

Universidade Federal do Rio de Janeiro



MODELOS DINÂMICOS PARA OBSERVAÇÕES BINÁRIAS COM
FUNÇÃO DE LIGAÇÃO ASSIMÉTRICA

Renata Souza Bueno

2011

MODELOS DINÂMICOS PARA OBSERVAÇÕES BINÁRIAS COM FUNÇÃO DE LIGAÇÃO ASSIMÉTRICA

Renata Souza Bueno

Dissertação submetida ao Corpo Docente do Instituto de Matemática - Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Orientador: Carlos Antonio Abanto-Valle

Rio de Janeiro

Junho de 2011

Modelos dinâmicos para observações binárias com função de
ligação assimétrica

Renata Souza Bueno

Dissertação submetida ao Corpo Docente do Instituto de Matemática - Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Aprovada por:

Prof^o. Carlos A. Abanto-Valle
DME - UFRJ - Orientador

Prof^o. Helio dos Santos Migon
DME - UFRJ

Prof^o. Víctor H. Lachos
IMECC - UNICAMP

Rio de Janeiro

Junho de 2011

À minha estrela guia, minha mãe, Dona Stella.

”Quanto mais aumenta nosso conhecimento, mais evidente fica nossa ignorância.”

John F. Kennedy

Agradecimentos

Primeiramente agradeço a minha mãe, pelo apoio incondicional. Sou grata por poder ter convivido com ela todos esses anos e receber seu amor e carinho. Ela sempre lutou para que eu tivesse condições de estudar e correr atrás dos meus sonhos. Sem ela eu não estaria aqui. Mãe, obrigada por tudo.

Ao meu exemplo, minha avó Edna, que sempre foi a minha referência de perseverança e de caráter. Obrigada por todo o carinho, confiança e cuidado comigo.

Agradeço ao meu pai todas as conversas e idas ao cinema, fundamentais para a minha formação. Ao meu irmão pelo companheirismo, por estar ao meu lado e por me dar o grande presente de ser tia. A minha tia Sandra, madrinha coruja, sempre torcendo por mim. A minha tia Regina, pelo carinho e pela ajuda. A minha tia Tania, que mesmo longe fisicamente, sempre se fez presente. Ao tio Mauro e à Manuela, pelas risadas e pelos conselhos. Meus queridos primos, Lucas e Gustavo. E minhas queridas primas, Bella e Bruneca, únicas e essenciais na minha vida. Bruneca, obrigada em especial pela ajuda no português. A minha prima Su, uma das minhas grandes incentivadoras. Aos meus tios-avós por toda a ajuda e torcida nessa caminhada.

Aos meus professores da graduação que foram fundamentais em toda a minha trajetória acadêmica. Aos professores do mestrado por todo o conhecimento e incentivo durante o curso. Agradeço, com carinho, à professora Alexandra pelos ensinamentos, conselhos e atenção que me deu durante esse caminho. Ao meu orientador, professor Carlos, pela dedicação, força, paciência e, por tudo que me ensinou e dividiu durante todo o trabalho.

As minhas amigas de turma que fizeram parte desta etapa comigo. Todos os momentos de estudo, todas as dúvidas, as alegrias e tristezas compartilhadas. As meninas foram fundamentais nesta trajetória, com elas tudo foi mais fácil. Agradeço à Pri por tornar

tudo mais engraçado, à Camilinha pela seriedade e por ser sempre a que me entendia quando estávamos estudando, à Carol por ter me aturado desde a graduação. Em especial, à Jones por ter se tornando tão importante nesta fase, por ter dividido as aflições, as risadas, por ter sido minha companheira em todos os momentos. À família LPGE, todo o convívio neste laboratório foi de extrema importância, tanto academicamente falando como pessoalmente. Agradeço com carinho especial à Sheilinha, Kelly Dance (parceira oficial de dança), Larisson, Patty e Joãozinho.

