



UFRJ

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

MODELOS BAYESIANOS ASSIMÉTRICOS

Ralph dos Santos Silva

2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



UFRJ

MODELOS BAYESIANOS ASSIMÉTRICOS

Ralph dos Santos Silva

Tese de Doutorado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

Orientadores: Helio S. Migon
Hedibert F. Lopes

Rio de Janeiro, dezembro de 2006.

MODELOS BAYESIANOS ASSIMÉTRICOS

Ralph dos Santos Silva

Orientadores: Helio S. Migon

Hedibert F. Lopes

Tese de Doutorado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

Presidente, Prof. Helio S. Migon
IM-UFRJ

Prof. Hedibert F. Lopes
GSB-UC, EUA

Prof. Dani Gamerman
IM-UFRJ

Prof. Nikolai V. Kolev
IME-USP

Prof. Fernando A. S. Moura
IM-UFRJ

Prof. Marcelo C. Medeiros
DE-PUC-RJ

Rio de Janeiro, dezembro de 2006.

Silva, Ralph dos Santos

Modelos Bayesianos Assimétricos/ Ralph dos Santos Silva - Rio de Janeiro: UFRJ/IM, 2006.

vi, 111f.: il.; 31cm.

Orientadores: Helio Santos Migon e Hedibert Freitas Lopes

Tese (doutorado) - UFRJ/IM/ Programa de Pós-graduação em Estatística, 2006.

Referências Bibliográficas: f.83-87.

1. Modelos assimétricos. 2. Análise Bayesiana. 3. MCMC. I. Migon, Helio S. e Lopes, Hedibert F. II. Universidade Federal do Rio de Janeiro, Instituto de Matemática. III. Título.

RESUMO

MODELOS BAYESIANOS ASSIMÉTRICOS

Ralph dos Santos Silva

Orientadores: Helio S. Migon

Hedibert F. Lopes

Resumo da Tese de Doutorado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

Neste trabalho propomos alguns modelos em distribuições assimétricas com estimação de todas as quantidades desconhecidas através do enfoque bayesiano. Primeiro estudamos uma classe de distribuições definidas nos reais positivos chamada de Gaussiana inversa generalizada estendida com aplicações a um modelo log-linear e outro de volatilidade estocástica. Na segunda parte introduzimos erros assimétricos (Azzalini, 1985) no modelo de longa dependência $ARFIMA(p, d, q)$ e usamos alguns critérios de comparação de modelos. Na terceira parte investigamos dependência e assimetria multivariada através do conceito de função cópula, com as distribuições marginais sendo assimétricas (Fernández e Steel, 1998). Alguns estudos de Monte Carlo ajudam a validar seleção de modelos em cópulas através do critério de informação do desvio. Seguimos com uma aplicação a dados de atuária. Na quarta parte estendemos o conceito de mistura em cópulas mostrando propriedades da distribuição e cópula resultantes, e como algumas medidas de dependência podem ser calculadas ou estimadas nestas misturas. Aplicações a dados de atuária e finanças ilustram o método e a proposta.

Palavras-chave: Cópulas, inferência bayesiana, distribuição assimétrica, Monte Carlo via cadeias de Markov, seleção de modelos.

ABSTRACT

BAYESIAN SKEW MODELS

Ralph dos Santos Silva

Advisors: Helio S. Migon

Hedibert F. Lopes

Abstract da Tese de Doutorado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

In this work we propose some statistical models with skew distributions and applications to real data sets. We make inference using a fully Bayesian approach estimating all unknown quantities in all models. First, we investigate the extended generalized inverse Gaussian distribution with applications to a log-linear and a stochastic volatility models. Second, we extend the long memory process ARFIMA(p, d, q) allowing the innovations to follow skew distributions (Azzalini, 1985). Third, we are interested in multivariate dependence and skewness through copula functions and skew distributions (Fernández e Steel, 1998) for the marginals. We deal with model selection in copulas using the deviance information criterion by Monte Carlo studies and follow with an application to actuarial data. Finally, we generalize some results found in the literature (Joe, 1997; Nelsen, 1999) about mixture in copulas. We show how some measures of dependence in the mixture can be computed, at least numerically. Applications in actuaries and finance illustrate the method.

Key-words: Bayesian inference; Copulas; Model selection; Monte Carlo Markov chain; Skew models.

PARA

CELSON FERREIRA DA SILVA

E

MARIA DA CONCEIÇÃO SANTOS DA SILVA

AGRADECIMENTOS

EM ESPECIAL A ROMY ELENA RODRÍGUEZ RAVINES.

AO MEU ORIENTADOR HELIO SANTOS MIGON;

AO MEU ORIENTADOR HEDIBERT FREITAS LOPES;

A TODOS OS PROFESSORES DO DME/IM/UFRJ; E

AO SUPORTE FINANCEIRO DADO PELA CAPES.

SUMÁRIO

Lista de Tabelas	iv
Lista de Figuras	vi
Capítulo 1: Introdução	1
1.1 Gaussiana inversa generalizada estendida	1
1.2 Modelos ARFIMA assimétricos	2
1.3 Copulas: inferência e seleção	2
1.4 Mistura em cópulas e assimetria	3
Capítulo 2: Gaussiana inversa generalizada estendida	4
2.1 Propriedades gerais	4
2.2 Modelos estatísticos com distribuição GIGE	6
2.2.1 Modelo log-linear	7
2.2.2 Modelo de volatilidade estocástica	9
2.3 Aplicações	11
2.3.1 Comparação de dois tratamentos	11
2.3.2 Índice BOVESPA	16
2.4 Discussão final	19
Capítulo 3: Modelos ARFIMA assimétricos	20
3.1 Propriedades gerais	21
3.2 Modelos ARFIMA(p, d, q)	26
3.2.1 Modelo ruído branco	28
3.2.2 Modelo ARMA(p, q)	34
3.2.3 Modelo ARFIMA($0, d, 0$)	36
3.3 Aplicações	37
3.3.1 Modelo ARMA(p, q): dados artificiais	37
3.3.2 Índice anual do preço do trigo	39
3.3.3 Modelo ARFIMA($0, d, 0$): dados artificiais	42
3.3.4 Nível mínimo do rio Nilo	43

3.4	Discussão final	45
Capítulo 4:	Cópuas: inferência e seleção	46
4.1	Propriedades gerais	47
4.2	Inferência e distribuições marginais	49
4.2.1	Distribuições marginais	52
4.2.2	Distribuição a priori	53
4.3	Aplicações	53
4.3.1	DIC e seleção de cópuas	54
4.3.2	Características gerais dos estudos de Monte Carlo	54
4.3.3	Marginais com distribuições normais	55
4.3.4	Marginais com distribuições exponenciais	57
4.3.5	Marginais com distribuições normais assimétricas	58
4.3.6	Requerimentos de indenizações de seguros	60
4.4	Discussão final	62
Capítulo 5:	Mistura em cópuas e assimetria	63
5.1	Propriedades gerais	63
5.2	Medidas de dependência	64
5.3	Mistura em cópuas	66
5.4	Inferência na distribuição a posteriori	72
5.5	Aplicações	73
5.6	Discussão final	79
Capítulo 6:	Conclusões e considerações finais	80
6.1	Gaussiana inversa generalizada estendida	80
6.2	Modelos ARFIMA assimétricos	80
6.3	Cópuas: inferência e seleção	81
6.4	Mistura em cópuas e assimetria	81
	Referências Bibliográficas	83
Apêndice A:	Função de Bessel do terceiro tipo	88
Apêndice B:	Priori de Jeffreys para dados i.i.d da GIGE	90

Apêndice C:	Priori de Jeffreys para dados i.i.d da NA	94
C.1	Aplicação: instituto Australiano de esportes	96
Apêndice D:	Critérios de seleção de modelos	99
D.1	AIC e BIC	99
D.2	DIC	99
D.3	EPD	100
Apêndice E:	Algumas distribuições importantes	101
Apêndice F:	Resultados de Monte Carlo para cópulas	105
F.1	Marginais com distribuições normais	105
F.2	Marginais com distribuições exponenciais	106
F.3	Marginais com distribuições normais assimétricas	107
Apêndice G:	Abreviações	108
Apêndice H:	Aspectos computacionais	109
H.1	Amostrador de Gibbs	109
H.2	Algoritmo de Metropolis-Hastings	109
H.3	Amostrador da fatia	110

LISTA DE TABELAS

2.1	Tempos de sobrevivência	11
2.2	Sumário da distribuição a posteriori do modelo (2.8)	12
2.3	Correlação estimada a posteriori do modelo (2.8)	12
2.4	Sumário da distribuição a posteriori do modelo (2.8) para $\lambda > 0$ e $\xi = 0$	15
2.5	Comparação de modelos log-linear GIGE e GG	15
2.6	Erro quadrático médio para o modelo log-linear GIGE	15
2.7	Sumário da distribuição a posteriori do modelo de V.E c/ erro normal	16
2.8	Sumário da distribuição a posteriori do modelo (2.10)	17
2.9	Comparação de modelos de vol. estocástica GIGE e GG	18
3.1	Resultados de simulação: dados i.i.d da normal assimétrica	31
3.2	Resultados de simulação: dados i.i.d da t-Student assimétrica	33
3.3	Resultados empíricos do modelo ARMA(p, q) NA e TA	38
3.4	Sumário da distribuição a posteriori do modelo ARMA(1,1) com erros NA	38
3.5	Sumário da distribuição a posteriori do modelo ARMA(1,1) com erros TA	39
3.6	Critérios de seleção de modelos aplicados ao índice de preço do trigo	40
3.7	Sumário da distribuição a posteriori do modelo ARMA(0,2) com erros TA	41
3.8	Resultados empíricos do modelo ARFIMA(0, d , 0) NA e TA	42
3.9	Sumário da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros NA	43
3.10	Sumário da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros TA	43
3.11	Critérios de seleção de modelos aplicados ao volume do rio Nilo	44
3.12	Sumário da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros TA	45
4.1	Algumas funções cópulas	48
4.2	τ de Kendall para as cópulas da tabela (4.1)	49
4.3	Densidades das cópulas da tabela (4.1)	51
4.4	Resultado de simulação: cópulas e marginais normais	55
4.5	Propriedades frequentista: cópulas e marginais normais	56
4.6	Resultado de simulação: cópulas e marginais exponenciais	57
4.7	Propriedades frequentista: cópulas e marginais exponenciais	58

4.8	Resultado de simulação: cópulas e marginais normais assimétricas	59
4.9	Propriedades frequentista: cópulas e marginais normais assimétricas	59
4.10	Comparação de modelo: pedidos de perda da companhia de seguro	61
4.11	Sumário da distribuição a posteriori do modelo de cópula Gumbel	61
5.1	Comparação de modelo: pedidos de perda da companhia de seguro	75
5.2	Sumário da distribuição a posteriori do modelo de cópula Gumbel	75
5.3	Comparação de modelo: índices de Amsterdã e Frankfurt	77
5.4	Sumário da distribuição a posteriori do modelo de cópula Student	77
5.5	Comparação de modelos: índices de Mexico e Argentina	78
5.6	Sumário da distribuição a posteriori do modelo de cópula Frank	79
C.1	Sumário da distribuição a posteriori de regressão NA.	97
C.2	Comparação de modelos na regressão NA.	97
F.1	Resultado de simulação: cópulas e marginais normais	105
F.2	Resultado de simulação: cópulas e marginais exponenciais	106
F.3	Resultado de simulação: cópulas e marginais normais assimétricas	107

LISTA DE FIGURAS

2.1	Densidades GIGE	7
2.2	Histogramas da amostra da posteriori: modelo (2.8)	13
2.3	Caracterização da posteriori conjunta do modelo (2.8)	14
2.4	Retornos IBOVESPA e volatilidades estimadas	16
2.5	Histogramas da posteriori do modelo (2.10)	17
2.6	Volatilidades estimadas do modelo (2.10)	18
3.1	Densidades normais assimétricas	22
3.2	Densidades t-Student assimétricas	25
3.3	Séries temporais artificiais dos processos ARFIMA	29
3.4	Função de verossimilhança de ν na distribuição TA	32
3.5	Índice anual de preço do trigo de 1500 a 1869.	39
3.6	Nível mínimo do rio Nilo	44
4.1	Contorno: cópulas Clayton e Frank e marginais normais padrão	46
4.2	Dados simulados e contorno para as cópulas da tabela (4.1)	50
4.3	Requerimentos de indenizações de seguros (Loss e ALAE).	60
5.1	Requerimentos de indenizações de seguros (Loss e ALAE)	74
5.2	Resíduos de GARCH(1,1) para os índices de Amsterdã e Frankfurt	76
5.3	Resíduos de GARCH(1,1) para os índices de Mexico e Argentina	78
C.1	Histograma dos resíduos do modelo de regressão	98

Capítulo 1

INTRODUÇÃO

Neste trabalho propomos e exploramos modelos estatísticos com distribuições assimétricas univariadas e multivariadas através do conceito de função cópula (Sklar, 1959; Schweizer e Sklar, 1983; Joe, 1997; Nelsen, 1999). Estimamos todas as quantidades desconhecidas dos modelos através do enfoque bayesiano. Na maioria dos casos apresentados as distribuições a posteriori têm formas fechadas desconhecidas, então técnicas de simulação Monte Carlo via cadeias de Markov (Gelman e Lopes, 2006) são utilizadas para obter amostras das respectivas distribuições a posteriori. Em vários casos utilizamos o amostrador da fatia (Neal, 2003) dentro do amostrador de Gibbs (Geman e Geman, 1984; Gelfand e Smith, 1990). Alguns métodos de seleção de modelos são utilizados em cada capítulo para determinar o melhor modelo ajustado, indicando os possíveis casos particulares e também na identificação dos modelos simulados nos estudos de Monte Carlo, principalmente os apresentados no capítulo 3. Destacamos o desvio preditivo esperado (EPD) (Gelfand e Ghosh, 1998), o critério de informação do desvio (DIC) (Spiegelhalter, Best, Carlin, e Linde, 2002) e suas alternativas dadas pelo critério de informação de Akaike (AIC) (Akaike, 1974), o critério de informação bayesiano (BIC) (Schwarz, 1978) e seus valores esperados EAIC e EBIC (veja Brooks (2002) para maiores detalhes), todos descritos no apêndice D. Este trabalho está dividido basicamente em contribuições por capítulo.

1.1 Gaussiana inversa generalizada estendida

No primeiro capítulo apresentamos uma distribuição que chamamos de Gaussiana inversa generalizada estendida (GIGE). Essa distribuição é uma extensão da distribuição dada em Jørgensen (1982). Trata-se de uma distribuição univariada contínua nos reais positivos de quatro parâmetros, sendo bastante flexível e tendo diversos casos particulares, entre eles a gama generalizada, a gama, a exponencial e a Weibul. Esta distribuição permite assimetrias a direita, e também a esquerda. Desenvolvemos a distribuição a priori de Jeffreys para o modelo de dados independentes e identicamente distribuídos. Com essa distribuição, propomos um modelo log-linear como extensão do trabalho de Achcar e Bolfarine (1986) e outro de volatilidade estocástica baseada nos retornos ao quadrado do índice BOVESPA onde comparamos os resultados com aqueles encontrados em Lopes e Migon (2002).

1.2 Modelos ARFIMA assimétricos

No segundo capítulo estendemos os modelos de longa dependência descritos através dos processos ARFIMA (p, d, q) (Granger, 1980; Hosking, 1981; Beran, 1994; Silva, 2003) permitindo que os mesmos tenham erros aleatórios assimétricos e em alguns casos também com caudas pesadas. Nos concentramos nas distribuições normal e t-Student assimétricas segundo Azzalini (1985) (veja também Azzalini e Dalla Valle (1996); Azzalini e Capitanio (1999); Azzalini e Capitanio (2003)). Em geral, essas classes de distribuições assimétricas possuem um parâmetro, ou um vetor de parâmetros no caso multivariado, que controla a assimetria. Pourahmadi (2007) apresenta alguns modelos de séries temporais multivariados usando distribuições assimétricas na mesma família acima, mas nenhuma estimação foi apresentada e citações indicam dificuldade da mesma. Lambert e Laurent (2002) propõem modelos conjuntos ARMA e generalizações do GARCH cujos erros seguem distribuições assimétricas propostas por Fernández e Steel (1998). No capítulo em questão, nos concentramos na apresentação dos modelos ARFIMA e seus casos particulares, com a estimação no enfoque bayesiano de todas as quantidades desconhecidas, especialmente através de técnicas de Monte Carlo via cadeias de Markov. Propomos também comparação de modelos através do EPD, DIC, AIC, BIC e duas outras alternativas, EAIC e EBIC. Além disso, obtemos a raiz quadrada do erro quadrático médio (REQM) e o erro absoluto médio (EAM). Seguimos com algumas aplicações a dados artificiais e a dados reais - a série anual do preço do trigo e do nível mínimo do rio Nilo.

1.3 Copulas: inferência e seleção

No terceiro capítulo abordamos dependência e assimetria multivariada através do conceito de função cópula. Por ter implicações práticas importantes, como por exemplo na modelagem de retornos financeiros, introduzimos distribuições marginais assimétricas. Nossa contribuição passa pela estimação no enfoque bayesiano de todas as quantidades desconhecidas dos modelos utilizando técnicas de simulação Monte Carlo via cadeias de Markov. Mostramos com estudos de Monte Carlo que o DIC, AIC, BIC, EAIC e EBIC conseguem identificar com alta probabilidade os modelos de cópulas simulados corretamente. Para a geração de dados artificiais de uma função cópula específica e suas marginais usamos técnicas de reamostragem por importância (SIR) (Gelman e Lopes, 2006), a menos que existam outros métodos de simulação direta como os de inversão. Para as distribuições a posteriori usamos o amostrador da fatia dentro do algoritmo de Gibbs. Os exemplos incluem as cópulas Clayton (Clayton, 1978), Frank (Frank, 1979), Gaussiana, Gumbel (Gumbel, 1960; Hougaard, 1986) e Student, e também distribuições marginais normais, exponenciais e normais assimétricas (Fernández e Steel, 1998). A estrutura de dependência das cópulas é definida através da medida não paramétrica τ de Kendall (Kruskal, 1958). Uma aplicação com dados de atuária ilustra o todo o problema.

1.4 Mistura em cópulas e assimetria

No quarto capítulo estendemos alguns resultados de mistura em cópulas encontrados na literatura, como por exemplo misturas contínuas baseadas em transformadas de Laplace dadas em Joe (1997) e Nelsen (1999), ou misturas discretas (Arakelian e Dellaportas, 2006). Propriedades das famílias de cópulas resultantes são mostradas principalmente em termos de medidas de dependência, com a τ de Kendall e a ρ de Spearman (Kruskal, 1958) sendo as duas principais. Para ilustrar o problema, escolhemos as cópulas apresentadas no capítulo 4 e também distribuições marginais assimétricas (Fernández e Steel, 1998). Mostramos que em alguns casos, dependendo da mistura, propriedades das distribuições marginais resultantes podem ser obtidas com relativa facilidade. Novamente, seguimos a abordagem bayesiana. As distribuições a posteriori dos parâmetros dos modelos - das distribuições marginais e das cópulas - são estimadas usando técnicas de simulação Monte Carlo via cadeias de Markov, em particular o amostrador da fatia dentro do algoritmo de Gibbs. Aplicações a dados de atuária e finanças ilustram a mistura contínua em cópulas e a estimação dos parâmetros. Tais aplicações mostram a aplicabilidade e generalidade das classes de cópulas resultantes.

No último capítulo apresentamos uma revisão geral das contribuições, as possíveis extensões e os projetos de trabalho em andamento e futuros em distribuições assimétricas, com destaque as versões e modelos multivariados.

Finalmente, os apêndices trazem revisões matemáticas, cálculos de prioris de Jeffreys, métodos de seleção de modelos, tabelas adicionais, algumas abreviações e resumos de técnicas de Monte Carlo via cadeias de Markov.

Capítulo 2

GAUSSIANA INVERSA GENERALIZADA ESTENDIDA

O objetivo aqui é investigar algumas propriedades de uma flexível classe de distribuição definida em \mathbb{R}^+ que permite assimetrias positivas e negativas dependendo dos valores dos parâmetros. Esta classe será chamada de Gaussiana inversa generalizada estendida (GIGE). Ela foi citada pela primeira vez em uma monografia (Jørgensen, 1982), porém suas propriedades até então não foram investigadas e nenhuma aplicação foi encontrada na literatura. Assim, neste trabalho investigamos algumas de suas propriedades e seguimos com algumas aplicações. Ela inclui distribuições conhecidas como casos particulares como por exemplo a Weibull. A distribuição a priori de Jeffreys no modelo de dados independentes e identicamente distribuídos é apresentada no apêndice B. Propomos um modelo log-linear e outro de volatilidade estocástica como extensões dos existentes na literatura atual. Na primeira aplicação, do modelo log-linear, comparamos nossos resultados com aqueles obtidos por Achcar e Bolfarine (1986) e concluímos que a GIGE ajusta melhor os dados ao custo de um parâmetro a mais. No segundo caso, estendemos o modelo univariado condicional de volatilidade estocástica considerando que os quadrados dos retornos do índice BOVESPA tenham distribuição GIGE. Comparamos os resultados com os obtidos por Lopes e Migon (2002) e uma vez mais concluímos que a GIGE explica melhor os dados e estima melhor as quantidades usualmente desejadas num modelo de volatilidade estocástica padrão, tais como log-volatilidades altamente correlacionadas e assimétricas.

2.1 Propriedades gerais

A distribuição GIGE é obtida como uma transformação poder da distribuição Gaussiana inversa generalizada (GIG) descrita em Jørgensen (1982). Mais precisamente, se W tem uma distribuição GIG com parâmetros λ , δ e γ , denotada aqui por $W \sim GIG(\lambda, \delta, \gamma)$, então sua função de densidade de probabilidade (f.d.p) é dada por

$$p_{gig}(w|\lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\gamma\delta)} w^{\lambda-1} \exp\left\{-\frac{1}{2}(\delta^2 w^{-1} + \gamma^2 w)\right\}, \quad (2.1)$$

para $w > 0$, onde $-\infty < \lambda < \infty$, $\delta > 0$, $\gamma > 0$ e K_ν é função de Bessel modificada do terceiro tipo com índice ν . A dependência dos parâmetros é a seguinte:

$$\begin{aligned} \delta &\geq 0, \quad \gamma > 0, \quad \text{se } \lambda > 0; \\ \delta &> 0, \quad \gamma > 0, \quad \text{se } \lambda = 0; \\ \delta &> 0, \quad \gamma \geq 0, \quad \text{se } \lambda < 0. \end{aligned}$$

Quando $\lambda > 0$ e $\delta = 0$, a f.d.p em (2.1) se reduz a f.d.p de uma distribuição gama, enquanto que f.d.p de uma distribuição inversa gama é obtida quando $\lambda < 0$ e $\gamma = 0$. Nosso teorema a seguir relacionamos a distribuição GIG com a GIGE e calculamos os seus momentos.

Teorema 2.1 *Se W tem distribuição GIG com parâmetros λ , δ e γ , então $Y = W^{1/\theta}$, $\theta > 0$, tem uma distribuição GIGE com parâmetros λ , δ , γ e θ , e função de densidade de probabilidade*

$$p_{gige}(y|\lambda, \delta, \gamma, \theta) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{\theta}{2K_\lambda(\gamma\delta)} y^{\lambda\theta-1} \exp\left\{-\frac{1}{2}(\delta^2 y^{-\theta} + \gamma^2 y^\theta)\right\}, \quad (2.2)$$

para $y > 0$, onde $-\infty < \lambda < \infty$, $\delta > 0$, $\gamma > 0$ e $\theta > 0$. Além disso, os momentos $\mu^p = E(Y^p|\lambda, \delta, \gamma, \theta)$ são dados por

$$\mu^p = \frac{K_{\lambda+\frac{p}{\theta}}(\gamma\delta)}{K_\lambda(\gamma\delta)} \left(\frac{\delta}{\gamma}\right)^{p/\theta}, \quad p \in \mathbb{R}$$

Prova: Sendo $Y = W^{1/\theta}$, é fácil checar que $p_{gige}(y|\lambda, \delta, \gamma, \theta) = p_{gig}(y^\theta|\lambda, \delta, \gamma)y^{\theta-1}$. Jørgensen (1982) mostra todos os momentos, $E(X^k)$ com $X \sim GIG(\lambda, \delta, \gamma)$, para $k \in \mathbb{R}$, logo temos $E(Y^r) = E(X^{r/\theta})$, portanto substituídos $k = r/\theta$ em todos os momentos da distribuição GIG.

Corolário 2.1 *A variância da GIGE é dada por*

$$\text{Var}(Y|\lambda, \delta, \gamma, \theta) = \left(\frac{\delta}{\gamma}\right)^{2/\theta} \left[\frac{K_{\lambda+\frac{2}{\theta}}(\gamma\delta)}{K_\lambda(\gamma\delta)} - \left(\frac{K_{\lambda+\frac{1}{\theta}}(\gamma\delta)}{K_\lambda(\gamma\delta)}\right)^2 \right]$$

Prova: A prova é imediata.

Introduzimos uma parametrização alternativa, que facilita o entendimento e comparação com outras distribuições, principalmente nas aplicações dos modelos log-linear e de volatilidade estocástica. Considerando, $\delta^2/2 = \xi^\theta$ e $\gamma^2/2 = \phi^\theta$, dizemos que $Y \sim GIGE(\lambda, \xi, \phi, \theta)$ quando sua f.d.p é dada por

$$p_{gige}(y|\lambda, \xi, \phi, \theta) = \left(\frac{\phi}{\xi}\right)^{\lambda\theta/2} \frac{\theta}{2K_\lambda(2(\xi\phi)^{\theta/2})} y^{\lambda\theta-1} \exp\left\{-\xi^\theta y^{-\theta} - \phi^\theta y^\theta\right\}, \quad (2.3)$$

para $y > 0$, onde $-\infty < \lambda < \infty$, $\xi > 0$, $\phi > 0$ e $\theta > 0$. Como $\delta = (2\xi^\theta)^{1/2}$ e $\gamma = (2\phi^\theta)^{1/2}$, então temos

$$\mu^p = \frac{K_{\lambda+\frac{p}{\theta}}(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} \left(\frac{\xi}{\phi}\right)^{p/2}, \quad p \in \mathbb{R} \quad (2.4)$$

e

$$\text{Var}(Y|\lambda, \xi, \phi, \theta) = \left(\frac{\xi}{\phi}\right) \left[\frac{K_{\lambda+\frac{2}{\theta}}(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} - \left(\frac{K_{\lambda+\frac{1}{\theta}}(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})}\right)^2 \right] \quad (2.5)$$

Algumas distribuições conhecidas são casos particulares da distribuição GIGE dada por (2.3). De acordo com o apêndice A, para $\lambda > 0$ e $\xi \rightarrow 0$, temos

$$\frac{1}{\xi^{\lambda\theta/2} K_\lambda(2(\xi\phi)^{\theta/2})} \simeq \frac{2\phi^{\lambda\theta/2}}{\Gamma(\lambda)} \quad (2.6)$$

que inserido em (2.3) resulta na seguinte f.d.p:

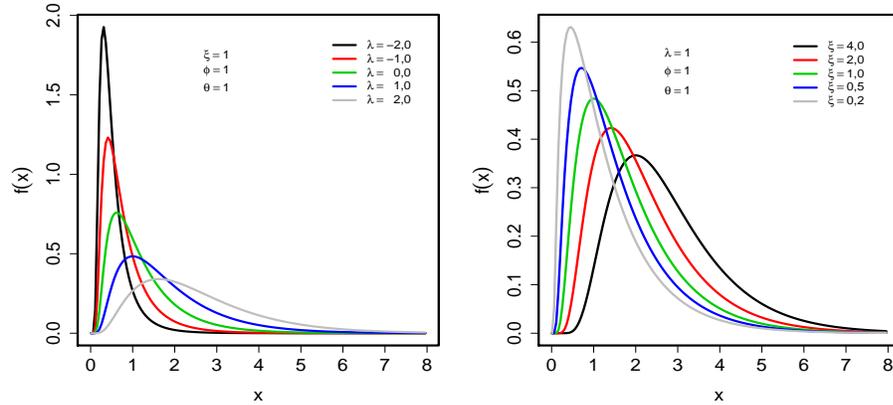
$$g(y|\lambda, \phi, \theta) = \frac{\theta}{\Gamma(\lambda)} \phi^{\lambda\theta} y^{\lambda\theta-1} \exp\{-\phi^\theta y^\theta\}, \quad (2.7)$$

para $y > 0$, que é a f.d.p de uma distribuição gama generalizada (GG) com parâmetros ϕ , λ e θ (Achcar e Bolfarine, 1986), denotadas aqui por $Y \sim GG(\phi, \lambda, \theta)$. Agora, é fácil verificar que as distribuições exponencial - $\mathcal{E}(\phi)$, Weibull - $\mathcal{W}(\phi, \theta)$ e gama - $G(\lambda, \phi)$ são casos particulares. Além disso, tomando $\theta = 2$ e $\lambda = 1/2$ temos a distribuição normal truncada $NT_{\mathbb{R}^+}(0, \phi^{-2})$ em \mathbb{R}^+ com escala ϕ^{-2} . Quando $\lambda < 0$ e $\phi \rightarrow 0$, temos a gama inversa generalizada e seus casos particulares. A prova é similar a dada acima.

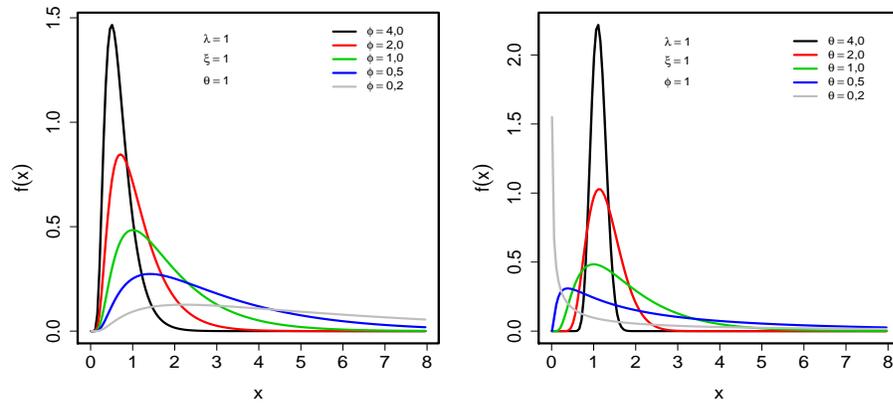
A figura 2.1 mostra o efeito da variação de um parâmetro quando os demais permanecem fixos. Note que quando λ ou ξ aumenta a variância também aumenta, enquanto que para ϕ ou θ crescente a variância diminui. Uma característica importante da distribuição GIGE é permitir assimetrias negativas para alguns valores dos parâmetros. Por exemplo, se $\lambda = 5$, $\xi = 0,5$, $\phi = 1$ e $\theta = 5$, temos um coeficiente de assimetria de -0,17.

2.2 Modelos estatísticos com distribuição GIGE

Como mencionamos na introdução, nosso objetivo é aplicar essa classe de distribuição mais flexível GIGE a dois modelos estatísticos de tal forma que possam ser tratados de maneira mais geral. No primeiro caso, um modelo de análise de sobrevivência, comparamos nossos resultados com aqueles obtidos por Achcar e Bolfarine (1986). No segundo caso, um modelo de volatilidade estocástica aplicado aos retornos ao quadrado do índice BOVESPA, comparamos os resultados com os obtidos por Lopes e Migon (2002). Por outro lado, apresentamos no apêndice B o desenvolvimento da distribuição a priori de Jeffreys do



(a) A densidade GIGE para λ variando. (b) A densidade GIGE para ξ variando.



(c) A densidade GIGE para ϕ variando. (d) A densidade GIGE para θ variando.

Figura 2.1: Densidades GIGE para diferentes valores dos parâmetros.

modelo para dados independentes e identicamente distribuídos da GIGE como alternativa a distribuição a priori própria e vaga. Em alguns casos a distribuição a priori de Jeffreys (Fonseca, 2004) pondera a função de verossimilhança de tal forma o comportamento da distribuição a posteriori permite a estimação coerente dos parâmetros sem a necessidade de informação subjetiva a priori. Assim, apresentamos a priori de Jeffreys para possíveis pesquisas futuras.

2.2.1 Modelo log-linear

Nosso objetivo aqui é aplicar a distribuição GIGE, uma classe mais geral e flexível de distribuições, num modelo de análise de sobrevivência e comparar os resultados com os obtidos por Achcar e Bolfarine (1986). Sejam y_1, \dots, y_n respostas relacionadas aos níveis x_1, \dots, x_n de uma certa covariável. Na modelagem abaixo, x_t é 0 ou 1, representando os grupos placebo e tratamento, respectivamente, enquanto

y_t são os tempos de sobrevivência. Achcar e Bolfarine (1986) modelaram $y_t|\phi_t, \lambda, \theta \sim GG(\lambda, \phi_t, \theta)$, onde $\log(\phi_t^{-1}) = \mu + \beta x_t$. Estendemos os seus trabalhos substituindo a distribuição GG com nossa distribuição GIGE, isto é,

$$y_t|\phi_t, \Theta_y \sim \text{GIGE}(\lambda, \xi, \phi_t, \theta) \quad (2.8)$$

com

$$h_t = \log(\phi_t^{-1}) = \mu + \beta x_t \quad (2.9)$$

e $\Theta_y = (\lambda, \xi, \theta)$. Quando $\xi \rightarrow 0$ recaímos no modelo log-linear de Achcar e Bolfarine (1986). Com isso temos mais flexibilidade na modelagem, podendo assim captar assimetrias negativas e em alguns casos caudas pesadas. A função de verossimilhança para $\Theta = (\Theta_y, \mu, \beta)$ é dada por

$$\begin{aligned} L(\Theta|\mathbf{y}, \mathbf{x}) &= \frac{\theta^n}{2^n \xi^{n\lambda/2}} \exp \left\{ -\frac{n\lambda\theta}{2}(\mu + \beta\bar{x}) - \xi^\theta \sum_{t=1}^n y_t^{-\theta} \right\} \\ &\times \exp \left\{ e^{-\theta\mu} \sum_{t=1}^n (e^{-\beta x_t} y_t)^\theta \right\} \prod_{t=1}^n \frac{y_t^{\theta\lambda-1}}{K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2))} \end{aligned}$$

com $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$ e $n\bar{x} = \sum_{t=1}^n x_t$. Completamos o modelo com a distribuição a priori dada por $\pi(\Theta) = \pi(\lambda)\pi(\xi)\pi(\theta)\pi(\mu)\pi(\beta)$, com $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$, $\xi \sim G(a_\xi, b_\xi)$, $\theta \sim G(a_\theta, b_\theta)$, $\mu \sim N(\mu_\mu, \sigma_\mu^2)$ e $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, onde $N(\mu, \sigma^2)$ representa a distribuição normal com média μ e variância σ^2 . Os hiperparâmetros $\mu_\lambda, \sigma_\lambda^2, a_\xi, b_\xi, a_\theta, b_\theta, \mu_\mu, \sigma_\mu^2, \mu_\beta$ e σ_β^2 são fixados em valores de tal forma que a variância a priori seja considerada grande e que assim garantam que a distribuição a priori seja relativamente vaga.

Inferência na distribuição a posteriori

Note que a distribuição a posteriori de Θ , $\pi(\Theta|\mathbf{y}, \mathbf{x}) \propto L(\Theta|\mathbf{y}, \mathbf{x})\pi(\Theta)$, não tem nenhuma forma conhecida e, como tem sido comum na literatura, a inferência na distribuição a posteriori é feita através de uma simulação Monte Carlo via cadeias de Markov (Gamerman e Lopes, 2006) que, após convergência, retorna valores amostrados da distribuição a posteriori. Sendo $\Theta_\lambda = (\xi, \theta, \mu, \beta)$, com $\Theta_\xi, \Theta_\theta, \Theta_\mu, \Theta_\beta$ definidos de maneira similar, as distribuições condicionais completas das componentes de Θ são dadas por:

$$\begin{aligned}
\pi(\lambda \mid \Theta_\lambda, \mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{n\lambda\theta}{2}(\mu + \beta\bar{x} + \log \xi) - \frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2 + \lambda\theta \sum_{t=1}^n \log y_t \right. \\
&\quad \left. - \sum_{t=1}^n \log K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2)) \right\} \\
\pi(\xi \mid \Theta_\xi, \mathbf{x}, \mathbf{y}) &\propto \exp \left\{ (a_\xi - 1 - n\lambda\theta/2) \log \xi - \xi^\theta \sum_{t=1}^n \log y_t^{-\theta} - b_\xi \xi \right. \\
&\quad \left. - \sum_{t=1}^n \log K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2)) \right\} \\
\pi(\theta \mid \Theta_\theta, \mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{n\lambda\theta}{2}(\mu + \beta\bar{x} + \log \xi) + \lambda\theta \sum_{t=1}^n \log y_t + (n - a_\theta - 1) \log \theta \right. \\
&\quad - b_\theta\theta - \xi^\theta \sum_{t=1}^n y_t^{-\theta} - e^{-\theta\mu} \sum_{t=1}^n (e^{-\beta x_t} y_t)^\theta \\
&\quad \left. - \sum_{t=1}^n \log K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2)) \right\} \\
\pi(\mu \mid \Theta_\mu, \mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{n\lambda\theta\mu}{2} - e^{-\theta\mu} \sum_{t=1}^n (e^{-\beta x_t} y_t)^\theta - \frac{1}{2\sigma_\mu^2}(\mu - \mu_\mu)^2 \right. \\
&\quad \left. - \sum_{t=1}^n \log K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2)) \right\} \\
\pi(\beta \mid \Theta_\beta, \mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{n\lambda\theta\beta\bar{x}}{2} - e^{-\theta\mu} \sum_{t=1}^n (e^{-\beta x_t} y_t)^\theta - \frac{1}{2\sigma_\beta^2}(\beta - \mu_\beta)^2 \right. \\
&\quad \left. - \sum_{t=1}^n \log K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2)) \right\}
\end{aligned}$$

As condicionais completas acima não têm nenhuma forma fechada conhecida, mas satisfazem certas condições para o uso do amostrador da fatia proposto por Neal (2003) dentro do amostrador de Gibbs.

