paralelamente, X_2 e X_3 possuem suas médias amostrais menores que as medianas amostrais, o que indica uma assimetria à esquerda para estas variáveis. Tais fatos são corroborados pelo modelo alfa normal assimétrico trivariado nas estimativas $\hat{\alpha}_1 > 0$, $\hat{\alpha}_2 < 0$ e $\hat{\alpha}_3 < 0$. Fato este que não ocorre para o modelo normal assimétrico trivariado.

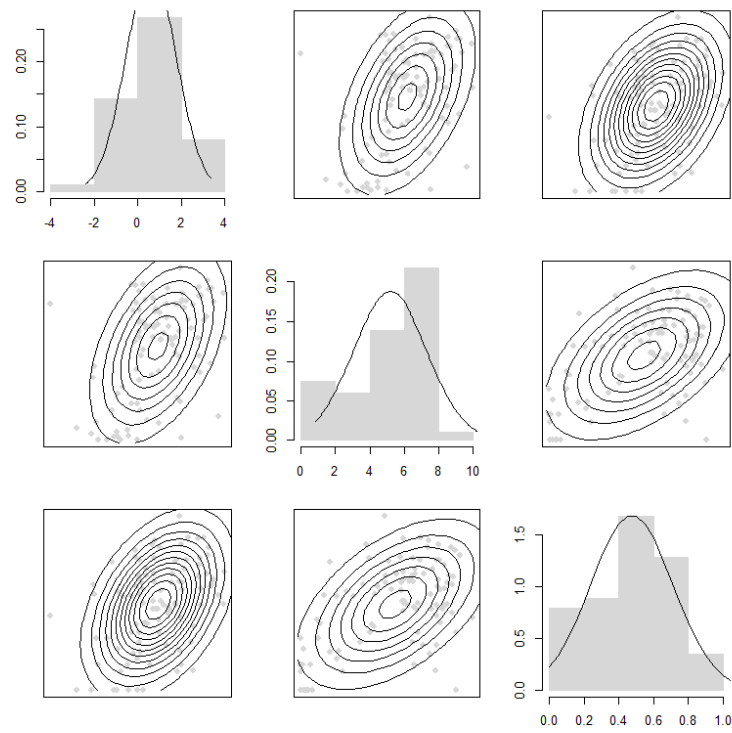


Figura 3.5: Gráfico de dispersão e distribuição normal multivariada ajustada.

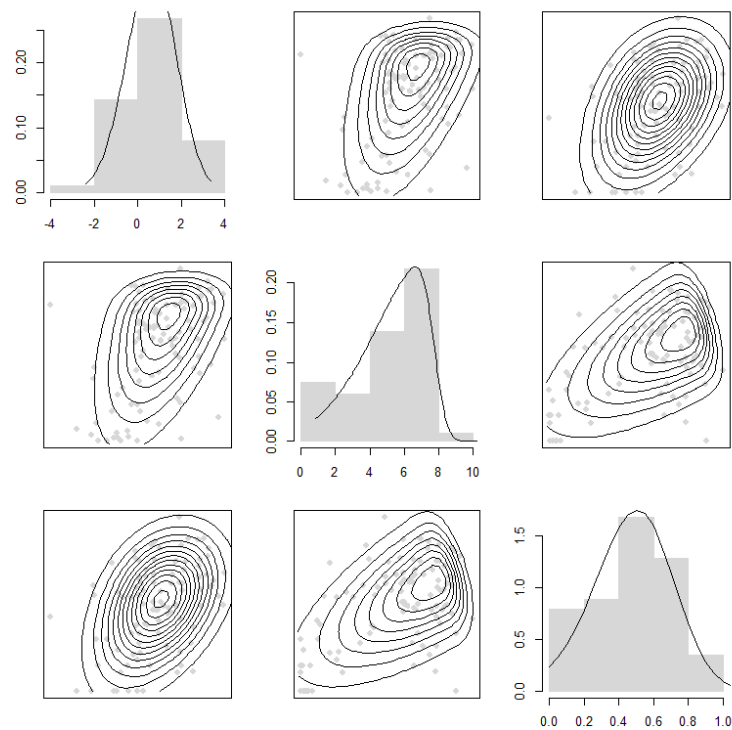


Figura 3.6: Gráfico de dispersão e distribuição normal assimétrica multivariada ajustada.

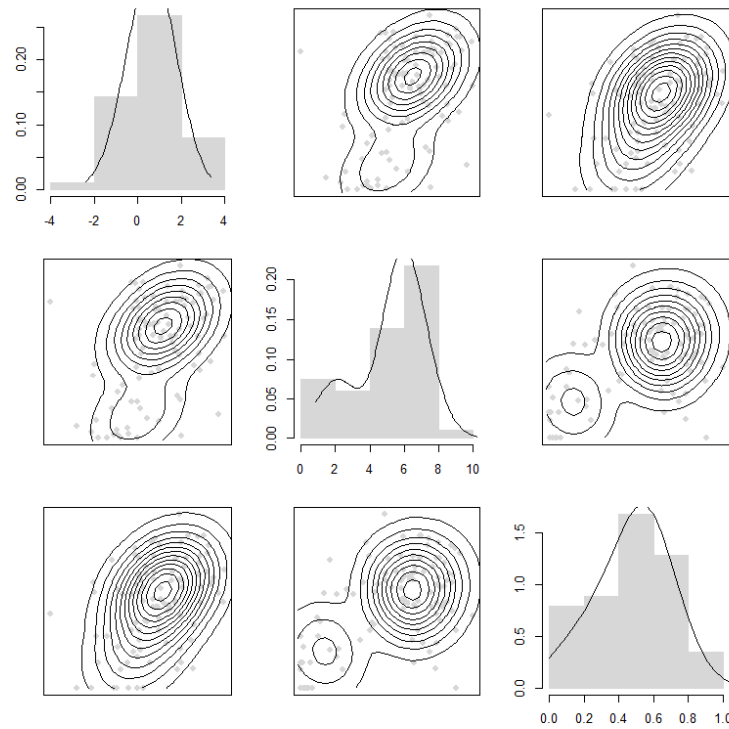


Figura 3.7: Gráfico de dispersão e distribuição alfa normal assimétrica multivariada ajustada.

3.5 Comentários Finais

Neste capítulo apresentamos a distribuição alfa normal assimétrica multivariada, sua função densidade de probabilidade, suas propriedades, comportamento gráfico, função geradora de momentos, os momentos principais - esperança e matriz de variância e covariância, seu modelo geral, as funções densidade de probabilidade marginal e condicional, bem como sua respectiva função de verossimilhança. Além disso, apresentamos sua aplicação em dados artificiais e reais, esta ultima envolvendo uma comparação com a distribuição normal clássica e com a distribuição normal assimétrica. Os procedimentos aqui apresentados apresentam-se como necessários para a modelagem de dados multivariados na presença de assimetria e/ou bimodalidade, bem como se sobrepõem a distribuição normal clássica multivariada e a distribuição normal assimétrica multivariada para o caso analisado. Vale salientar que com o crescimento do número p de variáveis, existe o rápido aumento do esforço computacional, uma vez que $p(3 + \frac{p-1}{2})$ deverão ser estimados. Por exemplo, se $p = 10$ iremos obter 75 parâmetros a serem estimados.

Capítulo 4

Redes de Classificação de K-dependência

No contexto de classificação, as redes de classificação podem ser vistas como estruturas particulares e também são conhecidas como classificadores bayesianos.

Uma rede de classificação é um grafo acíclico direcionado, em que os nós representam as variáveis aleatórias e os arcos as respectivas dependências condicionais entre as variáveis.

A Figura 4.1 exibe dois exemplos de redes probabilísticas.

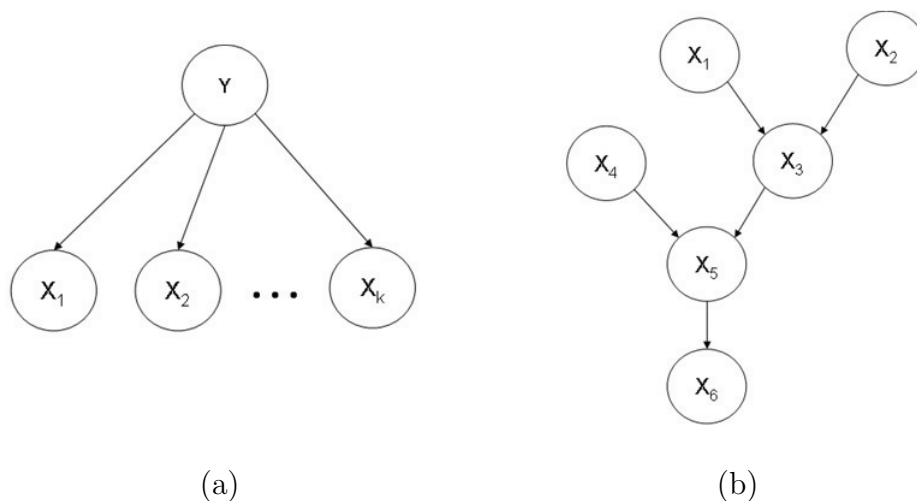


Figura 4.1: Em (a) apresentamos a estrutura tradicional da rede *naïve* Bayes e em (b) apresentamos uma rede probabilística com seis variáveis aleatórias.