Agradeço aos meus amigos que entenderam todas as ausências nos eventos e que sempre me apoiaram. A minha “best”, Luana, por sempre estar ao meu lado, pelo carinho e por me fazer tão bem. À Vanessa pela dedicação e pela torcida. À Fabi por seu companheirismo. À Ju, minha amiga VIP, pela força. À Alê e Michele pela alegria dos nossos encontros. À tia Neli por ser um anjinho na minha vida. Ao Fernando Pulgati, por me apresentar a pesquisa, pelos ensinamentos e pelo incentivo no ingresso ao mestrado.

Agradeço em especial a todas as pessoas que fazem cinema e teatro no Brasil, duas artes que me deixam extremamente maravilhada e não me deixam enlouquecer muito. Ao Chico Buarque, Tom Jobim e Vinicius de Moraes pelas suas belas canções que embalaram meus estudos. À cantora Pitty, por compor músicas libertadoras e por fazer shows que tem um poder renovador pra mim. E à escritora Martha Medeiros por todos os textos que me fizeram refletir, questionar e que me confortaram.

Agradeço a todos que fizeram parte deste trabalho, de forma direta ou indireta. Obrigada pela confiança.

Agradeço à CAPES pelo apoio financeiro.

Finalmente, agradeço aos professores Helio Migon e Víctor Lachos por aceitarem fazer parte da minha banca.

Resumo

O intuito deste trabalho é propor modelos que melhor ajustem uma série temporal de observações binárias. É natural assumir a distribuição Bernoulli para dados binários e que a relação entre a probabilidade de sucesso e os parâmetros é dada por uma função de ligação. Quando esta função possui uma forma simétrica em torno de 0.5 espera-se obter quantidades parecidas de “0” e “1” na amostra binária. Quando isto não ocorre, ligações simétricas não são apropriadas pois a inferência é sensível a escolha desta função e portanto podem ser estimados valores incorretos para os parâmetros.

Neste trabalho, será usada uma abordagem que envolve modelos dinâmicos binários com função de ligação assimétrica e a idéia de aumento de dados. Tal idéia permite modelar dados binários por meio de uma variável não observável. Através desta abordagem, a classe de modelos dinâmicos poderá utilizar a família de distribuições de mistura de escala normal assimétrica como função de ligação. Portanto, com a utilização de tais técnicas, serão elaborados modelos dinâmicos mais flexíveis que podem servir para ajustar dados com desvios na suposição de normalidade e/ou assimetria, acomodando assim séries binárias que possuam taxas diferentes de “0” e “1” na amostra. Quatro modelos serão propostos e toda a inferência será feita sob o enfoque Bayesiano. Um eficiente algoritmo de estimação baseado no método de Monte Carlo via cadeias de Markov (MCMC) será desenvolvido. Este trabalho inclui um exercício com dados gerados artificialmente e uma aplicação a dados reais.

Palavras Chaves: dados binários, mistura de escala normal assimétrica, modelos dinâmicos binários, aumento de dados, inferência Bayesiana, MCMC.

Abstract

In this work a new class of dynamic models is proposed that better fit series of binary observations. It is natural to assume Bernoulli distribution for binary data and that the relationship between the probability of success and the parameters is given by a link function. When this function has a symmetrical shape around 0.5, similar quantities to "0" and "1" are expected in the sample. Otherwise, symmetric links are not appropriate because since the inference is sensitive to the choice of this function and the parameters can be wrongly estimated.

In this work an approach that involves binary dynamic models with asymmetric link function and the idea of data augmentation will be used. This idea allows the modeling of binary data through a latent variable. Under the proposed framework, the class of dynamic models can use the scale mixture of skew normal distributions family as a link function. By using such techniques, will be formulated more flexible dynamic models that could adjust to data with deviations in normality and/or asymmetry, accommodating binary series that have different rates of "0" and "1" in the sample. Four models will be proposed and the inference will be made under the Bayesian approach. An efficient estimation algorithm based on Monte Carlo Markov chain (MCMC) will be developed. The work includes an exercise with artificial data and an application to real data.