2.2.2 Modelo de volatilidade estocástica

Seja z_t o valor observado de uma série temporal (financeira) no tempo t , para $t = 1, \dots, n$. No modelo de volatilidade estocástica padrão, $z_t \sim N(0, e^{\kappa_t})$, onde κ_t é o logaritmo da volatilidade no tempo t . Portanto, $y_t = z_t^2 \sim G(1/2, e^{-\kappa_t}/2)$. Da relação entre a gama, a gama generalizada e Gaussiana inversa generalizada estendida (seção 2.1), nota-se que $G(1/2, e^{-\kappa_t}/2) \equiv GIGE(1/2, \xi = 0, e^{-\kappa_t}/2, \theta = 1)$. Propomos que λ , ξ e θ sejam livres e modelamos y_t como uma distribuição GIGE não restrita, isto é, $y_t \mid \phi_t, \Theta_y \sim GIGE(\lambda, \xi, \phi_t, \theta)$ com $\Theta_y = (\lambda, \xi, \theta)$ e $\phi_t = e^{-ht}$, como antes. Assim, substituímos a equação (2.9) pelo conhecido e usual modelo autoregressivo de primeira ordem,

$$h_t | h_{t-1}, \Theta_h \sim N(\alpha + \beta(h_{t-1} - \alpha), \tau^2) \quad (2.10)$$

onde $\Theta_h = (\alpha, \beta, \tau^2)$. Como esperado, o modelo de volatilidade padrão é um caso particular onde $\Theta_y = (1/2, 0, 1)$ e $h_t = \kappa_t + \log 2$. Utilizamos a seguinte distribuição a priori para o vetor de parâmetros $\Theta = (\Theta_y, \Theta_h, h_0)$: $\lambda \sim N(\mu_\lambda, \tau_\lambda^2)$, $\xi \sim G(a_\xi, b_\xi)$, $\theta \sim G(a_\theta, b_\theta)$, $\alpha \sim N(\alpha_0, \tau_\alpha^2)$, $\beta \sim N(\beta_0, \tau_\beta^2)$, $\tau^{-2} \sim G(\nu_0/2, \nu_0 s_0^2/2)$ e $h_0 \sim N(m_0, C_0)$. Os hiperparâmetros μ_λ , τ_λ^2 , a_ξ , b_ξ , a_θ , b_θ , α_0 , τ_α^2 , β_0 , τ_β^2 , ν_0 , $\nu_0 s_0^2$, m_0 e C_0 são fixados em valores que garantem que a distribuição a priori seja relativamente vaga.

Inferência na distribuição a posteriori

Uma vez mais, uma simulação Monte Carlo via cadeias de Markov é utilizada para obter uma amostra da distribuição a posteriori dos parâmetros. Devido a linearidade e normalidade de $p(h_t | \Theta_h, h_{t-1})$, é simples amostrar da distribuição condicional completa de α , β , τ^2 e h_0 . Definamos $h = (h_0, \dots, h_n)$ e $y = (y_1, \dots, y_n)$. Logo,

- $\alpha | \beta, \tau^2, h \sim N(m; C)$ para $C = [\tau_\alpha^{-2} + \tau^{-2} n (1 - \beta)^2]^{-1}$ e $m = C[\tau_\alpha^{-2} \alpha_0 + \tau^{-2} (1 - \beta) \sum (h_t - \beta h_{t-1})]$,
- $\beta | \alpha, \tau^2, h \sim N(m; C)$ para $C = [\tau_\beta^{-2} + \tau^{-2} \sum (h_{t-1} - \alpha)^2]^{-1}$ e $m = C[\tau_\beta^{-2} \beta_0 + \tau^{-2} \sum (h_t - \alpha)(h_{t-1} - \alpha)]$,
- $h_0 | h_1, \Theta_h \sim N(m, C)$ para $C = (C_0^{-1} + \beta^2 \tau^{-2})^{-1}$ e $m = C[m_0 C_0^{-1} + \beta(h_1 - \alpha + \beta \alpha) \tau^{-2}]$
- $\tau^{-2} | \alpha, \beta, h \sim G(\nu_1/2, \nu_1 s_1^2/2)$ para $\nu_1 = \nu_0 + n$ e $\nu_1 s_1^2 = \nu_0 s_0^2 + \sum (h_t - \alpha - \beta(h_t - \alpha))^2$,

onde os somatórios são de $t = 1$ até $t = n$. Para λ , ξ e θ , uma maneira natural e simples, mas não necessariamente eficiente, possível de amostrá-los das densidades propostas univariadas $q_\lambda(\cdot)$, $q_\xi(\cdot)$ e $q_\theta(\cdot)$ com algum grau de proximidade, respectivamente a $\pi(\lambda | \xi, \theta, \Theta_h, h, y)$, $\pi(\xi | \lambda, \Theta_h, h, y)$ e $\pi(\theta | \lambda, \xi, \Theta_h, h, y)$. Ao invés, usaremos o amostrador da fatia, como no modelo log-linear apresentado na seção 2.2.1. Para $t = 1, \dots, n$ e $h_{-t} = (h_0, \dots, h_{t-1}, h_{t+1}, \dots, h_n)$, a distribuição condicional completa de h_t é proporcional a $p(y_t | h_t, \Theta_y) p(h_t | h_{t-1}, \Theta_h) p(h_{t+1} | h_t, \Theta_h)$, isto é,

$$\begin{aligned} p(h_t | h_{-t}, \Theta, \mathbf{y}) &\propto e^{-h_t \lambda \theta / 2} \frac{1}{K_\lambda(2(\xi e^{-h_t})^{\theta/2})} \exp \left\{ -y_t^\theta e^{-h_t \theta} \right\} \\ &\times \exp \left\{ -0.5 \tau^2 (1 - \beta^2) \left[h_t^2 - 2h_t \left(\alpha + 2\beta \frac{\bar{h}_t - \alpha}{1 - \beta^2} \right) \right] \right\} \end{aligned}$$

onde $\bar{h}_t = (h_{t+1} + h_{t-1})/2$.

2.3 Aplicações

Nesta seção apresentamos duas aplicações referentes aos modelos log-linear e de volatilidade estocástica com distribuição GIGE. Na primeira aplicação o modelo log-linear da seção 2.2.1 é ajustado para o problema de comparação de dois tratamentos. Na segunda aplicação o modelo de volatilidade estocástica da seção 2.2.2 é ajustado aos quadrados dos retornos diários do índice BOVESPA.

2.3.1 Comparação de dois tratamentos

Tomamos um caso especial do modelo log-linear (2.8) considerando uma variável indicadora (membro do grupo) x tal que $x = 0$ para tratamento 1 e $x = 1$ para tratamento 2. Geralmente, neste problema de duas amostras, nós temos aplicações de modelos de sobrevivência padrões. Alternativamente, propomos esta classe GIGE com mais flexibilidade para ajustar aos dados. Sejam y_1, \dots, y_{n_1} os tempos de sobrevivência de n_1 indivíduos do tratamento 1 e y_{n_1+1}, \dots, y_n os tempos de sobrevivência dos $n - n_1$ indivíduos do tratamento 2. Considere os tempos de sobrevivência dos pacientes de dois grupos de tratamento da tabela 2.1 (de Achcar e Bolfarine, 1986).

Tabela 2.1: Tempos de sobrevivência de pacientes de dois grupos de tratamento.

Tratamento 1	5	10	17	32	32	33	34	36	43	44
	44	48	48	61	64	65	65	66	67	68
	82	85	90	92	92	102	103	106	107	114
	114	116	117	124	139	142	143	151	158	195
Tratamento 2	20,9	32,2	33,2	39,4	40,0	46,8	54,3	57,3	58,0	59,7
	61,1	61,4	66,0	66,3	67,4	68,5	69,9	72,4	73,0	73,2
	88,7	89,3	91,6	93,1	94,2	97,7	101,6	101,9	107,6	108,0
	109,7	110,8	114,1	117,5	119,2	120,3	133,0	133,8	163,3	165,1

Considere $\mu_\lambda = \mu_\mu = \mu_\beta = 0$, $\sigma_\lambda^2 = 100$, $\sigma_\mu^2 = \sigma_\beta^2 = 1000$ e $a_\xi = b_\xi = a_\theta = b_\theta = 0,01$ de tal maneira que garantem que a distribuição a priori seja relativamente vaga. Agora, com todas as condicionais completas bem definidas usamos o amostrador da fatia dentro do amostrador de Gibbs - escrito em FORTRAN 90 com a biblioteca IMSL - com valores iniciais $\lambda^{(0)} = 1$, $\xi^{(0)} = 1$, $\mu^{(0)} = 0$, $\beta^{(0)} = 0$ e $\theta^{(0)} = 1$. Após 100.000 iterações do algoritmo, consideramos convergência em 10.000 iterações e tomamos uma observação para cada 15 amostradas, resultando numa amostra de tamanho

6.000 da distribuição a posteriori. Alguns dos parâmetros ainda apresentam autocorrelações significativas, mas que podem ser consideradas pequenas. A tabela 2.2 e a figura 2.2 trazem um sumário da distribuição a posteriori. O parâmetro $\sigma = 1/\theta$ é utilizado fins de comparação com os resultados obtidos em Achcar e Bolfarine (1986).

Tabela 2.2: Sumário da distribuição a posteriori do modelo log-linear com distribuição GIGE. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetro	Média	D.P	2,5%	Mediana	97,5%
λ	1,445	1,491	0,338	1,006	5,299
ξ	0,335	2,287	0,000	0,000	3,243
μ	4,318	0,778	2,235	4,547	4,988
β	-0,016	0,110	-0,228	-0,019	0,208
θ	2,263	0,955	0,778	2,156	4,487
σ	0,538	0,280	0,223	0,464	1,285

A matriz de correlação da amostra da distribuição a posteriori dos parâmetros é apresentada na tabela 2.3, na qual nota-se que λ , μ e θ são altamente correlacionados e além disso possuem uma dependência não linear como pode ser visto na figura 2.3. A variante do amostrador da fatia, que suprime o passeio aleatório, tem provado ser mais eficiente segundo nossos experimentos do que passos de passeio aleatório padrão quando estão envolvidos dependências não lineares na distribuição a posteriori dos parâmetros.

Tabela 2.3: Correlação estimada na distribuição a posteriori do modelo log-linear com distribuição GIGE.

Parâmetros	λ	ξ	μ	β	θ
λ	1,000	0,233	-0,987	0,244	-0,649
ξ	0,233	1,000	-0,235	0,041	-0,069
μ	-0,987	-0,235	1,000	-0,323	0,697
β	0,2443	0,041	-0,323	1,000	-0,284
θ	-0,649	-0,069	0,697	-0,284	1,000

Os resultados apresentados mostram que o modelo log-linear de Achcar e Bolfarine (1986) com distribuição gama generalizada e nosso modelo com distribuição GIGE são relativamente similares, do ponto

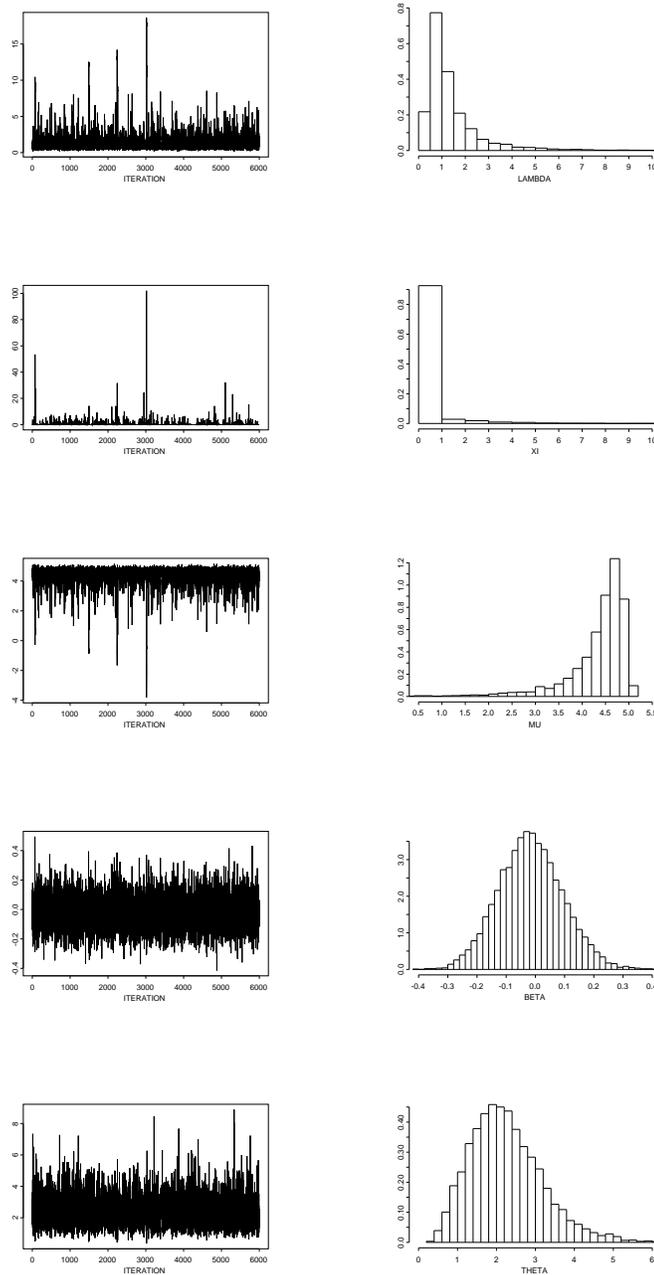
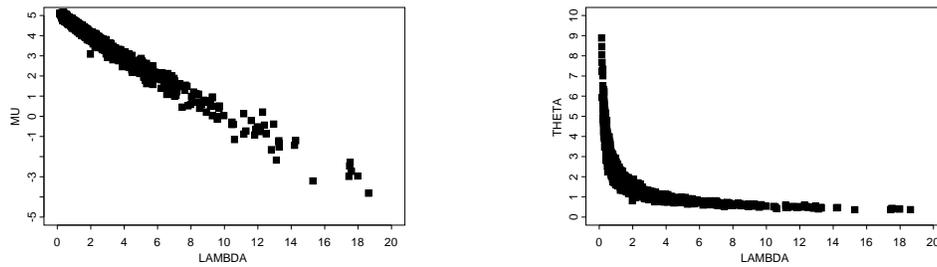
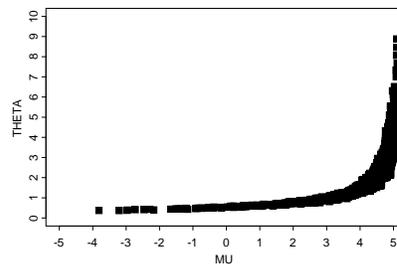


Figura 2.2: Modelo log-linear: caracterização através de histogramas da amostra da distribuição a posteriori de λ (acima), ξ , μ , β e θ (abaixo).

de vista da estimação (veja a tabela 2.4). Neste sentido os resultados encontrados aqui são similares aqueles apresentados em Achcar e Bolfarine (1986) com o aspecto adicional que a incerteza sobre ξ permite maior flexibilidade na modelagem. Entretanto, quando os dois modelos são checados utilizando



(a) Dispersão da amostra da posterior de (λ, μ) . (b) Dispersão da amostra da posterior de (λ, θ) .



(c) Dispersão da amostra da posterior de (μ, θ) .

Figura 2.3: Modelo log-linear: caracterização através da dispersão da distribuição a posterior conjunta de alguns parâmetros.

os critérios apresentados no apêndice D, o AIC e o BIC favorecem ao caso particular enquanto o DIC aponta como melhor ajuste nosso modelo estendido (veja tabela 2.5).

Erro quadrático médio

Na tabela 2.6 temos as médias dos parâmetros na distribuição a posteriori do modelo log-linear GIGE. Baseados nessas médias, calculamos o erro quadrático médio (EQM), $EQM = \sum_{i=1}^n r_i^2/n$, onde $r_i = (y_i - \mu_i)/\sigma_i$ para as médias μ_i e variâncias σ_i^2 como funções da média a posteriori dos parâmetros. A primeira linha da tabela corresponde a moda da distribuição a posteriori dada por Achcar e Bolfarine (1986) e o correspondente EQM. A segunda linha da tabela apresenta a média da distribuição a posteriori dada na tabela 2.2. As últimas três linhas são as médias a posteriori para o algoritmo de reamostragem por importância (SIR). Em SIR1, os parâmetros são gerados de propostas uniforme com tamanho unitário centrado na moda a posteriori encontrado em Achcar e Bolfarine (1986) que também são considerados a priori. Em SIR2, os parâmetros são gerados de propostas uniformes com tamanho unitário centrado

Tabela 2.4: Sumário da distribuição a posteriori do modelo log-linear com distribuição gama generalizada. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetro	Média	D.P	2,5%	Mediana	97,5%
λ	1,450	1,550	0,340	1,012	5,142
μ	4,319	0,788	2,334	4,538	4,996
β	-0,019	0,112	-0,231	-0,022	0,215
θ	2,268	0,970	0,796	2,139	4,558
σ	0,537	0,281	0,219	0,467	1,256

Tabela 2.5: Comparação de modelos: \mathcal{M}_1 : GIGE e \mathcal{M}_2 : GG. O fator de Bayes é $BF(\mathcal{M}_1, \mathcal{M}_2)=0,667$ e o pseudo fator de Bayes é $PBF(\mathcal{M}_1, \mathcal{M}_2)=0,984$.

Critério	GIGE (\mathcal{M}_1)	GG (\mathcal{M}_2)
AIC	835,74	834,06
BIC	847,65	843,59
DIC	4109,1	4119,7

Tabela 2.6: Erro quadrático médio para o modelo log-linear GIGE

	λ	ξ	μ	β	θ	EQM
AB	1,0000	0,0001	4,578	-0,03400	2,500	1,209
SLM	1,4450	0,3350	4,318	-0,01600	2,263	1,674
SIR1	0,7876	0,2500	4,682	-0,03457	2,506	0,974
SIR2	1,1210	0,4159	4,482	-0,01822	2,045	0,992
SIR3	1,1410	0,1147	4,470	-0,03715	2,004	0,991

na média a posteriori dado na tabela 2.2. SIR3 é como SIR2, mas as prioris são as mesmas apresentadas na seção anterior. As amostras são de tamanho 100.000 para ambos os passos de amostragem e reamostragem da distribuição a posteriori. As médias da distribuição a posteriori são próximas.

2.3.2 Índice BOVESPA

Para efeitos de comparação, vamos utilizar o índice BOVESPA, como apresentado em Lopes e Migon (2002). A série corresponde aos retornos, para quase 1.500 dias consecutivos, medidos diariamente pela bolsa de valores de São Paulo (veja figura 2.4). Para comparar, ajustamos um modelo AR(1) padrão para o logaritmo das volatilidades, isto é, $y_t|h_t \sim N(0, e^{h_t})$ e $h_t|h_{t-1}, \alpha, \beta, \tau^2 \sim N(\alpha + \beta(h_{t-1} - \alpha), \tau^2)$, onde $t = 1, \dots, n$ com $n = 1481$. Escolhemos distribuições a priori relativamente vagas para α , β e τ^2 . Mais especificamente, $\alpha \sim N(0, 6546; 10)$, $\beta \sim N(0, 9797; 10)$, $\tau^{-2} \sim G(a, b)$ com $a/b = 0,0167^{-1}$ e $a/b^2 = 10$, e $h_0 \sim N(0; 10)$. Os valores iniciais para α , β , τ^{-2} e h_0 foram $-0,6546$, $0,9797$, $0,0244$ e 0 , respectivamente. Depois 10.000 iterações do amostrador de Gibbs, tomamos uma observação a cada 20 das 100.000 seguintes amostradas, totalizando uma amostra de tamanho 5.000 da distribuição a posteriori. Um sumário da distribuição a posteriori para α , β e τ^{-2} são apresentados na tabela 2.7.

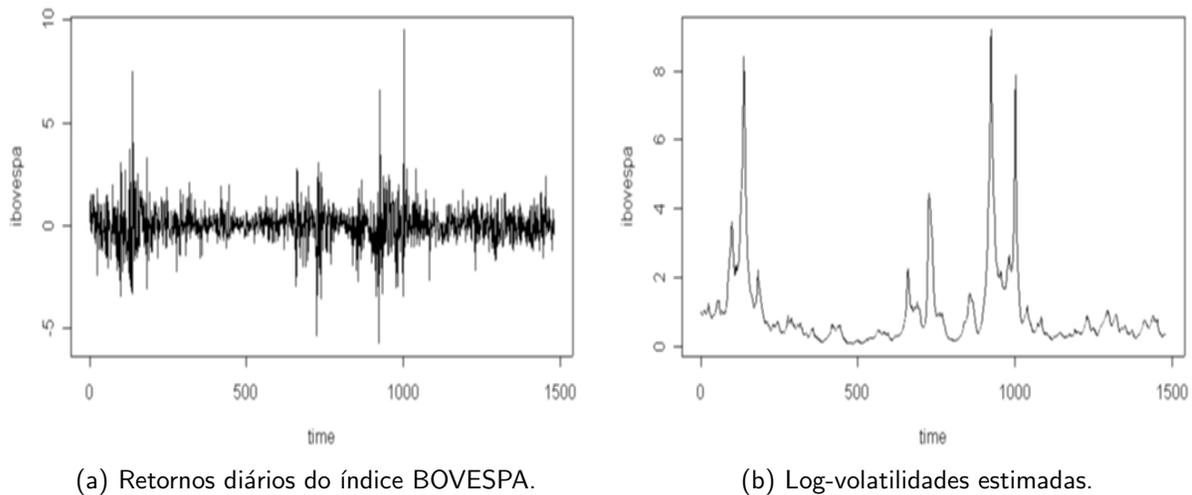


Figura 2.4: Esquerda: Retornos diários do índice BOVESPA (da bolsa de valores de $S\tilde{i}l_{\frac{1}{2}}$ Paulo). Direita: log-volatilidades estimadas pela média a posteriori.

Tabela 2.7: Modelo de volatilidade estocástica: sumário da distribuição a posteriori. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetro	Média	D.P	2,5%	Mediana	97,5%
α	-0,591	0,391	-1,144	-0,586	-0,009
β	0,976	0,007	0,961	0,976	0,988
τ^2	0,055	0,010	0,036	0,053	0,077

Aplicamos agora o modelo de volatilidade estocástica com distribuição GIGE ao quadrado dos índice de retornos definidos acima. Consideramos distribuição a priori relativamente vaga: $\alpha \sim N(0; 10^4)$, $\beta \sim N(0; 10^4)$, $\tau^{-2} \sim G(0, 01; 0, 01)$, e $h_0 \sim N(0; 10^4)$. Além disso, temos $\lambda \sim N(0; 100)$, $\xi \sim G(0, 01; 0, 01)$, e $\theta \sim G(0, 01; 0, 01)$. Os valores iniciais para α , β , τ^2 e h_0 foram 0, 0, 97, 1 e 0, respectivamente. Depois gerar 4 cadeias paralelas do amostrador de Gibbs com 20.000 iterações, tomamos uma observação a cada 10 amostradas das seguintes 10.000 geradas, totalizando uma amostra de tamanho 4.000 da distribuição a posteriori. Um sumário da distribuição a posteriori para α , β e τ^2 são apresentados na tabela 2.8 e na figura 2.5.

Tabela 2.8: Modelo de volatilidade estocástica: sumário da distribuição a posteriori. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetro	Média	D.P	2,5%	Mediana	97,5%
λ	0,439	0,108	0,282	0,419	0,720
$10^6 \xi$	0,149	0,351	0,0014	0,012	1,18
θ	1,140	0,203	0,758	1,136	1,550
α	0,206	0,356	-0,530	0,238	0,778
β	0,967	0,012	0,942	0,968	0,988
τ^2	0,070	0,022	0,031	0,068	0,119

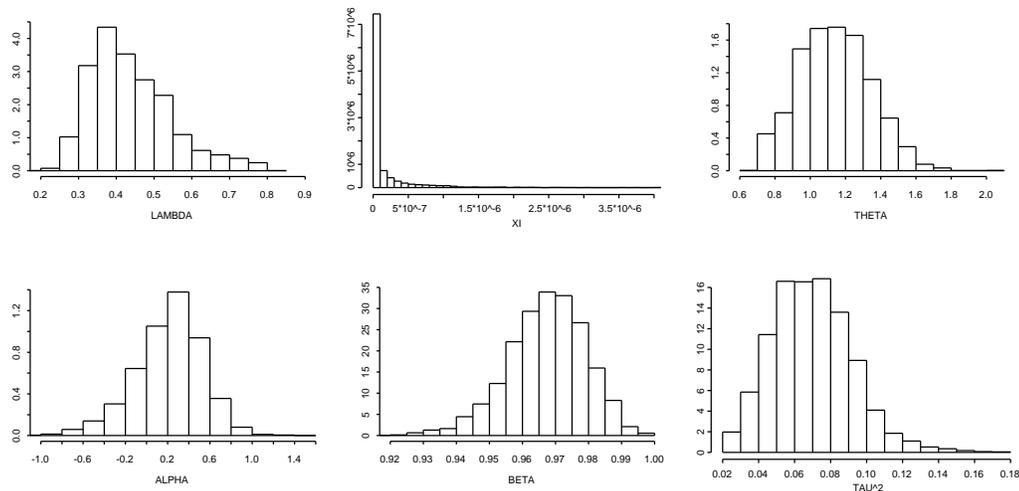


Figura 2.5: Modelo de volatilidade estocástica GIGE: caracterização através de histogramas da amostra da distribuição a posteriori de λ , ξ e θ (acima), α , β e τ^2 (abaixo).

A figura 2.6 exibe a média a posteriori de h_t , isto é, $E(h_t|y_1, \dots, y_n)$, para $t = 1, \dots, n$ para cada uma das 4 cadeias, onde percebe-se que não existe praticamente diferença entre elas. Além disso apresenta o mesmo comportamento como em Lopes e Migon (2002), isto é, volatilidades altas perto dos tempos 100, 700 e 1000. Visto que a média a posteriori de λ , ξ e θ não são significativamente diferentes de 0,5; 0 e 1,0, respectivamente, podemos argumentar que chegamos ao modelo usual de volatilidade estocástica. Quando os dois modelos são checados utilizando os critérios apresentados no apêndice D, todos os critérios apontam como melhor modelo o caso particular (veja tabela 2.9). Entretanto mais investigação se faz necessário devido as dependências existentes entre os parâmetros e como isso afeta esses critérios de seleção de modelos.

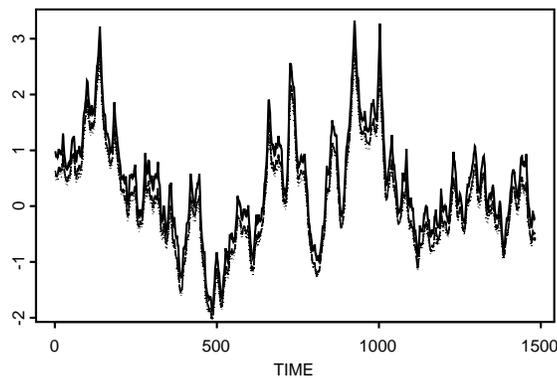


Figura 2.6: Modelo de volatilidade estocástica GIGE: $E(h_t|y_1, \dots, y_n)$, para $t = 1, \dots, n$ para as 4 cadeias geradas.

Tabela 2.9: Comparação de modelo: \mathcal{M}_1 : GIGE e \mathcal{M}_2 : GG. O fator de Bayes é $BF(\mathcal{M}_1, \mathcal{M}_2)=0,071$ e o pseudo fator de Bayes é $PBF(\mathcal{M}_1, \mathcal{M}_2)=0,069$.

Critério	GIGE (\mathcal{M}_1)	GG (\mathcal{M}_2)
AIC	4579,4	834,06
BIC	12461,2	11933,4
DIC	5857,7	5351,8

2.4 Discussão final

Uma classe de distribuição foi introduzida através da distribuição Gaussiana inversa generalizada. Esta classe é bastante flexível e inclui muitos casos particulares. Uma questão crucial na modelagem a ser melhor investigada é como a dependência dos parâmetros afeta ou é afetada pelos algoritmos usados. O modelo log-linear apresentado aqui reproduz os resultados encontrados em Achcar e Bolfarine (1986), porém mais investigação se faz necessário. Além disso, temos outras possibilidades para o modelo log-linear no parâmetro ξ , acomodando possíveis caudas pesadas. Igualmente, o modelo de volatilidade estocástica combinado com a GIGE teve boa performance e exibiu resultados similares aqueles encontrados em Lopes e Migon (2002). Para as aplicações, alguns critérios de seleção de modelos foram usados, dentre os quais destacamos o AIC, BIC e DIC, além do fator de Bayes e do pseudo fator de Bayes.

Capítulo 3

MODELOS ARFIMA ASSIMÉTRICOS

Tradicionalmente, modelos de longa dependência descritos através dos processos auto-regressivos fracionalmente integrados médias-móveis (ARFIMA) têm sido modelados com erros aleatórios independentes e identicamente distribuídos provenientes de alguma distribuição simétrica em torno de 0 (Hosking, 1981; Granger, 1980; Beran, 1994; Silva, 2003), sendo a principal delas a distribuição normal. Fatos estilizados em finanças garantem que retornos financeiros são ligeiramente assimétricos, possuem caudas pesadas e que em alguns casos as volatilidades apresentam longa dependência. Por isso, neste capítulo propomos a modelagem dos processos ARFIMA permitindo que a distribuição dos erros possam seguir uma distribuição assimétrica e também ter caudas pesadas. Azzalini (1985) propõe um maneira de obter classes de distribuições assimétricas, onde a função de densidade de probabilidade (f.d.p) da classe resultante é o produto de uma f.d.p simétrica em torno de 0 por uma função de distribuição acumulada (f.d.a) cuja derivada também é simétrica em torno de 0, sendo que o argumento desta f.d.a é multiplicada por um parâmetro que controla a assimetria da distribuição resultante. Aqui nos concentraremos em 2 casos: a normal e a t-Student assimétricas como as distribuições dos erros para o processo ARFIMA. Pourahmadi (2007) apresenta alguns modelos de séries temporais multivariados usando distribuições assimétricas (Azzalini e Dalla Valle, 1996; Azzalini e Capitanio, 1999; Azzalini e Capitanio, 2003), mas nenhuma estimação foi apresentada e citações indicam dificuldades de estimação dos parâmetros destes modelos no enfoque frequentista. Lambert e Laurent (2002) propõem modelos auto-regressivos médias-móveis (ARMA) e generalizados auto-regressivos condicionalmente heterocedásticos (GARCH) cujos erros seguem distribuições assimétricas propostas por Fernández e Steel (1998). Nos concentramos na apresentação dos modelos ARFIMA e seus casos particulares, com a estimação no enfoque bayesiano de todas as quantidades desconhecidas, inclusive das variáveis auxiliares. Alguns estudos de Monte Carlo ajudam a validar casos particulares da distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos. Inferimos nos modelos usando amostras das respectivas distribuições a posteriori obtidas através de métodos de Monte Carlo via cadeias de Markov (Geman e Lopes, 2006). Em especial, implementamos o amostrador de Gibbs (Geman e Geman, 1984; Gelfand e Smith, 1990) com passos de Metropolis-Hastings (Metropolis, Rosenbluth, Teller, e Teller, 1953; Hastings, 1970) para algumas condicionais completas e o amostrador da fatia (Neal, 2003) para outras. Propomos também comparações de modelos através dos critérios de informação do desvio (DIC) (Spiegelhalter et al., 2002), de Akaike (AIC) (Akaike, 1974), bayesiano (BIC) (Schwarz, 1978), e duas outras alternativas, EBIC e

EBIC (Brooks, 2002). Além disso, calculamos o desvio preditivo esperado (EPD) (Gelfand e Ghosh, 1998), a raiz quadrada do erro quadrático médio (REQM) e o erro absoluto médio (EAM).

3.1 Propriedades gerais

O seguinte lema, apresentado a mais de duas décadas atrás, é o ponto de partida deste capítulo na modelagem dos processos ARFIMA com erros tendo distribuições assimétricas.

Lema 3.1 *Seja f uma f.d.p simétrica em torno de 0, e G uma f.d.a absolutamente contínua tal que a derivada de G , $g \equiv G'$, é simétrica em torno de 0. Então*

$$2f(x)G(\lambda x), \quad x \in \mathbb{R},$$

é uma f.d.p para qualquer número real λ .

Prova: (Azzalini, 1985). Seja Y e X variáveis aleatórias com f.d.p f e g , respectivamente. Note que Y e X são simétricas em torno de 0, então λX também é simétrica em torno de 0. Assim, as funções características (veja James (1996, p. 223-233) $\Psi_Y(t)$ e $\Psi_{-\lambda X}(t)$ assumem valores reais para todo t real. Como Y e λX são independentes, então a função característica de $Y - \lambda X$ é o produto de duas funções características que assumem valores reais para todo t , portanto $\Psi_{Y-\lambda X}(t)$ é real para todo t . Logo $Y - \lambda X$ é simétrica em torno de 0. Assim,

$$\frac{1}{2} = P(Y - \lambda X < 0) = E_X(P(Y < \lambda x | X = x)) = \int_{-\infty}^{\infty} f(x)G(\lambda x)dx.$$

A partir do lema 3.1 pode ser proposta uma infinidade de distribuições assimétricas, mas a escolha mais usual é trabalhar com f.d.p e f.d.a que sejam mais conhecidas. A primeira escolha é trabalhar com f e G sendo a f.d.p e a f.d.a de uma variável aleatória normal padrão, respectivamente. A seguir definiremos as distribuições normal e t-Student assimétricas assim como algumas propriedades, cujos resultados e provas podem ser encontradas em Genton (2004) e suas referências.

Definição 3.1 *Se uma variável aleatória X tem f.d.p dada por*

$$\phi(x|\lambda, \mu, \sigma) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}$$

onde ϕ e Φ são a f.d.p e a f.d.a de uma variável aleatória normal padrão, respectivamente. Então dizemos que X tem distribuição normal assimétrica (NA) com parâmetros λ (assimetria), μ (posição) e σ (escala); para simplificar denotaremos por $X \sim NA(\lambda, \mu, \sigma)$ e no caso $\mu = 0$ e $\sigma = 1$ por $X \sim NA(\lambda)$.

A figura 3.1 mostra o comportamento da distribuição normal assimétrica, $NA(\lambda)$, para diferentes valores do parâmetro de assimetria λ com posição $\mu = 0$ e escala $\sigma = 1$ fixos. Observe que a medida que o parâmetro λ cresce o mesmo ocorre com a assimetria. Para valores negativos de λ a assimetria fica a esquerda.

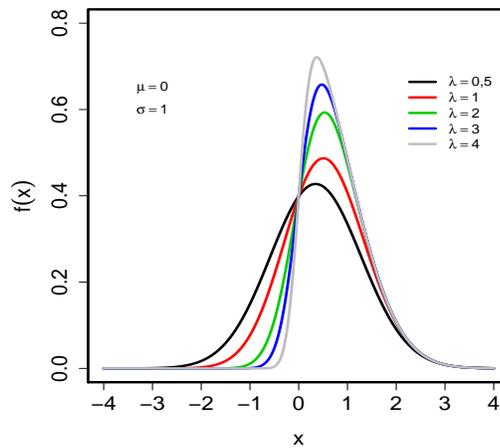


Figura 3.1: Densidades normais assimétricas.

Algumas propriedades da classe de distribuições normais assimétricas são descritas abaixo.

1. A densidade $NA(0)$ é a densidade $N(0, 1)$.
2. Quando $\lambda \rightarrow \infty$, $\phi(x|\lambda)$ tende para uma densidade normal truncada, $NT_{\mathbb{R}^+}(0, 1)$.
3. Se $X \sim NA(\lambda)$, então $X^2 \sim \chi_1^2$. Observe que a distribuição de X^2 não depende de λ . Assim, os momentos pares de X são iguais aos momentos pares de uma distribuição normal padrão.
4. Se $X \sim NA(\lambda)$, então $\mu + \sigma X \sim NA(\lambda, \mu, \sigma)$.
5. Se X é uma variável aleatória $N(0, 1)$, então

$$E(\Phi(hX + k)) = \Phi(k/\sqrt{1 + h^2})$$

para qualquer real h, k . Assim, a função geradora de momentos da distribuição normal assimétrica, $\text{NA}(\lambda, \mu, \sigma)$, é dada por

$$\Psi(t) = 2 \exp\left(\frac{(t - \mu)^2}{2\sigma^2}\right) \Phi\left(\frac{\lambda(t - \mu)}{\sigma\sqrt{1 + \lambda^2}}\right).$$

Este resultado é importante para o cálculo dos momentos ímpares de X .