A construção de uma rede de classificação está baseada no cálculo da distribuição de probabilidade *a posteriori* $P(Y|X)$, em que Y é a variável aleatória a ser classificada apresentando L possíveis valores (categorias) y_1, y_2, \dots, y_L e $X = (X_1, X_2, \dots, X_p)$ é um conjunto de p variáveis explicativas.

O método baseia-se em calcular a probabilidade de uma respectiva observação pertencer a cada uma das categorias, assim classifica a observação na categoria mais plausível. Se a classificação em foco for binária, podemos utilizar a curva ROC para inferir sobre a classificação.

Nesta Tese, consideramos a estrutura de redes de classificação de k -dependência, proposta por Sahami (1996). As redes de k -dependência consideram possíveis relações de dependência entre as variáveis explicativas. Desta forma, uma rede com k -dependência é uma rede de classificação que permite em sua estrutura que cada variável explicativa X_i possua no máximo k pais, isto é, máximo K variáveis explicativas. Em outras palavras, para cada variável explicativa X_i , $país_i$ é um conjunto com no máximo K outras variáveis explicativas para todo $i = 1, \dots, p$, sendo que uma estrutura que generaliza o popular classificador de *naïve* Bayes quando $k = 0$. A rede com 1-dependência (kDB1) possui a mesma estrutura que uma rede probabilística para classificação e bastante difundida na literatura, conhecida como *Tree Augmented Network* (TAN) (Friedman *et al.*, 1997). Além disso, k pode variar de 0 a $p - 1$, em que p é o número de variáveis explicativas consideradas.

Assumimos, nesta Tese, que o conjunto de variáveis explicativas são puramente contínuas e tradicionalmente modeladas via distribuição normal clássica, como evidenciado por Perez *et al.* (2006). Assim, para redes de classificação de k dependência (kDB), calculamos as probabilidades *a posteriori* dada por

$$P(Y = y_l | x_1, x_2, \dots, x_p) = \frac{P(Y = y_l) \prod_{i=1}^p f(x_i | país_i, y_l)}{\sum_j P(Y = y_j) \prod_{i=1}^p f(x_i | país_i, y_j)} \quad (4.1)$$

em que

$$f(x_i | país_i, y_k) \sim N(\mu_{i|país_i, y_k}, \sigma_{i|país_i, y_k}^2),$$

sendo $i = \{1, 2, \dots, p\}$ e $l = \{1, 2, \dots, L\}$, com $\mu_{i|país_i, y_l}$ e $\sigma_{i|país_i, y_l}^2$ a média e a variância da variável X_i condicionada à categoria y_l e dadas por (4.2) e (4.3), as quais são dadas por

$$\mu_{i|país_i, y_l} = \mu_{i|y_l} + \sum_{j=1}^{K_i} \frac{\sigma_{ij|y_l}}{\sigma_{j|y_l}^2} (x_j - \mu_{j|y_l}) \quad (4.2)$$

e

$$\sigma_{i|pais_i,y_l}^2 = \frac{|\sum_{X_i,pais_i|y_l}|}{|\sum_{pais_i|y_l}|}, \quad (4.3)$$

em que $\sum_{X_i,pais_i|y_l}$ é a matriz de variância e covariância entre a variável X_i e o conjunto de variáveis em $pais(X_i)$, ambos condicionados à categoria y_l . $\sum_{pais_i|y_l}$ é a matriz de variância do conjunto de variáveis em $pais(X_i)$ condicionado à categoria respectiva.

Considerando um conjunto de dados com uma variável de interesse e nove variáveis explicativas, exibimos as redes probabilísticas de k-dependência 0, 1, 2 na Figura 4.2, respectivamente.

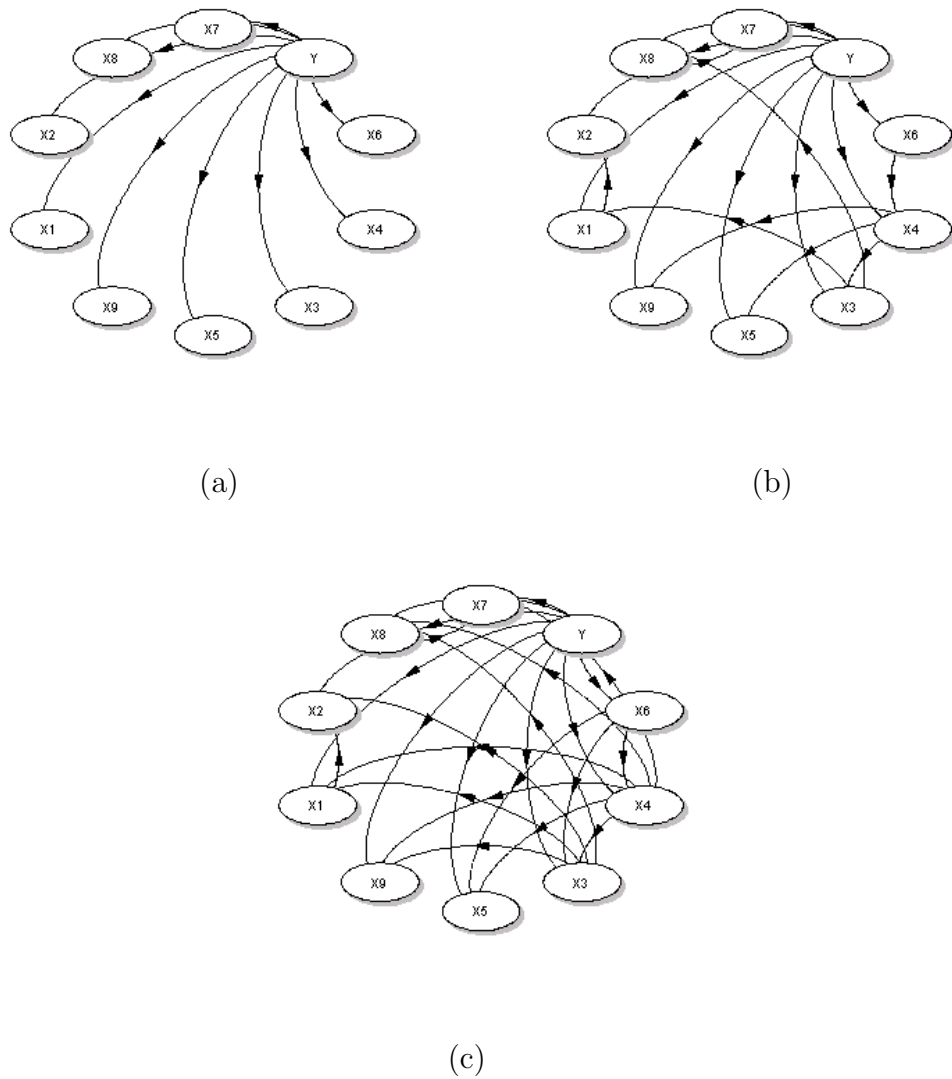


Figura 4.2: Exemplificação da estrutura das redes probabilísticas com: (a) 0-dependência, (b) 1-dependência e (c) 2-dependência.

A construção da estrutura de uma rede de k -dependência através de um conjunto de dados será apresentada na próxima seção.

4.1 Entropia e Informações Mútuas

Seja X_i uma variável explicativa e aleatória contínua assumindo valores x_i no espaço χ_i e com função densidade de probabilidade $f_i(x_i)$, com $i = 1, \dots, p$. Da mesma forma, seja Y uma variável aleatória discreta a ser classificada assumindo L possíveis valores em y_1, y_2, \dots, y_L . A informação mútua entre X_i e Y é dada por

$$\begin{aligned}
 I(X_i, Y) &= \sum_{k=1}^L \int_{\chi_i} p(x_i, y_k) \log \left(\frac{p(x_i, y_k)}{p(x_i) p(y_k)} \right) dx_i \\
 &= \sum_{k=1}^L p(y_k) \int_{\chi_i} p(x_i|y_k) \log \left(\frac{p(x_i|y_k)}{p(x_i)} \right) dx_i \\
 &= \sum_{k=1}^L p(y_k) \int_{\chi_i} p(x_i|y_k) \log(p(x_i|y_k)) dx_i - \sum_{k=1}^L \int_{\chi_i} p(y_k) p(x_i|y_k) \log(p(x_i)) dx_i \\
 &= \sum_{k=1}^L p(y_k) \int_{\chi_i} p(x_i|y_k) \log(p(x_i|y_k)) dx_i - \int_{\chi_i} p(x_i) \log(p(x_i)) dx_i \\
 &= H(X_i) - \sum_{k=1}^L p(y_k) H(X_i|y_k). \tag{4.4}
 \end{aligned}$$

A informação mútua entre X_i e X_j , com $i \neq j$, condicionalmente a Y é dada por

$$\begin{aligned}
 I(X_i, X_j|Y) &= \sum_{k=1}^L \int_{\chi_i} \int_{\chi_j} p(x_i, x_j, y_k) \log \left(\frac{p(x_i, x_j, y_k)}{p(x_i|y_k) p(x_j|y_k) p(y_k)} \right) dx_j dx_i \\
 &= \sum_{k=1}^L \int_{\chi_i} \int_{\chi_j} p(x_i, x_j|y_k) p(y_k) \log \left(\frac{p(x_i, x_j|y_k)}{p(x_i|y_k) p(x_j|y_k)} \right) dx_j dx_i \\
 &= \sum_{k=1}^L p(y_k) E \left[\log \left(\frac{p(x_i, x_j|y_k)}{p(x_i|y_k) p(x_j|y_k)} \right) | Y = y_k \right]. \tag{4.5}
 \end{aligned}$$

Considerando que as variáveis explicativas seguem uma distribuição normal bivariada, a informação mútua condicional é dada por (Perez *et al.*, 2006),

$$I(X_i, X_j|Y) = -\frac{1}{2} \sum_{k=1}^L p(y_k) \log(1 - \rho_k^2(X_i, X_j)). \quad (4.6)$$

Para a construção da estrutura de uma rede de k -dependência através de um conjunto de dados, Sahami (1996) propõe um Algoritmo, o qual utiliza-se das medidas de entropia e informação mútua apresentadas acima e exibido em 4.1.