Keywords: binary data, scale mixture of skew normal, dynamic models binary, data augmentation, Bayesian inference, MCMC.

Sumário

1	Introdução	1
2	Conceitos preliminares	4
2.1	Introdução	4
2.2	Inferência Bayesiana	4
2.2.1	Estimação pontual	6
2.2.2	Estimação por intervalo	7
2.3	Métodos de simulação estocástica	8
2.3.1	Algoritmo de Metropolis-Hastings	8
2.3.2	Amostrador de Gibbs	9
2.3.3	Diagnósticos de convergência	10
2.4	Critério de seleção de modelos	12
2.5	Modelos dinâmicos	13
2.5.1	Modelos lineares dinâmicos	14
2.5.2	Inferência nos modelos lineares dinâmicos	15
2.6	Modelos para dados binários	21
2.6.1	Modelos lineares generalizados	22
2.6.2	Modelos dinâmicos binários	23
2.7	Distribuição normal assimétrica	25
2.8	Distribuições de mistura de escala normal assimétrica	27

3 Modelos dinâmicos para observações binárias com função de ligação assimétrica	32
3.1 Introdução	32
3.2 Formulação do modelo	33
3.3 Modelos propostos	38
3.4 Procedimento de inferência	38
3.5 Estudo com dados artificiais	41
4 Aplicação empírica	51
4.1 Descrição dos dados	51
4.2 Modelos propostos e Procedimento de inferência	52
4.3 Resultados a <i>posteriori</i>	54
4.4 Comparação dos modelos	64
5 Conclusões e trabalhos futuros	67
A Cálculo das distribuições condicionais completas das variáveis latentes	69
B Definição da distribuição gaussiana inversa generalizada	72
C Cálculo das distribuições condicionais completas dos parâmetros	73
D Resultados a <i>posteriori</i> do modelo 2 para dados artificiais utilizando uma distribuição a <i>priori</i> não informativa para o parâmetro ν	78

Capítulo 1

Introdução

Uma variável aleatória é considerada binária ou dicotômica se só admite dois possíveis valores tais como sucesso ou fracasso, positivo ou negativo, correto ou incorreto. Este tipo de dados são comuns em aplicações na área social, médica, agrícola, genética entre outras. Geralmente atribui-se o valor “1” para o acontecimento de interesse chamado “sucesso” e o valor “0” para o acontecimento chamado de “fracasso”.

O objetivo deste trabalho é modelar séries binárias no tempo, propondo modelos mais flexíveis que possam acomodar melhor dados desta natureza.

É comum assumir que a variável aleatória binária segue uma distribuição Bernoulli(p), em que o parâmetro p denota a probabilidade de sucesso. A estimação deste parâmetro tem recebido uma grande atenção na literatura estatística. Usando o conceito dos modelos lineares generalizados (McCullagh e Nelder, 1989), assume-se que a relação entre a esperança da variável aleatória com possíveis covariáveis é dada por uma função de ligação. Entre as opções para função de ligação, é razoável adotar uma função de distribuição acumulada pois esta garante que a probabilidade de sucesso esteja no intervalo $[0,1]$. Na maioria dos casos esta função é simétrica em torno de $p = 0.5$. No entanto, segundo Chen et al. (1999) quando na amostra binária existem quantidades diferentes de “0” e “1”, funções de ligação simétricas não são indicadas.

A idéia proposta por este trabalho é utilizar uma família de distribuições mais geral como função de ligação, na qual é possível acomodar melhor amostras que possuam taxas diferentes de “0” e “1”.

Para modelar a série temporal binária será usada a classe de modelos dinâmicos binários como em Czado e Song (2008), tais modelos permitem explicitar efeitos de tendência, cíclico ou outras covariáveis que variam no tempo conservando o poder dos modelos dinâmicos na modelagem da dependência temporal da série dos dados. Em conjunto com esta classe de modelos será usada a idéia de aumento de dados, na qual assume-se que a variável binária é gerada através de uma variável aleatória contínua, não observável. Este fato possibilitará a não utilização da função de verossimilhança do tipo Bernoulli, no qual o processo de estimação dos parâmetros seria mais complicado.