6. Se $X \sim \text{NA}(\lambda, \mu, \sigma)$, então

$$\begin{aligned} E(X \mid \lambda, \mu, \sigma) &= \mu + \frac{\lambda\sigma}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}} \\ \text{Var}(X \mid \lambda, \mu, \sigma) &= \sigma^2 \left[1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2}\right] \\ E(X^3 \mid \lambda, \mu, \sigma) &= \mu^3 + 3\mu^3\sigma\delta\sqrt{\frac{2}{\pi}} + 3\mu\sigma^2 + 3\sigma^3\delta\sqrt{\frac{2}{\pi}} - \sigma^3\delta^3\sqrt{\frac{2}{\pi}} \\ \gamma &= \delta^3 \left[\frac{4}{\pi} - 1\right] \left[1 - \frac{2\delta^2}{\pi}\right]^{-3/2} \sqrt{\frac{2}{\pi}} \end{aligned} \tag{3.1}$$

onde γ é o coeficiente de assimetria com $-0,99527 < \gamma < 0,99527$ sendo $\delta = \lambda(1 + \lambda^2)^{-1/2}$.

7. Sejam U e V variáveis aleatórias independentes tal que $U \sim \text{N}(0, 1)$ e $V \sim \text{N}(0, 1)$. Então

$$X = \frac{\lambda}{\sqrt{1 + \lambda^2}}|U| + \frac{1}{\sqrt{1 + \lambda^2}}V \sim \text{NA}(\lambda).$$

Este resultado foi mostrado por Henze (1986). No mesmo artigo ele sugere um algoritmo, usado constantemente neste trabalho, para gerar valores da distribuição $\text{NA}(\lambda, \mu, \sigma)$ e conseqüentemente da distribuição t-Student assimétrica.

A seguir definiremos a distribuição t-Student assimétrica para justicar o que poderia ser a simples extensão do caso NA: $t_\nu(x|\lambda) = 2t_\nu(x)\mathbb{T}_\nu(\lambda x)$ para $\mu = 0$ e $\sigma = 1$. Para isso precisaremos dos lemas 3.2 e 3.3, cujas provas podem ser encontradas em Azzalini e Capitanio (2003). Como pode ser visto, a distribuição t-Student assimétrica surgirá como uma mistura no inverso da escala de uma distribuição NA com uma distribuição gama.

Lema 3.2 Se $V \sim \text{Gama}(\alpha, \beta)$, então para qualquer $a, b \in \mathbb{R}$

$$E(\Phi(a\sqrt{V} + b)) = P(T \leq a\sqrt{\alpha/\beta})$$

onde T tem distribuição t -Student não centrada com 2α graus de liberdade e parâmetro de posição $-b$.

Lema 3.3 Se $X|V = v \sim \text{NA}(\lambda, 0, v^{-1})$ e $V \sim \text{Gama}(\nu/2, \nu/2)$, então X tem f.d.p dada por

$$f_X(x) = 2t_\nu(x)\mathbb{T}_{\nu+1}\left(\lambda x \sqrt{\frac{\nu+1}{\nu+x^2}}\right)$$

onde t_ν e \mathbb{T}_ν denotam a f.d.p e a f.d.a de uma variável aleatória t -Student com ν graus de liberdade, respectivamente.

Definição 3.2 Se uma variável aleatória X tem f.d.p dada por

$$t_\nu(x|\lambda) = \frac{2}{\sigma} t_\nu(\omega)\mathbb{T}_{\nu+1}\left(\lambda\omega\sqrt{\frac{\nu+1}{\nu+\omega^2}}\right) \quad \text{com} \quad \omega = (x - \mu)/\sigma, \quad x \in \mathbb{R}$$

onde t_ν e $\mathbb{T}_{\nu+1}$ são a f.d.p e a f.d.a de variáveis aleatórias t -Student com ν e $\nu + 1$ graus de liberdade, respectivamente. Então dizemos que X tem distribuição t -Student assimétrica (TA) com parâmetros λ (assimetria), ν (graus de liberdade), μ (posição) e σ (escala); para simplificar denotaremos por $X \sim \text{TA}_\nu(\lambda, \mu, \sigma)$ e no caso $\mu = 0$ e $\sigma = 1$ por $X \sim \text{TA}_\nu(\lambda)$.

A figura 3.2 mostra o comportamento da distribuição t -Student assimétrica, $\text{TA}_\nu(\lambda)$, para diferentes valores do parâmetro de assimetria λ com posição $\mu = 0$, escala $\sigma = 1$ e graus de liberdade $\nu = 3$ fixos. Como esperado, a medida que o parâmetro λ cresce o mesmo ocorre com a assimetria.

Na distribuição t -Student assimétrica da definição 3.2, valem propriedades análogas a 1, 2 e 4 no caso da normal assimétrica. Essas propriedades podem ser facilmente provadas e algumas delas são descritas abaixo.

1. A densidade $\text{TA}_\nu(0)$ é a densidade $T_\nu(0, 1)$.
2. Quando $\lambda \rightarrow \infty$, $t_\nu(x|\lambda)$ tende para uma densidade t -Student truncada, $\text{TT}_{\mathbb{R}^+}(0, 1)$.

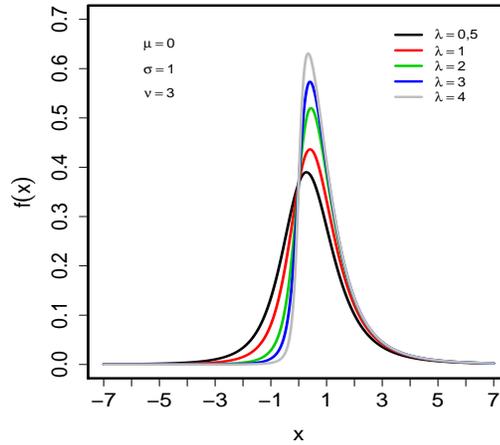


Figura 3.2: Densidades t-Student assimétricas.

3. Se $X \sim \text{TA}_\nu(\lambda)$, então $\mu + \sigma X \sim \text{TA}_\nu(\lambda, \mu, \sigma)$.

4. Se $X \sim \text{TA}_\nu(\lambda, \mu, \sigma)$, então

$$E(X | \lambda, \mu, \sigma) = \mu + \frac{\lambda\sigma}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}, \quad \nu > 1 \quad (3.2)$$

$$\text{Var}(X | \lambda, \mu, \sigma) = \sigma^2 \left[\frac{\nu}{\nu-2} - \frac{\nu}{\pi} \frac{\lambda^2}{1+\lambda^2} \frac{\Gamma^2(\frac{\nu-1}{2})}{\Gamma^2(\frac{\nu}{2})} \right], \quad \nu > 2$$

$$\gamma = \eta \left[\frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\nu}{\nu-2} + 2\eta^2 \right] \left[\frac{\nu}{\nu-2} - \eta^2 \right]^{-3/2}, \quad \nu > 3$$

onde γ é o coeficiente de assimetria sendo $\eta = (\nu/\pi)^{1/2} \Gamma((\nu-1)/2) \Gamma^{-1}(\nu/2) \delta$ e $\delta = \lambda(1+\lambda^2)^{-1/2}$.

Apesar de muitas das propriedades e definições acima já terem aparecido na literatura, julgamos necessário a apresentação neste trabalho para uma melhor compreensão das distribuições assimétricas (Azzalini, 1985; Azzalini e Capitanio, 2003) e da nossa proposta a seguir para modelos de longa dependência. Entretanto, ressaltamos que em nossos estudos poderíamos incluir outras distribuições assimétricas. Dentre algumas possibilidades destacamos f.d.p normal com f.d.a t-Student ou f.d.p t-Student com f.d.a normal padrão. Porém, retemos nossa atenção apenas nas duas subclasses definidas anteriormente: a normal assimétrica da definição 3.1 e a t-Student assimétrica da definição 3.2. Na modelagem dos processos de longa dependência descritos a seguir, as distribuições dos erros terão ou distribuição normal assimétrica ou t-Student assimétrica com devidas restrições na média e possivelmente na variância para satisfazer certas hipóteses do modelo.

3.2 Modelos ARFIMA(p, d, q)

Uma série temporal $\{y_t\}$ gerada por um processo auto-regressivo fracionalmente integrado médias-móveis, ARFIMA(p, d, q) com uma medida de posição α é descrita por

$$\Phi(L)(1-L)^d(y_t - \alpha) = \Theta(L)\epsilon_t \quad (3.3)$$

onde $\Phi(L) = 1 - \phi_1L - \phi_2L^2 - \phi_3L^3 - \dots - \phi_pL^p$ e $\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \theta_3L^3 + \dots + \theta_qL^q$ são polinômios em L , onde L é o operador de defasagem ($L^r y_t = y_{t-r}$, para $r = 1, 2, \dots$), p e q são números inteiros e d um número real entre $(-1/2, 1/2)$. Além disso as raízes do polinômio $\Phi(z) = 0$ e $\Theta(z) = 0$ devem estar todas fora do círculo unitário para garantir que o processo seja estacionário e inversível. Os erros ϵ_t são variáveis aleatórias independentes e identicamente distribuídas com média zero e variância constante. O operador de diferença fracionária d , $(1-L)^d$ em (3.3) é definido pela expansão binomial,

$$(1-L)^d = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^j (-L)^j = \sum_{j=1}^{\infty} \pi_j(d) L^j \quad (3.4)$$

onde

$$\pi_j(d) = \frac{\Gamma(j-d)}{\Gamma(j+d)\Gamma(-d)}$$

Geralmente, os erros ϵ_t têm distribuição normal com média 0 e a variância constante σ^2 , ou alguma outra distribuição simétrica como por exemplo a t-Student. Entretanto, Lambert e Laurent (2002) propõem para o modelo ARMA(p, q), um caso particular do modelo (3.3), erros assimétrico dados por Fernández e Steel (1998), com estimação pelo enfoque frequentista. Contudo, modelos de longa dependência, ARFIMA(p, d, q), com erros assimétricos (Azzalini, 1985; Azzalini e Capitanio, 2003) ainda não aparecem na literatura nem tampouco estimação bayesiana nesta classe de modelos ou em alguns de seus casos particulares, como por exemplo o modelo ARMA(p, q). Assim, propomos que os erros ϵ_t tenham distribuições NA ou TA dadas pelas definições 3.1 e 3.2, respectivamente. Logo, os parâmetros dos modelos a serem estimados são $\lambda, \alpha, \sigma, \Phi, \Theta$, onde $\Phi = (\phi_1, \phi_2, \dots, \phi_p)'$ e $\Theta = (\theta_1, \theta_2, \dots, \theta_q)'$, mais ν no caso TA. Porém, na nossa modelagem os parâmetros de posição μ das distribuições NA e TA são restritos a funções dos demais parâmetros, λ, σ e ν , de tal forma que os erros ϵ_t tenham média 0, isto é, tomamos $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}$ na distribuição NA de acordo com a equação (3.1) e $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}\Gamma((\nu-1)/2)\Gamma^{-1}(\nu/2)$ na distribuição TA de acordo com a equação (3.2). Além disso, a soma infinita em (3.4) é truncada nas m primeiras parcelas, resultando em aproximações das funções de verossimilhança dos casos NA e TA, respectivamente. Tal aproximação é justificada em Silva (2003) para o caso NA(0), com resultados que garantem que $m = 50$ produz bons resultados na estimação dos parâmetros.

Assim, a função de verossimilhança de um processo, $\{y_t\}$, ARFIMA(p, d, q) com erros tendo distribuição NA, baseada numa amostra de tamanho T , condicional aos valores iniciais latentes $E_0 = (\epsilon_0, \dots, \epsilon_{1-q})$ e $Y_0 = (y_0, \dots, y_{1-\max(p,m)})$, onde m é a ordem de aproximação da expansão (3.4), é dada por

$$L(\lambda, \alpha, \sigma, \Phi, d, \Theta, E_0, Y_0 | y_1, \dots, y_T) = \prod_{i=1}^T \frac{2}{\sigma} \phi\left(\frac{y_t - \mu_t}{\sigma}\right) \Phi\left(\lambda \frac{(y_t - \mu_t)}{\sigma}\right) \quad (3.5)$$

onde $\mu_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varphi_1(d) y_{t-1} + \dots + \varphi_m(d) y_{t-m} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} - \mu$, sendo $\mu = -\lambda \sigma (1 + \lambda^2)^{-1/2} (2/\pi)^{1/2}$. As variáveis auxiliares E_0 e Y_0 são desconhecidas e não observáveis, assim nosso vetor de parâmetros a ser estimado é $\Omega = (\lambda, \alpha, \sigma, \Phi, d, \Theta, E_0, Y_0)$. Logo, em nossa abordagem bayesiana, estimaremos todas essas quantidades, mas para isso precisamos definir a distribuição a priori de Ω . Em primeiro lugar, consideramos $\lambda, \alpha, \sigma, \Phi, d, \Theta, E_0$ e Y_0 independentes a priori. Para λ temos a distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos em (3.7), como veremos a seguir. Enquanto que $\alpha \sim N(a_\alpha, b_\alpha)$, $d \sim U(-1/2, 1/2)$ e $\sigma \sim IG(a_\sigma, b_\sigma)$ com valores para $a_\alpha, b_\alpha, a_\sigma$ e b_σ que garantem que a distribuição a priori seja relativamente vaga. Além disso, Φ e Θ têm distribuições uniformes nas respectivas regiões de estacionariedade e invertibilidade; $\epsilon_k \sim NA(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - q$ e $y_k \sim NA(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - \max(p, m)$. Combinando a distribuição a priori de Ω , que neste caso é própria, com a função de verossimilhança em (3.5), obtemos uma distribuição a posteriori própria com forma fechada desconhecida.

Para o processo ARFIMA(p, d, q) com erros tendo distribuição TA, a função de verossimilhança baseada numa amostra de tamanho T , condicional aos valores iniciais latentes $E_0 = (\epsilon_0, \dots, \epsilon_{1-q})$ e $Y_0 = (y_0, \dots, y_{1-\max(p,m)})$, onde m é a ordem de aproximação da expansão (3.4), é dada por

$$L(\lambda, \alpha, \sigma, \Phi, d, \Theta, E_0, Y_0 | y_1, \dots, y_T) = \prod_{i=1}^T \frac{2}{\sigma} t_\nu\left(\frac{y_t - \mu_t}{\sigma}\right) T_{\nu+1}\left(\lambda \frac{(y_t - \mu_t)}{\sigma}\right) \quad (3.6)$$

onde $\mu_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varphi_1(d) y_{t-1} + \dots + \varphi_m(d) y_{t-m} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} - \mu$, sendo $\mu = -\lambda \sigma (1 + \lambda^2)^{-1/2} (2/\pi)^{1/2} \Gamma((\nu - 1)/2) \Gamma^{-1}(\nu/2)$. Novamente, as variáveis auxiliares E_0 e Y_0 são desconhecidas e não observáveis, assim nosso vetor de parâmetros a ser estimado é $\Omega = (\lambda, \alpha, \sigma, \nu, \Phi, d, \Theta, E_0, Y_0)$. Consideramos $(\lambda, \nu), \alpha, \sigma, \Phi, d, \Theta, E_0$ e Y_0 independentes a priori. Para $(\lambda|\nu)$ temos a distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos em (3.8), como veremos a seguir. Enquanto que $\alpha \sim N(a_\alpha, b_\alpha)$, $d \sim U(-1/2, 1/2)$ e $\sigma \sim IG(a_\sigma, b_\sigma)$ com valores para $a_\alpha, b_\alpha, a_\sigma$ e b_σ que garantem que a distribuição a priori seja relativamente vaga. Para ν temos a distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos da t-Student, $TA_\nu(0)$, dada em (3.9), como veremos a seguir. Além disso, Φ e Θ têm distribuições uniformes nas respectivas regiões de estacionariedade e invertibilidade; $\epsilon_k \sim TA_\nu(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - q$ e $y_k \sim TA_\nu(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - \max(p, m)$.

Discutiremos na próxima seção condições na distribuição a priori de λ e ν para que a mesma seja própria e conseqüentemente a distribuição a posteriori.

Assim, as distribuições a posteriori dos modelos ARFIMA(p, d, q) com erros NA e TA são dados pelas combinações das respectivas funções de verossimilhança e distribuições a priori. Em nossos estudos, para inferir, estamos utilizando métodos de simulação Monte Carlo via cadeias de Markov, em especial o amostrador de Gibbs. Nos dois casos, as condicionais completas não possuem forma fechada conhecida e para cada uma delas é necessário o cálculo de praticamente toda função de verossimilhança, portando não se justifica a apresentação dessas expressões sendo assim omitidas neste trabalho.

Na figura 3.3 temos alguns exemplos de séries temporais artificiais dos processos ARFIMA(p, d, q) com erros NA a esquerda e TA a direita, assim como alguns casos particulares. Para os processos ARFIMA($0, d, 0$) usamos uma aproximação da ordem de 50.000 termos na expansão dada em (3.4) dispensando as 10.000 primeiras observações para diminuir o efeito dos valores iniciais, sendo a mesma idéia aplicada ao caso geral ARFIMA(p, d, q) com a introdução adicional de termos dependendo do coeficiente auto-regressivo ϕ e médias-móveis θ . Também para o processo ARMA(p, q) usamos idéias similares, onde dispensamos as 1.000 primeiras observações para diminuir o efeito das variáveis auxiliares iniciais. Este procedimento pode ser criticado por conta de ineficiências no processo de geração das séries artificiais. Entretanto, utilizamos este procedimento para ilustrar os processos ARFIMA com erros NA e TA, e seus casos particulares. Observe que escolhemos o grau de liberdade da TA igual a 5 porque queremos que as médias e variâncias sejam bem definidas e possamos apresentar de forma gráfica clara os processos, evitando pontos distantes da média e da posição α . Os valores dos coeficientes foram escolhidos ao acaso, sendo que $d = 0,40$ produz uma série temporal moderadamente persistente.

A seguir, consideramos alguns casos particulares do modelo ARFIMA(p, d, q) com erros NA e TA, e estudamos algumas de suas propriedades, por exemplo para compreender melhor o processo de estimação, a escolha da distribuição a priori, o efeito na distribuição a posteriori, escolha dos estimadores pontuais e como alguns critérios de seleção de modelos se comportam neste tipo de modelagem. O primeiro caso é o modelo chamado de ruído branco, que pode também ser visto como modelo para dados independentes e identicamente distribuídos, onde investigamos a distribuição a priori de Jeffreys. O segundo caso é o modelo ARMA(p, q) onde nos concentramos na estimação, seleção de modelos e aplicações. Por último, o modelo ARFIMA($0, d, 0$) com ênfase na estimação, seleção de modelos e aplicação a série do nível mínimo do rio Nilo.

3.2.1 Modelo ruído branco

O caso particular mais simples do modelo ARFIMA(p, d, q) dado em (3.3) é o ruído branco descrito por $p = d = q = 0$ e também $\alpha = 0$, resultando em $y_t = \epsilon_t$. Os dados neste caso são independentes e identicamente distribuídos nas famílias de distribuições NA da definição 3.1 e TA da definição 3.2.

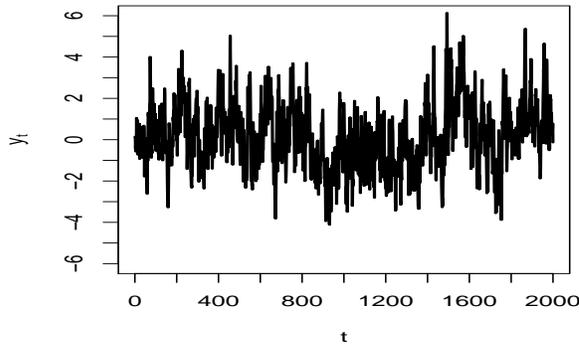
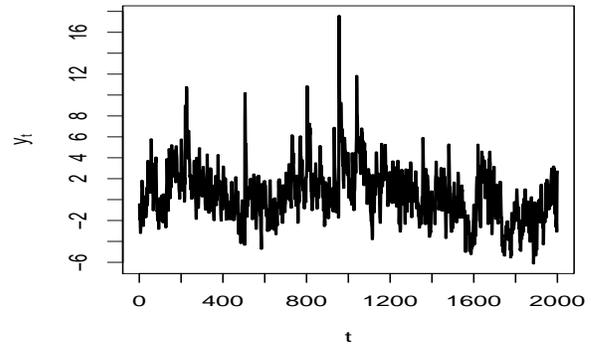
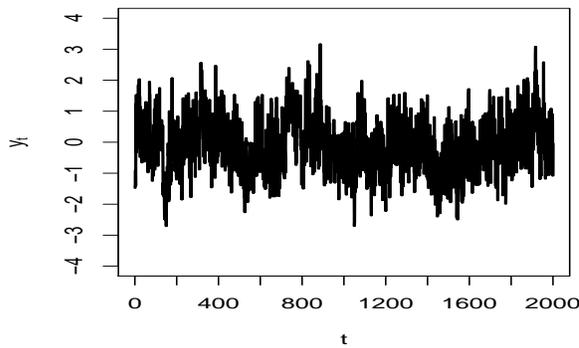
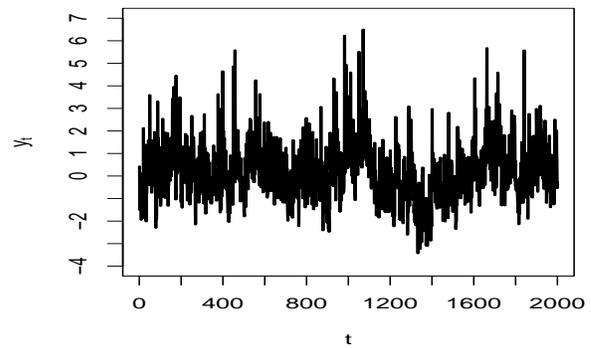
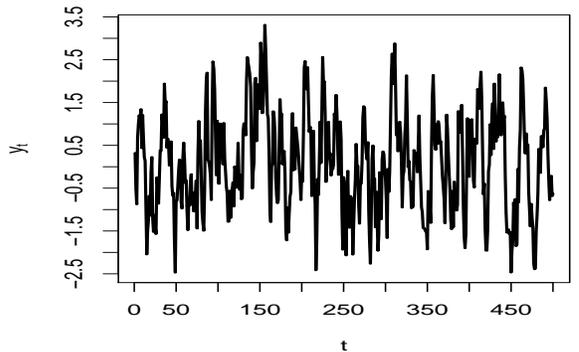
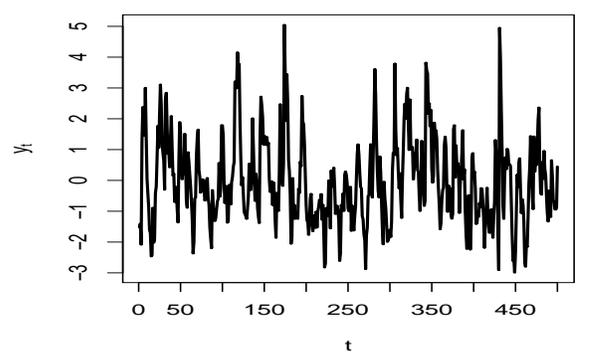
(a) ARFIMA(1, d, 1) $\phi = 0,5$, $d = 0,40$ e $\theta = 0,2$ (b) ARFIMA(1, d, 1) $\phi = 0,5$, $d = 0,40$ e $\theta = 0,2$ (c) ARFIMA(0, d, 0) $d = 0,40$ (d) ARFIMA(0, d, 0) $d = 0,40$ (e) ARMA(1, 1) $\phi = 0,7$ e $\theta = 0,2$ (f) ARMA(1, 1) $\phi = 0,7$ e $\theta = 0,2$

Figura 3.3: Séries temporais artificiais dos processos ARFIMA e casos particulares. A esquerda modelos com erros NA com $\alpha = 0$, $\sigma = 1$ e $\lambda = 2$, e a direita modelos com erros TA com $\alpha = 0$, $\sigma = 1$, $\lambda = 2$ e $\nu = 5$.

Nosso objetivo é calcular, revisar e estudar algumas propriedades das distribuições a priori de Jeffreys para as distribuições NA e TA, sem as respectivas restrições na média, isto é, nos parâmetros de posição μ , mas para estes modelos com as restrições de $\mu = 0$ e $\sigma = 1$, isto é, teremos uma amostra de dados independentes e identicamente distribuídos das famílias de distribuições NA e TA, e retemos nossa

atenção somente aos parâmetros de assimetria λ com os demais fixos e em outro caso para o grau de liberdade da t-Student com os demais fixos. Em alguns casos, é possível trabalhar com a distribuição a priori de Jeffreys para todos os parâmetros de forma conjunta. Entretanto, tal distribuição a priori conjunta apresenta alguns problemas devido a matriz de informação de Fisher ser singular para $\lambda \rightarrow 0$ (Azzalini, 1985). Detalhes desta priori para o caso NA e algumas outras discussões podem ser encontradas em Liseo e Loperfido (2004).

Seja $y = (y_1, y_2, \dots, y_T)$ uma amostra aleatória de uma variável aleatória NA com f.d.p dada na definição 3.1 com $\mu = 0$ e $\sigma = 1$. O logaritmo da função de verossimilhança é dada por $\ell(\lambda|y) = T \log(2) + \sum_{t=1}^T \log(\phi(y_t)) + \sum_{t=1}^T \log(\Phi(\lambda y_t))$. Conforme já sugerido na literatura (Liseo e Loperfido, 2004), concentraremos na distribuição a priori de Jeffreys que é dada por

$$p(\lambda) \propto [I(\lambda)]^{\frac{1}{2}} = \left(\int_{-\infty}^{\infty} 2y^2 \phi(y) \frac{\phi^2(\lambda y)}{\Phi(\lambda y)} dy \right)^{\frac{1}{2}}, \quad (3.7)$$

onde $I(\lambda)$ é a matriz de informação de Fisher dada por

$$I(\lambda) = E_Y(U^2(Y; \lambda)) = E_Y \left(\frac{\partial}{\partial \lambda} \log(f(y|\lambda)) \right)^2.$$

Liseo e Loperfido (2004) provaram que esta distribuição a priori é simétrica em torno de 0 e possui caudas da ordem $O(\lambda^{-\frac{3}{2}})$, o que implica que é uma distribuição a priori própria. Esta distribuição a priori necessita do cálculo numérico da integral, por exemplo usando a quadratura Gaussiana, para cada valor de λ . Esse cálculo pode ser implementado em algumas linguagens de programação desde que estejam disponíveis também a f.d.p e a f.d.a da normal padrão. Em nosso caso, usando a linguagem C e a biblioteca GSL (Galassi et al., 2005) foi necessário apenas a implementação da função integrando e chamar uma rotina de quadratura Gaussiana disponível. Assim, um estudo de Monte Carlo foi conduzido para avaliar o efeito desta distribuição a priori na posteriori de λ , mas precisamente na estimação do parâmetro em questão. Tomamos tamanhos de amostra 100 e 200, e geramos 10.000 amostras para cada uma delas com valores de $\lambda \in \{1/2, 1, 2, 3, 5\}$ e obtemos uma amostra a posteriori de tamanho 1.000 após 1.000 iterações do MCMC utilizando o amostrador da fatia (Neal, 2003). A tabela 3.1 contém um resumo da simulação, onde apresenta-se a média das médias a posteriori, a mediana das medianas a posteriori e a taxa de cobertura do intervalo de credibilidade de 0,95.

Os resultados da tabela 3.1 indicam que o melhor estimador do parâmetro de assimetria λ é a mediana a posteriori, o que equivale a utilizar-se a perda absoluta como a perda associada a estimação de λ . Além disso, de acordo com Liseo e Loperfido (2004) a média a posteriori somente existe para amostras cujo estimador de máxima verossimilhança seja finito. Entretanto, observe que as médias a posteriori também estão próximas dos valores reais e as taxas de cobertura estão próximas de 0,95 para ambos os casos, sendo que para estes tamanhos de amostra a probabilidade do estimador de máxima verossimilhança existir são altas, portanto podemos tomar a média a posteriori como estimador pontual.

Tabela 3.1: Resultados do estudo de Monte Carlo para amostras independentes e identicamente distribuídas da NA.

Amostra	$n = 100$			$n = 200$		
	Média	Mediana	Cobertura	Média	Mediana	Cobertura
$\lambda = 1/2$	0,507	0,495	0,949	0,503	0,499	0,950
$\lambda = 1$	1,021	0,995	0,944	1,009	0,997	0,949
$\lambda = 2$	2,086	1,994	0,945	2,042	1,997	0,945
$\lambda = 3$	3,210	3,004	0,945	3,083	2,995	0,944
$\lambda = 5$	5,775	5,032	0,944	5,251	4,993	0,948

A extensão desta distribuição a priori para o modelo de regressão pode ser encontrada no apêndice C, junto alguns comentários pertinentes e aplicação do mesmo.

Agora, seja $y = (y_1, y_2, \dots, y_T)$ uma amostra aleatória de uma variável aleatória TA com f.d.p dada na definição 3.2 com $\mu = 0$ e $\sigma = 1$. Novamente, o interesse é estimar o parâmetro de assimetria deste modelo. O logaritmo da função de verossimilhança é dada por $\ell(\lambda, \nu|y) = T \log(2) + \sum_{t=1}^T \log(t_\nu(y_t)) + \sum_{t=1}^T \log(\mathbb{T}_{\nu+1}(\lambda y_t (\nu+1)^{1/2} (\nu+y_t^2)^{-1/2}))$. Para algumas amostras o estimador de máxima de verossimilhança pode não existir e dependendo da escolha da distribuição a priori pode resultar numa distribuição a posteriori imprópria. Para tentar contornar este problema, sugerimos a distribuição a priori de Jeffreys para o parâmetro λ condicional a ν , que é dado por

$$p(\lambda|\nu) \propto \left(\int_{-\infty}^{\infty} 2y^2 \left(\frac{\nu+1}{\nu+y^2} \right) t_\nu(y) \frac{t_{\nu+1}^2 \left(\lambda y \sqrt{\frac{\nu+1}{\nu+y^2}} \right)}{\mathbb{T}_{\nu+1} \left(\lambda y \sqrt{\frac{\nu+1}{\nu+y^2}} \right)} dy \right)^{\frac{1}{2}}. \quad (3.8)$$

Lema 3.4 A distribuição a priori $p(\lambda|\nu)$ para a TA é simétrica em torno de 0.

Prova: Basta provar $l(\lambda)$ é simétrica em torno de 0. Temos

$$\begin{aligned} l(\lambda) &= \int_{-\infty}^{\infty} 2y^2 \left(\frac{\nu+1}{\nu+y^2} \right) t_\nu(y) \frac{t_{\nu+1}^2 \left(\lambda y \sqrt{\frac{\nu+1}{\nu+y^2}} \right)}{\mathbb{T}_{\nu+1} \left(\lambda y \sqrt{\frac{\nu+1}{\nu+y^2}} \right)} dy \\ &= \int_{-\infty}^{\infty} 2u^2 \left(\frac{\nu+1}{\nu+u^2} \right) t_\nu(u) \frac{t_{\nu+1}^2 \left(-\lambda u \sqrt{\frac{\nu+1}{\nu+u^2}} \right)}{\mathbb{T}_{\nu+1} \left(-\lambda u \sqrt{\frac{\nu+1}{\nu+u^2}} \right)} du = l(-\lambda) \end{aligned}$$

É fácil verificar que a distribuição a priori é decrescente em $|\lambda|$. Não existe ainda uma prova formal de que a distribuição seja própria, mas análises e comparações feitas com a distribuição t-Student, para

alguns valores do grau de liberdade e da escala, usando a distância de Kullback-Leibler apontem para uma distribuição a priori de Jeffreys própria.

Agora, precisamos definir a distribuição a priori para o grau de liberdade da TA. Em análises preliminares notou-se que a função de verossimilhança de ν , condicional aos demais parâmetros, pode ir para uma assíntota a medida que $\nu \rightarrow \infty$, podendo assim não existir o estimador de máxima verossimilhança. A figura 3.4 ilustra o problema, onde são apresentados duas funções de verossimilhança para ν para tamanhos de amostras 85 e 100 com os demais parâmetros fixos no valor gerado: $\lambda = 5$, $\mu = 0$ e $\sigma = 1$.

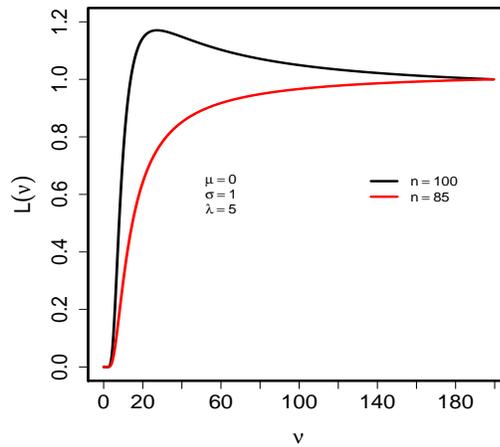


Figura 3.4: Função de verossimilhança de ν na distribuição TA.

Observe que a distribuição TA tem como caso particular ($\lambda = 0$) a distribuição t-Student com ν graus de liberdade. Como a distribuição TA apresenta as mesmas características em relação aos graus de liberdade de seu caso particular, então utilizaremos a priori de Jeffreys de seu caso particular que pode ser encontrado em Fonseca (2004), cujos resultados mostraram ser eficientes utilizando a perda absoluta, isto é, tomando-se a mediana a posteriori como estimador pontual para ν .

$$p(\nu) \propto \left[2 \left(\frac{2}{\nu+1} - \frac{\nu+2}{\nu(\nu+3)} \right) - \tau^{(2)} \left(\frac{\nu+1}{2} \right) + \tau^{(2)} \left(\frac{\nu}{2} \right) - \frac{2}{\nu} \right]^{\frac{1}{2}}, \quad (3.9)$$

onde $\tau^{(2)}(\cdot)$ é a função trigama. Logo a distribuição a priori conjunta é $p(\lambda, \nu) = p(\lambda|\nu)p(\nu)$. Um estudo de Monte Carlo foi conduzido para avaliar o efeito desta distribuição a priori na estimação dos parâmetros λ e ν com os demais fixos em $\mu = 0$ e $\sigma = 1$. Neste estudo tomamos 1.000 replicações de amostras de tamanho 100 para todas as combinações dos conjuntos $\lambda = \{1/2, 1, 2, 3, 5\}$ e $\nu = \{1, 3, 5, 10\}$. Para cada uma das amostras aleatórias da TA deste cruzamento obteve-se uma amostra de tamanho 1.000 da distribuição a posteriori após 1.000 iterações utilizando o amostrador da fatia dentro do amostrador de Gibbs. A tabela 3.2 contém um resumo da simulação, onde apresentamos a média das médias a

Tabela 3.2: Resultados do estudo de Monte Carlo para dados independentes e identicamente distribuídas da TA.

$n = 100$							
Valor real	Média	Mediana	Cobertura	Valor real	Média	Mediana	Cobertura
$\lambda = 1/2$	0,521	0,497	0,954	$\nu = 1$	0,950	1,029	0,927
$\lambda = 1$	1,050	1,004	0,927	$\nu = 1$	1,034	1,008	0,950
$\lambda = 2$	2,140	1,985	0,936	$\nu = 1$	1,035	0,999	0,947
$\lambda = 3$	3,379	2,962	0,941	$\nu = 1$	1,033	0,999	0,935
$\lambda = 5$	6,306	4,932	0,939	$\nu = 1$	1,028	1,001	0,945
$\lambda = 1/2$	0,509	0,492	0,947	$\nu = 3$	3,574	2,996	0,951
$\lambda = 1$	1,039	0,995	0,946	$\nu = 3$	3,455	3,031	0,940
$\lambda = 2$	2,111	2,024	0,954	$\nu = 3$	3,526	3,021	0,951
$\lambda = 3$	3,214	2,970	0,948	$\nu = 3$	3,526	3,013	0,949
$\lambda = 5$	5,812	4,905	0,947	$\nu = 3$	3,609	3,045	0,948
$\lambda = 1/2$	0,508	0,494	0,954	$\nu = 5$	9,677	4,880	0,951
$\lambda = 1$	1,037	0,997	0,938	$\nu = 5$	9,035	4,985	0,951
$\lambda = 2$	2,080	1,987	0,938	$\nu = 5$	9,216	5,092	0,967
$\lambda = 3$	3,217	2,974	0,960	$\nu = 5$	9,583	5,141	0,957
$\lambda = 5$	6,093	5,004	0,949	$\nu = 5$	9,393	5,087	0,946
$\lambda = 1/2$	0,515	0,501	0,936	$\nu = 10$	32,175	9,750	0,979
$\lambda = 1$	1,038	1,002	0,946	$\nu = 10$	31,567	9,828	0,967
$\lambda = 2$	2,137	2,028	0,949	$\nu = 10$	28,967	9,698	0,971
$\lambda = 3$	3,306	3,064	0,953	$\nu = 10$	32,687	9,888	0,966
$\lambda = 5$	6,021	5,044	0,949	$\nu = 10$	32,746	10,193	0,979

posteriori, a mediana das medianas a posteriori e a taxa de cobertura do intervalo de credibilidade de 0,95 para os dois parâmetros, onde cada linha da tabela corresponde a uma combinação entre λ e ν .