Algoritmo 4.1 Algoritmo para construção de uma rede de k -dependência.

1. Para cada variável X_i , calcule a medida de informação mútua, denotada por $\hat{I}(X_i, Y)$;
 2. Para cada par de variáveis explicativas, calcule a medida de informação mútua condicional, denotada por $\hat{I}(X_i, X_j|Y)$;
 3. Defina S como a lista de variáveis explicativas utilizadas, inicialmente considere S como vazio;
 4. Inicie a rede com a variável de classificação Y ;
 5. Repita até a lista S conter todas as variáveis explicativas:
 - (a) Selecione a variável explicativa X_{max} que ainda não está contida em S e que possua a maior medida $\hat{I}(X_{max}, Y)$;
 - (b) Adicione à rede a variável X_{max} ;
 - (c) Adicione um arco de Y para X_{max} ;
 - (d) Adicione $m = \min(|S|, K)$ arcos partindo das m variáveis explicativas X_j com o maior valor $\hat{I}(X_{max}, X_j|Y)$;
 - (e) Adicione X_{max} à lista S .
-

4.2 Rede KDB e Alfa Normal Assimétrica

Para considerar a distribuição alfa normal assimétrica para o algoritmo de construção da rede de k -dependência, necessitamos da entropia, informação mútua e informação mútua condicional para esta distribuição.

Proposição 14. *A entropia de Shannon para o modelo alfa normal assimétrico não pode ser expressa analiticamente.*

Prova. Considere $X \sim ANA(\alpha)$, por definição $H(X)$ é dada por,

$$\begin{aligned}
H(X) &= - \int_x f(x) \log(f(x)) dx \\
&= E \left[- \log \left(\frac{1 + (1 - \alpha x)^2}{2 + \alpha^2} \times \phi(x) \right) \right] \\
&= E \left[- \log(1 + (1 - \alpha x)^2) + \log(2 + \alpha^2) - \log(\phi(x)) \right] \\
&= H(x^*) + \log(2 + \alpha^2) - E \left[\log(1 + (1 - \alpha x)^2) \right],
\end{aligned}$$

em que $H(x^*)$ é a entropia de Shannon para uma variável $X^* \sim N(0,1)$ e o termo $E \left[\log(1 + (1 - \alpha x)^2) \right]$ não possui resolução analítica. \square

Este fenômeno se repete para o cálculo da informação mútua e informação mútua condicional, uma vez que estes podem ser expressos em função da entropia.

Para contornar este problema, utilizamos o método de Monte Carlo (Metropolis & Ulam, 1949; Rubinstein & Kroese, 2011), por meio da aproximação das expressões (4.4) e (4.5), para estimar $\hat{I}(X_i, Y)$ e $\hat{I}(X_i, X_j|Y)$ indicadas no Algoritmo 4.1.

Desta forma, a estrutura da rede de k -dependência pode ser estimada para o modelo alfa normal assimétrico, considerando-se o modelo multivariado proposto. Temos uma máxima k -dependência de ordem $p - 1$, considerando p como o número de variáveis, sendo que a distribuição alfa normal assimétrica multivariada é utilizada para obter a distribuição condicional de cada variável explicativa dados seus pais. Assim, a expressão (4.7) exibe o procedimento de cálculo da probabilidade *a posteriori* para o caso de p variáveis explicativas.

$$P(Y = y_l | x_1, \dots, x_p) = \frac{P(Y = y_l) \prod_{i=1}^p f(x_i | pais_i, y_l)}{\sum_j P(Y = y_j) \prod_{i=1}^p f(x_i | pais_i, y_j)}, \quad (4.7)$$

sendo

$$f(x_i | pais_i, y_k) \sim ANA(\alpha_{i y_k}, \mu_{i | y_k}, \sigma_{i | y_k}) \text{ se } pais_i = \emptyset,$$

ou, considerando $z_i = \frac{x_i - \mu_i}{\sigma_i}$

$$f(z_i | z_{pais_i}, \mathbf{y}_k) = \frac{(1 - \alpha'_i z_i)^2 + 2\alpha'_i z_i z'_{pais_i} \alpha_{pais_i} + (1 - \alpha'_{pais_i} z_{pais_i})^2}{K2} \phi(z_i | z_{pais_i}, \mathbf{y}_k), \quad (4.8)$$

se $país_i = \{x_j\}$, com $j \neq i$.

sendo,

$$K2 = (1 - \alpha'_i z_{país})^2 + 1 + \alpha'_i (\Sigma_{ii} - \Sigma_{i \times país} \Sigma_{país \times país}^{-1} \Sigma_{país \times i}) \alpha_i \\ + h_{país} (h_{país} + 2 (\alpha_{país}' z_{país} - 1)),$$

e

$$h_{país} = \alpha'_i \Sigma_{i \times país} \Sigma_{país \times país}^{-1} z_{país}.$$

Denominamos este classificador, o qual utiliza a expressão (4.8), de alfa normal assimétrico de k dependência (ANAKDB). Quando $k = 0$ este se reduz a versão alfa normal assimétrico *naïve* Bayes (ANANB).

4.3 Comentários Finais

Neste capítulo apresentamos as redes de classificação e sua construção para o caso da distribuição alfa normal assimétrica. Em suma, apresentamos os métodos alfa normal assimétrico *naïve* Bayes (ANANB) e alfa normal assimétrico de k -dependência (ANAKDB). Além disso, exibimos que a entropia de Shannon para a distribuição alfa normal assimétrica não pode ser escrita analiticamente. Maiores detalhes sobre redes probabilísticas podem ser encontrados em Jensen (1996) e Bishop (2006). Vale ressaltar que o tempo de processamento dos métodos aqui propostos irão depender do número de covariáveis em escopo, bem como do grau de k -dependência utilizada na rede. Durante o desenvolvimento destes métodos, notamos que a partir de 4 covariáveis, sendo $k = 3$, o processamento computacional demandará muito tempo.

Capítulo 5

Validação do modelo proposto para classificação

Neste capítulo realizamos uma avaliação comparativa entre as redes de k -dependência para o modelo proposto com relação aos métodos usuais de classificação.

Para verificar o desempenho de ambos os modelos, considere N classificações realizadas por um modelo qualquer dadas por $M = m_1, \dots, m_N$ e $D = d_1, \dots, d_N$ o valor real de cada observação. Assim, temos o objetivo de comparar M com D , isto é, comparar os valores preditos do modelo com os valores reais utilizados na predição. Assim, para o caso de classificação binária, a Tabela 5.1 apresenta a tabela de contingência 2×2 entre M e D , também conhecida com matriz de confusão.

Tabela 5.1: Matriz de confusão.

| | | M | |
|-----|---------|---------|---------|
| | | $\{1\}$ | $\{0\}$ |
| D | $\{1\}$ | VP | FN |
| | $\{0\}$ | FP | VN |

onde VP é o número de verdadeiros positivos, FP o número falsos positivos, FN é o número de falsos negativos e VN é o número de verdadeiros negativos. Naturalmente, temos que $VP + FP + FN + VN = N$.

Desta forma, utilizamos as seguintes métricas para avaliar a performance de M .

Nesta Tese, utilizamos as métricas abaixo para quantificar o desempenho do modelo M com base em amostras de teste D .

- *Accuracy (ACC)*: métrica comumente utilizada para interpretar a classificação geral. Baseia-se no cálculo da fração geral de acertos de um modelo, tanto para as classificações de indivíduos para a classe 1 quanto para as classificações de indivíduos para a classe 0. É definida como,

$$ACC = \frac{VP + VN}{VP + FP + VN + FN}.$$

- *Matthew's correlation coefficient (MCC)*: Baldi *et al.* (2000) métrica também utilizada para interpretar a classificação geral. Sua interpretação é semelhante ao coeficiente de correlação linear de Pearson: quanto mais próximo a 1, a classificação estará mais próxima de uma classificação perfeita; quanto próximo a 0, a classificação será completamente aleatória; quanto mais próximo a -1 , a classificação estará mais próxima de uma classificação totalmente inversa. A métrica *MCC* é definida por,

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}.$$

- *Precision (P)*: proporção de positivos corretos dentre os classificados como positivo, métrica também conhecida como valor preditivo positivo. A métrica *P* é definida por,

$$P = \frac{VP}{VP + FP}.$$

- *Recall (R)*: a fração real de positivos dentre os que são previstos como positivos, também conhecida como a sensibilidade. A métrica *R* é definida como,

$$R = \frac{VP}{VP + FN}.$$

- *F1-Score (F1)*: Esta métrica considera uma combinação entre *P* e *R* e pode ser utilizada para mensurar o desempenho geral do método de classificação (Powers, 2011). A métrica *F1* é definida como,

$$F1 = \frac{2 \times P \times R}{P + R}.$$

Quanto maior as métricas de *ACC*, *MCC*, e *F1-Score*, mais efetivo será o método como um classificador geral.