A idéia do aumento de dados permite modelar a equação de observação do modelo dinâmico utilizando a família de distribuições de mistura de escala normal assimétrica. Portanto, com esta abordagem, será proposta uma nova família paramétrica de ligações assimétricas para a probabilidade de sucesso. Com isso serão elaborados modelos dinâmicos mais flexíveis que possam acomodar desvios na suposição de normalidade especialmente para dados com forte assimetria e/ou caudas pesadas. Estes modelos se apresentam como uma alternativa robusta ao modelo normal que geralmente é usado na literatura.

Toda a análise será feita sob o enfoque Bayesiano. A distribuição a *posteriori* dos modelos estudados neste trabalho não possui forma analítica e portanto é necessário a utilização de métodos de simulação estocástica para a obtenção de amostras da distribuição a *posteriori*. Neste trabalho, será desenvolvido um eficiente, em termos computacionais, algoritmo de estimação baseado nos métodos de Monte Carlo via cadeias de Markov.

Seguindo esta abordagem serão propostos quatro modelos e será realizado um exercício com dados gerados artificialmente e uma aplicação a um conjunto de dados reais. Na aplicação a dados reais, além dos modelos propostos, o modelo proibito também será ajustado. Sendo assim, os ajustes dos modelos serão comparados e discutidos.

As principais contribuições deste trabalho incluem a proposta de modelos dinâmicos para dados binários com função de ligação sendo a função de distribuição na classe de misturas de escala normal assimétrica, o que possibilita maior flexibilidade no ajuste dos dados, a comparação dos modelos propostos com o modelo com função de ligação proibito, frequentemente utilizado na literatura estatística (Czado e Song, 2008), e a proposta de algoritmos eficientes computacionalmente para amostrar da distribuição a *posteriori* dos

parâmetros e quantidades latentes na classe de modelos estudadas.

A organização do trabalho é feita da seguinte forma, no Capítulo 2 será feita uma revisão de conceitos que serão usados ao longo de todo o trabalho, são eles, o conceito de inferência Bayesiana, o conceito dos métodos de simulação estocástica e o critério de seleção de modelos. Também neste capítulo são apresentados os conceitos de modelos dinâmicos e é feita uma revisão de modelos para dados binários. A distribuição normal assimétrica é apresentada e as distribuições de mistura de escala normal assimétrica são discutidas no fim do Capítulo 2. No Capítulo 3, a construção dos modelos dinâmicos para observações binárias com função de ligação assimétrica é discutida em detalhes e os modelos propostos são apresentados bem como o procedimento de estimação referente a estes modelos e um estudo com dados simulados. No Capítulo 4, os modelos propostos são ajustados a uma série de dados reais e comparados. Finalmente, o Capítulo 5 mostra a conclusão do trabalho e apresenta idéias para trabalhos futuros.

Capítulo 2

Conceitos preliminares

2.1 Introdução

Neste capítulo apresenta-se uma revisão dos conceitos que serão usados ao longo deste trabalho. Assim, na Seção 2.2 é feita uma revisão sobre o procedimento de inferência a partir do ponto de vista Bayesiano. Na Seção 2.3 são apresentados os métodos de simulação estocástica via cadeias de Markov utilizados na inferência Bayesiana, são eles, o algoritmo de Metropolis-Hastings e o amostrador de Gibbs. É realizada uma breve discussão de como monitorar a convergência das cadeias simuladas por estes métodos. A Seção 2.4 discute o critério de seleção de modelos. A classe geral dos modelos dinâmicos e o seu procedimento de inferência são apresentados na Seção 2.5. Uma revisão da modelagem estatística de dados binários é feita na seção 2.6, entre as abordagens descritas estão os modelos lineares generalizados e os modelos dinâmicos para observações binárias. A distribuição normal assimétrica é definida na Seção 2.7 bem como algumas de suas propriedades. Finalmente, na Seção 2.8 são apresentadas as distribuições de mistura de escala normal assimétrica.