Os resultados apresentados na tabela 3.2 indicam que os melhores estimadores pontuais dos parâmetros de assimetria λ e dos graus de liberdade ν são suas respectivas medianas a posteriori, o que equivale a

utilizar-se a perda absoluta como a perda associada a estimação de λ e ν . Tal resultado era esperado pela similaridade com a distribuição a priori de λ na distribuição NA dada na equação (3.7) e os resultados apresentados na tabela 3.1, assim como pelos resultados encontrados em Fonseca (2004) para a distribuição a priori de ν dada em (3.9). Porém, observe que as médias a posteriori para o parâmetro λ também estão próximas dos valores reais e as taxas de cobertura estão próximas de 0,95. Assim, podemos tomar a média a posteriori também como estimador pontual se considerarmos que para tamanhos de amostra 100 ou maior ela sempre existe. Por outro lado, a média a posteriori do parâmetro ν tem comportamento irregular, estando relativamente próximas para valores pequenos e longe para valores grandes, indicando cautela ao considerar a média a posteriori como estimador pontual, mesmo porque não existe garantias de que exista. Observe que apesar disso as taxas de coberturas para o parâmetro ν estão relativamente perto de 0,95.

3.2.2 Modelo ARMA(p,q)

Um outro caso particular do modelo (3.3) é o ARMA(p, q), isto é, o modelo ARFIMA($p, 0, q$). Este modelo pode ser descrito por

$$\Phi(B)(y_t - \alpha) = \Theta(B)\epsilon_t \quad (3.10)$$

com as mesmas condições anteriores de estacionariedade e invertibilidade para $\Phi(B)$ e $\Theta(B)$. Além disso, como no modelo geral, ϵ_t segue distribuições NA ou TA, ambas com média 0 e variância constante, com as devidas restrições nos parâmetros de posição como antes, isto é, $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}$ na distribuição NA de acordo com a equação (3.1) e $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}\Gamma((\nu - 1)/2)\Gamma^{-1}(\nu/2)$ na distribuição TA de acordo com a equação (3.2).

Assim, a função de verossimilhança de um processo, $\{y_t\}$, ARMA(p, q) com erros tendo distribuição NA, baseada numa amostra de tamanho T , condicional aos valores iniciais latentes $E_0 = (\epsilon_0, \dots, \epsilon_{1-q})$ e $Y_0 = (y_0, \dots, y_{1-p})$, é dada por

$$L(\lambda, \alpha, \sigma, \Phi, \Theta, E_0, Y_0 | y_1, \dots, y_T) = \prod_{i=1}^T \frac{2}{\sigma} \phi\left(\frac{y_t - \mu_t}{\sigma}\right) \Phi\left(\lambda \frac{(y_t - \mu_t)}{\sigma}\right) \quad (3.11)$$

onde $\mu_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} - \mu$, sendo $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}$. As variáveis auxiliares E_0 e Y_0 são desconhecidas e não observáveis, assim nosso vetor de parâmetros a ser estimado é $\Omega = (\lambda, \alpha, \sigma, \Phi, \Theta, E_0, Y_0)$. Logo, em nossa abordagem bayesiana, estimaremos todas essas quantidades, mas para isso precisamos definir a distribuição a priori de Ω . Em primeiro lugar, consideramos $\lambda, \alpha, \sigma, \Phi, \Theta, E_0$ e Y_0 independentes a priori. Para λ temos a distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos em (3.7). Enquanto que $\alpha \sim N(a_\alpha, b_\alpha)$ e $\sigma \sim IG(a_\sigma, b_\sigma)$ com valores para $a_\alpha, b_\alpha, a_\sigma$ e b_σ que garantem que a distribuição a priori seja relativa-

mente vaga. Além disso, Φ e Θ têm distribuições uniformes nas respectivas regiões de estacionariedade e invertibilidade; $\epsilon_k \sim \text{NA}(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - q$ e $y_k \sim \text{NA}(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - p$. Combinando a distribuição a priori de Ω , que neste caso é própria, com a função de verossimilhança em (3.11), obtemos uma distribuição a posteriori própria com forma fechada desconhecida.

Assim, para inferir nestes modelos, utilizaremos métodos de simulação Monte Carlo via cadeias de Markov (Gamerman e Lopes, 2006), em especial o amostrador de Gibbs (Geman e Geman, 1984; Gelfand e Smith, 1990). Para as condicionais completas dos parâmetros λ, α, σ , e as componentes de E_0 e Y_0 faremos uso do amostrador da fatia (Neal, 2003). As componentes de Φ são sorteadas em bloco pelo algoritmo de Metropolis-Hasting (Metropolis et al., 1953; Hastings, 1970) com propostas uniformes e uma validação da região de estacionariedade, sendo a mesma idéia aplicada ao vetor de parâmetros Θ .

Para o processo $\text{ARMA}(p, q)$ com erros tendo distribuição TA, a função de verossimilhança baseada numa amostra de tamanho T , condicional aos valores iniciais latentes $E_0 = (\epsilon_0, \dots, \epsilon_{1-q})$ e $Y_0 = (y_0, \dots, y_{1-p})$, é dada por

$$L(\lambda, \alpha, \sigma, \Phi, \Theta, E_0, Y_0 | y_1, \dots, y_T) = \prod_{i=1}^T \frac{2}{\sigma} t_{\nu} \left(\frac{y_t - \mu_t}{\sigma} \right) \mathbb{T}_{\nu+1} \left(\lambda \frac{(y_t - \mu_t)}{\sigma} \right) \quad (3.12)$$

onde $\mu_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} - \mu$, sendo $\mu = -\lambda \sigma (1 + \lambda^2)^{-1/2} (2/\pi)^{1/2} \Gamma((\nu - 1)/2) \Gamma^{-1}(\nu/2)$. Novamente, as variáveis auxiliares E_0 e Y_0 são desconhecidas e não observáveis, assim nosso vetor de parâmetros a ser estimado é $\Omega = (\lambda, \alpha, \sigma, \nu, \Phi, \Theta, E_0, Y_0)$. Consideramos (λ, ν) , α , σ , Φ , Θ , E_0 e Y_0 independentes a priori. Para $(\lambda | \nu)$ temos a distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos em (3.8). Enquanto que $\alpha \sim \text{N}(a_{\alpha}, b_{\alpha})$ e $\sigma \sim \text{IG}(a_{\sigma}, b_{\sigma})$ com valores para a_{α} , b_{α} , a_{σ} e b_{σ} que garantem que a distribuição a priori seja relativamente vaga. Para ν temos a distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos da t-Student, $\text{TA}_{\nu}(0)$, dada em (3.9). Embora não seja necessário, restringimos ν ao intervalo contínuo $(1, 100)$ com o limite inferior representando a existência da média de ϵ_t e o limite superior a aproximação do modelo NA. Além disso, Φ e Θ têm distribuições uniformes nas respectivas regiões de estacionariedade e invertibilidade; $\epsilon_k \sim \text{TA}_{\nu}(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - q$ e $y_k \sim \text{TA}_{\nu}(\lambda, \mu + \alpha, \sigma)$ independentes para $k = 0, \dots, 1 - p$. Novamente, a distribuição a posteriori neste caso tem forma fechada desconhecida, assim métodos de simulação Monte Carlo via cadeias de Markov serão utilizados como no modelo $\text{ARMA}(p, q)$ com erros NA, com mais o passo da geração de ν pelo amostrador da fatia no intervalo determinado acima.

Vale a pena notar que o caso particular do modelo $\text{ARMA}(p, q)$ com $q = 0$, isto é, o modelo auto-regressivo com erros assimétricos, não haveria a necessidade das variáveis auxiliares E_0 e Y_0 ao condicionar nas primeiras p observações do modelo, podendo assim ser tratado como um modelo de regressão e fazer o uso da priori de Jeffreys ou da sugestão dada no apêndice C.

3.2.3 Modelo ARFIMA(0, d , 0)

O modelo ARFIMA(0, d , 0) (Hosking, 1981) também é um caso particular do modelo (3.3), obtido ao considerarmos $p = q = 0$. Este modelo pode ser descrito por

$$(1 - B)^d(y_t - \alpha) = \epsilon_t \quad (3.13)$$

com as mesmas condições anteriores de estacionariedade e invertibilidade em d , $d \in (-1/2, 1/2)$. Os erros ϵ_t seguem distribuições NA ou TA, ambas com média 0 e variância constante, com as devidas restrições nos parâmetros de posição como antes, isto é, $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}$ na distribuição NA de acordo com a equação (3.1) e $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}\Gamma((\nu-1)/2)\Gamma^{-1}(\nu/2)$ na distribuição TA de acordo com a equação (3.2). De acordo com (Hosking, 1981), se $d > -1/2$, $\{y_t\}$ é inversível e possui representação auto-regressiva infinita dada por

$$(1 - B)^d y_t = \sum_{j=0}^{\infty} \varphi_j(d) y_{t-j} = \epsilon_t \quad (3.14)$$

onde

$$\varphi_0(d) = 1, \quad \varphi_1(d) = -d, \quad \varphi_2(d) = \frac{d(d-1)}{2!}, \quad \varphi_3(d) = -\frac{d(d-1)(d-2)}{3!}, \dots$$

Entretanto, o modelo ARFIMA(0, d , 0) também possui representação MA infinita desde que $d < 1/2$. Usando as representações AR e MA deste processo e aproximando a função de verossimilhança pelo truncamento da soma em (3.14), Silva (2003) mostrou que para o caso particular de erros normais, $NA(0, \mu, \sigma) \equiv N(\mu, \sigma)$, a representação AR é melhor se comparada a MA na estimação dos parâmetros, e que 50 termos no somatório são satisfatórios no sentido que consegue-se aproximar bem a função de verossimilhança e recuperar os valores dos parâmetros, pelo menos na estimação bayesiana. Assim, conjecturando-se que o mesmo seja válido para o modelo com erros NA e TA, também aproximamos a função de verossimilhança.

A função de verossimilhança aproximada de um processo, $\{y_t\}$, ARFIMA(0, d , 0) com erros NA, baseada numa amostra de tamanho T , condicional aos m valores iniciais, é dado por

$$L(\lambda, \alpha, \sigma, d | y_{m+1}, \dots, y_T) = \prod_{i=m+1}^T \frac{2}{\sigma} \phi\left(\frac{y_t - \mu_t}{\sigma}\right) \Phi\left(\lambda \frac{(y_t - \mu_t)}{\sigma}\right) \quad (3.15)$$

onde $\mu_t = \alpha + \varphi_1(d)y_{t-1} + \dots + \varphi_m(d)y_{t-m} - \mu$, sendo $\mu = -\lambda\sigma(1 + \lambda^2)^{-1/2}(2/\pi)^{1/2}$. Em geral, consideramos $m = 50$ em nossas aplicações. Falta agora, definir a distribuição a priori de $(\lambda, \alpha, \sigma, d)$. Como antes, λ , α , σ e d são independentes a priori com λ tendo distribuição a priori de Jeffreys para dados independentes e identicamente distribuídos em (3.7). Enquanto que $\alpha \sim N(a_\alpha, b_\alpha)$, $d \sim U(-1/2, 1/2)$ e $\sigma \sim IG(a_\sigma, b_\sigma)$ com valores para a_α , b_α , a_σ e b_σ que garantem que a distribuição a priori seja

relativamente vaga. A distribuição a posteriori resultante possui forma fechada desconhecida. Mais uma vez, métodos de simulação Monte Carlo via cadeias de Markov são utilizados com atenção especial ao amostrador de Gibbs e o amostrador da fatia para todas as condicionais completas.

Para o modelo ARFIMA(0, d , 0) com erros TA, as idéias se repetem com a aproximação da função de verossimilhança, a substituição das f.d.p da NA pela TA introduzindo assim o parâmetro ν representando os graus de liberdade, que por sua vez tem a mesma distribuição a priori do modelo ARMA(p , q) com erros TA. Novamente, utilizamos o amostrador da fatia dentro do amostrador de Gibbs para todas as condicionais completas incluindo a ν . Neste caso, a função de verossimilhança é dada por

$$L(\lambda, \alpha, \sigma, \nu, d | y_{m+1}, \dots, y_T) = \prod_{i=m+1}^T \frac{2}{\sigma} t_{\nu} \left(\frac{y_t - \mu_t}{\sigma} \right) \mathbb{T}_{\nu+1} \left(\lambda \frac{(y_t - \mu_t)}{\sigma} \right) \quad (3.16)$$

No caso geral, o modelo ARFIMA(p , d , q) é a combinação desses dois casos particulares, ARMA(p , q) e ARFIMA(0, d , q), assim como a estimação, isto é, a combinação de estimação desses dois modelos com a aproximação da função de verossimilhança na presença do termo fracionário $d \neq 0$. Assim, não estaremos considerando processo de estimação do modelo ARFIMA(p , d , q) mas seus dois casos particulares mais importantes ARMA(p , q) e ARFIMA(0, d , 0).

3.3 Aplicações

Nesta seção mostraremos algumas aplicações com dados artificiais e reais dos modelos ARFIMA(p , d , q) com erros NA e TA. Nos concentraremos nos dois casos particulares mais importantes, ARMA(p , q) e ARFIMA(0, d , 0), com ênfase na estimação bayesiana de todos os parâmetros, inclusive das variáveis auxiliares, e também em alguns critérios de seleção de modelos.

3.3.1 Modelo ARMA(p , q): dados artificiais

Nossa primeira aplicação é a dados artificiais do modelo ARMA(p , q) dado na equação (3.10) com erros NA e TA. Duas séries temporais artificiais de tamanho 300 foram geradas do modelo ARMA(1, 1) com erros NA e TA, respectivamente. Na tabela 3.3 temos vários critérios de comparação de modelos, dados no apêndice D, baseados numa amostra a posteriori de tamanho 1.000 tomadas 1 a cada 10 após 1.000 iterações do amostrador de Gibbs descrito anteriormente. A maioria dos critérios identificam os modelos corretamente, mas o DIC falha na identificação do modelo NA enquanto o EPD falha na identificação do modelo TA. Entretanto, não podemos generalizar esses resultados, havendo assim a necessidade de uma melhor investigação do comportamento de todos esses critérios na identificação dos modelos ARMA(p , q) com erros NA e TA, além da validação dos mesmos.

A tabela 3.4 traz um resumo da distribuição a posteriori do modelo ARMA(1, 1) com erros NA para os respectivos dados artificiais. Observe que para quase todos os parâmetros o intervalo de credibilidade

Tabela 3.3: Resultados empíricos do modelo ARMA(p, q) NA e TA para tamanhos de amostra $T = 300$

Dados	NA		TA	
	NA	TA	NA	TA
AIC	606,45	610,66	835,57	804,11
EAIC	612,78	612,62	843,30	806,47
BIC	624,97	632,88	854,09	826,33
EBIC	631,30	634,84	861,82	828,70
DIC	609,10	602,58	841,03	796,84
EPD	214,69	215,31	490,60	563,08
REQM	0,6852	0,6861	1,0472	1,0444
EAM	0,5358	0,5359	0,7841	0,7778

contém os valores verdadeiros e que as médias a posteriori assim como as medianas estão relativamente próximas desses valores verdadeiros, com exceção do parâmetro auto-regressivo ϕ . Portanto, os valores utilizados na geração foram basicamente recuperados no processo de estimação, mas ainda existe a necessidade de um estudo mais detalhado da estimação, embora vários resultados preliminares do modelo apontem para uma boa identificação dos parâmetros. As mesmas afirmações e conclusões valem para o modelo ARMA(1, 1) com erros TA cujo resumo da distribuição a posteriori é apresentado na tabela 3.5.

Tabela 3.4: Sumário da distribuição a posteriori do modelo ARMA(1,1) com erros NA. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetros	Real	Média	D.P	2,5%	Mediana	97,5%
α	0	-0,020	0,043	-0,100	-0,020	0,071
σ	1	1,014	0,065	0,889	1,011	1,145
λ	2	2,512	0,504	1,637	2,466	3,591
ϕ	0,7	0,797	0,038	0,720	0,798	0,874
θ	0,2	0,094	0,064	-0,031	0,092	0,217

Tabela 3.5: Sumário da distribuição a posteriori do modelo ARMA(1,1) com erros TA. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetros	Real	Média	D.P	2,5%	Mediana	97,5%
α	0	0,139	0,078	-0,006	0,133	0,310
σ	1	1,053	0,113	0,839	1,049	1,290
λ	2	2,310	0,636	1,279	2,230	3,760
ν	5	4,662	1,317	2,760	4,450	8,087
ϕ	0,7	0,680	0,044	0,590	0,680	0,761
θ	0,2	0,251	0,055	0,137	0,252	0,355

3.3.2 Índice anual do preço do trigo

Os modelos ARMA(p, q) dados pela equação (3.10) com erros NA e TA são ajustado para o índice de preço do trigo (preços Europeus) de 1500 a 1869 que corresponde a uma série temporal de 370 observações anuais. Os dados estão disponíveis de forma gratuita no sítio <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/>. A figura 3.5 mostra a série do índice de preço do trigo, onde parece exibir uma leve assimetria a direita. Análises preliminares indicaram uma dependência de ordem 2, podendo provavelmente ser descrita por um modelo MA(2) segundo análise gráfica das funções de autocorrelação e a sua versão parcial.

Com esta série, ajustamos alguns modelos ARMA(p, q) com erros NA e TA para alguns valores de p e q entre 0 e 3. Para cada combinação possível, obtivemos uma amostra da distribuição a posteriori de tamanho 1.000 tomadas 1 a cada 20 após 1.000 iterações do amostrador de Gibbs, com detalhes descritos

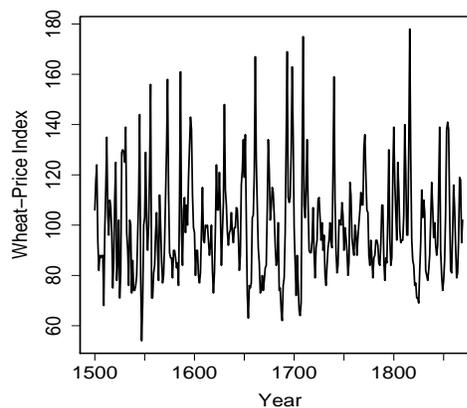


Figura 3.5: Índice anual de preço do trigo de 1500 a 1869.

Tabela 3.6: Critérios de seleção de modelos: ARMA(p, q) com erros NA e TA aplicados ao índice de preço do trigo

ARMA	AIC	EAIC	BIC	EBIC	DIC	EPD	REQM	EAM
Erros com distribuição normal assimétrica								
(1,0)	3111,47	3116,18	3127,13	3131,84	3112,89	211311	17,041	12,902
(2,0)	3094,87	3099,48	3114,44	3119,05	3094,09	193842	16,245	12,234
(3,0)	3094,86	3101,20	3118,34	3124,68	3095,55	193561	16,241	12,250
(0,1)	3094,87	3099,21	3110,53	3114,86	3095,54	281870	20,366	15,705
(0,2)	3093,07	3098,16	3112,63	3117,73	3093,25	287288	20,372	15,725
(0,3)	3094,63	3100,93	3118,11	3124,41	3095,24	286855	20,395	15,681
(1,1)	3093,04	3098,31	3112,61	3117,88	3093,59	245975	18,752	14,348
(2,1)	3094,39	3100,59	3117,87	3124,07	3094,79	225497	17,523	13,332
(1,2)	3095,00	3100,71	3118,48	3124,19	3094,42	317277	20,883	16,134
(2,2)	3096,51	3102,53	3123,90	3129,93	3094,56	275538	17,986	13,708
Erros com distribuição t-Student assimétrica								
(1,0)	3092,85	3098,43	3112,42	3118,00	3094,01	231782	16,962	12,812
(2,0)	3073,22	3079,65	3096,71	3103,13	3074,07	211421	16,175	12,165
(3,0)	3074,55	3081,75	3101,95	3109,15	3074,95	211763	16,128	12,155
(0,1)	3082,15	3086,69	3101,72	3106,26	3081,23	300129	20,366	15,657
(0,2)	3072,60	3078,16	3096,08	3101,64	3071,73	323408	20,485	15,769
(0,3)	3074,87	3081,32	3102,26	3108,71	3073,77	319606	20,373	15,656
(1,1)	3073,60	3078,88	3097,08	3102,36	3072,17	248411	18,075	13,737
(2,1)	3074,71	3081,49	3102,10	3108,88	3074,26	235601	17,051	12,944
(1,2)	3074,62	3081,05	3102,01	3108,44	3073,48	325529	20,223	15,545
(2,2)	3076,13	3083,36	3107,44	3114,67	3074,60	317107	19,192	14,667

anteriormente. Os valores iniciais considerados foram 99 para μ , 1 para σ , 3 para ν no caso de erros TA, 0 para λ e valores perto de 0 para os coeficientes em Φ e Θ . Na tabela 3.6 apresentamos vários critérios

de seleção de modelos (veja o apêndice D). Baseados nesta informação e nas distribuições a posteriori de todos os modelos, consideramos o modelo ARMA(0,2) com erros TA como tendo melhor ajuste, embora o EPD, REQM e EAM não indiquem tal modelo. Outros modelos ARMA(p, q) para diferentes ordens de p e q não são apresentados pois em alguns dos casos abaixo já temos parâmetros com massa de probabilidade concentrada em 0, como por exemplo o modelo ARMA(3,0) onde o coeficiente ϕ_3 tem essa características a posteriori.

A tabela 3.7 traz um resumo da distribuição a posteriori do modelo ARMA(0,2) com erros TA. Embora não tenha havido uma justificativa formal para o ajuste de modelos ARMA(p, q) com erros assimétrico e com caudas pesadas, verificamos pelas distribuição a posteriori do parâmetro λ a presença de assimetria, da posteriori do parâmetro ν que os erros possuem caudas pesadas e da posteriori do parâmetro σ uma escala diferente da unitária. Além disso, temos a distribuição a posteriori dos coeficientes θ_1 e θ_2 concentrados em regiões acima de 0, indicando alta probabilidades dos mesmos serem diferentes de 0 e portanto consideramos estatisticamente significativos. Logo, com esta aplicação ao índice de preço anual do trigo, mostramos que nosso modelo geral ARMA(p, q) com erros TA conseguiu capturar características importantes da série como assimetria e as caudas pesadas, numa modelagem bayesiana onde todas as quantidades desconhecidas, incluindo as variáveis auxiliares, são estimadas conjuntamente. Os critérios de seleção de modelos ajudaram na escolha do melhor ajuste. Assim, consideramos importante essa classe de modelos.

Tabela 3.7: Sumário da distribuição a posteriori do modelo ARMA(0,2) com erros TA. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetros	Média	D.P	2,5%	Mediana	97,5%
α	99,740	1,615	96,760	99,693	102,890
σ	15,985	1,858	12,576	15,879	19,873
λ	1,494	0,399	0,733	1,486	2,311
ν	5,053	3,861	2,870	4,574	8,828
θ_1	0,636	0,051	0,540	0,635	0,735
θ_2	0,165	0,048	0,075	0,167	0,260

3.3.3 Modelo ARFIMA(0, d , 0): dados artificiais

Como no modelo ARMA(p, q), apresentamos duas séries temporais artificiais do modelo ARFIMA(0, d , 0) dado na equação (3.13) com erros NA e TA, respectivamente. Lembramos que no amostrador de Gibbs utilizamos o amostrador da fatia para todas as condicionais completas dos parâmetros e que a aproximação auto-regressiva utilizada na função de verossimilhança foi da ordem 50. Na tabela 3.8 temos vários critérios de comparação de modelos, dados no apêndice D, baseados numa amostra a posteriori de tamanho 1.000 tomadas 1 a cada 5 após 1.000 iterações do amostrador de Gibbs descrito anteriormente. Mais uma vez os critérios identificam os modelos corretamente, mas o DIC falha na identificação do modelo NA enquanto o EPD falha na identificação do modelo TA. Entretanto, não podemos generalizar esses resultados, sendo necessário mais investigação.

Tabela 3.8: Resultados empíricos do modelo ARFIMA(0, d , 0) NA e TA para tamanhos de amostra $T = 1.000$

Dados	NA		TA	
	NA	TA	NA	TA
AIC	2020,89	2021,84	2307,39	2284,45
EAIC	2024,80	2025,71	2311,40	2289,18
BIC	2040,31	2046,11	2326,81	2308,73
EBIC	2044,22	2049,98	2330,82	2313,45
DIC	2020,70	2019,58	2307,40	2283,90
EPD	967,16	969,47	1373,89	1385,99
REQM	0,7136	0,7139	0,8567	0,8552
EAM	0,5632	0,5633	0,6466	0,6432

A tabela 3.9 traz um resumo da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros NA para os respectivos dados artificiais. Observe que para todos os parâmetros o intervalo de credibilidade contém os valores verdadeiros e que tanto as médias quanto as medianas a posteriori estão relativamente próximas desses valores verdadeiros. Mesmo utilizando uma aproximação da função de verossimilhança, conseguimos estimar relativamente bem os parâmetros principalmente o de longa dependência, d . Portanto, os valores utilizados na geração foram basicamente recuperados no processo de estimação, mas ainda existe a necessidade de um estudo mais detalhado da estimação, embora vários resultados preliminares do modelo apontam para uma boa identificação dos parâmetros. Afirmarções e conclusões equivalentes valem para o modelo ARFIMA(0, d , 0) com erros TA cujo resumo da distribuição a poste-

riori é apresentado na tabela 3.10. A exceção dos graus de liberdade ν , cujos intervalo de credibilidade, média e mediana a posteriori indicam sobre-estimação.

Tabela 3.9: Sumário da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros NA. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetros	Real	Média	D.P	2,5%	Mediana	97,5%
α	0	0,028	0,022	-0,019	0,029	0,071
σ	1	1,027	0,040	0,946	1,026	1,103
λ	2	2,094	0,252	1,649	2,077	2,613
d	0,4	0,396	0,025	0,349	0,397	0,443

Tabela 3.10: Sumário da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros TA. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetros	Real	Média	D.P	2,5%	Mediana	97,5%
α	0	-0,016	0,027	-0,067	-0,016	0,037
σ	1	0,991	0,072	0,848	0,991	1,127
λ	2	1,823	0,299	1,290	1,814	2,424
ν	5	8,760	2,919	5,030	8,039	15,436
d	0,4	0,412	0,024	0,365	0,411	0,459

3.3.4 Nível mínimo do rio Nilo

Nesta aplicação queremos verificar se a série temporal do nível mínimo do rio Nilo apresenta alguma assimetria, além da longa dependência já identificada na literatura (Beran, 1994; Silva, 2003). Para isso, aplicamos o modelo ARFIMA(0, d , 0) dado na equação (3.13) com erros NA e TA a série temporal do nível mínimo rio Nilo. Os dados foram coletados de 722 a 1284, como mostra a figura 3.6, medidos no Roda Gauche perto do Cairo. Esta série temporal pode exibir uma longa dependência como em Beran (1994) e Silva (2003), que ajustam um modelo ARFIMA(0, d , 0) com erros normais.

Mais uma vez, lembramos que no amostrador de Gibbs utilizamos o amostrador da fatia para todos as condicionais completas dos parâmetros e que a aproximação auto-regressiva utilizada na função de

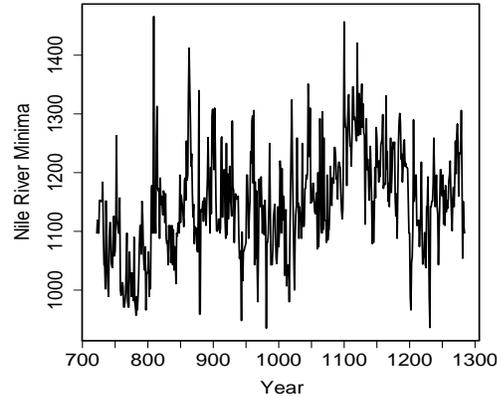


Figura 3.6: Nível mínimo do rio Nilo para os anos de 722 a 1284 medidos no Roda Gauche perto do Cairo.

verossimilhança foi da ordem 50.

Na tabela 3.11 temos vários critérios de comparação de modelos, dados no apêndice D, baseados numa amostra a posteriori de tamanho 1.000 tomadas 1 a cada 5 após 1.000 iterações do amostrador de Gibbs. Basicamente estamos decidindo entre erros NA e TA. Observa-se que o modelo TA tem melhor ajuste segundo a maioria dos critérios, com exceção do EPD. Então, tomando-se o modelo ARFIMA(0, d , 0) com erros TA como melhor ajuste, apresentamos na tabela 3.12 um sumário da distribuição a posteriori, onde destacamos a presença de assimetria descrita pela média a posteriori do parâmetro λ assim como seu intervalo de credibilidade, e também a distribuição a posteriori dos graus de liberdade ν indicando que o modelo possui de certa forma caudas pesadas. Vale a pena ressaltar a estimação da escala σ diferente de 1, sendo sua presença importante no modelo. Por último, e não menos importante, é a distribuição a posteriori do parâmetro de longa dependência d , com média a posteriori e intervalo de credibilidade apontando a presença de longa memória de acordo com a literatura (Beran, 1994; Silva, 2003).

Tabela 3.11: Critérios de seleção de modelos: ARFIMA(0, d , 0) com erros NA e TA aplicados volume mínimo do rio Nilo

Erros	AIC	EAIC	BIC	EBIC	DIC	EPD	REQM	EAM
NA	6887,38	6891,49	6905,05	6909,15	6887,59	5580385	67,600	50,833
TA	6865,36	6870,06	6887,44	6892,15	6864,77	5634570	67,289	50,486

Tabela 3.12: Sumário da distribuição a posteriori do modelo ARFIMA(0, d , 0) com erros TA. D.P, 2,5% e 97,5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetros	Média	D.P	2,5%	Mediana	97,5%
α	1148,3	2,7564	1143,1	1148,2	1154,0
σ	62,408	5,8328	52,397	61,853	75,514
λ	0,6979	0,2762	0,1623	0,7014	1,2360
ν	6,7717	2,3533	4,0221	6,2972	12,609
d	0,4424	0,0316	0,3840	0,4426	0,5040

3.4 Discussão final

A classe de distribuições proposta por Azzalini (1985) permite modelar assimetrias ao custo adicional de um parâmetro. A introdução de erros normais e t-Student assimétrico nos modelos ARFIMA(p, d, q) permitem modelagens mais flexíveis de séries temporais reais, em especial dos casos particulares ARMA(p, q) e ARFIMA(0, $d, 0$), com a distinção entre assimetria ou não governada por um único parâmetro em questão, λ . A introdução de variáveis auxiliares facilitam a construção do modelo e as mesmas são estimadas na distribuição a posteriori junto com os demais parâmetros através de métodos de simulação Monte Carlo via cadeias de Markov. Em especial usamos o algoritmo de Gibbs com passos de Metropolis-Hastings para as condicionais completas dos coeficientes auto-regressivos e médias-móveis, respectivamente, e pelo amostrador da fatia para as condicionais completas dos demais parâmetros. As distribuições a priori de Jeffreys apresentadas fornecem bons resultados no que diz respeito a estimação dos parâmetros dos modelos. A utilização de alguns critérios de seleção de modelos, entre eles o DIC, ajudam na decisão do melhor ajuste, isto é, na escolha entre erros normais ou t-Student assimétrico assim como da ordem de p e q .

Capítulo 4

CÓPULAS: INFERÊNCIA E SELEÇÃO

Neste capítulo investigamos alguns critérios de seleção de modelos aplicados a funções de dependência, também conhecidas como funções cópulas ou simplesmente cópulas, e algumas marginais assimétricas. Uma cópula é uma função de distribuição acumulada (f.d.a) multivariada cujas marginais têm distribuições uniforme no intervalo $(0,1)$. Existem diversas cópulas com diferentes estruturas de dependência, as quais servem para unir possivelmente diferentes distribuições marginais, formando assim uma distribuição multivariada com determinada estrutura de dependência e marginais. Como uma cópula é uma f.d.a, em alguns casos temos a função de densidade de probabilidade (f.d.p) de uma cópula que é dada pela derivadas mistas nos argumentos, chamada também de densidade cópula. A figura 4.1 mostra as curvas de contorno das f.d.p de duas distribuições bivariadas dadas por duas diferentes cópulas, chamadas de Clayton (Clayton, 1978) e Frank (Frank, 1979), com ambas as distribuições marginais sendo normais padrão. Observe que a estrutura de dependência e assimetria são diferentes, entretanto todas as distribuições marginais são iguais.

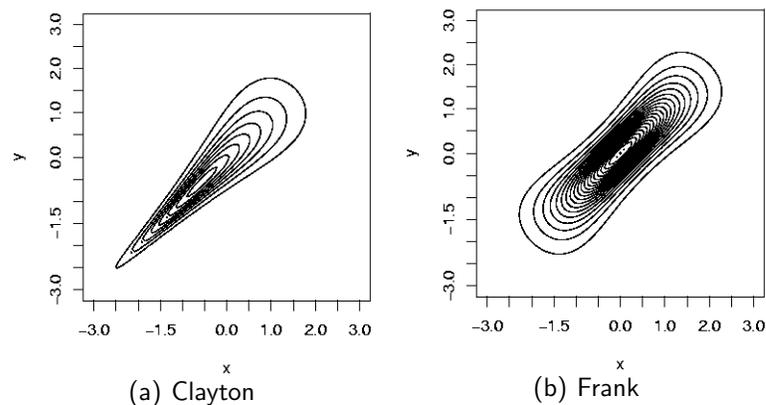


Figura 4.1: Contorno das distribuições bivariadas dadas pela combinação de cópulas Clayton e Frank com distribuições marginais normais padrão e uma medida de dependência τ de Kendall igual a $2/3$.

Nossas contribuições são a estimação conjunta de todos os parâmetros dos modelos pelo enfoque bayesiano, evitando assim a estimação em dois passos, e o uso do critério de informação do desvio (DIC), dado no apêndice D, para identificar o melhor ajuste dentre os 6 modelos simulados em nossos estudos de Monte Carlo. Com esse estudo de Monte Carlo queremos mostrar que o DIC e outros critérios de

seleção de modelos baseados na função desvio identificam o modelo simulado com alta probabilidade. Para a simulação dos dados artificiais de uma cópula específica e suas marginais usamos técnicas de reamostragem por importância (SIR), ou métodos de simulação direta como os de inversão. Inferência na distribuição a posteriori é feita através de uma amostra da mesma, a qual é obtida usando métodos de simulação Monte Carlo via cadeias de Markov, em especial o amostrador da fatia (Neal, 2003) dentro do algoritmo de Gibbs (Geman e Geman, 1984; Gelfand e Smith, 1990).

4.1 Propriedades gerais

Funções de dependência, também conhecidas como cópulas, são f.d.a multivariadas cujas marginais têm distribuição uniforme no intervalo (0,1). Formalmente, uma cópula C é dada por

$$C(u_1, \dots, u_d) = P(U_1 \leq u_1, \dots, U_d \leq u_d) \quad (4.1)$$

onde $U_i \sim U(0, 1)$ para $i = 1, 2, \dots, d$. O teorema de Sklar (Sklar, 1959) relaciona as cópulas com suas marginais. Algumas cópulas importantes são:

- $C^-(u_1, \dots, u_d) = \max(u_1 + \dots + u_d - d + 1, 0)$, que não é uma cópula para $d \geq 3$. Para $d = 2$, C^- é definida como a cópula contra-monotônica e representa dependência negativa perfeita entre as variáveis;
- $C^\perp(u_1, \dots, u_d) = \prod_{i=1}^d u_i$, é definida como a cópula produto e representa independência entre as variáveis; e
- $C^+(u_1, \dots, u_d) = \min(u_1, \dots, u_d)$, é definida como a cópula comonotônica e representa dependência positiva perfeita entre as variáveis.

Além disso, como consequência do teorema de Sklar, pode ser provado que para qualquer cópula $C(u_1, \dots, u_d)$,

$$C^-(u_1, \dots, u_d) \leq C(u_1, \dots, u_d) \leq C^+(u_1, \dots, u_d),$$

que são os limites de Fréchet-Hoeffding. Maiores detalhes sobre cópulas e suas propriedades podem ser encontradas no livros de Schweizer e Sklar (1983), Joe (1997) e Nelsen (1999).

Na tabela 4.1 apresentamos 6 famílias de cópulas bivariadas usadas em nosso estudo empírico. Todas elas são cópulas com um único parâmetro, porque consideramos os graus de liberdade da cópula Student como conhecido. Elas serão denotadas por $C(u, v | \theta)$. Note que a maioria dos parâmetros das cópulas têm domínios diferentes, mas poderíamos usar a abordagem de Huard, Évin, e Favre (2006) e definir uma transformação de θ tal que todas as famílias de cópulas tivessem o mesmo domínio de valores de θ

e assim definir uma mesma distribuição a priori. Entretanto, apesar de tal abordagem ser possível, nem todas as cópulas têm o mesmo comportamento e graus de dependência, não importa qual medida de dependência seja usada. Logo, ficamos com a parametrização da tabela 4.1 que é usualmente encontrada na literatura e definimos distribuições a priori própria e relativamente vagas para todas elas.

Tabela 4.1: Definição de algumas cópulas usadas nesta tese.

Cópula	$C(u, v \theta)$	$\theta \in \Omega$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$
Frank	$-\frac{1}{\theta} \log \left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right)$	$\mathbb{R} \setminus \{0\}$
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp \left\{ \frac{2\theta st - s^2 - t^2}{2(1-\theta^2)} \right\} ds dt$	$[-1, 1]$
Gumbel	$\exp \left\{ -[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta} \right\}$	$[1, \infty)$
Cauda-pesada	$u + v - 1 + [(1-u)^{-1/\theta} + (1-v)^{-1/\theta} - 1]^{-\theta}$	$(0, \infty)$
Student t	$\int_{-\infty}^{T_v^{-1}(u)} \int_{-\infty}^{T_v^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2}) \nu \pi \sqrt{1-\theta^2}} \left(1 + \frac{s^2 + t^2 - 2\theta st}{\nu(1-\theta^2)} \right) ds dt$	$[-1, 1]$

Sabe-se que o coeficiente de correlação linear pode não ser uma boa medida para identificar e medir dependência entre variáveis aleatórias. Entretanto, existem muitas outras medidas de dependência que possuem melhores propriedades quando comparadas ao coeficiente de correlação linear (Joe, 1997; Nelsen, 1999). A medida não paramétrica entre variáveis aleatórias τ de Kendall associada a cópulas é provavelmente a mais conhecida e usada dentre elas.