A seguir são exibidos estes resultados para o conjunto de dados artificiais e reais.

5.1 Dados artificiais

A base de dados artificiais foi gerada por meio de uma modificação do procedimento Breiman (1998), o qual se baseou na normalidade das variáveis explicativas. Para este estudo, uma base total de 1 000 observações foi gerada a partir de

$$X|Y = 1 \sim ANAM \left(\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} \frac{a}{2} \\ a \\ \frac{a}{2} \end{bmatrix} \right)$$

e

$$X|Y = 0 \sim ANAM \left(\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 1 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} a \\ 2a \\ a \end{bmatrix} \right).$$

Para esta comparação, consideramos $a = -1, 0$ e 1 , bem como o procedimento de validação cruzada *5-fold*, o qual é uma generalização de outros métodos de validação, sendo baseado na divisão aleatória do conjunto de dados em 5 subconjuntos (20% da base original). Cada subconjunto é usado como um conjunto de teste para o modelo de ajuste considerando os outros 4 subconjuntos (80% da base original) como conjunto de treinamento (Mitchell, 1997).

Os resultados dos métodos propostos foram também comparados com alguns métodos comuns na área de classificação, sendo estes: GNB = *naïve Bayes* utilizando a distribuição normal clássica; SVM = *support vector machine* com *kernel radial basis function*; LR = regressão logística e LDA = análise discriminante linear. Maiores detalhes sobre estes métodos e suas aplicações para o contexto de classificação podem ser encontrados em Louzada *et al.* (2016).

Todo este processo foi repetido 30 vezes, a Tabela 5.1 apresenta as médias e os desvios padrões das métricas ACC , MCC e $F1$. As Figuras 5.1-5.5 e 5.6 exibem os gráficos *boxplots* para cada medida e para cada técnica.

Tabela 5.2: Médias e desvios padrões das medidas ACC , MCC , P , R e $F1$ para as diferentes técnicas de classificação

| a | ANAKDB | ANANB | GNB | SVM | LR | LDA |
|------------|---------------|---------------|---------------|---------------|---------------|---------------|
| <i>ACC</i> | | | | | | |
| 0 | 0.751 (0.014) | 0.749 (0.017) | 0.748 (0.018) | 0.747 (0.015) | 0.749 (0.017) | 0.749 (0.018) |
| -1 | 0.752 (0.013) | 0.736 (0.024) | 0.689 (0.019) | 0.731 (0.010) | 0.713 (0.023) | 0.704 (0.025) |
| 1 | 0.753 (0.013) | 0.736 (0.017) | 0.702 (0.025) | 0.735 (0.013) | 0.682 (0.022) | 0.677 (0.023) |
| <i>MCC</i> | | | | | | |
| 0 | 0.499 (0.038) | 0.499 (0.045) | 0.497 (0.047) | 0.496 (0.040) | 0.501 (0.045) | 0.502 (0.048) |
| -1 | 0.501 (0.036) | 0.468 (0.067) | 0.378 (0.039) | 0.464 (0.032) | 0.417 (0.046) | 0.401 (0.052) |
| 1 | 0.502 (0.035) | 0.476 (0.052) | 0.396 (0.051) | 0.473 (0.034) | 0.357 (0.042) | 0.348 (0.044) |
| <i>P</i> | | | | | | |
| 0 | 0.755 (0.011) | 0.748 (0.016) | 0.749 (0.017) | 0.748 (0.015) | 0.746 (0.020) | 0.749 (0.020) |
| -1 | 0.774 (0.051) | 0.731 (0.029) | 0.654 (0.023) | 0.741 (0.026) | 0.691 (0.030) | 0.679 (0.033) |
| 1 | 0.748 (0.026) | 0.711 (0.015) | 0.702 (0.016) | 0.754 (0.025) | 0.693 (0.021) | 0.689 (0.020) |
| <i>R</i> | | | | | | |
| 0 | 0.746 (0.024) | 0.754 (0.026) | 0.748 (0.025) | 0.748 (0.026) | 0.758 (0.030) | 0.752 (0.029) |
| -1 | 0.749 (0.023) | 0.741 (0.034) | 0.706 (0.027) | 0.714 (0.024) | 0.739 (0.046) | 0.738 (0.058) |
| 1 | 0.758 (0.036) | 0.797 (0.026) | 0.703 (0.028) | 0.704 (0.025) | 0.655 (0.022) | 0.646 (0.027) |
| <i>F1</i> | | | | | | |
| 0 | 0.749 (0.009) | 0.751 (0.015) | 0.747 (0.016) | 0.747 (0.012) | 0.751 (0.016) | 0.749 (0.018) |
| -1 | 0.742 (0.021) | 0.735 (0.030) | 0.718 (0.024) | 0.726 (0.022) | 0.724 (0.031) | 0.720 (0.032) |
| 1 | 0.755 (0.016) | 0.751 (0.013) | 0.702 (0.018) | 0.725 (0.012) | 0.672 (0.017) | 0.666 (0.018) |

Através destes resultados, além da verificação da qualidade de estimação dos classificadores propostos, notamos que, na ausência de assimetria ($a = 0$), os métodos de classificação são bastante similares, possuindo suas métricas muito próximas, bem como apresentando distribuições semelhantes. Porém, na presença de assimetria e/ou bimodalidade multivariada ($a = -1$ ou $a = 1$), os métodos propostos são mais adequados, em especial o método ANAKDB o qual está baseado na distribuição desenvolvida. Notamos também que, neste caso, o classificador ANANB possui melhor capacidade preditiva que os métodos em comparação.

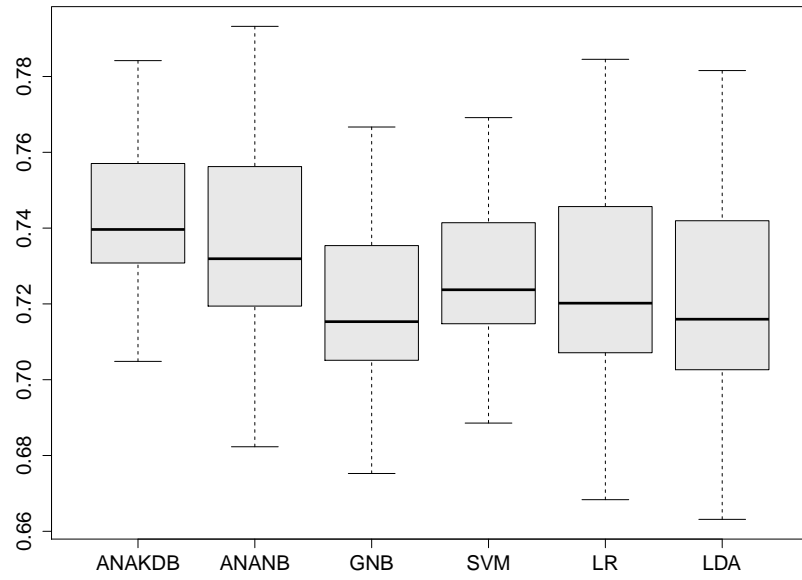


Figura 5.1: Distribuição das métrica F1 para os métodos em comparação quando $a = -1$.

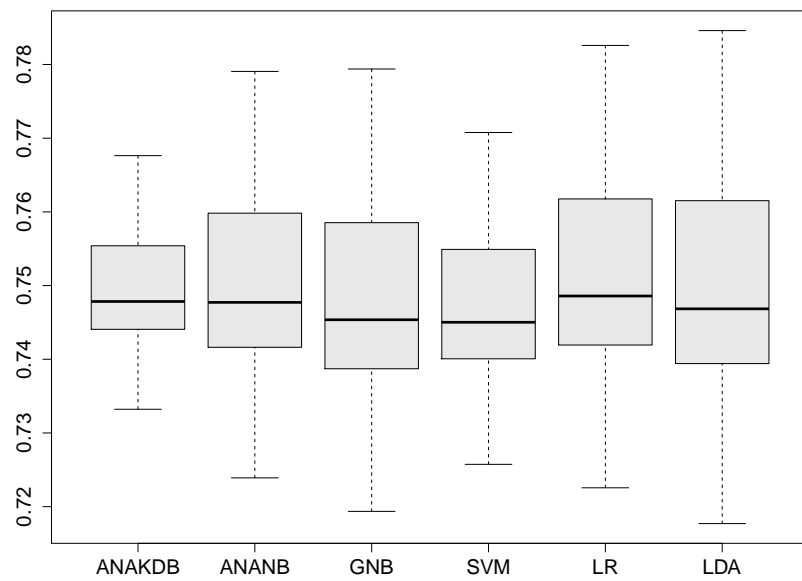


Figura 5.2: Distribuição das métrica F1 para os métodos em comparação quando $a = 0$.