2.2 Inferência Bayesiana

Nesta seção, serão introduzidas as idéias básicas da inferência Bayesiana. Diferentemente da inferência clássica, a inferência Bayesiana trata os parâmetros de um modelo

como variáveis aleatórias, com isso proporciona ao modelo muito mais versatilidade.

Seja $\theta \in \Theta$ o parâmetro de interesse, em que Θ denota o espaço paramétrico. A incerteza inicial acerca de θ é quantificada pela densidade $p(\theta)$, a qual é denominada distribuição a *priori* de θ e está associada ao grau de incerteza antes de observar uma amostra de dados.

Após observar uma amostra x de um vetor aleatório X relacionado com θ através da distribuição amostral $p(X | \theta)$, o conhecimento a respeito desse parâmetro aumenta. A regra de atualização utilizada para quantificar este aumento de informação é dada pelo Teorema de Bayes. O teorema é descrito pela seguinte expressão:

$$p(\theta | x) = \frac{p(x | \theta)p(\theta)}{\int_{\Theta} p(x | \theta)p(\theta)d\theta} . \quad (2.1)$$

A função $l(\theta; x) = p(x | \theta)$ é conhecida como função de verossimilhança e representa a informação que vem dos dados. Quando combinada com a distribuição a *priori*, obtém-se a distribuição a *posteriori* de θ , $p(\theta | x)$. A forma usual do Teorema de Bayes é

$$p(\theta | x) \propto l(\theta; x)p(\theta). \quad (2.2)$$

Esta forma simplificada do Teorema de Bayes será útil em problemas que envolvam estimação de parâmetros, já que o denominador é apenas uma constante normalizadora. Na maior parte das aplicações de interesse, a integral do denominador não possui forma analítica fechada e sua avaliação usando métodos numéricos em altas dimensões torna-se impraticável. Para aproximar a distribuição a *posteriori* usa-se, frequentemente, métodos de simulação estocástica.

Se após observar $X = x$, o interesse é prever uma quantidade X^* , também relacionada com θ , e descrita probabilisticamente por $p(x^* | \theta)$ então

$$p(x^* | x) = \int_{\Theta} p(x^*, \theta | x)d\theta = \int_{\Theta} p(x^* | \theta, x)p(\theta | x)d\theta = \int_{\Theta} p(x^* | \theta)p(\theta | x)d\theta \quad (2.3)$$

descreve a distribuição preditiva de x^* . A última igualdade se deve a independência entre X e X^* condicionado em θ . Finalmente, segue da última equação que

$$p(x^* | x) = E_{\theta|x}[p(x^* | \theta)].$$

Através da inferência Bayesiana, a estimação e a previsão não se limitam a um único valor, mas são representadas por uma distribuição (a *posteriori*, no caso de estimação, e preditiva, no caso de previsão).

Muitas vezes, existe o interesse de se obter uma informação resumida a respeito do parâmetro. O caso mais simples de sumarização é a estimação pontual, onde toda a informação presente é resumida através de um único número. O outro caso é através do intervalo de credibilidade, que expressa de forma probabilística a pertinência ou não de θ ao intervalo. Estes dois casos serão discutidos a seguir. Para um tratamento aprofundado dos conceitos da inferência Bayesiana ver Migon e Gamerman (1999).