Sejam (X_i, Y_i) e (X_j, Y_j) vetores aleatórios independentes e identicamente distribuídos ($i \neq j$), então

$$\tau = P((X_i - X_j)(Y_i - Y_j) > 0) - P((X_i - X_j)(Y_i - Y_j) < 0).$$

Observe que τ é a probabilidade de concordância menos a probabilidade de discordância entre dois vetores aleatórios independentes (Kruskal, 1958). Para cópulas, o τ de Kendall é dado por (Nelsen, 1999)

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 = 1 - 4 \int_0^1 \int_0^1 \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) du dv. \quad (4.2)$$

Tabela 4.2: τ de Kendall e seus respectivos domínios para as cópula usadas neste tese.

Cópula	$\tau = g(\theta)$	$\tau \in \Omega$
Clayton	$\frac{\theta}{\theta + 2}$	$(0, 1]$
Frank	$1 - \frac{4}{\theta} \left(1 - \int_0^\theta \frac{s}{\exp(s) - 1} ds \right)$	$[-1, 1] \setminus \{0\}$
Gaussian	$\frac{2}{\pi} \arcsin \theta$	$[-1, 1]$
Gumbel	$1 - \theta^{-1}$	$[0, 1]$
Cauda-pesada	$\frac{1}{2\theta + 1}$	$(0, 1]$
Student t	$\frac{2}{\pi} \arcsin \theta$	$[-1, 1]$

A tabela 4.2 mostra as expressões do τ de Kendall para as 6 famílias de cópulas apresentadas na tabela 4.1. A maior parte delas possuem expressões simples, exceto a cópula Frank que pode ser calculada numericamente com facilidade. Estas medidas τ de Kendall são usadas para definir as estruturas de dependência das cópulas e conseqüentemente os valores dos parâmetros θ para cada uma delas. Na figura 4.2 pode ser visto como cópulas distintas resultam em distribuições conjuntas totalmente diferentes, mas com as mesmas distribuições marginais e medidas de dependência τ de Kendall. Algumas cópulas têm estruturas de dependência similares mesmo para alguma medida de dependência igual ou extrema, como por exemplo as cópulas Gaussiana e t-Student com valores altos para o grau de liberdade. Logo, a escolha da cópula tem papel importante na modelagem dos dados, que em alguns casos podem apresentar dificuldades. Por outro lado, as distribuições marginais também são importantes para a modelagem dos dados visto que as cópulas somente conseguem capturar a estrutura de dependência entre as marginais, mas não as propriedades das marginais em si tal como assimetria ou caudas pesadas. Na próxima seção apresentamos algumas distribuições marginais a serem usadas em nosso estudo empírico e nas aplicações.

4.2 Inferência e distribuições marginais

Em nossa abordagem bayesiana, estamos interessados na distribuição a posteriori conjunta dos parâmetros condicional aos dados. Formalmente, precisamos $\pi(\Psi) \propto L(\Psi)p(\Psi)$ onde Ψ é um vetor de parâmetros incluindo os da cópula e das distribuições marginais, enquanto $\pi(\theta, \Psi)$, $L(\theta, \Psi)$ e $p(\theta, \Psi)$ são a distribuição

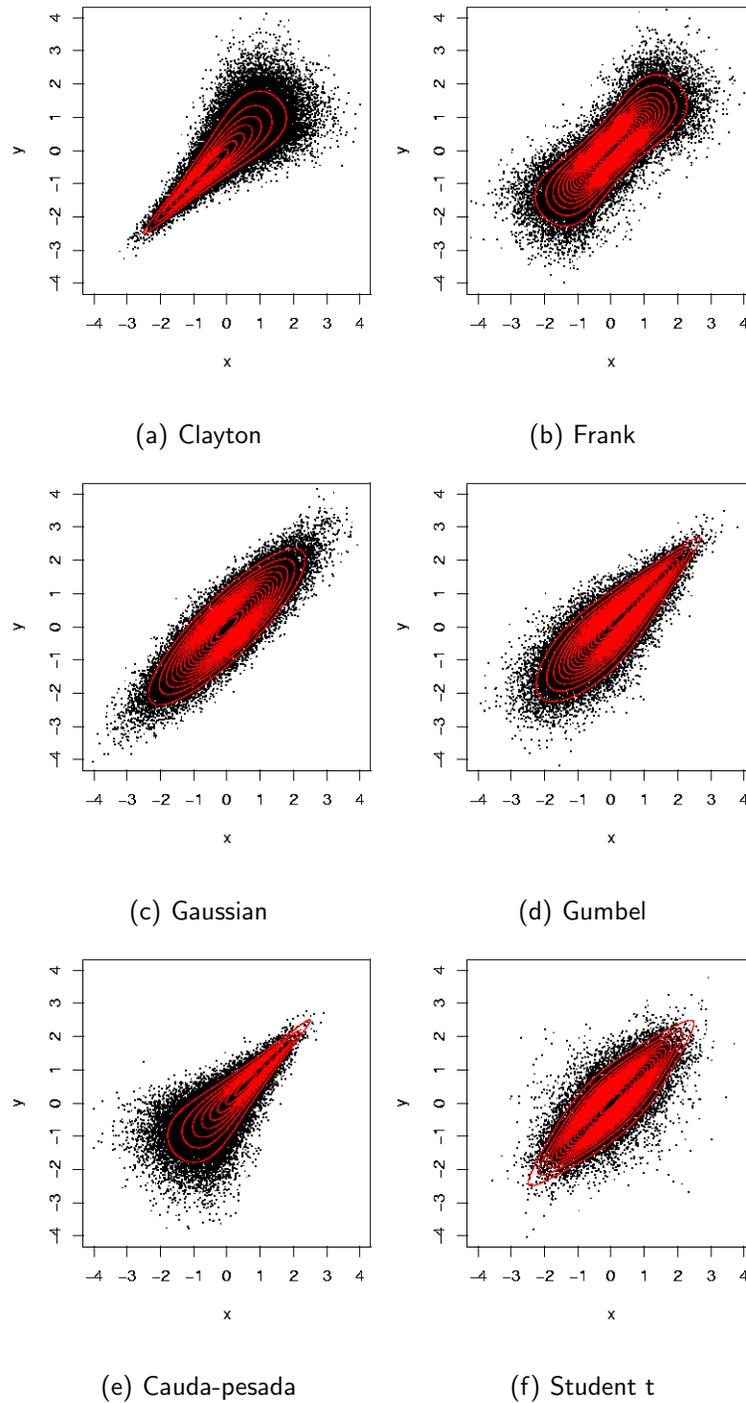


Figura 4.2: Dados simulados e contorno para as 6 famílias de cópulas com distribuições marginais normais padrão e τ de Kendall igual a $2/3$

a posteriori, a função de verossimilhança e a distribuição a priori, respectivamente. Nos concentraremos nas distribuições bivariadas dadas pelas cópulas apresentadas na tabela 4.1. A f.d.p de uma variável aleatória contínua bivariada (X_1, X_2) é dada por

$$h(x_1, x_2 | \theta, \Psi) = c(H_1(x_1 | \Psi), H_2(x_2 | \Psi) | \Psi)h_1(x_1 | \Psi)h_2(x_2 | \Psi) \quad (4.3)$$

onde h_i , $i = 1, 2$ e H_j , $j = 1, 2$ representam respectivamente as f.d.p e as f.d.a das distribuições marginais. Como C é uma f.d.a, então c , chamada de densidade cópula, é a f.d.p calculada como a derivada mista em x_1 e x_2 . Na tabela 4.3 pode ser encontrada as 6 densidades cópulas das famílias dadas na tabela 4.1. Entretanto, devemos lembrar que nem todas as cópulas conhecidas possuem uma densidade cópula, como por exemplo a cópula comonotônica que não possui derivada mista.

Seja $((x_{11}, x_{21}), \dots, (x_{1n}, x_{2n}))$ uma amostra aleatória de tamanho $n \geq 1$ de dados i.i.d da distribuição bivariada com f.d.p dadas por (4.3). Logo, a função de verossimilhança é dada por

$$L(\Psi) = \prod_{i=1}^n c(H_1(x_{1i} | \Psi), H_2(x_{2i} | \Psi) | \Psi)h_1(x_{1i} | \Psi)h_2(x_{2i} | \Psi) \quad (4.4)$$

Tabela 4.3: Definição das densidades cópulas usadas nesta trabalho. $\phi_{(X,Y)}(\cdot)$, $\phi(\cdot)$, $\Phi(\cdot)$, $t_{(X,Y),\nu}(\cdot)$, $t_\nu(\cdot)$ e $T_\nu(\cdot)$ são as f.d.p bivariada com coeficiente de correlação θ , a univariada e a f.d.a de variáveis aleatórias normais padrão, e t-Student padrão, respectivamente.

Cópula	$c(u, v \theta) = \frac{\partial^2 C(u, v \theta)}{\partial u \partial v}$
Clayton	$(1 + \theta)(uv)^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-(1/\theta+2)}$
Frank	$\frac{-\theta e^{-\theta(u+v)}(e^{-\theta} - 1)}{[e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v} + e^{-\theta}]^2}$
Gaussian	$\frac{\phi_{(X,Y)}(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi_X(\Phi^{-1}(u))\phi_Y(\Phi^{-1}(v))}$
Gumbel	$(uv)^{-1}(-\ln(u))^{\theta-1}(-\ln(v))^{\theta-1}(\theta - 1 + [(-\log u)^\theta + (-\log v)^\theta]^{1/\theta})$ $\times [(-\log u)^\theta + (-\log v)^\theta]^{1/\theta-2} \exp \left\{ -[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta} \right\}$
Cauda-pesada	$(1 + 1/\theta)[(1 - u)(1 - v)]^{-(1/\theta+1)}[(1 - u)^{-1/\theta} + (1 - v)^{-1/\theta} - 1]^{-(\theta+2)}$
Student	$\frac{t_{(X,Y),\nu}(T_\nu^{-1}(u), T_\nu^{-1}(v))}{t_{X,\nu}(T_\nu^{-1}(u))t_{Y,\nu}(T_\nu^{-1}(v))}$

4.2.1 Distribuições marginais

Agora, precisamos definir as distribuições marginais para X_1 e X_2 . Para nossos estudos empíricos escolhemos 3 diferentes distribuições marginais, que podem ser usados em muitas aplicações ou pelo menos são capazes de capturar importantes características das distribuições marginais, como por exemplo assimetrias. Nossa primeira escolha é a distribuição normal, por ser conhecida e encontrada em diversas aplicações na literatura. Assim, se X tem distribuição normal com média μ e variância σ^2 , sua f.d.p é dada por

$$\phi(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

e sua f.d.a por

$$\Phi(x|\mu, \sigma^2) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(z - \mu)^2\right\} dz.$$

Nossa segunda escolha é a distribuição exponencial, muitas vezes encontrada em modelos de análise de sobrevivência. Logo, se X tem distribuição exponencial com média $1/\lambda$, então sua f.d.p é dada por

$$h(x|\lambda) = \lambda \exp\{-\lambda x\}$$

e sua f.d.a por

$$H(x|\lambda) = 1 - \exp\{-\lambda x\}.$$

A última escolha é uma versão assimétrica da distribuição normal. Aqui, seguimos a abordagem de Fernández e Steel (1998) para introduzir assimetria em distribuições com f.d.p simétricas, como por exemplo a distribuição normal. A idéia principal é dada por

$$h(x|\gamma) = \frac{2\gamma}{\gamma^2 + 1} [g(x\gamma)I_{(-\infty, 0)} + g(x\gamma^{-1})I_{[0, \infty)}] \quad (4.5)$$

onde $g(\cdot)$ pode ser qualquer f.d.p simétrica em torno de 0 e $\gamma \in (0, \infty)$. Os momentos de ordem k são dados por

$$EX^k = \frac{\gamma^{2(k+1)} + (-1)^k}{\gamma^k(\gamma^2 + 1)} \mu_k$$

onde

$$\mu_k = \int_0^\infty t^k 2g(t) dt.$$

Veja o apêndice E onde apresentamos algumas características das distribuições normal e t-Student truncadas, e seus respectivos momentos μ_k para $k = 1, 2, 3, 4$. Assim, escolhemos $g(\cdot)$ como a f.d.p

de uma distribuição normal padrão e introduzimos mais dois parâmetros: um de posição, μ , e outro de escala, σ^2 . Entretanto, vale ressaltar que poderíamos ter distribuições marginais com caudas pesadas.

4.2.2 Distribuição a priori

A especificação do modelo é completada definindo a distribuição a priori dos parâmetros, o da cópula e os das distribuições marginais. Todos eles são considerados independentes e com distribuições conhecidas como a normal, gama e inversa gama, mas com valores dos hiperparâmetros que garantem que a distribuição a priori resultante seja relativamente vaga. Por exemplo, temos distribuição normal para os parâmetros de posição tanto da distribuição normal quanto da normal assimétrica com média 0 e variância 10^6 , distribuição a priori gama com média 1 e variância 10^4 para os parâmetros de assimetria e inversa gama para os parâmetros de escala. Além disso, distribuições a priori normal, gama ou gama truncada com variância da ordem de 10^4 são escolhidas para os parâmetros das cópulas, dependendo exclusivamente do domínio de cada uma delas, exceto para as cópulas Gaussiana e Student onde distribuições uniforme no intervalo (0,1) são escolhidas. Para a cópula Student, fixamos os graus de liberdade em 4 para todas as aplicações, incluindo os estudos de Monte Carlo. A única razão para isto é considerar que todas as cópulas possuem um único parâmetro.

Como as cópulas e suas f.d.p têm geralmente formas fechadas não usuais, como por exemplo as dadas nas tabelas 4.1 e 4.3, as distribuições a posteriori dos modelos também não possuem formas fechadas conhecidas, então métodos de Monte Carlo via cadeias de Markov são utilizados para obter amostras das distribuições a posteriori. Em especial, escolhemos o amostrador de Gibbs e na maioria dos casos usamos o amostrador da fatia para cada condicional completa, exceto quando algumas delas possuem forma fechada conhecida. Estas escolhas conduzem a uma taxa de convergência rápida e produz uma amostra da posteriori com relativa baixa autocorrelação para nossos dados artificiais e para os dados reais na aplicação da seção 4.3.

4.3 Aplicações

Nesta seção apresentamos alguns estudos de Monte Carlo para mostrar a performance do DIC e suas alternativas na seleção de modelos ou cópulas, considerando um conjunto finito de modelos. Começamos com os dados artificiais em 3 estudos de Monte Carlo com as 6 famílias de cópulas apresentadas na tabela 4.1 e as 3 distribuições marginais dadas na seção 4.2. Nesses estudos, ambas as marginais pertencem a mesma classe de distribuição, isto é, ambas são normais, exponenciais ou normais assimétricas. Finalmente, aplicamos os mesmos modelos de cópulas com marginais normais assimétricas aos dados de atuária de requerimentos de indenizações de seguros (Frees e Valdez, 1998).

4.3.1 DIC e seleção de cópulas

Como mencionamos, seleção de modelos em cópulas pode ser uma tarefa difícil pois algumas cópulas apresentam comportamentos similares, como os visto na figura 4.2. As duas primeiras escolhas em muitos casos são o critério de informação de Akaike (AIC) e o critério de informação bayesiano (BIC). Como alternativa, Chen e Fan (2005) apresentam o teste da razão da pseudo-verossimilhança, Genest, Quessy, e Rémilliard (2005) propõem um teste de ajuste baseado no processo de τ de Kendall, enquanto que Dobrić e Schmidt (2004) apresentam um teste de ajuste baseado no teste qui-quadrado p dimensional. Outros testes podem ser encontrados em Evin (2004) e Fermanian (2005). Do ponto de vista bayesiano, probabilidades a posteriori, ou pesos, podem ser calculados para cada modelo (veja por exemplo Arakelian e Dellaportas (2006)). Huard et al. (2006) propõem um método para selecionar a cópula mais provável dado um conjunto finito delas, mas a abordagem deles implica expressar os dados em percentis, o que por sua vez implica em transformação dos dados. Além disso, probabilidades a priori devem ser dadas para cada modelo em questão. Como alternativa a todos esses métodos descritos acima, queremos mostrar através de estudos de Monte Carlo que o DIC e suas alternativas AIC, EAIC, BIC e EBIC calculadas no enfoque bayesiano (veja o apêndice D para uma revisão destes métodos) funcionam razoavelmente bem como métodos de seleção de modelos em cópulas. Sua implementação computacional é fácil e nenhuma transformação dos dados se faz necessária. Por outro lado, todas as famílias de cópulas devem possuir suas respectivas densidades bem definidas.

4.3.2 Características gerais dos estudos de Monte Carlo

Antes de começar a falar dos resultados dos estudos de Monte Carlo propriamente dito, descrevemos as quantidades e as características de nossa simulação tanto da geração dos dados artificiais quanto da distribuição a posteriori. Primeiro, todos os algoritmos e resultados foram programados e obtidos usando linguagem de programação C com a biblioteca científica GNU (Galassi et al., 2005), um gerador de números aleatórios chamado *Mersenne Twister - MT19937* (Matsmoto e Nishimura, 1998) e o programa estatístico R (R Development Core Team, 2006), tudo em plataforma GNU/Linux com processador pentium IV 1.6Ghz e 512Mb de memória RAM. Para a cópula Gaussiana com marginais normais, a amostra foi obtida através da distribuição normal bivariada com o algoritmo disponível no R. Para a cópula Frank usamos o método da inversão descrito por Nelsen (1999). Para todos os outros casos, usamos o algoritmo SIR (Gammerman e Lopes, 2006) com uma taxa de reamostragem de 5%. Por exemplo, resultados dos dados artificiais simulados das 6 famílias de cópulas com marginais normais usando essas técnicas podem ser vista na figura 4.2. Para a inferência na distribuição a posteriori e a seleção de modelos, o amostrador de Gibbs foi usado e testado para todos os casos com uma rápida taxa de convergência mesmo para valores iniciais longe dos usados no processo de geração dos dados artificiais. Assim, em nosso estudo de simulação, geramos 1.500 iterações do algoritmo de Gibbs,

dispensamos as 500 primeiras iterações e ficamos com uma amostra de tamanho 1.000 para todos os casos. As autocorrelações das amostras foram pequenas em todos os casos. Além disso, consideramos 500 replicações de dados para 3 tamanhos de amostras, $n = \{100, 200, 500\}$, e 2 valores para a medida de dependência τ de Kendall, $\tau = \{1/3, 2/3\}$. Abaixo, apresentamos os resultados principais baseados no DIC (veja o apêndice D) junto com outros detalhes. Resultados baseados no AIC, EAIC, BIC e EBIC conduzem as mesmas conclusões, como podem ser visto nas tabelas F.1, F.2 e F.3 do apêndice F.

4.3.3 Marginais com distribuições normais

O primeiro estudo de simulação foi conduzido com as 6 famílias de cópulas da tabela 4.1 e marginais normais. Os tempos de computação para todas as 500 replicações foram 18, 36 e 90 horas para as amostras de tamanho 100, 200 e 500, respectivamente, e para ambas as medidas τ de Kendall escolhidas. A tabela 4.4 mostra um sumário dos resultados da simulação pelo número de identificação correta do modelo de cópulas usando DIC para as 500 replicações. Resultados similares baseados no AIC, EAIC, BIC e EBIC podem ser visto na tabela F.1 do apêndice F.

Tabela 4.4: Número de identificações corretas entre as 500 replicações para o modelo de cópulas com distribuições marginais normais.

Cópula	τ	1/3			2/3			
		n	100	200	500	100	200	500
Clayton			457	488	500	497	500	500
Frank			312	384	475	471	494	500
Gaussian			252	372	462	427	475	499
Gumbel			191	319	448	426	481	500
Cauda-pesada			366	407	473	469	497	500
Student			317	413	488	387	455	496

Observe em primeiro lugar que a medida que o tamanho das amostras cresce temos melhores identificações dos modelos baseadas no DIC. Assim, acreditamos que a medida que o tamanho da amostra cresce a probabilidade de identificação correta dos modelos se aproxime de 1. Entretanto, vale a pena ressaltar que mesmo para tamanhos de amostra $n = 500$ as cópulas Gaussiana e Student identificam uma a outra com certa frequência porque a primeira pode ser vista como caso particular da segunda, podendo apresentar comportamento semelhantes. Contudo, em nosso exemplos fixamos o grau de liberdade da cópula Student em 4, para que fosse considerado diferente da cópula Gaussiana. O mesmo

comportamento ocorre com a medida τ de Kendall, isto é, valores altos conduzem a uma melhor identificação do modelo pois o comportamento da estrutura de dependência fica bem definida neste caso. Além disso, algumas cópulas têm comportamentos similares quando τ se aproxima de 0, o que pode levar a uma identificação errônea do modelo de cópulas, mesmo para amostras relativamente grandes. Assim, a facilidade com que os modelos de cópulas são identificados depende diretamente da presença de cópulas similares, do grau de dependência das marginais e do tamanho da amostra. Cópulas que são similares geralmente possuem pesos iguais, sendo sua identificação difícil. Por outro lado, cópulas com comportamentos bem diferentes das outras são facilmente identificadas, por exemplo a cópula Clayton. A figura 4.2 mostra como a cópula Clayton de alguma maneira possui comportamento diferente das demais. Isto explica porque em todos os casos, a Clayton é a cópula com maior número correto de identificações. Logo, o método de seleção de modelos DIC, e suas alternativas AIC, EAIC, BIC e EBIC, têm vantagens interessantes sobre alguns testes estatísticos comuns na literatura. Ele é fácil de implementar numericamente, e a julgar pelas nossas simulações, provêm identificações confiáveis, mesmo para tamanhos de amostra pequeno. Além disso, pode ser aplicado a qualquer família de cópulas, desde que possuam suas respectivas densidades de cópulas de tal forma que possam ser calculadas numericamente.

Na tabela 4.5 apresentamos a média e o erro quadrático médio em parênteses de todas as 500 replicações, considerando somente os modelos simulados corretamente identificados, para os dados artificiais com $n = 500$ e $\tau = 2/3$, tomando como estimativas pontuais a média a posteriori dos parâmetros

Tabela 4.5: Propriedades frequentista dos modelos corretamente identificados - marginais normais.

Cópula	θ	τ	μ_X	σ_X^2	μ_Y	σ_Y^2
Clayton	4,0400 (0,0853)	0,6674 (0,0003)	-0,0008 (0,0021)	1,0092 (0,0033)	0,0005 (0,0020)	1,0145 (0,0032)
Frank	10,037 (0,2741)	0,6657 (0,0002)	-0,0008 (0,0019)	1,0022 (0,0037)	-0,0002 (0,0017)	1,0015 (0,0033)
Gaussian	0,8643 (0,0001)	0,6651 (0,0002)	0,0042 (0,0017)	1,0062 (0,0036)	0,0018 (0,0017)	1,0065 (0,0043)
Gumbel	3,0027 (0,0219)	0,6654 (0,0003)	-0,0008 (0,0018)	1,0035 (0,0036)	-0,0005 (0,0020)	1,0050 (0,0036)
Cauda-pesada	0,2527 (0,0003)	0,6650 (0,0002)	-0,0005 (0,0021)	0,9980 (0,0028)	-0,0021 (0,0020)	0,9999 (0,0031)
Student	0,8634 (0,0002)	0,6642 (0,0003)	-0,0047 (0,0020)	1,0059 (0,0040)	-0,0037 (0,0021)	1,0054 (0,0039)

de cada modelo. Observe que os parâmetros θ , μ_X , σ_X , μ_X e σ_X são bem estimados em todos os casos no sentido que eles estão próximos aos valores usados na geração dos dados artificiais. O mesmo acontece para os outros casos, mas com erro quadrático médio maior para as amostra de tamanho menores.

4.3.4 Marginais com distribuições exponenciais

Um outro estudo de Monte Carlo, talvez tão importante quanto o primeiro devido sua aplicação em modelos de análise de sobrevivência, foi conduzido com as mesmas 6 cópulas da tabela 4.1 e com ambas as marginais tendo distribuições exponenciais. Os tempos de computação para todas as 500 replicações foram em média 12, 24 e 55 horas para as amostras de tamanho 100, 200 e 500, respectivamente, e ambas as medidas τ de Kendall escolhidas. A tabela 4.6 traz os resultados da simulação pelo número correto de identificação dos modelos usando DIC. Resultados similares baseados no AIC, EAIC, BIC e EBIC podem ser visto na tabela F.2 do apêndice F.

Tabela 4.6: Número de identificações corretas entre as 500 replicações para o modelo de cópulas com distribuições marginais exponenciais.

Cópula	τ	1/3			2/3			
		n	100	200	500	100	200	500
Clayton			447	491	500	500	500	500
Frank			312	395	480	471	497	500
Gaussian			293	377	466	423	468	499
Gumbel			189	313	449	427	486	500
Cauda-pesada			396	430	471	486	500	500
Student			329	423	487	355	442	496

Como no primeiro caso, observe que a medida que o tamanho da amostra cresce temos uma melhor identificação dos modelos baseados no DIC. Uma vez mais as cópulas Gaussiana e Student identificam uma a outra com certa frequência. O mesmo acontece com o τ de Kendall, a medida que aumenta a dependência temos uma melhor identificação dos modelos. A cópula Clayton de novo é a com maior número correto de identificações. A julgar por esses resultados, o método de seleção de modelos DIC, e suas alternativas AIC, EAIC, BIC e EBIC, conseguem identificar razoavelmente bem os modelos simulados.

Na tabela 4.7 apresentamos a média e o erro quadrático médio com as mesmas especificações da simulação anterior. Novamente, os resultados mostram que os parâmetros dos modelos são bem

Tabela 4.7: Propriedades frequentista dos modelos corretamente identificados - marginais exponenciais.

Cópula	θ	τ	λ_X	λ_Y
Clayton	4,0098 (0,0402)	0,6665 (0,0001)	1,0045 (0,0015)	1,0042 (0,0014)
Frank	10,035 (0,2224)	0,6658 (0,0002)	1,0000 (0,0017)	1,0001 (0,0016)
Gaussian	0,8635 (0,0001)	0,6639 (0,0001)	1,0035 (0,0020)	1,0031 (0,0020)
Gumbel	2,9933 (0,0171)	0,6646 (0,0002)	1,0076 (0,0019)	1,0085 (0,0020)
Cauda-pesada	0,2528 (0,0003)	0,6650 (0,0003)	1,0043 (0,0019)	1,0050 (0,0019)
Student	0,8626 (0,0001)	0,6628 (0,0002)	1,0006 (0,0020)	1,0005 (0,0019)

identificados. O mesmo acontece nos outros casos, mas com um erro quadrático médio maior para tamanhos de amostra menores.

4.3.5 Marginais com distribuições normais assimétricas

Agora, consideramos marginais com distribuições normais assimétricas ao invés de normais ou exponenciais. Estas distribuições são importantes em aplicações reais onde existe a necessidade de medir e testar a assimetria, como por exemplo na análise de retornos financeiros. Mesmo neste caso onde temos mais parâmetros a serem estimados, os resultados conduzem praticamente as mesmas conclusões das duas primeiras. Os tempos de computação para todas as 500 replicações foram 23, 44 e 103 horas para as amostras de tamanho 100, 200 e 500, respectivamente, e para ambos valores de τ de Kendall. A tabela 4.8 mostra os resultados da simulação pelo número de identificação correta dos modelos usando DIC. Resultados similares baseados no AIC, EAIC, BIC e EBIC podem ser visto na tabela F.3 do apêndice F.

Note que temos resultados similares as simulações anteriores, isto é, a medida que o tamanho da amostra cresce temos melhor identificação dos modelos, as cópulas Gaussiana e Student identificam uma a outra com certa frequência e a medida que o τ de Kendall aumenta temos também melhor identificação

Tabela 4.8: Número de identificações corretas entre as 500 replicações para o modelo de cópulas com distribuições marginais normais assimétricas.

Cópula	τ	1/3			2/3			
		n	100	200	500	100	200	500
Clayton			379	480	500	459	499	500
Frank			306	400	476	440	490	500
Gaussian			235	364	483	344	461	494
Gumbel			182	287	432	317	455	499
Cauda-pesada			372	414	462	479	488	500
Student			280	409	492	329	439	497

dos modelos. Assim, o método DIC é ainda capaz de identificar a cópula correta com alta probabilidade mesmo para distribuições marginais assimétricas com 3 parâmetros. Apresentamos a média e o erro quadrático médio com as mesmas especificações da tabela 4.9. Observe que os parâmetros são bem

Tabela 4.9: Propriedades frequentista dos modelos corretamente identificados. Marginais normais assimétricas.

Cópula	θ	τ	μ_X	σ_X^2	γ_X	μ_Y	σ_Y^2	γ_Y
Clayton	4,062 (0,085)	0,669 (0,000)	-0,003 (0,009)	1,006 (0,004)	1,506 (0,007)	-0,001 (0,008)	1,005 (0,004)	1,502 (0,007)
Frank	10,05 (0,259)	0,666 (0,000)	0,001 (0,013)	1,000 (0,007)	1,511 (0,014)	0,004 (0,011)	0,996 (0,007)	1,509 (0,012)
Gaussian	0,865 (0,000)	0,665 (0,000)	-0,009 (0,009)	0,994 (0,005)	1,514 (0,007)	-0,014 (0,011)	0,989 (0,006)	1,523 (0,000)
Gumbel	3,021 (0,022)	0,667 (0,000)	0,001 (0,011)	1,002 (0,006)	1,507 (0,009)	0,006 (0,011)	1,001 (0,007)	1,503 (0,010)
Cauda-pesada	0,249 (0,000)	0,668 (0,000)	0,002 (0,011)	1,003 (0,008)	1,515 (0,010)	-0,007 (0,010)	1,002 (0,007)	1,522 (0,010)
Student	0,863 (0,000)	0,663 (0,000)	0,000 (0,012)	0,993 (0,006)	1,512 (0,008)	0,003 (0,010)	0,995 (0,005)	1,507 (0,007)

identificados. O mesmo acontece para os outros casos, mas com erro quadrático médio maior para amostra de tamanhos menores.

Todos os resultados baseados no DIC são corroborados pelo AIC, EAIC, BIC e EBIC, conforme resultados apresentados no apêndice F. Também, os parâmetros dos modelos foram recuperados em média, mesmo para amostras de tamanho pequeno.

4.3.6 Requerimentos de indenizações de seguros

A figura 4.3 mostra o conjunto de dados correspondentes aos requerimentos de indenizações de seguros com 1.500 observações de despesas ajustadas de perdas alocadas (ALAE) e pagamentos de indenizações (*Loss*), ambos na escala logarítmica. O sumário estatísticos dessas duas medidas mostram que ambas parecem ser assimétricas para a esquerda com um coeficiente de correlação de 0,425 e a medida τ de Kendall igual a 0,302. Uma melhor descrição dos dados pode ser encontrada em Frees e Valdez (1998). Para tentar capturar e medir a dependência entre essas duas variáveis, assim como a assimetria em suas marginais, consideramos as 6 famílias de cópulas da tabela 4.1 com ambas as marginais sendo normais assimétricas.

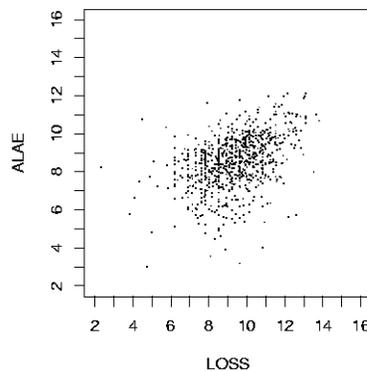


Figura 4.3: Requerimentos de indenizações de seguros (Loss e ALAE).

A tabela 4.10 traz os valores do DIC e os outros critérios para todos os modelos de cópulas ajustados. Para o AIC, EAIC, BIC e EBIC o número de parâmetros foi fixado em 7, com 3 de cada marginal mais 1 da cópula. Assim, baseado nessas medidas, vemos que a cópula Gumbel tem um melhor ajuste quando comparado com as outras 5 famílias de cópulas. Claramente, a cópula Clayton é a que tem o pior ajuste quando comparada as outras, pela estrutura de dependência que o dado exhibe. Observe, entretanto, que o modelo de cópula cauda-pesada tem valores próximos a Gumbel visto que ambas tentam capturar estruturas similares de dependência.

Podemos ver na tabela 4.11 um sumário estatístico da distribuição a posteriori para o modelo de cópula Gumbel com marginais normais assimétricas. Observe que a variável *Loss* tem alguma probabili-

Tabela 4.10: DIC, pd, AIC, EAIC, BIC e EBIC para os 6 modelos de cópulas para os dados dos pedidos de perda da companhia de seguro.

Cópula	DIC	pd	AIC	EAIC	BIC	EBIC
Clayton	5376	6,91	5376	5383	5413	5420
Frank	5175	7,08	5175	5182	5212	5219
Gaussian	5174	7,03	5174	5181	5211	5218
Gumbel	5128	7,22	5127	5134	5164	5172
Cauda-pesada	5139	7,07	5139	5146	5176	5183
Student	5189	6,89	5190	5196	5227	5234

dade de ser simétrica, mas a variável ALAE tem alta probabilidade de ser assimétrica a esquerda tendo-se como base o intervalo de credibilidade do parâmetro γ e sua média abaixo de 1. O parâmetro da cópula θ por sua vez está concentrado em valores ao redor de 1.4 indicando alguma dependência positiva entre as variáveis *Loss* e ALAE. A distribuição a posteriori da medida τ de Kendall da cópula Gumbel pode ser sumarizada por 0,306, 0,016, 0,280, 0,305 e 0,334, correspondendo a média a posteriori, desvio padrão, o percentil de 2.5%, a mediana e o percentil de 97.5%, respectivamente. Logo, a cópula Gumbel foi capaz de identificar uma dependência significativa entre as variáveis *Loss* e ALAE. Além disso, o modelo com as marginais normais assimétricas foi capaz de identificar e descrever a assimetria das variáveis. O DIC e suas alternativas AIC, EAIC, BIC e EBIC conduzem aos mesmos resultados na seleção da cópula com melhor ajuste, neste caso a cópula Gumbel.

Tabela 4.11: Sumário da distribuição a posteriori do modelo de cópula Gumbel baseado numa amostra de tamanho 1000. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetro	Loss					ALAE				
	Média	D.P	2,5%	Mediana	97,5%	Média	D.P	2,5%	Mediana	97,5%
μ	9,475	0,109	9,294	9,474	9,654	8,895	0,070	8,776	8,896	9,011
σ^2	2,723	0,101	2,566	2,720	2,896	1,956	0,074	1,832	1,957	2,072
γ	0,960	0,036	0,901	0,959	1,020	0,841	0,025	0,802	0,841	0,883
θ	1,442	0,034	1,389	1,439	1,501	-	-	-	-	-

4.4 Discussão final

Neste capítulo mostramos que o DIC produz bons resultados quando aplicado a seleção de modelos em cópulas, isto é, identifica os modelos simulados com alta probabilidade, dependendo apenas do tamanho da amostra e do grau de dependência entre as marginais. Além disso, pode ser utilizado numa grande variedade de aplicações sem qualquer tipo de estimação em 2 passos ou a necessidade de definir probabilidades a priori para cada modelo. Outros critérios baseados na função desvio, AIC, EAIC, BIC e EBIC, também concordam com o DIC tendo, portanto, performance similar. Todos estes métodos são fáceis de implementar em algoritmos de Monte Carlo via cadeias de Markov. Nossas implementações mostraram ter rápida taxa de convergência. Entretanto, o DIC e os outros critérios somente podem ser calculados quando as densidades cópulas existem, implicando que a f.d.p conjunta exista. Todo esse estudo pode ser considerado um passo inicial para uma melhor compreensão de cópulas e das idéias a serem apresentadas no próximo capítulo. Todos as rotinas e os códigos em C e R deste capítulo estão disponíveis sob requisição.

Capítulo 5

MISTURA EM CÓPULAS E ASSIMETRIA

A objetivo neste capítulo é mostrar como obter novas cópulas através de misturas baseadas em cópulas e suas distribuições marginais, e como estender alguns resultados encontrados na literatura. Tais resultados são importantes e motivados por aplicações em modelos hierárquicos. Nestes modelos geralmente é necessário o conhecimento das distribuições resultantes de níveis superiores da hierarquia após a integração dos níveis inferiores. Além disso, a inclusão de misturas discretas permite modelar diversas estruturas de dependência, sendo assim importante explorar algumas de suas propriedades. Propriedades das cópulas resultantes da mistura são mostradas através de algumas medidas de dependência. Avaliações numéricas dessas medidas são dadas através de integração Monte Carlo. Joe (1997) apresenta resultados de mistura baseado em distribuições condicionais e transformadas de Laplace, enquanto Nelsen (1999) apresenta uma mistura nos parâmetros das cópulas. Seguimos utilizando as mesmas 6 famílias de cópulas da tabela 4.1. Para as marginais, escolhemos distribuições assimétricas (Fernández e Steel, 1998) dadas por (4.5). Um exemplo didático com a cópula Clayton com marginais exponenciais e a cópula produto com marginais gama ilustra o procedimento. Nas aplicações utilizamos as 6 cópulas da tabela 4.1 mais a cópula produto, onde a mistura é feita na escala da distribuição normal assimétrica com uma inversa gama, obtendo assim uma t-Student assimétrica, além de outra cópula.