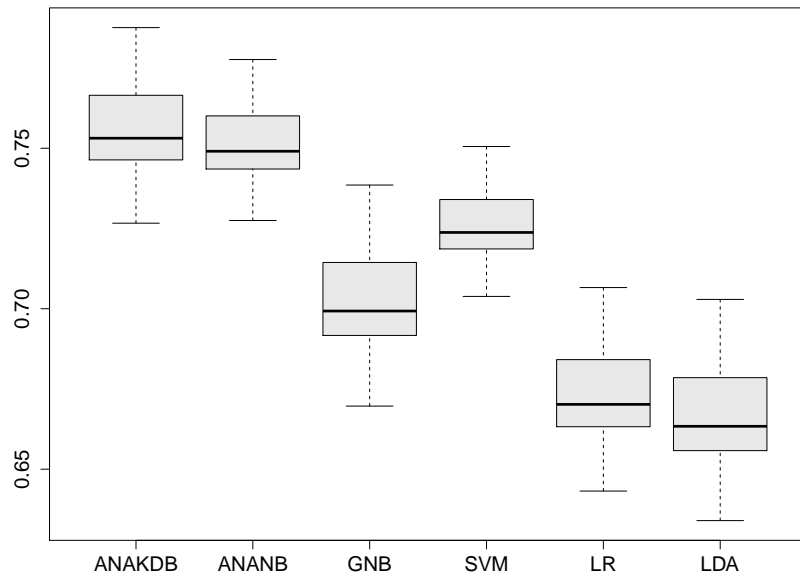


Figura 5.3: Distribuição das métrica F1 para os métodos em comparação quando $a = 1$.

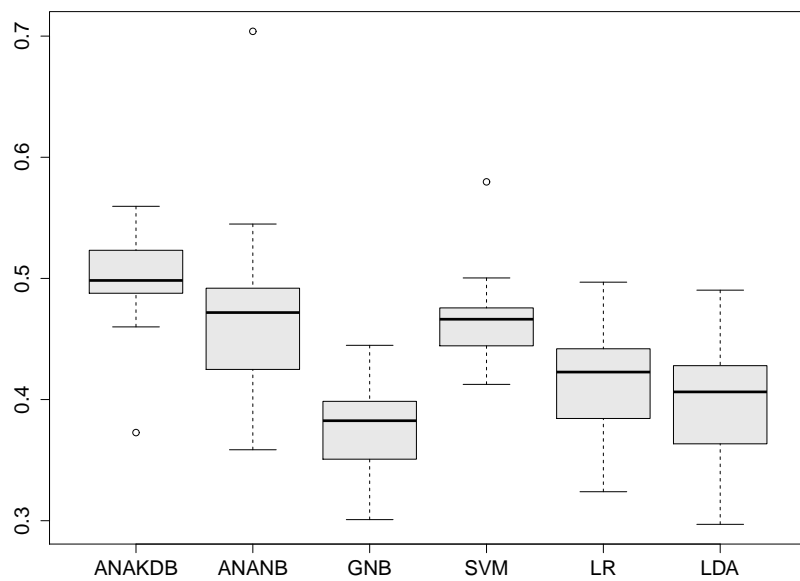


Figura 5.4: Distribuição das métrica MCC para os métodos em comparação quando $a = -1$.

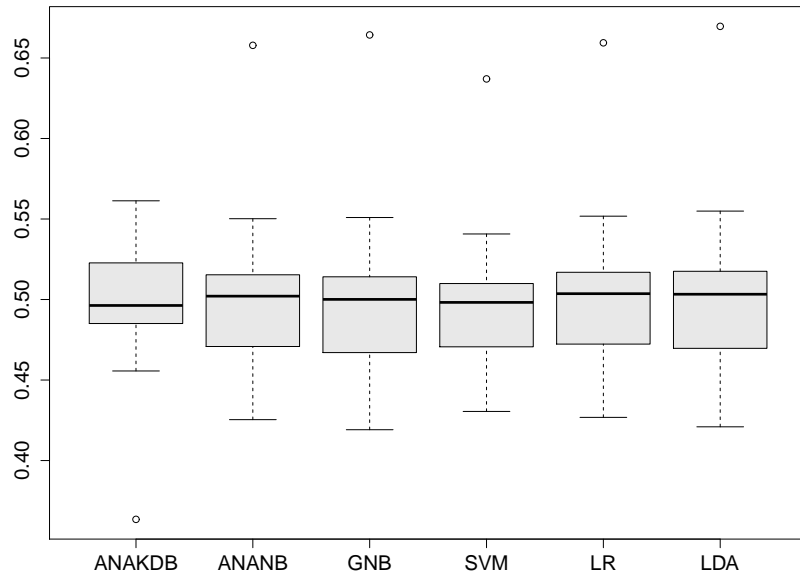


Figura 5.5: Distribuição das métrica MCC para os métodos em comparação quando $a = 0$.

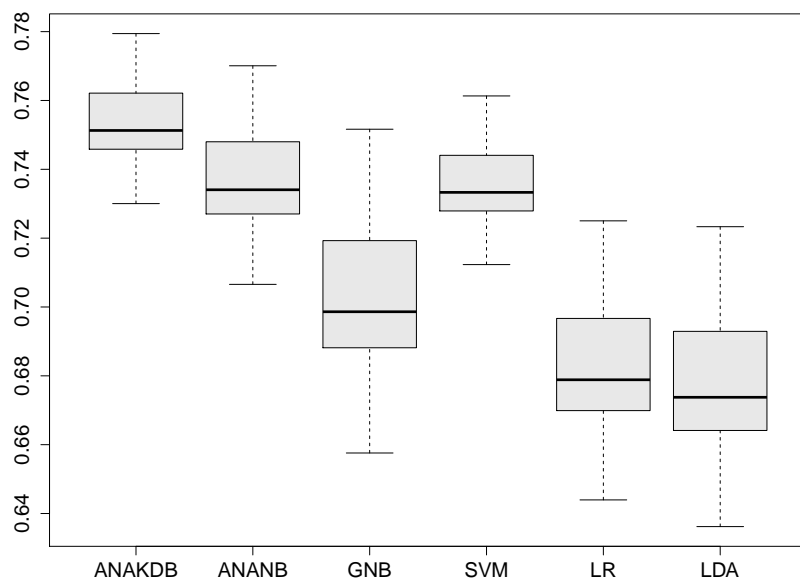


Figura 5.6: Distribuição das métrica MCC para os métodos em comparação quando $a = 1$.

5.2 Dados reais

Nesta seção apresentamos o classificador desenvolvido em aplicação a três conjuntos de dados diferentes, os quais são descritos abaixo.

HABERMAN Este conjunto de dados contém casos de um estudo que foi realizado entre 1958 e 1970, na Universidade de Chicago Hospital Billings referente a sobrevivência de pacientes que haviam sido operados devido a câncer de mama. O conjunto de dados possui quatro variáveis e 306 observações, sendo destas 225 (74%) ocorrência de morte. As variáveis observadas foram Y = ocorrência de morte após 5 anos da data de cirurgia, X_1 = Idade do paciente no momento da operação, X_2 = Ano em que do paciente realizou a operação (Ano - 1900), X_3 = Número de gânglios axilares positivos detectados.

NORMTEMP Conjunto de dados publicados em Mackowiak *et al.* (1992) obtidos no portal da revista científica 'Journal of Statistics Education' (<http://www.amstat.org/publications/jse/datasets/>) a qual levanta a discussão sobre a presença ou ausência de normalidade dos dados (Shoemaker, 1996). Este conjunto possui três variáveis e 130 observações, sendo balanceado (50%) entre o sexo masculino e feminino. Assim, possui as variáveis Y = Gênero (Masculino ou Feminino), X_1 = Temperatura do corpo em graus *Fahrenheit* e X_2 = Batimento cardíaco por minuto.

HEART Conjunto de dados referente a características de saúde de paciente com presença ou ausência de alguma doença no coração. O conjunto possui 270 observações e 14 variáveis, sendo que para este estudo consideramos quatro variáveis: Y = presença (45%) de doença no coração, X_1 = Idade, X_2 = pressão arterial em repouso, X_3 = Nível colesterol em mg/dl e X_4 = frequência cardíaca máxima atingida.

Os conjuntos de dados HABERMAN e HEART estão disponíveis no repositório de dados da Universidade da Califórnia Irvine (<http://archive.ics.uci.edu/ml/>).

As Tabelas 5.3-5.2 exibem as estatísticas básicas das variáveis explicativas de cada conjunto de dados.