2.2.1 Estimação pontual

Seja uma amostra aleatória X_1, \dots, X_n tomada de uma distribuição com função (densidade) de probabilidade $p(x | \theta)$ onde o valor do parâmetro θ é desconhecido. Assim, o valor de θ deve ser estimado a partir dos valores observados na amostra. Seja $\delta(\mathbf{X})$ um estimador para θ e $\delta(\mathbf{X}) \in \Theta$. Na estimação pontual, o objetivo é a minimização de uma função perda $L(\delta(\mathbf{X}); \theta)$. A função perda pode ser interpretada como a punição por escolher um estimador $\delta(\mathbf{X})$ quando θ é o verdadeiro valor do parâmetro. Para cada possível valor de θ e cada possível estimativa $a \in \Theta$ será associada uma perda, de modo que, quanto maior a distância entre a e θ , maior o valor da perda. Neste caso, a perda esperada a *posteriori* é dada por

$$E[L(a; \theta) | x] = \int_{\Theta} L(a, \theta) p(\theta | x) d\theta,$$

e a regra de Bayes consiste em escolher a estimativa que minimiza esta perda esperada. Os estimadores obtidos minimizando esta perda esperada são denominados estimadores de Bayes.

Dentre as principais funções de perda, a mais utilizada é a função perda quadrática, definida como $L(a, \theta) = (a - \theta)^2$. A perda quadrática é às vezes criticada por penalizar demais o erro de estimação. A função de perda absoluta, definida como $L(a, \theta) = |a - \theta|$, introduz punições que crescem linearmente com o erro de estimação. Para reduzir ainda mais o efeito de erros de estimação grandes pode-se considerar funções que associam

uma perda fixa a um erro cometido, não importando sua magnitude. Tal função perda é denominada perda zero-um e, é definida como

$$L(a, \theta) = \begin{cases} 1 & \text{se } |a - \theta| > \epsilon \\ 0 & \text{se } |a - \theta| \leq \epsilon \end{cases}$$

para todo $\epsilon > 0$. Os estimadores associados às perdas quadrática, absoluta e zero-um, são a média, a mediana e a moda a *posteriori*, respectivamente.

2.2.2 Estimação por intervalo

Um inconveniente da estimação pontual é que ela não informa sobre a precisão da estimativa e restringe toda informação presente na distribuição a *posteriori* a um único valor numérico. Uma forma de contornar este problema é através do cálculo de intervalos de credibilidade, que são intervalos para o parâmetro que concentram a maior massa de probabilidade. Os intervalos de credibilidade, ao contrário dos intervalos de confiança frequentistas, são calculados de forma natural através da distribuição a *posteriori* do parâmetro em questão. Será discutido aqui o conceito de intervalo de credibilidade baseado na distribuição a *posteriori*.

Definição 1. *Uma região $C \subset \Theta$ é uma região de credibilidade para θ se $P(\theta \in C) \geq 1 - \alpha$. Neste caso, $1 - \alpha$ é chamado nível de credibilidade. No caso escalar, a região C é usualmente dada por um intervalo da forma $[c_1, c_2]$.*

Note que a definição expressa de forma probabilística a pertinência ou não de θ ao intervalo. O tamanho do intervalo traz informações sobre a dispersão de θ . Assim, quanto menor o intervalo, mais concentrada está a distribuição deste parâmetro, quanto maior, menos concentrada está a distribuição. Além disso, a exigência de que a probabilidade da definição 1 possa ser maior do que o nível de credibilidade é essencialmente técnica, pois deseja-se que o intervalo seja o menor possível, o que em geral implica em usar uma igualdade. No entanto, a desigualdade será útil se θ tiver uma distribuição discreta onde nem sempre é possível satisfazer a igualdade.

Uma característica importante é que os intervalos de credibilidade são invariantes sob transformações 1 a 1, $\phi(\theta)$. Ou seja, se $C = [a, b]$ é um intervalo de credibilidade

100(1 - α)% para θ então $[\phi(a), \phi(b)]$ é um intervalo de credibilidade 100(1 - α)% para $\phi(\theta)$.

2.3 Métodos de simulação estocástica

Na prática a distribuição a *posteriori* do vetor paramétrico dificilmente possui forma analítica conhecida. Através da utilização de métodos de simulação estocástica é possível obter amostras, de forma aproximada, da distribuição a *posteriori*. Entre eles, estão os métodos de Monte Carlo via cadeias de Markov (MCMC).