A literatura de cópulas com métodos bayesianos de estimação ainda é pequena. Algumas referências nesta área incluem Nicoloutsopoulos (2005), Lopes e Ausin (2006), Huard et al. (2006), Arakelian e Dellaportas (2006) e Pitt, Chan, e Kohn (2006). Uma outra contribuição neste trabalho é apresentada no capítulo 4: a estimação de todas as quantidades desconhecidas dos modelos através do enfoque bayesiano. Aplicamos técnicas de Monte Carlo via cadeias de Markov na distribuição a posteriori. Seleção de modelo é feita através do DIC e suas alternativas AIC, EAIC, BIC e EBIC. Aplicações reais com dados de atuária e de finanças ajudam a ilustrar nossa proposta.

5.1 Propriedades gerais

Como na seção 4.1, cópulas são funções de distribuições multivariadas em $I^d \equiv [0, 1]^d$ com todas as marginais sendo uniformes no intervalo (0,1). Além disso, se $X \equiv (X_1, \dots, X_d) \in \mathbb{R}^d$ tem distribuição multivariada contínua com $H(x_1, \dots, x_d) \equiv P(X_1 \leq x_1, \dots, X_d \leq x_d)$, então pelo teorema de Sklar (Sklar, 1959) existe uma única função cópula $C : [0, 1]^d \rightarrow [0, 1]$ de H tal que $H(x_1, \dots, x_d) =$

$C(H_1(x_1), \dots, H_d(x_d))$ para todo $x \equiv (x_1, \dots, x_d) \in \mathbb{R}^d$ e marginais contínuas, onde $H_i(x_i) \equiv P(X_i \leq x_i)$, $x_i \in \mathbb{R}$ para $i = 1, 2, \dots, d$. Por outro lado, se C é uma cópula em I^d e $H_1(x_1), \dots, H_d(x_d)$ são funções de distribuição acumuladas (f.d.a) em \mathbb{R} , então $H(x_1, \dots, x_d) = C(H_1(x_1), \dots, H_d(x_d))$ é uma f.d.a em \mathbb{R}^d com distribuições marginais unidimensionais $H_1(x_1), \dots, H_d(x_d)$.

5.2 Medidas de dependência

Cópulas apresentam diversas estruturas de assimetria e dependência. Uma maneira de caracterizá-las é através de medidas de dependência. Na seqüência apresentamos algumas medidas de dependência encontradas na literatura. Essas medidas serão utilizadas na seção 5.3 para a caracterização das cópulas resultantes das misturas.

τ de Kendall

Vimos na seção 4.1 que a medida de dependência τ de Kendall é uma alternativa ao coeficiente de correlação. A medida τ de Kendall é dada em (4.2). Ela está limitada ao intervalo $[-1, 1]$. Entretanto, a maioria das funções cópulas não têm forma fechada para a medida τ de Kendall. Algumas exceções são Clayton, Gumbel, Gaussian e Student (veja a tabela 4.2). Por outro lado, a cópula Frank não possui forma fechada simples sendo necessário avaliações numéricas da mesma. Esta é a principal medida utilizada em nossas aplicações.

ρ de Spearman

Outra medida de dependência importante é a ρ de Spearman (Kruskal, 1958). Como a medida τ de Kendall, a ρ de Spearman pode ser interpretada como a probabilidade de concordância menos a probabilidade de discordância para dois vetores aleatórios. Ela está limitada ao intervalo $[-1, 1]$. Ela pode ser escrita em termos da função cópula por

$$\rho = 12 \int_I \int_I C(u, v) du dv - 3 = 12 \int_{I^2} uv dC(u, v) - 3. \quad (5.1)$$

Entretanto, para muitas cópulas as medidas de ρ de Spearman não têm expressões em forma fechada, sendo necessário avaliações numéricas.

Índices de dependência de cauda

Os índices de cauda bivariados, chamados índice de cauda inferior e superior, são definidos através de probabilidades condicionais nas caudas dos quadrantes inferior esquerdo e superior direito, respectivamente. Eles estão relacionados a eventos extremos. O índice de cauda inferior é dado por

$$\lambda_L = \lim_{u \rightarrow 0} P(U < u \mid V < u) = \lim_{u \rightarrow 0} \frac{C(u, u)}{u} \quad (5.2)$$

e o índice de cauda superior é dado por

$$\lambda_U = \lim_{u \rightarrow 0} P(U > u \mid V > u) = \lim_{u \rightarrow 0} \frac{1 - 2u - C(u, u)}{1 - u}. \quad (5.3)$$

Outros índices de dependência

Schweizer e Wolff (1981) mostram que qualquer distância L_p padronizada entre as superfícies $z = C(u, v)$ e $z = uv$ deve resultar em uma medida não paramétrica simétrica de dependência, isto é,

$$D(q) = \left[\kappa_q \int_I \int_I |C(u, v) - uv|^q dudv \right]^{1/q} \quad (5.4)$$

onde $1 \leq q < \infty$ e κ_q é uma constante tal que $D(q) = 1$ quando $C = C^-$ ou $C = C^+$. Por exemplo quando $q = 1$, $D(1)$ resulta na medida chamada de Schweizer e Wolff σ dada por

$$\sigma = D(1) = 12 \int_I \int_I |C(uv) - uv| dudv \quad (5.5)$$

e $q = 2$ resulta na raiz quadrada no índice de dependência Φ dado por

$$\Phi = D(2) = \left[90 \int_I \int_I |C(uv) - uv|^2 dudv \right]^{1/2}. \quad (5.6)$$

Outras duas medidas de dependência importantes e interessantes são Gini γ e Blomqvist β (veja Nelsen (1999)). A primeira pode ser vista como a diferença padronizada do valor esperado da distância L_1 de $z = uv$ para C^- e C^+ que é dada por

$$\gamma = 2 \int_{I^2} (|u + v - 1| - |u - v|) dC(u, v) = 4 \int_I [C(u, 1 - u) + C(u, u)] du - 2 \quad (5.7)$$

e a segunda pode ser vista como a probabilidade de concordância menos a probabilidade de discordância de um vetor aleatório e um ponto fixo, neste caso a mediana. Essa medida também é chamada de coeficiente de correlação medial e é dada por

$$\beta = P[(X - \text{med}(X))(Y - \text{med}(Y)) > 0] - P[(X - \text{med}(X))(Y - \text{med}(Y)) < 0] \quad (5.8)$$

onde $\text{med}(X)$ e $\text{med}(Y)$ são as medianas de X e Y , respectivamente.

5.3 Mistura em cópulas

Nesta seção apresentamos a idéia de mistura em cópulas, como os índices de dependência descritos na seção 5.1 são representados e sugestões de como avaliá-los numericamente usando integração Monte Carlo. Para facilitar a notação, os parâmetros das cópulas serão omitidos.

Sejam X e Y vetores aleatórios. F representará a f.d.a de X dado Y e suas marginais univariadas serão denotadas por F_i , $i = 1, \dots, d$. G e H representarão as f.d.a de Y e X , respectivamente, e suas marginais univariadas serão denotadas por G_j , $j = 1, \dots, p$ e H_i , $i = 1, \dots, d$.

Teorema 5.1 *Seja $X \equiv (X_1, \dots, X_d)$ e $Y \equiv (Y_1, \dots, Y_m)$ vetores aleatórios tal que $X | Y$ tenha f.d.a dada por $C_{X|Y}(F_1(x_1|y), \dots, F_d(x_d|y))$ e Y tenha f.d.a dada por $C_Y(G_1(y_1), \dots, G_p(y_p))$, então X tem f.d.a dada por*

$$C_X(H_1(x_1), \dots, H_d(x_d)) = \int_{\mathbb{R}^p} C_{X|Y}(F_1(x_1|y), \dots, F_d(x_d|y)) dC_Y(G_1(y_1), \dots, G_p(y_p))$$

onde $x \equiv (x_1, \dots, x_d) \in \mathbb{R}^d$ e $y \equiv (y_1, \dots, y_p) \in \mathbb{R}^p$. Além disso, C_X é uma função cópula.

Prova: Segue do teorema de Sklar (Sklar, 1959) e de $H(x) = \int_{\mathbb{R}^p} F(x|y) dG(y)$. Para verificar que C_X é de fato uma função cópula, mostre que $\lim_{u_i \rightarrow 0} C_X(u_1, \dots, u_i, \dots, u_d) = 0$ e $C_X(1, \dots, 1, u_i, 1, \dots, d) = u_i$ para $i = 1, \dots, d$ usando o fato de $C_{X|Y}$ ser uma cópula. As demais propriedades são dadas de forma análoga.

O teorema 5.1 diz que mistura em cópulas resulta em outra cópula. Ainda é possível estender este resultado fazendo mistura nos parâmetros da cópula, C_X ou de maneira equivalente em $C_{X|Y}$ e C_Y . O resultado do teorema 5.1 inclui mistura discretas e/ou contínuas. Além disso, $C_{X|Y}$ e C_Y podem ser quaisquer cópulas. Elas são capazes de determinar a estrutura de assimetria e dependência da cópula resultante C_X , enquanto a escolha das distribuições marginais na mistura são capazes de determinar as características e propriedades das distribuições marginais de X . Um modelo hierárquico em cópulas pode ser uma aplicação da mistura onde a distribuição resultante é requisitada depois da integração de níveis mais baixos da hierarquia.

Teorema 5.2 *Qualquer distribuição marginal de ordem inferior ($k < d$) pode ser obtida como uma mistura da marginal $C_{X|Y}$ de mesma dimensão e da cópula C_Y , isto é,*

$$C_X(H_{s_1}(x_{s_1}), \dots, H_{s_t}(x_{s_t})) = \int_{\mathbb{R}^p} C_{X|Y}(F_{s_1}(x_{s_1}|y), \dots, F_{s_t}(x_{s_t}|y)) dC_Y(G_1(y_1), \dots, G_p(y_p))$$

onde $s_1, \dots, s_t \in \{1, \dots, d\}$.

Prova: Tome os limites $x_\ell \rightarrow \infty$ para todos $\ell \neq s_1, \dots, s_t$, troque a ordem dos limites e integrais visto que são todos bem definidos, então aplique a propriedade de marginalização de uma cópula, $C_{X|Y}$, para chegar ao resultado.

O teorema 5.2 tem alguns casos particulares. Um deles é descrito no seguinte corolário.

Corolário 5.1 *Se $d = p$ e X_i somente depende de Y_i , que por sua vez é independente de todos os outros X_l , $l \neq i$, então*

$$H_i(x_i) = \int_{\mathbb{R}} F_i(x_i|y_i) dG_i(y_i)$$

Na verdade, podemos ver Y_i como um vetor aleatório com as mesmas condições de independência. O corolário 5.1 é importante pois na maioria das aplicações ele indica quais serão as distribuições marginais resultantes. Por exemplo, sejam X_i uma variável aleatória (v.a) normal com média 0 e variância Y_i , e Y_i com distribuição inversa gama com parâmetros $\nu/2$ e $\nu/2$. Então, a distribuição marginal resultante será uma v.a t-Student com ν graus de liberdade, não importa quais cópulas estejam envolvidas na mistura. Tal resultado não depende das cópulas $C_{X|Y}$ e C_Y , mas pode certamente determinar algumas características da cópula C_X . Os resultados mostram a forma geral da mistura, as f.d.a e suas marginais, mas características importantes das cópulas são dadas através das medidas de dependência.

Abaixo, mostramos como os índices de dependência apresentados na seção 5.1 ficam após a mistura e maneiras possíveis de avaliá-las numericamente, especialmente usando integração Monte Carlo.

Teorema 5.3 *A medida τ de Kendall para a cópula resultante C_X , $d = 2$, é dada por*

$$\tau_X = 4 \int_{\mathbb{R}^2} \int_{\mathbb{R}^p} C_{X|Y}(F_{s_1}(x_{s_1}|y), F_{s_2}(x_{s_2}|y)) dC_Y(G_1(y_1), \dots, G_p(y_p)) dC_X(H_{s_1}(x_{s_1}), H_{s_2}(x_{s_2})) - 1$$

onde $s_1, s_2 \in \{1, \dots, d\}$.

Prova: Aplique a definição da medida τ de Kendall em (4.2) para C_X e substitua o integrando em C_X por sua expressão dada no teorema 5.1.

Na maioria dos casos τ_X não tem forma fechada mas pode ser avaliada numericamente, usando por exemplo integração Monte Carlo como abaixo.

$$\tau_X \simeq \frac{4}{L} \sum_{\ell=1}^L \left[\frac{4}{K} \sum_{k=1}^K C_{X|Y}(F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)}), F_{s_2}(x_{s_2}^{(\ell)}|y^{(k)})) \right] - 1. \quad (5.9)$$

com $y^{(k)}$ e $(x_{s_1}^{(\ell)}, x_{s_2}^{(\ell)})$ sendo independentes para todos k e ℓ . Uma abordagem simples neste caso seria a versão amostral da τ de Kendall, que para amostras relativamente pequenas produz boas aproximações. Essa medida é a preferida nas aplicações da seção 5.5.

Teorema 5.4 A medida ρ de Spearman para a cópula resultante C_X , $d = 2$, é dada por

$$\rho_X = 12 \int_{\mathbb{R}^2} H_{s_1}(x_{s_1}) H_{s_2}(x_{s_2}) C_X(H_{s_1}(x_{s_1}), H_{s_2}(x_{s_2})) - 3$$

com

$$H_{s_t}(x_{s_t}) = \int_{\mathbb{R}^p} F_{s_t}(x_{s_t}|y) dC_Y(G_1(y_1), \dots, G_p(y_p)), \quad t = 1, 2$$

onde $s_1, s_2 \in \{1, \dots, d\}$.

Prova: Aplique a definição da medida ρ de Spearman em (5.1) para C_X e substitua H_{s_t} por sua expressão dada no teorema 5.2.

Na maioria dos casos ρ_X não tem forma fechada conhecida mas pode ser avaliada numericamente, usando por exemplo integração Monte Carlo como abaixo.

$$\rho_X \simeq \frac{12}{L} \sum_{\ell=1}^L \left[H_{s_1}(x_{s_1}^{(\ell)}) H_{s_2}(x_{s_2}^{(\ell)}) \right] - 3 \quad (5.10)$$

se H_{s_1} e H_{s_2} estão disponíveis em forma fechada, caso contrário

$$\rho_X \simeq \frac{12}{L} \sum_{\ell=1}^L \left[\frac{1}{K} \sum_{k=1}^K F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)}) \right] \left[\frac{1}{J} \sum_{j=1}^J F_{s_2}(x_{s_2}^{(\ell)}|y^{(j)}) \right] - 3 \quad (5.11)$$

com $y^{(j)}$, $y^{(k)}$ e $(x_{s_1}^{(\ell)}, x_{s_2}^{(\ell)})$ sendo independentes para todos j , k e ℓ . Novamente uma abordagem mais simples seria o cálculo da versão amostral da medida ρ de Spearman, que para amostras relativamente pequenas produz boas aproximações.

Teorema 5.5 *Os índices de cauda inferior e superior da cópula resultante C_X , $d = 2$, são dados respectivamente por*

$$\lambda_L = \lim_{z \rightarrow -\infty} \frac{1}{H_{s_t}(z)} \int_{\mathbb{R}^p} C_{X|Y}(F_{s_1}(z|y), F_{s_2}(z|y)) dC_Y(G_1(y_1), \dots, G_p(y_p))$$

e

$$\lambda_U = \lim_{z \rightarrow \infty} \frac{1}{1 - H_{s_t}(z)} \left[1 - 2H_{s_t}(z) - \int_{\mathbb{R}^p} C_{X|Y}(F_{s_1}(z|y), F_{s_2}(z|y)) dC_Y(G_1(y_1), \dots, G_p(y_p)) \right]$$

com H_{s_t} como no teorema 5.4.

Prova: Aplique a definição do índice de cauda inferior em (5.2) para C_X e substitua H_{s_t} por sua expressão dada no teorema 5.2. Faça o mesmo para o índice de cauda superior.

Uma vez mais, ambos λ_L e λ_U em geral não têm forma fechada. Uma aproximação para o valor de λ_L pode ser obtida através de uma regressão local usando métodos de mínimos quadrados e tomando o intercepto como valor aproximado, isto é, seja $y_r = \alpha + \xi_1 z_r + \xi_2 z_r^2$ uma regressão linear local com z_r sendo uma seqüência decrescente e

$$y_r \simeq \left[\frac{1}{K} \sum_{k=1}^K F_{s_t}(z_r | y^{(k)}) \right]^{-1} \left[\frac{1}{J} \sum_{j=1}^J C_{X|Y}(F_{s_1}(z_r | y^{(j)}), F_{s_2}(z_r | y^{(j)})) \right] \quad (5.12)$$

com $y^{(j)}$ e $y^{(k)}$ sendo independentes para todos j e k , então $\lambda_L \simeq \hat{\alpha}$ onde $\hat{\alpha}$ é a estimativa de mínimos quadrados da regressão acima. A mesma idéia pode ser aplicada ao índice de cauda superior λ_U .

Teorema 5.6 *A medida $D(q)$, conseqüentemente a medida Schweizer e Wolff σ e o índice de dependência Φ^2 , para a cópula resultante C_X , $d = 2$, é dada por*

$$D_X(q) = \left[\kappa_q \int_{\mathbb{R}} \int_{\mathbb{R}} |C_X(H_{s_1}(x_{s_1}), H_{s_2}(x_{s_2})) - H_{s_1}(x_{s_1})H_{s_2}(x_{s_2})|^q dH_{s_1}(x_{s_1})dH_{s_2}(x_{s_2}) \right]^{1/q}$$

com C_X como no teorema 5.1 e H_{s_t} como no teorema (5.4), sendo $\sigma_X = D_X(1)$ com $\kappa_1 = 12$ e

$\Phi_X^2 = (D_X(2))^2$ com $\kappa_2 = 90$.

Prova: A prova é bem simples. Apenas aplique a definição de $D(q)$ em (5.4) e substitua as quantidades apropriadas.

Uma aproximação numérica por integração Monte Carlo para estas quantidades é dada por

$$D_X(q) \simeq \left[\frac{\kappa_q}{L} \sum_{\ell=1}^L \left| \frac{1}{K} \sum_{k=1}^K C_{X|Y}(F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)}), F_{s_2}(x_{s_2}^{(\ell)}|y^{(k)})) - \frac{1}{J} \sum_{j=1}^J F_{s_1}(x_{s_1}^{(\ell)}|y^{(j)}) \frac{1}{M} \sum_{m=1}^M F_{s_2}(x_{s_2}^{(\ell)}|y^{(m)}) \right|^q \right]^{1/q} \quad (5.13)$$

com $y^{(j)}$, $y^{(k)}$, $y^{(m)}$ e $(x_{s_1}^{(\ell)}, x_{s_2}^{(\ell)})$ sendo independentes para todos os j , k , ℓ e m . As aproximações para Schweizer e Wolff σ_X e o índice de dependência Φ_X^2 são obtidas diretamente.

Teorema 5.7 *A medida Gini γ para a cópula resultante C_X , $d = 2$, é dada por*

$$\gamma_X = 4 \int_{\mathbb{R}} \int_{\mathbb{R}^p} [C_{X|Y}(F_{s_1}(x_{s_1}|y), 1 - F_{s_1}(x_{s_1}|y)) + C_{X|Y}(F_{s_1}(x_{s_1}|y), F_{s_1}(x_{s_1}|y))] dC_Y(G_1(y_1), \dots, G_p(y_p)) dH_1(x_{s_1}) - 2$$

com H_{s_1} como no teorema (5.4).

Prova: Aplique a definição da medida Gini γ em (5.7) e substitua C_X como no teorema 5.1.

Mais uma vez, uma aproximação numérica para γ_X baseado em integração Monte Carlo é dada por

$$\gamma_X \simeq \frac{4}{L} \sum_{\ell=1}^L \frac{1}{K} \sum_{k=1}^K C_{X|Y}(F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)}), 1 - F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)})) + C_{X|Y}(F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)}), F_{s_1}(x_{s_1}^{(\ell)}|y^{(k)})) - 2 \quad (5.14)$$

com $y^{(k)}$ e $(x_{s_1}^{(\ell)}, x_{s_2}^{(\ell)})$ sendo independentes para todos k e ℓ .

A medida Blomqvist β é uma medida simples no sentido que não necessita nenhuma maneira ou fórmula especial de reescreve-la em termos de C_X ou qualquer outra expressão que envolva X e Y e suas marginais. Sua aproximação numérica por Monte Carlo é simples e dada por

$$\beta_X \simeq \frac{1}{L} \sum_{\ell=1}^L \mathbb{I}_{\{(x_{s_1}^{(\ell)} - \text{med}(x_{s_1}))(x_{s_2}^{(\ell)} - \text{med}(x_{s_2})) < 0\}} - \frac{1}{K} \sum_{k=1}^K \mathbb{I}_{\{(x_{s_1}^{(k)} - \text{med}(x_{s_1}))(x_{s_2}^{(k)} - \text{med}(x_{s_2})) > 0\}} \quad (5.15)$$

onde \mathbb{I}_A é uma função indicadora que retorna 1 quando a desigualdade em A for verdadeira e 0 caso contrário. Além disso, $\text{med}(x_{s_1})$ e $\text{med}(x_{s_2})$ são as medianas amostrais de x_{s_1} e x_{s_2} , respectivamente.

No exemplo 5.1 queremos ilustrar a idéia da mistura em cópulas, como podemos obter as distribuições marginais e quais os resultados de todas essas medidas de dependência.

Exemplo 5.1 *Sejam X_i e Y_i , $i = 1, 2$ variáveis aleatórias tal que $X_i|Y_i \sim \text{Exponencial}(y_i)$ com média $E(X_i|Y_i = y_i) = 1/y_i$ e $Y_i \sim \text{Gama}(\lambda_i, \eta_i)$ com média $E(Y_i) = \lambda_i/\eta_i$ e variância $V(Y_i) = \lambda_i/\eta_i^2$. É fácil provar que X_i tem f.d.p $h_i(x_i) = \lambda_i \eta_i^{\lambda_i} / (\eta_i + x_i)^{\lambda_i+1}$ e f.d.a $H_i(x_i) = 1 - \eta_i^{\lambda_i} / (\eta_i + x_i)^{\lambda_i}$. Além disso, consideremos $C_{X|Y}$ como a cópula Clayton dada na tabela 4.1 cuja f.d.p pode ser encontrada na tabela 4.3 e C_Y a cópula produto C^\perp dada na seção 4.1. Fazendo $\lambda_1 = \lambda_2 = 2$, $\eta_1 = \eta_2 = 2$, $\theta = 2$, e usando o algoritimo SIR (Gamerman e Lopes, 2006) com taxa de reamostragem de 5% para gerar valores de $X|Y$, então*

1. $\tau_X^S \simeq 0,328$ e $\tau_X^F \simeq 0,334$ onde S corresponde a versão amostral da medida τ de Kendall para uma amostra de tamanho 5.000 e F a aproximação em (5.9) com $L = 10.000$ e $K = 50.000$. Observe que τ_X^S dá o mesmo resultado com um tamanho de amostra menor. A medida τ de Kendall para a cópula Clayton sem mistura é 0,50, com amostra gerada usando a média de y_i , $E(Y_i) = 1$ para $i = 1, 2$.
2. $\rho_X^S \simeq 0,473$, $\rho_X^H \simeq 0,482$ e $\rho_X^F \simeq 0,477$ onde S corresponde a versão amostral da medida Spearman ρ para uma amostra de tamanho 5.000, H a aproximação em (5.10) com $L = 10.000$, e F a aproximação em (5.11) com $L = 10.000$ e $J = K = 50.000$. Observe que ρ_X^S dá o mesmo resultado com um tamanho de amostra menor. A medida ρ de Spearman para a cópula Clayton sem mistura é 0,68, com amostra gerada usando a média de y_i , $E(Y_i) = 1$ para $i = 1, 2$.
3. $\lambda_L \simeq 0,532$ usando a idéia de modelo de regressão local com a seqüência z_r indo para 0 e y_r dado por (5.12). O índice de cauda inferior para a cópula Clayton sem mistura é aproximadamente 0,707, com amostra gerada usando a média de y_i , $E(Y_i) = 1$ para $i = 1, 2$.

4. $\sigma_X \simeq 0,474$ e $\Phi^2 \simeq 0,205$ usando (5.13) com $L = 10.000$ e $J = K = M = 50.000$. A medida Schweizer e Wolff σ e o índice de dependência Φ para a cópula Clayton sem mistura são aproximadamente 0,68 e 0,43, respectivamente, com amostra gerada usando a média de y_i , $E(Y_i) = 1$ para $i = 1, 2$.
5. $\gamma_X \simeq 0,442$ usando (5.14) com $L = 10.000$ e $K = 50.000$. A medida Gini γ para a cópula Clayton sem mistura é aproximadamente 0,57, com amostra gerada usando a média de y_i , $E(Y_i) = 1$ para $i = 1, 2$.
6. $\beta_X \simeq 0,330$ usando (5.15) com $L = K = 20.000$. A medida Blomqvist β para a cópula Clayton sem mistura é aproximadamente 0,51, com amostra gerada usando a média de y_i , $E(Y_i) = 1$ para $i = 1, 2$.

Agora, pode ser visto que todas as medidas indicam que a dependência entre as duas variáveis X_1 e X_2 decresce quando comparada as medidas de dependência da cópula Clayton sem mistura. Uma plausível explicação para este fato é a escolha de C_Y como sendo a cópula produto, isto é, marginais independentes. Uma outra possibilidade para a mistura seria a utilização da cópula Clayton tanto em $C_{X|Y}$ quanto em C_Y . Lembre-se que neste caso o algoritimo SIR pode ser usado em dois passos para gerar uma amostra da distribuição da mistura resultante. Uma outra característica importante não mostrada aqui é a mistura discreta, especialmente quando Y é uma v.a com distribuição discreta. O cálculo numérico das medidas de dependência mostradas acima podem demandar relativo esforço computacional. Atenção especial deve ser dada as versões amostrais das medidas τ de Kendall, utilizada em nossas aplicações da seção 5.5, e ρ de Spearman. Na próxima seção mostramos como inferir usando a abordagem bayesiana usando a estrutura dada pela mistura.

5.4 Inferência na distribuição a posteriori

Neste capítulo consideramos a abordagem bayesiana, onde todas as quantidades desconhecidas são estimadas a partir da distribuição a posteriori. A exceção é para a cópula Student, onde os graus de liberdade são dados e fixos. Formalmente, $\pi(\Psi) \propto L(\Psi)p(\Psi)$ onde Ψ é um vetor de parâmetros incluindo os da cópula e das distribuições marginais, enquanto $\pi(\Psi)$, $L(\Psi)$ e $p(\Psi)$ são a distribuição a posteriori, a função de verossimilhança e a distribuição a priori, respectivamente. Para facilitar o entendimento e

expor melhor as idéias, nos concentramos daqui em diante em distribuições e cópulas bivariadas. Seja (X_1, X_2) uma v.a contínua bivariada com cópula C_X . Logo, sua f.d.p conjunta é dada por

$$h(x_1, x_2 | \Psi) = c_X(H_1(x_1 | \Psi), H_2(x_2 | \Psi)|\theta)h_1(x_1 | \Psi)h_2(x_2 | \Psi) \quad (5.16)$$

onde h_i e H_i como antes representam a f.d.p e a f.d.a da distribuição marginal, respectivamente. A densidade cópula, c_X , é a f.d.p calculada como a derivada mista de C_X , se existir, em x_1 e x_2 . Veja, por exemplo, na tabela 4.3 a f.d.p das 6 famílias de cópulas apresentadas na tabela 4.1.

Seja $((x_{11}, x_{21}), \dots, (x_{1n}, x_{2n}))$ uma amostra de tamanho $n \geq 1$ de dados i.i.d da distribuição conjunta bivariada com f.d.p dada por (5.16), então

$$L(\Psi) = \prod_{i=1}^n c_X(H_1(x_{1i} | \Psi), H_2(x_{2i} | \Psi)|\theta)h_1(x_{1i} | \Psi)h_2(x_{2i} | \Psi) \quad (5.17)$$

A especificação do modelo fica completa após se atribuir uma distribuição a priori para o vetor de parâmetros Ψ . Todos os parâmetros são considerados independentes a priori e todos com distribuições própria e relativamente vagas. Seguimos com especificações semelhantes a apresentada na seção 4.2.2. No caso do modelo de mistura, $L(\Psi)$ vem da distribuição de $X|Y$, enquanto $Y|\Psi$ funciona como uma distribuição a priori $p(Y|\Psi)$ e a priori $p(\Psi)$ completa a especificação do modelo.

Como as funções cópulas geralmente não possuem formas simples, a distribuição a posteriori em geral não estará disponível em forma fechada conhecida. Inferência na distribuição a posteriori é facilitada por métodos de Monte Carlo via cadeias de Markov. Em especial, o amostrador de Gibbs foi escolhido para as aplicações deste capítulo e para a maioria das condicionais completas utilizamos o amostrador da fatia, exceto quando algumas das condicionais completas possuem forma fechada conhecida e fácil de amostrar. Estas escolhas, como nas aplicações do capítulo 4, induzem uma taxa de convergência rápida e com autocorrelações amostrais pequenas para as amostras das aplicações na seção 5.5.

5.5 Aplicações

Nossa principal meta nesta seção é mostrar a aplicabilidade da mistura em cópulas, apresentada na seção 5.3, a conjuntos de dados reais. Outras duas características importantes são a abordagem bayesiana e a seleção de modelos via DIC. Os dados de atuária de requerimentos de indenizações de seguros (Frees e Valdez, 1998) são ajustados ao modelo de mistura. Além disso, dois conjuntos de dados financeiros são analisados, os índices diários de mercado de Amsterdã e Frankfurt e os do México e Argentina (veja Hürliman (2004); Roch e Alegre (2006) para algumas aplicações em cópulas com índices de mercado.)

Devido a fatos estilizados em finanças, estamos interessados numa modelagem com distribuições assimétricas e com caudas pesadas. Para isso, usamos a abordagem de Fernández e Steel (1998) dada

em (4.5). Em especial, para ilustrar a idéia principal da mistura apresentada neste capítulo, temos a distribuição normal assimétrica dada por

$$f_i(x_i|\mu_i, y_i, \gamma_i) = \frac{2\gamma_i}{\gamma_i^2 + 1} \left[\phi\left(\frac{x_i - \mu_i}{\gamma_i\sqrt{y_i}}\right) I_{[0, \infty)} + \phi\left(\frac{\gamma_i(x_i - \mu_i)}{\sqrt{y_i}}\right) I_{(-\infty, 0)} \right] \quad (5.18)$$

que é usada como distribuições marginais de $C_{X|Y}$ em todas as 3 aplicações com $C_{X|Y}$ sendo uma das 6 famílias de cópulas da tabela 4.1. Aqui ϕ corresponde a f.d.p de uma distribuição normal padrão e dado Y_i , y_i representa a escala da distribuição. Além disso, Y_i tem distribuição inversa gama com parâmetros $\eta_i/2$ e $\eta_i\lambda_i^2/2$ (veja Fernández e Steel (1998) para maiores detalhes deste tipo de mistura). Entretanto, tomamos C_Y simplesmente como a cópula produto. Observe que X_i tem distribuição t-Student assimétrica. Este modelo é capaz de capturar estruturas mais simples dos dados com seus casos particulares. Outras escolhas seriam possíveis para as distribuições marginais e para as cópulas. Para todas as aplicações geramos 11.000 iterações do algoritmo de Gibbs, descartamos as 1.000 primeiras iterações e tomamos 1 valor a cada 10 amostrados, totalizando uma amostra de tamanho 1.000 das distribuições a posteriori de cada modelo. O parâmetro ν da cópula Student foi fixado em 4 para todas as aplicações.

Requerimentos de indenizações de seguros

A figura 5.1 mostra o conjunto de dados correspondentes aos requerimentos de indenizações de seguros com 1.500 observações de despesas ajustadas de perdas alocadas (ALAE) e pagamentos de indenizações (*Loss*), ambos na escala logarítmica. O sumário estatístico dessas duas medidas mostram que ambas parecem ser assimétricas a esquerda com um coeficiente de correlação de 0,425 e a medida τ de Kendall igual a 0,302. Uma melhor descrição dos dados pode ser encontrada em Frees e Valdez (1998).

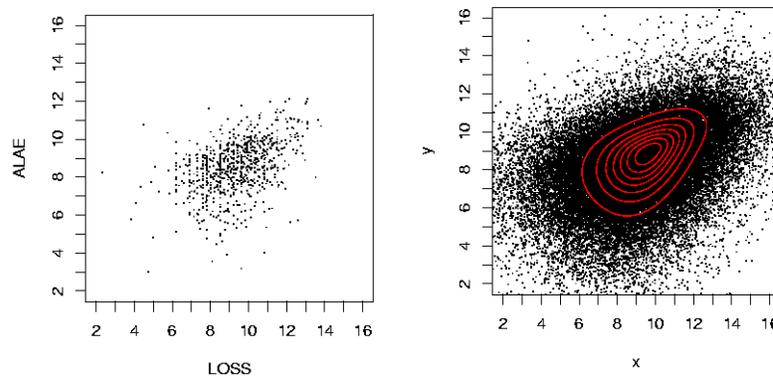


Figura 5.1: Requerimentos de indenizações de seguros (Loss e ALAE). Dados reais na esquerda e simulados na direita.

A tabela 5.1 apresenta o DIC e os outros critérios para todos os modelos de mistura em cópulas. Para o AIC, EAIC, BIC e EBIC fixamos em 11 o número de parâmetros, correspondendo a 3 de cada marginal normal assimétrica, 2 de cada inversa gama e 1 da cópula. A cópula Gumbel tem melhor ajuste do que as outras, embora tenha tido valores próximos a cópula cauda-pesada.

Tabela 5.1: DIC, pd, AIC, EAIC, BIC e EBIC para os 6 modelos de cópulas para os dados dos pedidos de perda da companhia de seguro.

Cópula	DIC	pd	AIC	EAIC	BIC	EBIC
Clayton	5375	6,77	5383	5390	5442	5449
Frank	5175	7,07	5183	5190	5242	5249
Gaussian	5173	6,78	5182	5189	5240	5247
Gumbel	5127	7,01	5135	5142	5194	5201
Cauda-pesada	5139	7,08	5147	5154	5205	5212
Student	5189	6,85	5198	5205	5256	5263

Na tabela 5.2 temos o sumário da distribuição a posteriori para o modelo de cópula Gumbel. Observe que ambas as variáveis tem alta probabilidade de serem assimétricas a esquerda dados pelo intervalo de credibilidade e média abaixo de 1, além disso ambas possuem caudas pesadas. O parâmetro da cópula, θ , está concentrado em torno de 1,4 indicando dependência positiva entre as variáveis. A figura 5.1 mostra 50.000 pares de valores do modelo cópula Gumbel e o contorno usando o estimador de Rao-

Tabela 5.2: Sumário da distribuição a posteriori do modelo de cópula Gumbel baseado numa amostra de tamanho 1000. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Parâmetro	Loss					ALAE				
	Média	D.P	2,5%	Mediana	97,5%	Média	D.P	2,5%	Mediana	97,5%
μ	9,481	0,099	9,316	9,486	9,644	8,895	0,076	8,772	8,893	9,026
λ^2	2,504	1,516	0,407	2,279	5,397	1,810	1,187	0,325	1,644	3,943
γ	0,957	0,033	0,907	0,955	1,012	0,840	0,026	0,797	0,841	0,883
η	9,276	10,37	2,150	5,098	32,44	7,873	8,955	2,147	4,623	26,07
θ	1,442	0,033	1,387	1,441	1,498	-	-	-	-	-

Blackwell (Gamerman e Lopes, 2006). A mesma amostra foi usada para estimar um τ de Kendall igual a 0,290. O modelo geral capturou características importante dos dados como a dependência entre elas, as assimetrias e caudas pesadas em ambas as marginais.

Índices de Amsterdã e Frankfurt

Os dados correspondem a 2733 retornos diários dos índices de Amsterdã e Frankfurt de 10 de julho de 1987 a 31 de dezembro de 1997 (Franses e van Dijk, 2000). Estamos interessados somente em estudar a estrutura de dependência, por isso ajustamos primeiro modelos GARCH(1,1) com erros normais para ambos os índices e aplicamos o modelo de mistura nos resíduos. Testes estatísticos mostraram que todos os parâmetros estimados foram estatisticamente significativos. A figura 5.2 mostra a dispersão dos resíduos dos modelos GARCH(1,1) para os retornos de Amsterdã e Frankfurt.

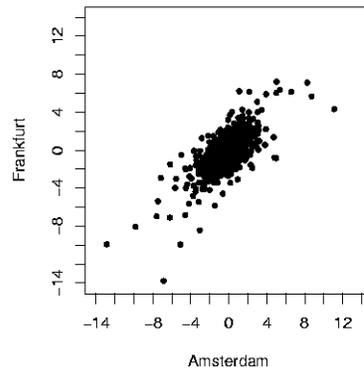


Figura 5.2: Resíduos do ajuste de GARCH(1,1) para ambos índices de Amsterdã e Frankfurt.