Tabela 5.3: Estatísticas descritivas para o conjunto HABERMAN.

| Variável | Média | Mediana | Desvio Padrão | Assimetria | Curtose |
|-------------|--------|---------|---------------|------------|---------|
| n=360 | | | | | |
| X1 | 52.458 | 52 | 10.803 | 0.145 | -0.615 |
| X2 | 62.853 | 63 | 3.249 | 0.078 | -1.132 |
| X3 | 4.026 | 1 | 7.19 | 2.955 | 11.426 |
| Y=1 , n=225 | | | | | |
| X1 | 52.018 | 52 | 11.012 | 0.07 | -0.808 |
| X2 | 62.862 | 63 | 3.223 | 0.107 | -1.152 |
| X3 | 2.791 | 0 | 5.87 | 3.632 | 16.864 |
| Y=0, n=81 | | | | | |
| X1 | 53.679 | 53 | 10.167 | 0.475 | -0.204 |
| X2 | 62.827 | 63 | 3.342 | 0.006 | -1.138 |
| X3 | 7.457 | 4 | 9.186 | 2.022 | 5.675 |

Tabela 5.4: Estatísticas descritivas para o conjunto NORMTEMP.

| Variável | Média | Mediana | Desvio Padrão | Assimetria | Curtose |
|------------|--------|---------|---------------|------------|---------|
| n=130 | | | | | |
| X1 | 98.249 | 98.3 | 0.733 | -0.004 | 0.648 |
| X2 | 73.762 | 74 | 7.062 | -0.174 | -0.53 |
| Y=1 , n=65 | | | | | |
| X1 | 98.105 | 98.1 | 0.699 | -0.204 | -0.513 |
| X2 | 73.369 | 73 | 5.875 | -0.049 | -0.468 |
| Y=0, n=65 | | | | | |
| X1 | 98.394 | 98.4 | 0.743 | 0.094 | 1.331 |
| X2 | 74.154 | 76 | 8.105 | -0.28 | -0.822 |

Tabela 5.5: Estatísticas descritivas para o conjunto HEART.

| Variável | Média | Mediana | Desvio Padrão | Assimetria | Curtose |
|-------------|---------|---------|---------------|------------|---------|
| n=270 | | | | | |
| X1 | 54.433 | 55 | 9.109 | -0.162 | -0.575 |
| X2 | 131.344 | 130 | 17.862 | 0.715 | 0.855 |
| X3 | 249.659 | 245 | 51.686 | 1.171 | 4.726 |
| X4 | 149.678 | 153.5 | 23.166 | -0.522 | -0.145 |
| Y=1 , n=120 | | | | | |
| X1 | 56.592 | 58 | 8.116 | -0.533 | 0.02 |
| X2 | 134.442 | 130 | 19.095 | 0.867 | 0.809 |
| X3 | 256.467 | 255.5 | 47.969 | 0.289 | 0.406 |
| X4 | 138.858 | 141.5 | 23.131 | -0.239 | -0.399 |
| Y=0, n=150 | | | | | |
| X1 | 52.707 | 52 | 9.51 | 0.153 | -0.634 |
| X2 | 128.867 | 130 | 16.458 | 0.405 | 0.17 |
| X3 | 244.213 | 236 | 54.019 | 1.747 | 7.464 |
| X4 | 158.333 | 161 | 19.283 | -0.664 | 0.371 |

As aplicações foram realizadas através do método de validação cruzada *5-fold*, utilizando 30 repetições. A Tabela 5.2 apresenta médias e desvios padrões das medidas *ACC*, *MCC*, *R*, *P* e *F1* para os três conjuntos de dados e as diferentes técnicas de classificação.

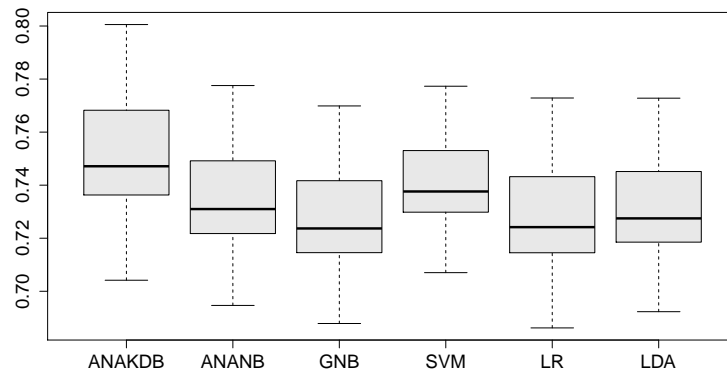
Através destes resultados notamos que as redes de classificação desenvolvidas possuem capacidade preditiva satisfatória em comparação com os demais métodos, pois em na maior parte dos experimentos possuem - em média - a maior capacidade preditiva. Analisando conjuntamente as métricas, o método ANAKDB mostra-se superior para os conjuntos HABERMAN e HEART. Para o conjunto HABERMAN, o método ANAKDB possui o melhor desempenho considerando as métricas *ACC* (0.750), *R* (0.945), *P* (0.789) e *F1* (0.849). Para o conjunto HEART, o método ANAKDB possui o maior médio desempenho para todas as medidas analisadas. Apesar da maior aproximação das métricas para o conjunto de dados NORMTEMP, o método ANANB mostra-se superior, em especial para a métrica *F1* (0.556). Vale salientar que para o conjunto NORTEMP, ANAKDB possui a maior precisão ($P = 0.551$) e ANANB a maior sensibilidade ($R = 0.605$). Além disso, para estas aplicações o método SVM mostra-se também com alta capacidade preditiva, porém levemente inferior.

As Figuras 5.7-5.9 apresentam, respectivamente, os *boxplots* da distribuição das medidas gerais para os conjuntos HABERMAN, NORMTEMP e HEART. Os *boxplots* também evidenciam, no geral, o maior desempenho de ANAKDB para o conjunto HABERMAN, o maior desempenho de ANANB para o conjunto HEART e a aproximação das técnicas para o conjunto HEART.

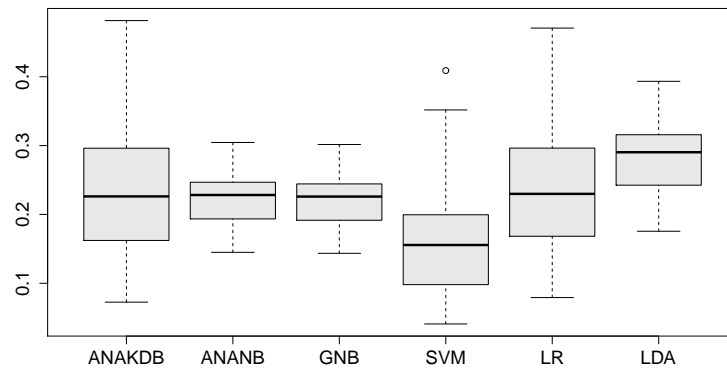
Estes resultados também são validados pelas Tabelas descritivas já exibidas em 5.3-5.2, as quais demonstram uma forte assimetria, em especial em X_3 , para o conjunto HABERMAN, uma moderada assimetria para o conjunto HEART e fraca assimétrica para as covariáveis do conjunto NORMTEMP.

Tabela 5.6: Médias e desvios padrões das medidas ACC , MCC , R , P e $F1$ para os três conjuntos de dados e as diferentes técnicas de classificação.

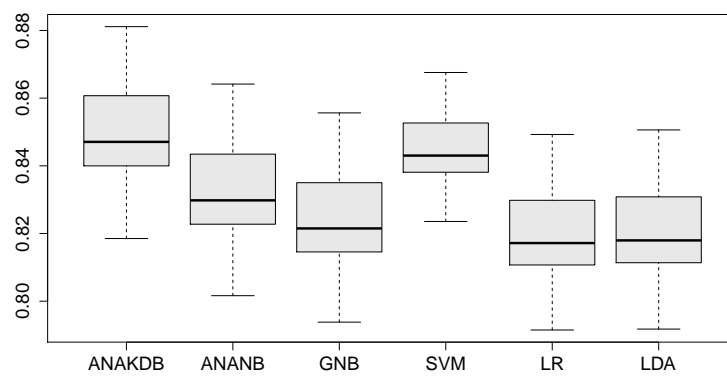
| Conjunto | Técnica | ACC | MCC | R | P | $F1$ |
|----------|---------|---------------|---------------|---------------|---------------|---------------|
| HABERMAN | ANAKDB | 0.750 (0.026) | 0.236 (0.103) | 0.945 (0.023) | 0.783 (0.031) | 0.849 (0.017) |
| | ANANB | 0.734 (0.022) | 0.225 (0.043) | 0.890 (0.031) | 0.772 (0.024) | 0.832 (0.017) |
| | GNB | 0.726 (0.022) | 0.222 (0.043) | 0.882 (0.030) | 0.764 (0.024) | 0.823 (0.017) |
| | SVM | 0.740 (0.019) | 0.163 (0.087) | 0.937 (0.027) | 0.754 (0.018) | 0.844 (0.012) |
| | LR | 0.727 (0.023) | 0.238 (0.099) | 0.897 (0.027) | 0.746 (0.017) | 0.819 (0.015) |
| | LDA | 0.730 (0.022) | 0.285 (0.059) | 0.890 (0.032) | 0.751 (0.011) | 0.820 (0.016) |
| HEART | ANAKDB | 0.765 (0.034) | 0.527 (0.077) | 0.771 (0.082) | 0.73 (0.064) | 0.74 (0.043) |
| | ANANB | 0.746 (0.058) | 0.491 (0.135) | 0.718 (0.095) | 0.716 (0.08) | 0.709 (0.072) |
| | GNB | 0.739 (0.057) | 0.486 (0.134) | 0.711 (0.093) | 0.709 (0.079) | 0.702 (0.071) |
| | SVM | 0.753 (0.04) | 0.5 (0.1) | 0.736 (0.066) | 0.718 (0.053) | 0.722 (0.049) |
| | LR | 0.71 (0.053) | 0.446 (0.144) | 0.702 (0.077) | 0.658 (0.073) | 0.678 (0.071) |
| | LDA | 0.713 (0.052) | 0.451 (0.143) | 0.702 (0.077) | 0.663 (0.071) | 0.68 (0.07) |
| NORMTEMP | ANAKDB | 0.535 (0.076) | 0.176 (0.132) | 0.489 (0.149) | 0.551 (0.111) | 0.498 (0.111) |
| | ANANB | 0.521 (0.112) | 0.217 (0.179) | 0.605 (0.127) | 0.517 (0.102) | 0.556 (0.111) |
| | GNB | 0.516 (0.111) | 0.216 (0.178) | 0.599 (0.125) | 0.512 (0.101) | 0.551 (0.109) |
| | SVM | 0.502 (0.121) | 0.153 (0.129) | 0.483 (0.135) | 0.500 (0.128) | 0.493 (0.130) |
| | LR | 0.537 (0.090) | 0.171 (0.160) | 0.540 (0.081) | 0.544 (0.120) | 0.538 (0.094) |
| | LDA | 0.537 (0.090) | 0.171 (0.160) | 0.540 (0.081) | 0.544 (0.120) | 0.538 (0.094) |