A idéia central do método MCMC é construir uma cadeia de Markov que seja fácil de ser simulada e tenha distribuição de equilíbrio dada pela distribuição de interesse. No caso da inferência Bayesiana, o objetivo é obter uma amostra da distribuição a *posteriori* e calcular estimativas amostrais de características desta distribuição através de técnicas de simulação iterativa, baseadas em cadeias de Markov.

A seguir, são apresentados dois métodos MCMC: o algoritmo de Metropolis-Hastings e o amostrador de Gibbs.

2.3.1 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis foi apresentado inicialmente por Metropolis et al. (1953) e generalizado por Hastings (1970) resultando no algoritmo de Metropolis-Hastings. A seguir descreve-se o método.

Assuma que $\pi(\theta)$ é a distribuição de interesse, que não possui uma forma analítica fechada, para qual se deseja gerar amostras sucessivas. Suponha que a cadeia esteja no estado θ e um valor θ' é gerado de uma distribuição proposta $q(\cdot | \theta)$, conhecida como densidade de transição. O novo valor θ' é aceito com probabilidade

$$\alpha(\theta, \theta') = \min \left\{ 1, \frac{\pi(\theta')q(\theta | \theta')}{\pi(\theta)q(\theta' | \theta)} \right\}. \quad (2.4)$$

Este mecanismo de atualização garante a convergência da cadeia para a distribuição de equilíbrio, que neste caso é a distribuição a *posteriori*. Uma característica importante é que só se precisa conhecer $\pi(\theta)$ parcialmente, isto é, a menos de uma constante, já

que neste caso a probabilidade (2.4) não se altera. Isto é fundamental em aplicações Bayesianas onde não se conhece completamente a distribuição a *posteriori*. O algoritmo 1 apresenta os passos do método.

Algoritmo 1. Algoritmo de Metropolis-Hastings:

1. Inicialize o contador de iterações $t = 0$ e especifique um valor inicial $\theta^{(0)}$.
2. Gere um novo valor θ' da distribuição $q(\cdot | \theta^{(t)})$.
3. Calcule a probabilidade de aceitação $\alpha(\theta^{(t)}, \theta')$ e gere $u \sim U(0, 1)$.
4. Se $u \leq \alpha(\theta^{(t)}, \theta')$ então aceite o novo valor e faça $\theta^{(t+1)} = \theta'$, caso contrário, rejeite e faça $\theta^{(t+1)} = \theta^{(t)}$.
5. Incremente o contador de t para $t + 1$ e volte ao passo 2 até a convergência ser alcançada.

A distribuição proposta pode depender do estado atual da cadeia, por exemplo, $q(\cdot | \theta)$ poderia ser uma distribuição normal centrada em θ caso este parâmetro estivesse definido em \mathbb{R} . É importante ressaltar que a eficiência do método está diretamente ligada à escala da distribuição proposta. Caso a variância da distribuição proposta seja muito pequena, a cadeia de Markov irá convergir lentamente, uma vez que seus incrementos serão pequenos. Se a variância for grande, a taxa de rejeição dos valores propostos será alta e a cadeia tenderá a não se mover. Para maiores detalhes veja Gamerman e Lopes (2006).

2.3.2 Amostrador de Gibbs

O amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, no qual os elementos de θ são atualizados um de cada vez (ou em blocos), tomando a distribuição condicional completa como proposta e probabilidade de aceitação igual a 1. A distribuição condicional completa é a distribuição condicional aos demais parâmetros do modelo. No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, isto é, não existe mecanismo de aceitação-rejeição. O algoritmo foi introduzido a comunidade estatística por Gelfand e Smith (1990).

Assuma que a distribuição de interesse seja $\pi(\theta)$ em que $\theta = (\theta_1, \dots, \theta_d)'$. Cada componente θ_i pode ser um escalar, um vetor ou uma matriz. Considere também que as distribuições condicionais completas $\pi_i(\theta_i) = \pi_i(\theta_i | \theta_{-i})$, $i = 1, \dots, d$ e $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)'$ são conhecidas e pode-se amostrar delas.