A tabela 5.3 mostra o DIC e os outros critérios. Para o AIC, EAIC, BIC e EBIC fixamos em 11 o número de parâmetros, correspondendo a 3 de cada marginal normal assimétrica, 2 de cada inversa gama e 1 da cópula. A cópula Student tem melhor ajuste do que as outras, embora apresente valores próximos a Gaussiana. Vale a pena lembrar que o grau de liberdade da cópula Student está fixo em 4.

Na tabela 5.4 temos o sumário da distribuição a posteriori para o modelo de cópula Student. Observe que ambas as variáveis tem alta probabilidade de serem assimétricas a esquerda dados pelo intervalo de credibilidade e média abaixo de 1, além disso ambas possuem cauda pesada. O parâmetro da cópula, θ , está concentrado em torno de 0,7 indicando dependência positiva entre os dois mercados. O modelo geral capturou características importante dos dados como a dependência entre elas, as assimetrias e caudas pesadas em ambas as marginais.

Tabela 5.3: DIC, pd, AIC, EAIC, BIC e EBIC para os 6 modelos de cópula para os resíduos de GARCH(1,1) para os índices de Amsterdã e Frankfurt.

Cópula	DIC	pd	AIC	EAIC	BIC	EBIC
Clayton	6284	6,90	6293	6300	6358	6365
Frank	5669	6,94	5677	5684	5742	5749
Gaussian	5533	6,48	5542	5548	5607	5613
Gumbel	5710	6,85	5719	5726	5784	5791
Cauda-pesada	6376	6,89	6384	6391	6449	6456
Student	5517	6,55	5526	5533	5591	5598

Tabela 5.4: Sumário da distribuição a posteriori do modelo de cópula Student baseado numa amostra de tamanho 1000. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Ψ	Amsterdã					Frankfurt				
	Média	D.P	2.5%	Median	97.5%	Média	D.P	2.5%	Median	97.5%
μ	0,126	0,034	0,068	0,125	0,182	0,103	0,040	0,038	0,104	0,168
λ^2	1,309	0,874	0,208	1,149	3,060	1,591	1,067	0,280	1,402	3,540
γ	0,899	0,016	0,872	0,900	0,924	0,918	0,016	0,893	0,917	0,945
η	8,533	9,898	2,128	4,640	31,55	8,260	9,109	2,164	4,510	28,58
θ	0,708	0,013	0,686	0,709	0,728	-	-	-	-	-

Índices de Mexico e Argentina

Nesta aplicação queremos mostrar que dados de mesma natureza podem ter estruturas de dependência diferentes. Os dados correspondem a 3132 retornos diários dos índices de Mexico e Argentina de 22 de dezembro de 1992 a 22 de dezembro de 2004 (Arakelian e Dellaportas, 2006). Estamos interessados somente em estudar a estrutura de dependência, por isso ajustamos primeiro modelos GARCH(1,1) com erros normais para ambos os índices e aplicamos o modelo de mistura nos resíduos. Todos os parâmetros estimados foram estatisticamente significativos. A figura 5.3 mostra a dispersão dos resíduos dos modelos GARCH(1,1) para os retornos de Mexico e Argentina.

A tabela 5.5 mostra o DIC e os outros critérios. Para o AIC, EAIC, BIC e EBIC fixamos em 11

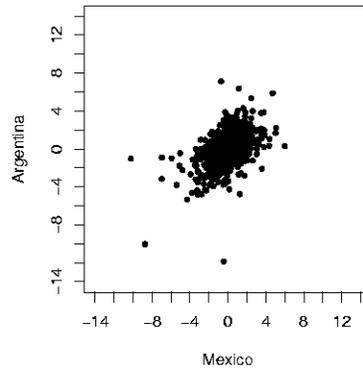


Figura 5.3: Resíduos do ajuste de GARCH(1,1) para ambos índices de Mexico e Argentina.

o número de parâmetros, correspondendo a 3 de cada marginal normal assimétrica, 2 de cada inversa gama e 1 da cópula. O modelo de cópula Frank tem melhor ajuste do que os outros, embora tenha apresentado valores próximos ao Gaussiano.

Tabela 5.5: DIC, pd, AIC, EAIC, BIC e EBIC para os 6 modelos de cópula para os resíduos de GARCH(1,1) para os índices de Mexico e Argentina.

Cópula	DIC	pd	AIC	EAIC	BIC	EBIC
Clayton	5792	7,03	5800	5807	5866	5873
Frank	5501	7,18	5509	5516	5576	5583
Gaussian	5538	6,96	5547	5554	5613	5620
Gumbel	5685	7,07	5692	5699	5759	5766
Cauda-pesada	5930	6,98	5938	5945	6004	6011
Student	5586	7,09	5594	5601	5660	5668

Na tabela 5.6 temos o sumário da distribuição a posteriori para o modelo de cópula Frank. Algumas características dadas pela estimação são posição fora do $(0,0)$, parâmetros de escala perto de 1, assimetrias e caudas pesadas para as marginais. Usando dados artificiais do modelo de cópula Frank, tomando a média a posteriori dos parâmetros, temos um τ de Kendall estimado em 0,407. O modelo capturou características importante dos dados como a dependência, as assimetrias e as caudas pesadas em ambas as marginais.

Tabela 5.6: Sumário da distribuição a posteriori do modelo de cópula Frank baseado numa amostra de tamanho 1000. D.P, 2.5% e 97.5% são o desvio padrão e percentis da distribuição a posteriori dos parâmetros, respectivamente.

Ψ	Mexico					Argentina				
	Média	D.P	2.5%	Median	97.5%	Média	D.P	2.5%	Median	97.5%
μ	0,104	0,032	0,050	0,105	0,158	0,074	0,029	0,028	0,076	0,121
λ^2	1,089	0,713	0,238	0,965	2,358	1,038	0,636	0,218	0,930	2,252
γ	0,917	0,017	0,888	0,916	0,945	0,935	0,016	0,910	0,935	0,961
η	7,782	8,513	2,165	4,526	25,28	8,605	9,967	2,142	4,661	31,56
θ	4,220	0,152	3,967	4,223	4,465	-	-	-	-	-

5.6 Discussão final

Neste capítulo apresentamos um método para a construção de novas famílias de cópulas a partir das existentes. O método consiste em mistura de cópulas através de suas distribuições marginais, estendendo resultados encontrados na literatura. Revisamos algumas medidas de dependência e mostramos como elas podem ser calculadas no modelo de mistura. Integração Monte Carlo é usada para todas as medidas e também as versões amostrais das medidas não paramétricas τ de Kendall e ρ de Spearman. A inferência no modelo de mistura em cópulas foi facilitada pela abordagem bayesiana usando técnicas de Monte Carlo via cadeias de Markov. Um exemplo com variáveis aleatórias contínuas ilustra a idéia completa da mistura em cópulas. Além disso, conjuntos de dados reais são usados para mostrar a aplicabilidade do modelo de mistura em cópulas.

Capítulo 6

CONCLUSÕES E CONSIDERAÇÕES FINAIS

Neste trabalho propomos alguns modelos em distribuições assimétricas e dependência através de cópulas. Algumas contribuições apresentadas nos capítulos anteriores são análise bayesiana dos modelos, estimação de todas as quantidades desconhecidas através métodos de simulação Monte Carlo via cadeias de Markov e o uso de alguns métodos de seleção de modelos.

6.1 Gaussiana inversa generalizada estendida

No capítulo 2, apresentamos extensões do modelo log-linear e de volatilidade estocástica através da distribuição Gaussiana inversa generalizada estendida. Para as aplicações, alguns critérios de seleção de modelos foram usados, dentre os quais temos o AIC, BIC e DIC. Agora, estamos investigando também a classe de distribuição a priori de Jeffreys para os parâmetros da distribuição GIGE e a respectiva distribuição a posteriori. O apêndice B mostra o cálculo da distribuição a priori de Jeffreys para o caso de dados independentes e identicamente distribuídos. Mesmo para este caso mais simples, várias derivadas e integrais devem ser avaliadas numericamente. Os resultados aparecerão em trabalhos futuros pois ainda estão sob investigação.

Além disso, estamos verificando como os modelos alternativos diferem um do outro, possíveis pontos de massa na distribuição a priori de alguns parâmetros, por exemplo para o parâmetro ξ que para ambos os modelos log-linear e de volatilidade estocástica parecem próximos de 0. Por outro lado, o modelo de volatilidade estocástica com GIGE abre novas áreas de pesquisa visto que os parâmetros múltiplos da distribuição GIGE são relacionados a variabilidade dos dados.

6.2 Modelos ARFIMA assimétricos

No capítulo 3, a introdução de erros normais e t-Student assimétricos no modelo ARFIMA(p, d, q) permitiram maior flexibilidade na modelagem de séries temporais reais. Nesta proposta, trabalhamos com a classe de distribuições assimétricas propostas por Azzalini (1985). A introdução de variáveis auxiliares facilitaram a construção do modelo e as mesmas foram estimadas na distribuição a posteriori junto com os demais parâmetros através de métodos de simulação Monte Carlo via cadeias de Markov. Estudos indicaram que algumas distribuições a priori de Jeffreys de alguns casos particulares do modelo

forneceram bons resultados na estimação, sendo portanto utilizadas no modelo mais geral. A utilização de alguns critérios de seleção de modelos ajudaram na decisão do melhor ajuste, isto é, na escolha entre erros normais ou t-Student assimétricos, assim como a escolha da ordem de p e q .

Ainda estamos investigando detalhes de estimação nesta classe de modelos em geral. Nosso objetivo é trabalhar com distribuições multivariadas, por exemplo em modelos ARFIMA multivariados com distribuições assimétricas. Outras possíveis extensões incluem o cálculo da distribuição a priori de Jeffreys no caso multivariado, a introdução de análise fatorial nestes modelos de séries temporais com erros assimétricos, a introdução de modelos de séries temporais não-lineares com erros assimétricos e a comparação destes modelos com outros tipos de distribuições assimétricas como por exemplo os propostos por Fernández e Steel (1998).

6.3 Cópulas: inferência e seleção

No capítulo 4 estudamos alguns modelos em cópulas com estimação conjunta de todos os parâmetros através do enfoque bayesiano, evitando assim estimação em 2 passos. Mostramos que o DIC é um bom critério de seleção de modelo quando aplicado a cópulas. Além disso, pode ser utilizado numa grande variedade de aplicações sem qualquer tipo de estimação em 2 passos ou a necessidade de definir probabilidades a priori para cada modelo. Outros critérios baseados na função desvio - AIC, EAIC, BIC e EBIC - também concordam com o DIC, e tiveram, portanto a mesma performance. Todos eles são fáceis de implementar em métodos de simulação Monte Carlo via cadeias de Markov como o proposto, o que por sua vez mostrou ter uma rápida taxa de convergência. Entretanto, o DIC e os outros critérios somente podem ser calculados quando as densidades cópulas existem, implicando que a f.d.p conjunta exista. Assim, os estudos do capítulo 4 formaram uma base de compreensão de cópulas, tipos de dependência, distribuições marginais e aspectos de inferência necessários para o desenvolvimento do capítulo 5.

6.4 Mistura em cópulas e assimetria

No capítulo 5 apresentamos um método para construção de novas cópulas a partir das existentes. O método consiste em mistura de cópulas através de suas distribuições marginais, estendendo resultados encontrados na literatura. Revemos algumas medidas de dependência e como elas podem ser calculadas no modelo de mistura. Integração Monte Carlo pode ser usada para todas as medidas e também as versões amostrais das medidas não paramétricas τ de Kendall e ρ de Spearman. A inferência no modelo de mistura em cópulas foi facilitada pela abordagem bayesiana usando técnicas de simulação Monte Carlo via cadeias de Markov. Um exemplo com variáveis aleatórias contínuas ilustrou a idéia completa da mistura em cópulas. Além disso, conjuntos de dados reais foram usados para mostrar a aplicabilidade do modelo de mistura em cópulas.

O estudo de misturas discretas e contínuas, assim como nos parâmetros das cópulas, estão sob investigação. Além disso, para as aplicações reais modelos de volatilidade estocástica e hierárquicos em cópulas com mistura estão em nossos planos.

Dos planos de trabalho propostos acima, estamos dando maior ênfase aos modelos multivariados e algumas possíveis extensões de cada modelo discutido em cada capítulo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Achcar, J., e Bolfarine, H. (1986). The log-linear model with a generalized gamma distribution for the error: a Bayesian approach. *Statistics and Probability Letters*, 4, 325–332.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 255–723.
- Arakelian, V., e Dellaportas, P. (2006). Contagion tests via copula threshold models. *Submetido*.
- Azzalini, A. (1985). A class of distributions wich includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.
- Azzalini, A., e Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of Royal Statistical Society, Series B*, 61, 579–602.
- Azzalini, A., e Capitanio, A. (2003). Distributions generated by perturbations of symmetry with emphasis on a multivariate skew t distribtuion. *Journal of Royal Statistical Society, Series B*, 65, 367–389.
- Azzalini, A., e Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726.
- Beran, J. (1994). *Statistics for long memory process*. New York: Chapman and Hall.
- Brooks, S. P. (2002). Discussion of the paper Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64, 616–618.
- Chen, X., e Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *La Revue Canadienne de Statistique*, 33(3), 389–414.
- Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141–151.
- Cook, R., e Weisberg, S. (1994). *An introduction to regression graphics*. New York: John Wiley and

Sons.

- Dobrić, J., e Schmidt. (2004). *Testing goodness of fit for parametric families of copulas - application to financial data*. Cologne, Germany.
- Evin, G. (2004). *Choix de la meilleure famille de copule en hydrologie*. France.
- Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, 95, 119–152.
- Fernández, C., e Steel, M. (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441), 359–371.
- Fonseca, T. (2004). *Análise bayesiana de referência para a classe de distribuições hiperbólicas generalizadas*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
- Frank, M. J. (1979). On the simultaneous associative of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*, 19, 194–226.
- Franses, P., e van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge University Press: Cambridge.
- Frees, E. W., e Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2, 1–25.
- Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., et al. (2005). *GNU scientific library reference manual - revised second edition*. (ISBN 0-9541617-3-4)
- Gamerman, D., e Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman & Hall/CRC.
- Gelfand, A., e Ghosh, S. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85, 1–11.
- Gelfand, A., e Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

- Geman, S., e Geman, D. (1984). Stochastic relaxation, gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Genest, C., Quessy, J.-F., e Rémilliard, B. (2005). Goodness-of-fit procedures for copula models based on probability integral transformation. *Scandinavian Journal of Statistics*, 32. (in press)
- Genton, M. (2004). *Skew-elliptical distributions and their applications* (1 ed.). New York: Chapman and Hall.
- Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.
- Gumbel, E. J. (1960). Distributions des valeurs extrême en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9, 171–173.
- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.
- Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13, 271–275.
- Hosking, J. (1981). Fractional differencing. *Biometrika*, 68(1), 165–176.
- Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73, 671–678.
- Huard, D., Évin, G., e Favre, A. (2006). Bayesian copula selection. *Computational Statistics and Data Analysis*, 51(2), 809–822.
- Hürliman, W. (2004). Fitting bivariate cumulative returns with copulas. *Computational Statistics and Data Analysis*, 45, 355–372.
- James, B. (1996). *Probabilidade um curso em nível intermediário* (2 ed.). Rio de Janeiro: IMPA.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall/CRC.
- Jørgensen, S. (1982). *Statistical properties of the generalized inverse Gaussian distribution*. New York: Springer-Verlag.

- Kruskal, W. (1958). Ordinal measure of association. *Journal of the American Statistical Association*, 53, 814–861.
- Lambert, P., e Laurent, S. (2002). *Modelling skewness dynamics in series of financial data using skewed location-scale distributions*. Louvain, Belgium. (submetido)
- Liseo, B., e Loperfido, N. (2004). Default Bayesian analysis of the skew-normal distribution. *Journal of Statistical Planning and Inference*, 136(2), 373–389.
- Lopes, H., e Ausin, M. (2006). *Time-varying joint distribution through copulas* (Rel. Téc.). University of Chicago, Chicago, USA. (Submetido)
- Lopes, H., e Migon, H. (2002). Comovements and contagion in emergent markets: stock indexes volatilities. In A. Carriquiry, C. Gatsonis, A. Gelman, D. Higdon, R. Kass, e D. Pauler (Eds.), *Case studies in bayesian statistics, volume vi* (p. 285-300). Springer-Verlag: New York.
- Matsmoto, M., e Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1), 3-30.
- Metropolis, N., Rosenbluth, A., Teller, A., e Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.
- Neal, R. (2003). Slice sampling (*with discussion*). *Annals of Statistics*, 31, 705-767.
- Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer-Verlag.
- Nicoloutsopoulos, D. (2005). *Parametric and Bayesian non-parametric estimation of copulas*. Tese de Doutorado, University College London.
- Pitt, M., Chan, D., e Kohn, R. (2006). Efficient bayesian inference for gaussian copula regression models. *Biometrika*, 93(3), 537–554.
- Pourahmadi, M. (2007). Skew-normal ARMA models with nonlinear heteroscedastic predictors. *Communications in Statistics*, 36(9). (in press)

- R Development Core Team. (2006). *R: A language and environment for statistical computing*. Vienna, Austria. (ISBN 3-900051-07-0)
- Roch, O., e Alegre, A. (2006). Testing the bivariate distribution of daily equity returns using copulas. An application to the Spanish stock market. *Computational Statistics and Data Analysis*, 51(2), 1312-1329.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Schweizer, B., e Sklar, A. (1983). *Probabilistic metric spaces*. Amsterdam: North-Holland.
- Schweizer, B., e Wolff, E. (1981). On nonparametric measures of dependence for random variables. *Annals of Statistics*, 9, 870–885.
- Silva, R. (2003). *Modelos bayesianos de longa dependência com erros hiperbólicos generalizados*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
- Sklar, A. (1959). Fonctions de répartition et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.
- Spiegelhalter, D., Best, N., Carlin, B., e Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society, Serie B*, 64, 583–639.
- Tierney. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22, 1701–1786.

Apêndice A

FUNÇÃO DE BESSEL DO TERCEIRO TIPO

A função de Bessel modificada do terceiro tipo com índice $\lambda \in \mathbb{R}$, denotada por $K_\lambda(\cdot)$, é uma função integral, que dentre as várias representações, tem-se

$$K_\lambda(u) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp \left[-\frac{1}{2} u (y + y^{-1}) \right] dy$$

Algumas propriedades desta função são:

- $K_\lambda(u) = \frac{\pi I_{-\lambda}(u) - I_\lambda(u)}{2 \operatorname{sen}(\lambda\pi)}$ onde $I_\lambda(\cdot)$ é a função de Bessel modificada do primeiro tipo.
- $K_\lambda(u) = K_{-\lambda}(u)$
- $K_{\lambda+1}(u) = \frac{2\lambda}{u} K_\lambda(u) + K_{\lambda-1}(u)$
- $K_{n+\frac{1}{2}}(u) = \sqrt{\frac{\pi}{2u}} e^{-u} \left[1 + \sum_{i=1}^n \frac{(n+i)!}{(n-i)!i!} (2u)^{-i} \right]$, se $\lambda = n + \frac{1}{2}$ e $n = 0, 1, 2, \dots$
- $K_{-\frac{1}{2}}(u) = K_{\frac{1}{2}}(u) = \sqrt{\frac{\pi}{2u}} e^{-u}$ quando $n = 0$ na relação acima.

Em casos limites onde $u \rightarrow 0$ tem-se

- $K_\lambda(u) \simeq \frac{1}{2} \Gamma(\lambda) \left(\frac{u}{2} \right)^{-\lambda}$, se $\lambda \geq 0$
- $K_\lambda(u) \simeq \frac{1}{2} \Gamma(-\lambda) \left(\frac{u}{2} \right)^\lambda$, se $\lambda < 0$
- $K_0(u) \simeq -\ln(u)$.

No caso que $u \rightarrow \infty$ tem-se

- $K_\lambda(u) \simeq \sqrt{\frac{\pi}{2u}} e^{-u}$.

Para $\lambda \rightarrow \infty$ tem-se a seguinte relação assintótica

- $K_\lambda(u) \simeq 2^\lambda \lambda^{\lambda-\frac{1}{2}} e^{-\lambda} u^{-\lambda\sqrt{\frac{\pi}{2}}}$ que pode ser vista como análoga a fórmula de Stirling para a função gama.

- $K'_\lambda(u) = -\frac{1}{2} [K_{\lambda-1}(u) + K_{\lambda+1}(u)] = -\frac{\lambda}{u} K_\lambda(u) + K_{\lambda-1}(u)$
- $\left. \frac{\partial}{\partial \lambda} K_\lambda(u) \right|_{\lambda=0} = 0$
- $\left. \frac{\partial}{\partial \lambda} K_\lambda(u) \right|_{\lambda=n} = \frac{n!(u/2)^{-n}}{2} \sum_{k=0}^{n-1} \frac{(u/2)^k K_k(u)}{(n-k)k!}$
- $\frac{\partial}{\partial \lambda} K_\lambda(u) = \frac{\pi}{2} \csc(\lambda\pi) \left[\frac{\partial}{\partial \lambda} I_{-\lambda}(u) - \frac{\partial}{\partial \lambda} I_\lambda(u) \right] - \pi \cot(\lambda\pi) K_\lambda(u)$,
para $\lambda \neq 0, \pm 1, \pm 2, \dots$
- $\left. \frac{\partial}{\partial \lambda} I_\lambda(u) \right|_{\lambda=0} = -K_0(u)$
- $(-1)^n \left[\left. \frac{\partial}{\partial \lambda} I_\lambda(u) \right|_{\lambda=n} \right] = -K_n(u) + \frac{n!(u/2)^{-n}}{2} \sum_{k=0}^{n-1} (-1)^k \frac{(u/2)^k I_k(u)}{(n-k)k!}$
- $\frac{\partial}{\partial \lambda} I_\lambda(u) = I_\lambda(u) \ln(u/2) - (z/2)^\lambda \sum_{k=0}^{\infty} \frac{\psi(\lambda+k+1)}{\Gamma(\lambda+k+1)} \frac{(z/2)^{2k}}{k!}$, onde
 $\psi(u) = \frac{\partial}{\partial u} \ln(\Gamma(u))$ é a função digama.

Apêndice B

**PRIORI DE JEFFREYS PARA DADOS INDEPENDENTES E IDENTICAMENTE
DISTRIBUÍDOS DA GIGE**

Nesta seção estamos interessados em desenvolver a distribuição a priori de Jeffreys, para isso precisaremos calcular função escore $U = U(\lambda, \xi, \phi, \theta)$ e o valor esperado $E(UU')$. Considere, então, uma amostra aleatória (X_1, \dots, X_n) provenientes da distribuição GIGE com f.d.p dada por (2.3). Baseado nesta amostra obtemos a função de verossimilhança e seu respectivo logaritmo que é dado por

$$\begin{aligned} \ell(\lambda, \xi, \phi, \theta) = & \frac{n\lambda\theta}{2} (\log(\phi) - \log(\xi)) + n \log(\theta) - n \log(2) - n \log \left(K_\lambda(2(\xi\phi)^{\theta/2}) \right) \\ & + (\lambda\theta - 1) \sum_{i=1}^n \log(x_i) - \xi^\theta \sum_{i=1}^n x_i^{-\theta} - \phi^\theta \sum_{i=1}^n x_i^\theta \end{aligned} \quad (\text{B.1})$$

Neste caso, precisamos somente um valor x da amostra para desenvolver a distribuição a priori de Jeffreys. Abaixo, descrevemos a função escore, omitindo alguns cálculos envolvidos. Definimos $k_\lambda(x) = \partial K_\lambda(x)/\partial \lambda$ que deve ser calculada numericamente pois não existe forma fechada conhecida. $K'(z)$ é a derivada com respeito a z dada no apêndice A.

$$U_1(X; \lambda) = \frac{\partial}{\partial \lambda} \ell(\lambda, \xi, \phi, \theta) = \theta \left[\log(X) - \left(\frac{k_\lambda(2(\xi\phi)^{\theta/2})}{\theta K_\lambda(2(\xi\phi)^{\theta/2})} + \frac{1}{2} (\log(\xi) - \log(\phi)) \right) \right]$$

$$U_2(X; \xi) = \frac{\partial}{\partial \lambda} \ell(\lambda, \xi, \phi, \theta) = -\frac{\lambda\theta}{2\xi} - \theta\phi^{\theta/2}\xi^{\theta/2-1} \frac{K'_\lambda(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} - \theta\xi^{\theta-1} X^{-\theta}$$

$$U_3(X; \phi) = \frac{\partial}{\partial \lambda} \ell(\lambda, \xi, \phi, \theta) = \frac{\lambda\theta}{2\phi} - \theta\xi^{\theta/2}\phi^{\theta/2-1} \frac{K'_\lambda(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} - \theta\phi^{\theta-1} X^\theta$$

$$\begin{aligned} U_4(X; \theta) = & \frac{\partial}{\partial \lambda} \ell(\lambda, \xi, \phi, \theta) = \frac{\lambda}{2} (\log(\phi) - \log(\xi)) + \frac{1}{\theta} + \lambda \log(X) \\ & - \xi^{\theta/2}\phi^{\theta/2} (\log(\xi) + \log(\phi)) \frac{K'_\lambda(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} \\ & - \xi^\theta \log(\xi) X^{-\theta} + \xi^\theta X^{-\theta} \log(X) - \phi^\theta \log(\phi) X^\theta - \phi^\theta X^\theta \log(X) \end{aligned}$$

Como $E(U_1(X; \lambda)) = 0$, temos

$$E(\log(X)) = \frac{k_\lambda(2(\xi\phi)^{\theta/2})}{\theta K_\lambda(2(\xi\phi)^{\theta/2})} + \frac{1}{2} (\log(\xi) - \log(\phi)) \quad (\text{B.2})$$

Definimos abaixo algumas relações e redefiniremos as respectivas derivadas parciais. Sejam

$$\begin{aligned}
 \rho &= 2(\xi\phi)^{\theta/2}; \\
 a_1 &= \frac{\theta}{2}(\log(\phi) - \log(\xi)) - \frac{k_\lambda(\rho)}{K_\lambda(\rho)}; \\
 a_2 &= -\frac{\lambda\theta}{2\xi} - \theta\phi^{\theta/2}\xi^{\theta/2-1}\frac{K'_\lambda(\rho)}{K_\lambda(\rho)}; \\
 b_2 &= -\theta\xi^{\theta-1}; \\
 a_3 &= \frac{\lambda\theta}{2\phi} - \theta\xi^{\theta/2}\phi^{\theta/2-1}\frac{K'_\lambda(\rho)}{K_\lambda(\rho)}; \\
 b_3 &= -\theta\phi^{\theta-1}; \\
 a_4 &= \frac{\lambda}{2}(\log(\phi) - \log(\xi)) + \frac{1}{\theta} - \xi^{\theta/2}\phi^{\theta/2}(\log(\xi) + \log(\phi))\frac{K'_\lambda(\rho)}{K_\lambda(\rho)}; \\
 b_4 &= -\xi^\theta \log(\xi); \quad e \\
 c_4 &= -\phi^\theta \log(\phi).
 \end{aligned}$$

Logo,

$$\begin{aligned}
 U_1(X; \lambda) &= a_1 + \theta \log(X); \\
 U_2(X; \xi) &= a_2 + b_2 X^{-\theta}; \\
 U_3(X; \phi) &= a_3 + b_3 X^\theta; \quad e \\
 U_4(X; \theta) &= a_4 + \lambda \log(X) + b_4 X^{-\theta} + \xi^\theta X^{-\theta} \log(X) + c_4 X^\theta - \phi^\theta X^\theta \log(X).
 \end{aligned}$$

A distribuição a priori de Jeffreys é definida como $p^J(\lambda, \xi, \phi, \theta) \propto \sqrt{|E(UU')|}$ onde $U = U(\lambda, \xi, \phi, \theta) = (U_1(X; \lambda), U_2(X; \xi), U_3(X; \phi), U_4(X; \theta))'$, lembrando que $E(UU')$ na verdade é a matriz de Fisher. Assim, necessitaremos calcular os valores esperados $E(U_i U_j)$ para $i, j = 1, 2, 3, 4$ e $i \leq j$. Segue-se que

$$\begin{aligned}
 E(U_1^2) &= a_1^2 + 2\theta a_1 E(\log(X)) + \theta^2 E(\log(X))^2 \\
 E(U_2^2) &= a_2^2 + 2a_2 b_2 E(X^{-\theta}) + b_2^2 E(X^{-2\theta}) \\
 E(U_3^2) &= a_3^2 + 2a_3 b_3 E(X^\theta) + b_3^2 E(X^{2\theta})
 \end{aligned}$$

$$\begin{aligned}
E(U_4^2) &= a_4^2 + \lambda^2 E(\log(X))^2 + b_4^2 E(X^{-2\theta}) + \xi^{2\theta} E(X^{-2\theta} (\log(X))^2) + c_4^2 E(X^{2\theta}) \\
&\quad + \phi^{2\theta} E(X^{2\theta} (\log(X))^2) + 2a_4 \lambda E(\log(X)) + 2a_4 b_4 E(X^{-\theta}) \\
&\quad + 2a_4 \xi^\theta E(X^{-\theta} \log(X)) + 2a_4 c_4 E(X^\theta) - 2a_4 \phi^\theta E(X^\theta \log(X)) \\
&\quad + 2b_4 \lambda E(X^{-\theta} \log(X)) + 2\lambda \xi^\theta E(X^{-\theta} (\log(X))^2) \\
&\quad + 2c_4 \lambda E(X^\theta \log(X)) - 2\lambda \phi^\theta E(X^\theta (\log(X))^2) \\
&\quad + 2b_4 \xi^\theta E(X^{-2\theta} \log(X)) + 2b_4 c_4 - 2b_4 \phi^\theta E(\log(X)) \\
&\quad + 2c_4 \xi^\theta E(\log(X)) - 2\xi^\theta \phi^\theta E(\log(X))^2 - 2c_4 \phi^\theta E(X^{2\theta} \log(X)) \\
E(U_1 U_2) &= a_1 a_2 + a_1 b_2 E(X^{-\theta}) + a_2 \theta E(\log(X)) + b_2 \theta E(X^{-\theta} \log(X)) \\
E(U_1 U_3) &= a_1 a_3 + a_1 b_3 E(X^\theta) + a_3 \theta E(\log(X)) + b_3 \theta E(X^\theta \log(X)) \\
E(U_1 U_4) &= a_1 a_4 + a_1 \lambda E(\log(X)) + a_1 b_4 E(X^{-\theta}) + a_1 \xi^\theta E(X^{-\theta} \log(X)) \\
&\quad + a_1 c_4 E(X^\theta) - a_1 \phi^\theta E(X^\theta \log(X)) + a_4 \theta E(\log(X)) \\
&\quad + \lambda \theta E(\log(X))^2 + b_4 \theta E(X^{-\theta} \log(X)) + \theta \xi^\theta E(X^{-\theta} (\log(X))^2) \\
&\quad + c_4 \theta E(X^\theta \log(X)) - \theta \phi^\theta E(X^\theta (\log(X))^2) \\
E(U_2 U_3) &= a_2 a_3 + a_2 b_3 E(X^\theta) + a_3 b_2 E(X^{-\theta}) + b_2 b_3 \\
E(U_2 U_4) &= a_2 a_4 + a_2 \lambda E(\log(X)) + a_2 b_4 E(X^{-\theta}) + a_2 \xi^\theta E(X^{-\theta} \log(X)) \\
&\quad + a_2 c_4 E(X^\theta) - a_2 \phi^\theta E(X^\theta \log(X)) + a_4 b_2 E(X^{-\theta}) + b_2 b_4 E(X^{-2\theta}) \\
&\quad + b_2 \xi^\theta E(X^{-2\theta} \log(X)) + b_2 c_4 - b_2 \phi^\theta E(\log(X)) + b_2 \lambda E(X^{-\theta} \log(X)) \\
E(U_3 U_4) &= a_3 a_4 + a_3 \lambda E(\log(X)) + a_3 b_4 E(X^{-\theta}) + a_3 \xi^\theta E(X^{-\theta} \log(X)) \\
&\quad + a_3 c_4 E(X^\theta) - a_3 \phi^\theta E(X^\theta \log(X)) + a_4 b_3 E(X^\theta) + b_3 \lambda E(X^\theta \log(X)) \\
&\quad + b_3 b_4 + b_3 \xi^\theta E(\log(X)) + b_3 c_4 E(X^{2\theta}) - b_3 \phi^\theta E(X^{2\theta} \log(X))
\end{aligned}$$

Agora, precisamos calcular algumas combinações de valores esperados da potência de X e seu logaritmo. Alguns desses valores esperados podem ser encontrados de forma fechada enquanto outros somente poderão ser resolvidos numericamente. Os valores esperados da forma $E(X^p)$ são dados por (2.4) enquanto que $E(\log(X))$ é dado por (B.2). Abaixo segue-se os demais valores esperados, sendo que alguns são definidos por meio de integrais que podem ser facilmente calculada por métodos numéricos. Definimos $R_\lambda(z) = K_{\lambda+1}(z)/K_\lambda(z)$.

$$\begin{aligned}
E(\log(X))^2 &= \left(\frac{\phi}{\xi}\right)^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\log(x))^2 x^{\lambda\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^\theta \log(X)) &= \left(\frac{\xi}{\phi}\right)^{\frac{\theta}{2}} R_\lambda(\rho) \left[\frac{k_{\lambda+1}(\rho)}{\theta K_{\lambda+1}(\rho)} + \frac{1}{2}(\log(\xi) - \log(\phi)) \right] \\
E(X^{-\theta} \log(X)) &= \left(\frac{\phi}{\xi}\right)^{\frac{\theta}{2}} \frac{1}{R_{\lambda-1}(\rho)} \left[\frac{k_{\lambda-1}(\rho)}{\theta K_{\lambda-1}(\rho)} + \frac{1}{2}(\log(\xi) - \log(\phi)) \right] \\
E(X^\theta (\log(X))^2) &= \left(\frac{\phi}{\xi}\right)^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\log(x))^2 x^{(\lambda+1)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^{-\theta} (\log(X))^2) &= \left(\frac{\phi}{\xi}\right)^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\log(x))^2 x^{(\lambda-1)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^{2\theta} \log(X)) &= \left(\frac{\xi}{\phi}\right)^\theta \frac{K_{\lambda+2}(\rho)}{K_\lambda(\rho)} \left[\frac{k_{\lambda+2}(\rho)}{\theta K_{\lambda+2}(\rho)} + \frac{1}{2}(\log(\xi) - \log(\phi)) \right] \\
E(X^{-2\theta} \log(X)) &= \left(\frac{\phi}{\xi}\right)^\theta \frac{K_{\lambda-2}(\rho)}{K_\lambda(\rho)} \left[\frac{k_{\lambda-2}(\rho)}{\theta K_{\lambda-2}(\rho)} + \frac{1}{2}(\log(\xi) - \log(\phi)) \right] \\
E(X^{2\theta} (\log(X))^2) &= \left(\frac{\phi}{\xi}\right)^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\log(x))^2 x^{(\lambda+2)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^{-2\theta} (\log(X))^2) &= \left(\frac{\phi}{\xi}\right)^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\log(x))^2 x^{(\lambda-2)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx
\end{aligned}$$

Com os valores esperados calculados acima, a matriz de Fisher pode ser calculada e conseqüentemente a distribuição a priori de Jeffreys.

Apêndice C

**PRIORI DE JEFFREYS PARA DADOS INDEPENDENTES E IDENTICAMENTE
DISTRIBUÍDOS DA NORMAL ASSIMÉTRICA**

Liseo e Loperfido (2004) desenvolveram a distribuição a priori de referência para os 3 parâmetros da distribuição normal assimétrica (NA), calculando para isso a matriz de informação de Fisher e em seguida a distribuição a priori de Jeffreys. A idéia é sempre trabalhar com a distribuição a priori de Jeffreys conjunta para todos os parâmetros, então como extensão desenvolvemos a distribuição a priori de Jeffreys para o modelo de regressão na posição da distribuição NA. Sejam $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ uma amostra aleatória onde y_i denota a variável resposta. O modelo de regressão linear é dado por:

$$y_i \sim NA(\mathbf{x}_i\boldsymbol{\beta}, \sigma, \lambda), \quad \lambda \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^p, \sigma \in \mathbb{R}^+, \quad i = 1 \dots, n \quad (\text{C.1})$$

onde $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ é o vetor de regressoras e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ é o vetor de coeficientes. Assim, o logaritmo da função de verossimilhança do modelo de regressão NA supondo uma amostra de tamanho n é dada por

$$\ell(\lambda, \sigma, \boldsymbol{\beta}) = -\frac{n}{2} \log(2/\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i\boldsymbol{\beta})^2 + \sum_{i=1}^n \log \Phi \left(\lambda \frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma} \right). \quad (\text{C.2})$$

Portanto, temos a função escore dada por

$$U_1(Y; \lambda) = \frac{\partial}{\partial \lambda} \ell(\lambda, \sigma, \boldsymbol{\beta}) = \sum_{i=1}^n Z_i \frac{\phi(\lambda Z_i)}{\Phi(\lambda Z_i)} \quad (\text{C.3})$$

$$U_2(Y; \sigma) = \frac{\partial}{\partial \sigma} \ell(\lambda, \mu, \sigma) = \sum_{i=1}^n \frac{Z_i^2}{\sigma} - \frac{n}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^n Z_i \frac{\phi(\lambda Z_i)}{\Phi(\lambda Z_i)} \quad (\text{C.4})$$

$$U_3(Y; \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\mu}} \ell(\lambda, \mu, \sigma) = \frac{1}{\sigma} \sum_{i=1}^n Z_i \mathbf{x}_i' - \frac{\lambda}{\sigma} \sum_{i=1}^n \frac{\phi(\lambda Z_i)}{\Phi(\lambda Z_i)} \mathbf{x}_i' \quad (\text{C.5})$$

onde $Z_i = \frac{Y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}$.