(a)

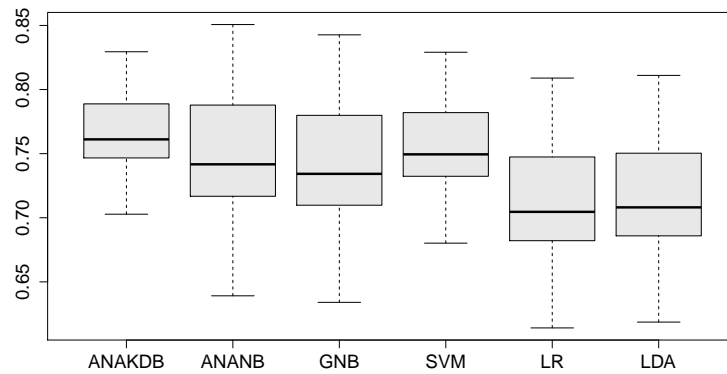


(b)

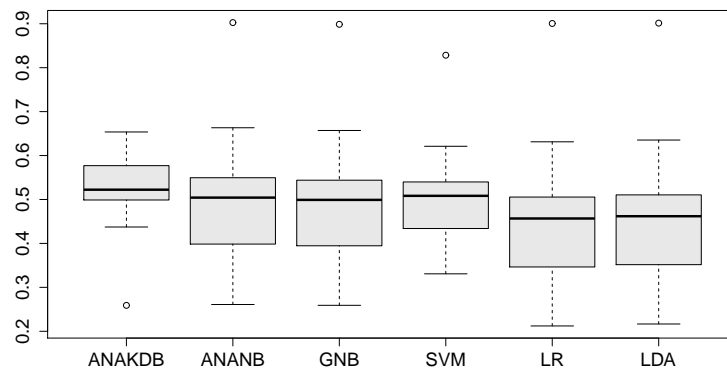


(c)

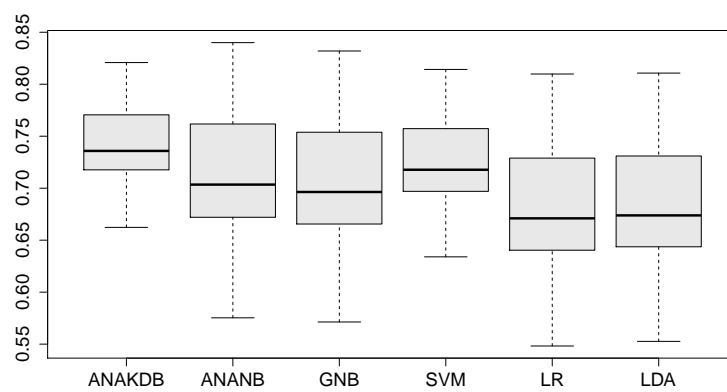
Figura 5.7: Distribuição das métricas (a) ACC , (b) MCC e (c) $F1$ para o conjunto de dados HABERMAN.



(a)

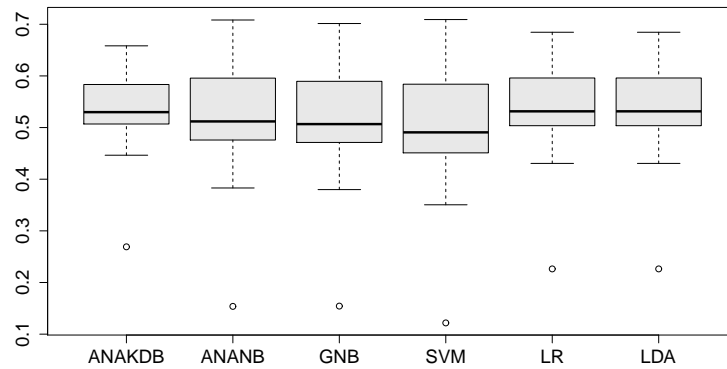


(b)

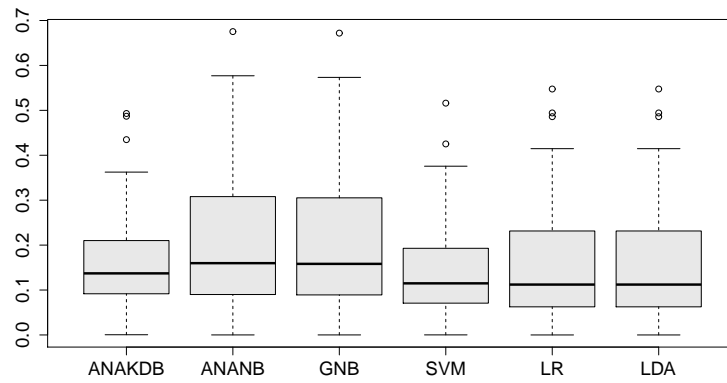


(c)

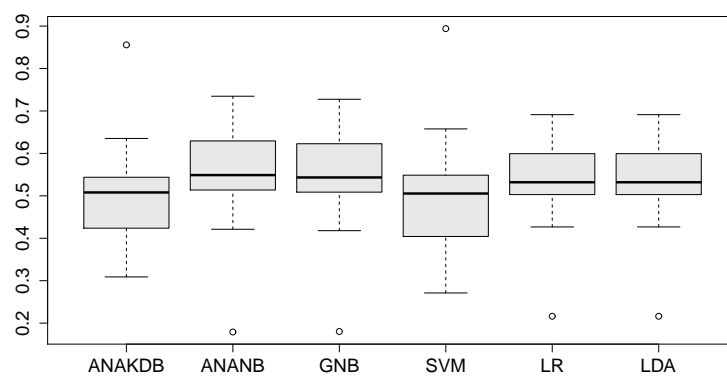
Figura 5.8: Distribuição das métricas (a) ACC , (b) MCC e (c) $F1$ para o conjunto de dados HEART.



(a)



(b)



(c)

Figura 5.9: Distribuição das métricas (a) ACC , (b) MCC e (c) $F1$ para o conjunto de dados NORMTEMP.

5.3 Comentários Finais

Neste capítulo apresentamos a validação das redes de classificação de k -dependência para o caso da distribuição alfa normal assimétrica multivariada, a qual se baseou em um estudo de simulação e na aplicação do método para três conjuntos de dados reais. Para todos os casos foi possível verificar para os métodos propostos um desempenho preditivo satisfatório, em comparação com os métodos tradicionalmente utilizados.

Capítulo 6

Considerações Finais

Nesta Tese, propomos uma nova distribuição de probabilidade denominada distribuição alfa normal assimétrica multivariada, a qual foi obtida pela extensão do modelo univariado proposto por Elal-Olivero (2010), bem como apresentamos suas propriedades, funções marginais e condicionais e procedimentos de inferência.

Tal distribuição mostra-se estatisticamente eficiente para análise de dados na presença de assimetria e/ou bimodalidade, as quais estão fortemente inseridas na análise de dados de problemas reais. Além disso, a distribuição desenvolvida possui como caso particular a distribuição normal multivariada e, para os casos analisados, se sobrepõe à distribuição normal assimétrica multivariada, a qual não é capaz de tratar dados bimodais.

Para o contexto de classificação, apresentamos dois novos modelos de *data mining*, sendo estes redes mais flexíveis para dados contínuos e capaz de captar assimetrias e/ou bimodalidades. Notamos, por meio dos procedimentos de simulação realizados e da aplicação a um conjunto de dados reais, que esta nova proposição é significativamente mais apropriada que a modelagem usual em comparação com a distribuição normal clássica, tendo um ganho satisfatório de desempenho, além de ser igualmente adequada para conjuntos de dados sem a presença de assimetria.

Além disso, as comparações exibiram que existe um maior desempenho preditivo desta metodologia na presença de bimodalidade e assimetria, em comparação com os métodos SVM, regressão logística, análise discriminante.