Definimos o vetor $\mathbf{U}(Y, \lambda, \sigma, \boldsymbol{\beta}) = (U_1(Y; \lambda), U_2(Y; \sigma), U_3(Y; \boldsymbol{\beta}))'$ e calculamos a matriz de Fisher por $\mathbf{I}(\lambda, \sigma, \boldsymbol{\beta}) = E_Y[\mathbf{U}(Y, \lambda, \sigma, \boldsymbol{\beta})\mathbf{U}'(Y, \lambda, \sigma, \boldsymbol{\beta})]$. Usando a transformação de variáveis $z_i = \frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}$ com $dz = dy_i/\sigma$ e que se $Z \sim NA(\lambda) \Rightarrow Z^2 \sim \chi_{(1)}^2$ temos

$$\begin{aligned}
E[U_1^2] &= \sum_{i=1}^n \int_{-\infty}^{\infty} 2z_i^2 \frac{\phi(\lambda z_i)}{\Phi(\lambda z)} \phi(z_i) dz_i = na_2(\lambda) \\
E[U_2^2] &= \frac{1}{\sigma^2} \sum_{i=1}^n \int_{-\infty}^{\infty} \left[z_i^4 - 4z_i^2 + 3 + \lambda^2 z_i^2 \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \right] 2\phi(\lambda z_i) \Phi(\lambda z_i) dz_i \\
&= \frac{n}{\sigma^2} [E(Z^4) - 4E(Z^2) + 3 + \lambda^2 a_2(\lambda)] = \frac{n}{\sigma^2} (2 + \lambda^2 a_2(\lambda)) \\
E[U_3^2] &= \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}'_i \mathbf{x}_i) \int_{-\infty}^{\infty} \left[z_i^2 - 2\lambda z_i \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} + \lambda^2 \frac{\phi^2(\lambda z_i)}{\Phi^2(\lambda z_i)} \right] 2\phi(\lambda z_i) \Phi(\lambda z_i) dz_i \\
&= \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}'_i \mathbf{x}_i) [E(Z^2) + \lambda^2 a_0(\lambda)] = \frac{1}{\sigma^2} (1 + \lambda^2 a_0(\lambda)) \sum_{i=1}^n (\mathbf{x}'_i \mathbf{x}_i) \\
E[U_1 U_2] &= \frac{1}{\sigma} \sum_{i=1}^n \int_{-\infty}^{\infty} \left[z_i^3 - z_i - \lambda z_i^2 \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \right] 2\phi(\lambda z) \phi(z_i) dz_i = -\frac{n\lambda}{\sigma} a_2(\lambda) \\
E[U_1 U_3] &= \frac{1}{\sigma} \sum_{i=1}^n \mathbf{x}_i \int_{-\infty}^{\infty} \left[z_i^2 - \lambda z_i \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \right] 2\phi(\lambda z_i) \phi(z_i) dz_i \\
&= -\frac{\lambda}{\sigma} \left[\sqrt{\frac{2}{\pi}} (1 + \lambda^2)^{-3/2} - \lambda a_1(\lambda) \right] \sum_{i=1}^n \mathbf{x}_i \\
E[U_2 U_3] &= \frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \int_{-\infty}^{\infty} \left[z_i^3 - z_i - 2\lambda z_i^2 \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} + \lambda \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} + \lambda^2 z_i \frac{\phi^2(\lambda z_i)}{\Phi^2(\lambda z_i)} \right] 2\phi(z_i) \Phi(\lambda z_i) dz_i \\
&= \frac{1}{\sigma^2} \left[\sqrt{\frac{2}{\pi}} \frac{\lambda(1 + 2\lambda^2)}{(1 + \lambda^2)^{3/2}} + \lambda^2 a_1(\lambda) \right] \sum_{i=1}^n \mathbf{x}_i
\end{aligned}$$

onde

$$a_j(\lambda) = \int_{-\infty}^{\infty} 2z^j \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} \phi(z) dz, \quad \text{para } j = 0, 1, 2$$

Segue-se que

$$\mathbf{I}(\lambda, \sigma, \boldsymbol{\beta}) = \begin{bmatrix} na_2(\lambda) & -\frac{\lambda}{\sigma} a_2(\lambda) & -\frac{n\lambda}{\sigma} \left[\sqrt{\frac{2}{\pi}} (1 + \lambda^2)^{-3/2} - \lambda a_1(\lambda) \right] \sum_{i=1}^n \mathbf{x}_i \\ \dots & \frac{n}{\sigma^2} (2 + \lambda^2 a_2(\lambda)) & \frac{1}{\sigma^2} \left[\sqrt{\frac{2}{\pi}} \frac{\lambda(1 + 2\lambda^2)}{(1 + \lambda^2)^{3/2}} + \lambda^2 a_1(\lambda) \right] \sum_{i=1}^n \mathbf{x}_i \\ \dots & \dots & \frac{1}{\sigma^2} (1 + \lambda^2 a_0(\lambda)) \sum_{i=1}^n (\mathbf{x}'_i \mathbf{x}_i) \end{bmatrix}$$

Logo, a distribuição a priori de Jeffreys é dada por

$$p(\lambda, \sigma, \mu) \propto \frac{1}{\sigma^{2p}} |A(\lambda)|^{1/2} \quad (\text{C.6})$$

onde

$$A(\lambda) = \begin{bmatrix} a_2(\lambda) & -\lambda a_2(\lambda) & -\sqrt{\frac{2}{\pi}}(1 + \lambda^2)^{-3/2} - \lambda a_1(\lambda) \sum_{i=1}^n \mathbf{x}_i \\ \dots & 2 + \lambda^2 a_2(\lambda) & \sqrt{\frac{2}{\pi}} \frac{\lambda(1+2\lambda^2)}{(1+\lambda^2)^{3/2}} + \lambda^2 a_1(\lambda) \sum_{i=1}^n \mathbf{x}_i \\ \dots & \dots & (1 + \lambda^2 a_0(\lambda)) \sum_{i=1}^n (\mathbf{x}'_i \mathbf{x}_i) \end{bmatrix}$$

Então podemos pensar na distribuição a priori de Jeffreys como o produto de 3 funções: $g_1(\sigma) = 1/\sigma^2$, $g_2(\lambda) = |A(\lambda)|^{1/2}$ e $g_3(\mu) = 1$, podendo facilmente avaliar o comportamento da mesma. Vale ressaltar que a distribuição a priori de Jeffreys para λ dada pela função de verossimilhança perfilada com $\mu = 0$ e $\sigma = 1$ é $p(\lambda) \propto |a_2(\lambda)|^{1/2}$.

A distribuição a priori de referência, de Jeffreys apresentada em Liseo e Loperfido (2004) e a de Jeffreys para o modelo de regressão apresentam bimodalidade quando olhado para a distribuição marginal de λ , com função de densidade de probabilidade igual a 0 em $\lambda = 0$. Este resultado é devido a matriz de Fisher ser singular no ponto $\lambda = 0$. Esperávamos a que a distribuição a priori favorecesse a modelos simétricos ($\lambda = 0$) assim como a distribuição a priori de Jeffreys de λ (somente). Diversas reparametrizações foram estudadas, sendo que a maior parte levava aos mesmos problemas e outras transferiam o problema para a função de verossimilhança. A solução encontrada até o presente momento foi considerar a distribuição a priori de λ somente com uma distribuição imprópria para os coeficientes β e a escala σ , isto é, $p(\beta, \sigma) \propto 1/\sigma^{2p}$ onde p é o número de variáveis regressoras.

C.1 Aplicação: instituto Australiano de esportes

Aplicamos o modelo de regressão linear (C.1) aos dados de 102 atletas homens e 100 mulheres coletados pelo instituto Australiano de esportes (AIS) (Cook e Weisberg, 1994). Nos definimos os seguintes modelos alternativos:

$$(M_1) \quad BMI_i \sim \text{NA}(\beta_1 + \beta_2 LBM_i, \sigma, \lambda).$$

$$(M_2) \quad BMI_i \sim \text{NA}(\beta_1 + \beta_2 LBM_i + \beta_3 SEXO_i, \sigma, \lambda).$$

$$(M_3) \quad BMI_i \sim \text{NA}(\beta_1 + \beta_2 LBM_i + \beta_3 SEXO_i + \beta_4 LBM_i SEXO_i, \sigma, \lambda).$$

$$(M_4) \quad BMI_i \sim \text{NA}(\beta_1 + \beta_2 LBM_i + \beta_4 LBM_i SEXO_i, \sigma, \lambda).$$

onde LBM é a massa corporal magra, BMI é o índice de massa corporal - peso/(altura)² - e SEX é uma variável binária indicando 0 para mulheres e 1 para homens. Geramos um MCMC para obter uma amostra da distribuição a posteriori. Especificamente, utilizamos o amostrados da fatia (Neal, 2003) para as condicionais completas de λ e σ , e Metropolis-Hasting para os β . Um sumário da distribuição a posteriori baseados em uma amostra de tamanho 20.000 tomadas após 5.000 iterações do MCMC é

dada pela tabela C.1 onde todos os critérios de seleção de modelos são dados na tabela C.2 de acordo com o apêndice D. Escolhemos o modelo M_3 como o melhor ajuste para os dados. Tomamos a média a posteriori como estimador pontual dos parâmetros, de tal forma que os critérios DIC e EPD estivessem bem definidos.

Tabela C.1: Média a posteriori e desvio padrão entre parênteses dos modelos alternativos de regressão NA.

Parâmetro	M_1	M_2	M_3	M_4
λ	4,286 (0,838)	2,818 (0,522)	2,998 (0,588)	2,820 (0,520)
σ	3,123 (0,173)	2,706 (0,157)	2,684 (0,153)	2,667 (0,151)
β_1	11,288 (0,053)	8,641 (0,053)	5,818 (0,050)	7,172 (0,049)
β_2	0,143 (0,001)	0,206 (0,001)	0,257 (0,001)	0,233 (0,001)
β_3	-	-2,076 (0,053)	2,734 (0,051)	-
β_4	-	-	-0,078 (0,001)	-0,036 (0,001)

Tabela C.2: Comparação de modelos alternativos de regressão NA.

Critério	M_1	M_2	M_3	M_4
AIC	827	803	798	800
BIC	841	820	817	820
DIC	824	798	791	793
EPD	1608	1314	1276	1280

Todos os parâmetros tem baixa probabilidade de serem iguais a 0 e a figura C.1 exibe uma distribuição assimétrica para os resíduos do modelo de regressão. A média a posteriori do coeficiente de assimetria é $\gamma_1 = 0,653$ com intervalo de credibilidade de 95% (0,468;0,801). Concluimos, então, que o modelo de regressão possui uma certa assimetria para a direita.

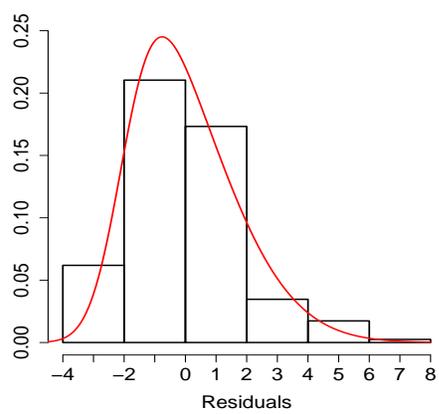


Figura C.1: Histograma dos resíduos do modelo de regressão.

Apêndice D

CRITÉRIOS DE SELEÇÃO DE MODELOS

Nesta seção vamos fazer uma breve revisão dos critérios de adequação, determinação e/ou seleção de modelos. Temos a classe de modelos alternativos \mathcal{M}_i , $i = 1, 2, \dots, q$, com uma função de verossimilhança bem definida para cada um, $p_i(\mathbf{z}|\Theta)$, $i = 1, 2, \dots, q$, onde Θ representa o vetor de parâmetros.

D.1 AIC e BIC

A escolha entre o melhor modelo alternativo é tradicionalmente feito pelo uso do critério de informação que penaliza a função de verossimilhança, com o AIC (Akaike, 1974),

$$AIC(\mathcal{M}_i) = -2 \log p_i(\mathbf{z}|\hat{\Theta}_i) + 2d_i \quad (\text{D.1})$$

e o BIC (Schwarz, 1978),

$$BIC(\mathcal{M}_i) = -2 \log p_i(\mathbf{z}|\hat{\Theta}_i) + \log(n)d_i \quad (\text{D.2})$$

sendo os mais populares. Aqui, d_i e $\hat{\Theta}_i$ são, respectivamente, o número de parâmetros e o estimador de máxima verossimilhança sob o modelo i , para $i = 1, 2$. Um dos maiores problemas com AIC/BIC é definir p_i , pois nem sempre é fácil onde a distribuição a priori implicitamente reduz o *número efetivo de parâmetros* através da dependência entre eles.

D.2 DIC

Spiegelhalter et al. (2002) desenvolveram um critério de informação onde o número efetivo de parâmetros é definido como $d_i = \bar{D}_i - D(\tilde{\Theta}_i)$, onde $D(\Theta_i) = -2 \ln p(\mathbf{z}|\Theta_i)$ é a função de *desvio*, $\tilde{\Theta}_i = E(\Theta_i|\mathbf{z}, \mathcal{M}_i)$ e $\bar{D}_i = E(D(\Theta_i)|\mathbf{z}, \mathcal{M}_i)$. O *Critério de Informação do Desvio*, ou simplesmente DIC, é definido como $DIC(\mathcal{M}_i) = -2 \log p_i(\mathbf{z}|\tilde{\Theta}_i) + 2d_i$, com \bar{D}_i e d_i medindo o ajuste e a complexidade do modelo, respectivamente. O DIC generaliza o AIC e baixos valores do DIC significa melhor ajuste. Em termos computacionais, o DIC é atrativo pois pode ser facilmente incorporado em algoritmos de Monte Carlo via cadeias de Markov. Para propostas computacionais, o DIC pode ser reescrito como $DIC(\mathcal{M}_i) = 2\bar{D}_i - D(\tilde{\Theta}_i)$. Sob o modelo \mathcal{M}_i , suponha que $\{\Theta_i^{(1)}, \dots, \Theta_i^{(M)}\}$ corresponde a uma amostra de $\pi(\Theta_i|\mathbf{z}, \mathcal{M}_i)$, isto é, a distribuição a posteriori dos parâmetros. Assim, o DIC pode ser estimado como,

$$DIC(\mathcal{M}_i) \approx -\frac{4}{M} \sum_{j=1}^M \log p_i(z|\Theta_i^{(j)}) + 2 \log p_i(z|\hat{\Theta}_i) \quad (\text{D.3})$$

com $\hat{\Theta}_i = \frac{1}{M} \sum_{j=1}^M \Theta_i^{(j)}$ uma estimativa de $\tilde{\Theta}_i$.

Entretanto, podemos usar a mesma idéia do DIC e calcular $(AIC_j^{(1)}, \dots, AIC_j^{(N)})$ e $(BIC_j^{(1)}, \dots, BIC_j^{(N)})$ para toda amostra da distribuição a posteriori do modelo \mathcal{M}_j e suas respectivas médias $EAIC(\mathcal{M}_j) \approx \frac{1}{N} \sum_{k=1}^N AIC_j^{(k)}$ e $EBIC(\mathcal{M}_j) \approx \frac{1}{N} \sum_{k=1}^N BIC_j^{(k)}$. Observe que em ambos os casos precisamos definir o número de parâmetros dos modelos (veja Brooks (2002) para maiores detalhes).

D.3 EPD

O último critério usado para comparação de modelos é o proposto por Gelfand e Ghosh (1998). Este critério, o *Desvio Preditivo Esperado* (EPD), é obtido como minimização da perda a posteriori de um dado modelo, \mathcal{M}_i . Quando a perda do erro ao quadrado é considerada, o EPD pode ser calculado explicitamente. Neste caso, $EPD = \sum_{i=1}^n \sigma_i^2 + \frac{k}{k+1} \sum_{i=1}^n (\mu_i - y_i)^2$, onde μ_i e σ_i^2 são a média e a variância da distribuição preditiva, respectivamente ($\mu_i = E[Y_{i,rep}|y_{i,obs}]$ e $\sigma_i^2 = Var[Y_{i,rep}|y_{i,obs}]$). O modelo que minimiza este critério é selecionado.

Apêndice E

ALGUMAS DISTRIBUIÇÕES IMPORTANTES

Lema E.1 Se X é uma variável aleatória normal com média μ e variância σ^2 , isto é, $X \sim N(\mu; \sigma^2)$.

Então, $Y = |X - \mu|$ tem função de densidade dada por $f(y) = 2\frac{1}{\sigma}\phi\left(\frac{y}{\sigma}\right)\mathbb{I}_{\{y>0\}}$ onde $\phi(\cdot)$ é a função de densidade da normal padrão.

Prova: Temos

$$F_Y(y) = P(|X - \mu| \leq y) = F_X(y + \mu) - F_X(-y + \mu).$$

Segue-se que

$$\begin{aligned} f_Y(y) &= f_X(y + \mu) + f_X(-y + \mu) = \frac{1}{\sigma}\phi\left(\frac{y}{\sigma}\right) + \frac{1}{\sigma}\phi\left(\frac{-y}{\sigma}\right) \\ &= 2\frac{1}{\sigma}\phi\left(\frac{y}{\sigma}\right)\mathbb{I}_{\{y>0\}} \end{aligned}$$

pois $\phi(-x) = \phi(x)$.

Lema E.2 Se X é uma variável aleatória normal com média μ e variância σ^2 , isto é, $X \sim N(\mu; \sigma^2)$.

Então, $Y = (X - \mu)\mathbb{I}_{\{X - \mu > 0\}}$ tem função de densidade dada por $f(y) = 2\frac{1}{\sigma}\phi\left(\frac{y}{\sigma}\right)\mathbb{I}_{\{y>0\}}$ onde $\phi(\cdot)$ é a função de densidade da normal padrão.

Prova: Temos

$$\begin{aligned} F_Y(y) &= P(X - \mu \leq y | X - \mu > 0) = \frac{P(\mu \leq X \leq y + \mu)}{P(X > \mu)} \\ &= 2[F_X(y + \mu) - F_X(\mu)] \end{aligned}$$

Segue-se que

$$f_Y(y) = 2f_X(y + \mu) = 2\frac{1}{\sigma}\phi\left(\frac{y}{\sigma}\right)\mathbb{I}_{\{y>0\}}.$$

Lema E.3 *Seja Y uma variável aleatória contínua com função de densidade dada em (E.1) ou (E.2)*

com $\sigma = 1$. Então, $EY = \sqrt{\frac{2}{\pi}}$, $EX^2 = 1$, $EY^3 = \sqrt{\frac{2^3}{\pi}}$, e $EX^4 = 3$.

Prova: Temos

$$\begin{aligned} EY &= \int_0^{\infty} \frac{2y}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{1}{2} e^{-\frac{u}{2}} du = \sqrt{\frac{2}{\pi}} \\ EY^2 &= \int_0^{\infty} \frac{2y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = -\frac{2ye^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \Big|_0^{\infty} + 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1 \\ EY^3 &= \int_0^{\infty} \frac{2y^3}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{u}{2} e^{-\frac{u}{2}} du = \sqrt{\frac{2^3}{\pi}} \\ EY^4 &= \int_0^{\infty} \frac{2y^4}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} u^{\frac{5}{2}-1} e^{-\frac{u}{2}} du = 3 \end{aligned}$$

Segue-se que $\text{Var}Y = \left(1 - \frac{2}{\pi}\right)$ e o resultado geral (σ qualquer) pode facilmente ser provado.

Lema E.4 *Se X é uma variável aleatória t -Student posição μ , escala σ^2 e ν graus de liberdade, isto é,*

$X \sim t_{\nu}(\mu; \sigma^2)$. Então, $Y = |X - \mu|$ tem função de densidade dada por $f(y) = 2\frac{1}{\sigma} t_{\nu}\left(\frac{y}{\sigma}\right) \mathbb{I}_{\{y>0\}}$ onde

t_{ν} é a função de densidade da t -Student com posição 0, escala 1 e ν graus de liberdade.

Prova: Temos

$$F_Y(y) = P(|X - \mu| \leq y) = F_X(y + \mu) - F_X(-y + \mu).$$

Segue-se que

$$\begin{aligned} f_Y(y) &= f_X(y + \mu) + f_X(-y + \mu) = \frac{1}{\sigma} t_{\nu}\left(\frac{y}{\sigma}\right) + \frac{1}{\sigma} t_{\nu}\left(\frac{-y}{\sigma}\right) \\ &= 2\frac{1}{\sigma} t_{\nu}\left(\frac{y}{\sigma}\right) \mathbb{I}_{\{y>0\}} \end{aligned}$$

pois $t_{\nu}(-x) = t_{\nu}(x)$.

Lema E.5 *Se X é uma variável aleatória t -Student com posição μ , escala σ^2 e ν graus de liberdade,*

isto é, $X \sim t_{\nu}(\mu; \sigma^2)$. Então, $Y = (X - \mu)\mathbb{I}_{\{X-\mu>0\}}$ tem função de densidade dada por $f(y) =$

$2\frac{1}{\sigma} t_{\nu}\left(\frac{y}{\sigma}\right) \mathbb{I}_{\{y>0\}}$ onde t_{ν} é a função de densidade da t -Student com posição 0, escala 1 e ν graus de

liberdade.

Prova: Temos

$$\begin{aligned} F_Y(y) &= P(X - \mu \leq y | X - \mu > 0) = \frac{P(\mu \leq X \leq y + \mu)}{P(X > \mu)} \\ &= 2[F_X(y + \mu) - F_X(\mu)] \end{aligned}$$

Segue-se que

$$f_Y(y) = 2f_X(y + \mu) = 2\frac{1}{\sigma}t_\nu\left(\frac{y}{\sigma}\right)\mathbb{I}_{\{y>0\}}.$$

Lema E.6 *Seja Y uma variável aleatória contínua com função de densidade dada em (E.4) ou (E.5) com*

$$\begin{aligned} \sigma = 1. \text{ Então, } EY &= \Gamma_\nu \times \frac{2\nu^{\frac{1}{2}}}{\nu - 1} \text{ para } \nu > 1, EY^2 = \frac{\nu}{(\nu - 2)^2} \text{ para } \nu > 2, EY^3 = \Gamma_\nu \times \frac{4\nu^{\frac{3}{2}}}{(\nu - 1)(\nu - 3)} \\ \text{para } \nu > 3 \text{ e } EY^4 &= \frac{3\nu^2\Gamma(\frac{\nu-4}{2})}{4\Gamma(\frac{\nu}{2})} \text{ para } \nu > 4, \text{ onde } \Gamma_\nu = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})}. \end{aligned}$$

Prova: Temos

$$\begin{aligned} EY &= \Gamma_\nu \nu^{\nu/2} \int_0^\infty 2y(\nu + y^2)^{-\frac{(\nu+1)}{2}} dy \\ &= \Gamma_\nu \nu^{\nu/2} \frac{2}{\nu - 1} \left[-(\nu + y^2)^{-\frac{(\nu-1)}{2}} \right]_0^\infty = \Gamma_\nu \times \frac{2\nu^{\frac{1}{2}}}{\nu - 1}. \\ EY^2 &= \Gamma_\nu \nu^{\nu/2} \int_0^\infty 2y^2(\nu + y^2)^{-\frac{(\nu+1)}{2}} dy \\ &= \Gamma_\nu \nu^{\nu/2} \frac{2}{\nu - 1} \left(-\frac{y(\nu + y^2)^{-\frac{(\nu-1)}{2}}}{\nu - 2} \right) \Big|_0^\infty + \int_0^\infty (\nu + y^2)^{-\frac{(\nu-1)}{2}} dy \\ &= \frac{\Gamma_\nu \nu^{\frac{1}{2}}}{\nu - 1} \int_0^\infty 2 \left(1 + \frac{1}{\nu - 2} \left(\frac{y}{\sqrt{\frac{\nu}{\nu-2}}} \right)^2 \right)^{-\frac{((\nu-2)+1)}{2}} dy = \frac{\nu}{(\nu - 2)^2}. \end{aligned}$$

$$\begin{aligned}
\mathbf{E}Y^3 &= \Gamma_\nu \nu^{\frac{\nu}{2}} \int_0^\infty 2y^3 (\nu + y^2)^{-\frac{(\nu+1)}{2}} dy \\
&= \Gamma_\nu \nu^{\frac{\nu}{2}} \int_0^\infty u (\nu + u)^{-\frac{(\nu+1)}{2}} du \\
&= \Gamma_\nu \nu^{\frac{\nu}{2}} \frac{2}{\nu - 1} \left(-u (\nu + u)^{-\frac{(\nu-1)}{2}} \right]_0^\infty + \int_0^\infty (\nu + u)^{-\frac{\nu-1}{2}} du \Big) \\
&= \Gamma_\nu \times \frac{4\nu^{\frac{3}{2}}}{(\nu - 1)(\nu - 3)}. \\
\mathbf{E}Y^4 &= \Gamma_\nu \nu^{\frac{\nu}{2}} \int_0^\infty 2y^4 (\nu + y^2)^{-\frac{(\nu+1)}{2}} dy = \Gamma_\nu \nu^{\frac{\nu}{2}} \int_0^\infty u^{\frac{3}{2}} (\nu + u)^{-\frac{(\nu+1)}{2}} du \\
&= \Gamma_\nu \nu \int_0^\infty \left(\frac{u}{\nu}\right)^{\frac{3}{2}} \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+1)}{2}} du = \Gamma_\nu \nu^2 \int_0^\infty w^{\frac{3}{2}} (1 + w)^{-\frac{(\nu+1)}{2}} dw \\
&= \Gamma_\nu \nu^2 \int_0^1 s^{\frac{5}{2}-1} (1 - s)^{\frac{(\nu+4)}{2}-1} ds = \frac{3\nu^2 \Gamma(\frac{\nu-4}{2})}{4\Gamma(\frac{\nu}{2})}
\end{aligned}$$

O resultado geral (σ qualquer) pode facilmente ser provado.

Apêndice F

RESULTADOS DE MONTE CARLO PARA CÓPULAS

F.1 Marginais com distribuições normais

Tabela F.1: Número de identificações corretas usando o AIC, EAIC, BIC e EBIC entre as 500 replicações para o modelo de cópulas com distribuições marginais normais.

Cópula	τ n	1/3			2/3		
		100	200	500	100	200	500
AIC							
Clayton		461	490	500	497	500	500
Frank		324	385	476	473	495	500
Gaussian		238	368	464	425	474	499
Gumbel		238	330	449	432	481	500
Cauda-pesada		238	330	449	432	481	500
Student		310	410	489	393	458	496
EAIC							
Clayton		457	488	500	497	500	500
Frank		316	388	475	472	494	500
Gaussian		248	371	464	427	474	499
Gumbel		213	326	448	428	481	500
Cauda-pesada		355	404	471	467	497	500
Student		313	413	489	392	457	496
BIC							
Clayton		461	490	500	497	500	500
Frank		324	385	476	473	495	500
Gaussian		238	368	464	425	474	499
Gumbel		239	330	449	432	481	500
Cauda-pesada		342	403	470	466	496	500
Student		310	410	489	393	458	496
EBIC							
Clayton		457	488	500	497	500	500
Frank		316	388	475	472	494	500
Gaussian		248	371	464	427	474	499
Gumbel		213	326	448	428	481	500
Cauda-pesada		355	404	471	467	497	500
Student		313	413	489	392	457	496

F.2 Marginais com distribuições exponenciais

Tabela F.2: Número de identificações corretas usando o AIC, EAIC, BIC e EBIC entre as 500 replicações para o modelo de cópulas com distribuições marginais exponenciais.

Cópula	τ n	1/3			2/3		
		100	200	500	100	200	500
AIC							
Clayton		449	490	500	500	500	500
Frank		329	396	479	472	497	500
Gaussian		294	371	467	420	468	499
Gumbel		226	330	451	429	486	500
Cauda-pesada		383	428	471	483	500	500
Student		325	421	486	351	441	495
EAIC							
Clayton		446	490	500	500	500	500
Frank		322	396	480	472	497	500
Gaussian		292	374	466	424	468	499
Gumbel		205	321	450	429	486	500
Cauda-pesada		388	430	471	485	500	500
Student		327	423	486	354	442	496
BIC							
Clayton		449	490	500	500	500	500
Frank		329	396	479	472	497	500
Gaussian		294	371	467	420	468	499
Gumbel		226	330	451	429	486	500
Cauda-pesada		383	428	471	483	500	500
Student		325	421	486	350	441	495
EBIC							
Clayton		446	490	500	500	500	500
Frank		322	396	480	472	497	500
Gaussian		292	374	466	424	468	499
Gumbel		205	321	450	429	486	500
Cauda-pesada		388	430	471	485	500	500
Student		327	423	486	354	442	496

F.3 Marginais com distribuições normais assimétricas

Tabela F.3: Número de identificações corretas usando o AIC, EAIC, BIC e EBIC entre as 500 replicações para o modelo de cópulas com distribuições marginais normais assimétricas.

Cópula	τ n	1/3			2/3		
		100	200	500	100	200	500
AIC							
Clayton		436	481	500	492	500	500
Frank		307	402	476	449	494	500
Gaussian		257	368	482	394	461	494
Gumbel		241	317	435	353	461	500
Cauda-pesada		320	409	462	455	485	500
Student		323	409	493	367	442	496
EAIC							
Clayton		432	481	500	497	500	500
Frank		328	402	476	466	494	500
Gaussian		262	367	482	398	460	494
Gumbel		222	307	433	352	459	500
Cauda-pesada		370	412	461	474	488	500
Student		318	409	493	367	441	497
BIC							
Clayton		436	481	500	492	500	500
Frank		308	402	476	449	494	500
Gaussian		257	368	482	394	461	494
Gumbel		241	317	435	353	461	500
Cauda-pesada		320	409	462	455	485	500
Student		323	409	493	367	442	496
EBIC							
Clayton		432	481	500	497	500	500
Frank		328	402	476	466	494	500
Gaussian		262	367	482	398	460	494
Gumbel		222	307	433	352	459	500
Cauda-pesada		370	412	461	474	488	500
Student		318	409	493	367	441	497

Apêndice G

ABREVIACÕES

<i>AIC</i>	abreviação do inglês de <i>Akaike's information criterion</i>
<i>ARMA</i>	abreviação do inglês de <i>autoregressive moving average</i>
<i>ARFIMA</i>	abreviação do inglês de <i>autoregressive fractionally integrated moving average</i>
<i>BIC</i>	abreviação do inglês de <i>Bayesian information criterion</i>
<i>EAIC</i>	abreviação do inglês de <i>expected AIC</i> (Brooks, 2002)
<i>EAM</i>	erro absoluto médio
<i>EBIC</i>	abreviação do inglês de <i>expected BIC</i> (Brooks, 2002)
<i>EPD</i>	abreviação do inglês de <i>expected predictive deviance</i>
<i>f.d.p</i>	função de densidade de probabilidade
<i>f.d.a</i>	função de distribuição acumulada
<i>GARCH</i>	abreviação do inglês de <i>generalized autoregressive conditional heteroscedastic</i>
<i>GG</i>	gama generalizada
<i>GIGE</i>	Gaussiana inversa generalizada estendida
<i>IG</i>	inversa gama
<i>NA</i>	normal assimétrica
<i>NT</i>	normal truncada
<i>REQM</i>	raíz quadrada do erro quadrático médio
<i>TA</i>	t-Student assimétrica
<i>TT</i>	t-Student truncada

Apêndice H

ASPECTOS COMPUTACIONAIS

H.1 Amostrador de Gibbs

O amostrador de Gibbs (Geman e Geman, 1984; Gelfand e Smith, 1990) permite obter amostras das distribuições conjunta e marginais a posteriori, obtendo-se desta maneira um resumo estatísticos de tais distribuições, utilizando a estrutura Markoviana. Suponha que a distribuição de $p(\Theta | D) = p(\theta_1, \theta_2, \dots, \theta_n | D)$, ou seja, $(\Theta | Y_T)$ tem distribuição n -dimensional, e D é toda a informação disponível. A distribuição de $p_j(\theta_j | \{\theta_i\}_{j \neq i}, D)$ é chamada de condicional completa de θ_j . Tais condicionais completas são sempre proporcionais ao produto da função de verossimilhança pela distribuição a priori, isto é, $p_j(\theta_j | \{\theta_i\}_{j \neq i}, D) \propto p(Y_T | \Theta)p(\Theta) \forall j$. Dado o conjunto de valores iniciais $(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)})$, as amostras são obtidas da seguinte maneira:

$$1) \theta_1^{(k)} \sim p_1(\theta_1 | \theta_2^{(k-1)}, \dots, \theta_n^{(k-1)}, D)$$

$$2) \theta_2^{(k)} \sim p_2(\theta_2 | \theta_1^{(k)}, \theta_3^{(k-1)} \dots \theta_n^{(k-1)}, D)$$

$$\vdots$$

$$n) \theta_n^{(k)} \sim p_n(\theta_n | \theta_1^{(k)}, \dots, \theta_{n-1}^{(k)}, D)$$

- Repita passos 1, 2, \dots , n para $k = 1, 2, 3, \dots$

Sob certas condições de regularidade (Gelfand e Smith, 1990; Tierney, 1994) a cadeia de Markov acima converge para uma distribuição estacionária e esta é a distribuição a posteriori de $p(\theta_1, \theta_2, \dots, \theta_n | D)$.

H.2 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings foi desenvolvido por Metropolis et al. (1953) e depois generalizado por Hastings (1970). Em Tierney (1994) encontra-se uma boa exposição teórica deste algoritmo. O algoritmo depende de uma densidade proposta $q(\theta, \xi)$ onde

$$\int_{R_\xi} q(\theta, \xi) d\xi = 1.$$

Denotando-se $U(0, 1)$ como uma distribuição uniforme no intervalo $(0, 1)$, tem-se uma versão bem geral do algoritmo de Metropolis-Hastings para amostrar da distribuição a posteriori de $p(\theta | D)$

0) Escolha um valor inicial θ_0 e faça $i = 0$

1) Gere um valor candidato θ^* da proposta $q(\theta_i, \cdot)$ e u de $U(0, 1)$

2) Faça
$$\begin{cases} \theta_{i+1} = \theta^* & , \text{ se } u \leq a(\theta_i, \theta^*) \\ \theta_{i+1} = \theta_i & , \text{ caso contrário} \end{cases}$$

onde a probabilidade de aceitação $a(\theta, \xi) = \min \left\{ 1, \frac{p(\xi | D)q(\xi, \theta)}{p(\theta | D)q(\theta, \xi)} \right\}$

3) Faça $i=i+1$ e volte para o passo 1.

Aconselha-se calcular a probabilidade de aceitação através de seu logaritmo para evitar problemas de precisão numérica.

H.3 Amostrador da fatia

O amostrador da fatia - traduzido do inglês *slice sampler* - é um método que serve para gerar valores de uma variedade de distribuições, em especial densidades que estejam num intervalo fechado e limitado. A idéia básica é gerar valores da distribuição uniforme definida pela região abaixo da função de densidade ao qual se deseja amostrar, e considerar apenas as coordenadas horizontais. Uma cadeia de Markov que converge para esta distribuição uniforme pode ser construída alternando-se valores na direção vertical e na horizontal definindo a fatia pela posição vertical atual. Para isso se introduz variáveis auxiliares e então usa-se o amostrador de Gibbs para a área abaixo da densidade. Suponha que se deseja gerar valores de uma densidade $p(x)$, $x \in A \subseteq \mathbb{R}$. Para isso considera-se a região bi-dimensional abaixo de $p(x)$ ou de $g(x) = cp(x)$, e então:

- Introduz-se uma variável auxiliar z com $z | x \sim U(0, g(x))$;
- Segue-se que a distribuição conjunta de (z, x) é uniforme na região $\{(z, x) : 0 \leq z \leq g(x)\}$ com densidade

$$f(z, x) = \begin{cases} \frac{1}{c} & , \text{ se } 0 \leq z \leq g(x) \\ 0 & , \text{ caso contrário} \end{cases}$$

- A distribuição condicional de $x | z$ $f(x | z) \propto f(z, x) = \begin{cases} \frac{1}{c} & , \text{ se } 0 \leq z \leq g(x) \\ 0 & , \text{ caso contrário} \end{cases}$, ou seja, $(x | z) \sim U(S(z))$, onde $S(z) = \{x : g(x) \geq z\}$.

Assim, $S(x)$ é a união dos intervalos que constituem a fatia através da densidade definida por z . A obtenção de $S(x)$ na maioria das vezes não é simples. A estrutura do modelo acima nos conduz a simular tais valores usando o amostrador de Gibbs, onde as condicionais completas são uniformes.

- Gere $z^{(i)} \sim U(0, g(x^{(i-1)}))$
- Gere $x^{(i)} \sim U(S(z^{(i)}))$ onde $S(z^{(i)}) = \{x : g(x) \geq z^{(i)}\}$

Dentre as vantagens desse método estão: aplica-se a várias distribuições, não necessita especificar uma densidade proposta como no Metropolis-Hastings e somente utiliza a distribuição uniforme para gerar os valores. A principal desvantagem é que a determinação de $S(z)$ pode ser difícil. O método é detalhado em Neal (2003), onde se propõe uma maneira de solucionar a determinação de $S(z)$ e a generalização do algoritmo.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)