Ao mesmo tempo em que os procedimentos aqui desenvolvidos permeiam as mais variadas aplicações, outros diversos desenvolvimentos teóricos poderão dar continuidade ao

trabalho desenvolvido. Em especial a criação de implementações (pacotes) para os métodos desenvolvidos, a estimação Bayesiana para os parâmetros da distribuição alfa normal multivariada, verificação de valores extremos, métodos computacionais para acelerar o tempo de estimação dos parâmetros, métodos de regressão, desenvolvimento de novas distribuições de probabilidade, novas estruturas de redes probabilísticas e até mesmo novos procedimentos de classificação utilizando a distribuição proposta.

Referências Bibliográficas

- Aguilera, P., Fernández, A., Fernández, R., Rumí, R. & Salmerón, A. (2011). Bayesian networks in environmental modelling. *Environmental Modelling & Software*, **26**(12), 1376–1388.
- Arellano-Valle, R. B., Gómez, H. W. & Quintana, F. A. (2004). A new class of skew normal distributions. *Communications in Statistics - Theory and Methods*, **33**(7), 1465–1480.
- Arellano-Valle, R. B., Cortes, M. A. & Gomez, H. W. (2010). An extension of the epsilon skew normal distribution. *Communications in Statistics - Theory and Methods*, **39**(5), 912–922.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.
- Azzalini, A. & Bowman, A. (1990). A look at some data on the old faithful geyser. *Applied Statistics*, pages 357–365.
- Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(3), 579–602.
- Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**(4), 715–726.
- Bahrami, W., Agahi, H. & Rangin, H. (2009). A two-parameter balakrishnan skew-normal distribution. *J. Statist. Res. Iran*, **6**, 231–242.
- Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, **16**(5), 412–424.

- Bishop, C. M. (2006). Pattern recognition. *Machine Learning*.
- Bodén, M., Dellaire, G., Burrage, K. & Bailey, T. L. (2010). A bayesian network model of proteins' association with promyelocytic leukemia (pml) nuclear bodies. *Journal of Computational Biology*, **17**(4), 617–630.
- Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**(1), 99–113.
- Breiman, L. (1998). Arcing classifier (with discussion and a rejoinder by the author). *The annals of statistics*, **26**(3), 801–849.
- Brent, R. P. (1971). An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, **14**(4), 422–425.
- Burkom, H., Elbert, Y., Ramac-Thomas, L., Cuellar, C. & Hung, V. (2013). Refinement of a population-based bayesian network for fusion of health surveillance data. *Online Journal of Public Health Informatics*, **5**(1).
- Chan, T. U., Hart, B. T., Kennard, M. J., Pusey, B. J., Shenton, W., Douglas, M. M., Valentine, E. & Patel, S. (2012). Bayesian network models for environmental flow decision making in the daly river, northern territory, australia. *River Research and Applications*, **28**(3), 283–301.
- Cheng, I., Chen, F., Rodrigues, S. D., Pañella, O. G., Vicent, L. & Basu, A. (2010). Intelligent games for education-an intention monitoring approach based on dynamic bayesian network. In *Eurographics 2010-Education Papers*, pages 25–32. The Eurographics Association.
- Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods and Applications*, **14**(3), 331–341.
- de Campos, L. M. & Romero, A. E. (2009). Bayesian network models for hierarchical text classification from a thesaurus. *International journal of approximate reasoning*, **50**(7), 932–944.

- Elal-Olivero, D. (2010). Alpha-skew-normal distribution. *Proyecciones (Antofagasta)*, **29**(3), 224–240.
- Everitt, B. S., Hothorn, T., Hothorn, M. T. & Everitt, C. (2014). Package HSAUR3.
- Friedman, N., Geiger, D. & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning*, **29**(2-3), 131–163.
- Gómez, H. W., Salinas, H. S. & Bolfarine, H. (2006). Generalized skew-normal models: properties and inference. *Statistics*, **40**(6), 495–505.
- Gui, W., Chen, P.-H. & Wu, H. (2012). A symmetric component alpha normal slash distribution: Properties and inferences. *Journal of Statistical Theory and Applications*, **12**(1), 55–66.
- Gupta, A. K., González-Farías, G. & Domínguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, **89**(1), 181–190.
- Hallin, M. & Ley, C. (2012). Skew-symmetric distributions and fisher information—a tale of two densities. *Bernoulli*, **18**(3), 747–763.
- Handam, A. (2012). A note on generalized alpha-skew-normal distribution. *International Journal of Pure and Applied Mathematics*, **74**(4), 491–496.
- Harandi, S. S. & Alamatsaz, M. (2013). Alpha-skew-laplace distribution. *Statistics and Probability Letters*, **83**, 774–782.
- Härdle, W. (1991). *Smoothing techniques: with implementation in S*. Springer.
- Henningsen, A. & Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in r. *Computational Statistics*, **26**(3), 443–458.
- Hothorn, T. & Everitt, B. S. (2014). *A handbook of statistical analyses using R*. CRC press.
- Jamalizadeh, A., Behboodian, J. & Balakrishnan, N. (2008). A two-parameter generalized skew-normal distribution. *Statistics and Probability Letters*, **78**(13), 1722–1726.
- Jensen, F. V. (1996). *An introduction to Bayesian networks*, volume 210. UCL press London.

- Justel, A., Peña, D. & Zamar, R. (1997). A multivariate kolmogorov-smirnov test of goodness of fit. *Statistics and Probability Letters*, **35**(3), 251–259.
- Kotecha, J. H. & Djuric, P. M. (1999). Gibbs sampling approach for generation of truncated multivariate gaussian random variables. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 3, pages 1757–1760. IEEE.
- Ley, C. & Paindaveine, D. (2010). On the singularity of multivariate skew-symmetric models. *Journal of Multivariate Analysis*, **101**(6), 1434–1444.
- Louzada, F. & Ara, A. (2012). Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool. *Expert Systems with Applications*, **39**(14), 11583–11592.
- Louzada, F., Ara, A. & Fernandes, G. B. (2016). Classification methods applied to credit scoring: A systematic review and overall comparison. *arXiv preprint arXiv:1602.02137*.
- Mackowiak, P. A., Wasserman, S. S. & Levine, M. M. (1992). A critical appraisal of 98.6 f, the upper limit of the normal body temperature, and other legacies of carl reinhold august wunderlich. *Jama*, **268**(12), 1578–1580.
- Madden, M. G. (2009). On the classification performance of tan and general bayesian networks. *Knowledge-Based Systems*, **22**(7), 489–495.
- Mayor, M. & Frei, P.-Y. (2003). *New worlds in the cosmos: the discovery of exoplanets*. Cambridge University Press.
- McAssey, M. P. (2013). An empirical goodness-of-fit test for multivariate distributions. *Journal of Applied Statistics*, **40**(5), 1120–1131.
- Metropolis, N. & Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, **44**(247), 335–341.
- Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
- Mitchell, T. M. (2006). *The discipline of machine learning*. Carnegie Mellon University, School of Computer Science, Machine Learning Department.

- Murakami, Y. & Mizuguchi, K. (2010). Applying the naïve bayes classifier with kernel density estimation to the prediction of protein–protein interaction sites. *Bioinformatics*, **26**(15), 1841–1848.
- Nagarajan, R., Scutari, M. & Lèbre, S. (2013). *Bayesian Networks in R with Applications in Systems Biology*. Springer.
- Orun, A. & Seker, H. (2012). Development of a computer game-based framework for cognitive behaviour identification by using bayesian inference methods. *Computers in Human Behavior*, **28**(4), 1332–1341.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- Peng, Y., Wang, G., Kou, G. & Shi, Y. (2011). An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing*, **11**(2), 2906–2915.
- Perez, A., Larranaga, P. & Inza, I. (2006). Supervised classification with conditional gaussian networks: Increasing the structure complexity from naive bayes. *International Journal of Approximate Reasoning*, **43**(1), 1–25.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, **2**(1), 37–63.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rubinstein, R. Y. & Kroese, D. P. (2011). *Simulation and the Monte Carlo method*, volume 707. Wiley. com.
- Rubio, A. & Gámez, J. A. (2011). Flexible learning of k-dependence bayesian network classifiers. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 1219–1226. ACM.
- Sahami, M. (1996). Learning limited dependence bayesian classifiers. In *KDD*, pages 335–338.

- Seber, G. (1977). *F. linear regression analysis*.
- Shoemaker, A. L. (1996). What's normal?-temperature, gender, and heart rate. *Journal of Statistics Education*, **4**(2).
- Stewart, G. W. (1998). *Matrix Algorithms: Basic Decompositions*. SIAM (Society for industrial and applied mathematics).
- Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Müller, A. & Candrian, G. (2013). Machine learning approach for classification of adhd adults. *International Journal of Psychophysiology*.
- Toral, R. & Chakrabarti, A. (1993). Generation of gaussian distributed random numbers by using a numerical inversion method. *Computer physics communications*, **74**(3), 327–334.
- Witten, I. H., Frank, E. & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*. Elsevier.
- Xu, J., Chang, H., Chen, C., Kaiser, W. & Pottie, G. (2013). Context-driven, prescription based personal activity classification: Methodology, architecture and end-to-end implementation.
- Xueling, W. (2010). Study of bayesian network classification models and its application in credit scoring [j]. *Computer & Digital Engineering*, **8**, 030.
- Zhang, M.-L., Peña, J. M. & Robles, V. (2009). Feature selection for multi-label naive bayes classification. *Information Sciences*, **179**(19), 3218–3229.