O problema a ser resolvido é gerar uma amostra de $\pi(\theta)$ quando os esquemas de geração direta são custosos, complexos, ou simplesmente impossíveis. O algoritmo 2 descreve os passos do amostrador de Gibbs, no qual fornece um esquema de geração alternativa baseado em sucessivas gerações das distribuições condicionais completas.

Algoritmo 2. Algoritmo de Gibbs:

1. Inicialize o contador de iterações da cadeia $j = 1$ e especifique valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})'$;
2. Obtenha os novos valores $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_d^{(j)})'$ de $\theta^{(j-1)}$ através de gerações sucessivas de valores

$$\begin{aligned} \theta_1^{(j)} &\sim \pi_1(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_d^{(j-1)}) \\ \theta_2^{(j)} &\sim \pi_2(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_d^{(j-1)}) \\ &\vdots \\ \theta_d^{(j)} &\sim \pi_d(\theta_d | \theta_1^{(j)}, \dots, \theta_{d-1}^{(j)}); \end{aligned}$$

3. Atualize o contador j para $j+1$ e volte ao passo 2 até a convergência ser alcançada.

Este algoritmo é extremamente útil quando a forma das distribuições condicionais completas são conhecidas. Uma introdução ao amostrador de Gibbs é dado por exemplo em Gamerman e Lopes (2006). A seguir será visto como analisar a convergência das cadeias geradas pelos métodos de simulação estocástica.

2.3.3 Diagnósticos de convergência

Os métodos de MCMC são uma ótima ferramenta para resolução de muitos problemas práticos na análise Bayesiana. Na utilização destes métodos o grande objetivo é obter

uma amostra, após a convergência, que se aproxime da distribuição a *posteriori*. É uma prática comum descartar as amostras geradas antes da convergência e essa quantidade é chamada de aquecimento ou *burn-in* e não é fixa. Espera-se que depois deste período de aquecimento a cadeia tenha esquecido os valores iniciais e convergido para a distribuição de interesse. Para evitar a autocorrelação das amostras geradas, é comum descartar alguns valores obtidos após a convergência. Logo, são guardados apenas os valores entre um intervalo específico de iterações, que é chamado de *thin*. Para maiores detalhes, ver Gamerman e Lopes (2006).

Existem diversos métodos para avaliação de convergência das cadeias simuladas. Nenhum deles, entretanto, é conclusivo. Estes fornecem apenas indícios de convergência. Assim, para que se tenha confiança de que as cadeias geradas tenham, de fato, atingido a distribuição estacionária de interesse, o ideal é que esta conclusão seja tomada com base em mais de um procedimento. Somente após esta confirmação pode-se formar a amostra da distribuição a *posteriori* dos parâmetros do modelo. Existem várias formas de se abordar o problema da convergência da cadeia.

Uma das formas é o modo informal, que é baseado em uma inspeção gráfica, onde analisa-se a trajetória de uma ou mais cadeias em períodos distintos de tempo e é verificado se convergem para o mesmo ponto de estabilidade. Desta forma, procura-se estudar as propriedades estatísticas da série a partir de simulações da cadeia de maneira empírica. Técnicas gráficas podem ser ilusórias indicando uma constância que pode não ser tão evidente sob outra escala. Além disso, muitas cadeias podem apresentar um comportamento similar ao da convergência sem que esta tenha sido atingida. Por isso, técnicas visuais devem ser analisadas com cautela.

A outra forma de se analisar a convergência em métodos de MCMC é utilizando alguns critérios mais formais, como o método proposto por Geweke (1992). Este sugere um procedimento para teste de convergência a partir da avaliação de médias ergódicas de uma única cadeia gerada, com base na idéia de que, após convergência, diferentes intervalos da cadeia gerada devam apresentar comportamentos semelhantes. Seja uma cadeia gerada com um número n suficientemente grande de iterações. A idéia é testar a igualdade das médias \bar{x}_1 e \bar{x}_2 , calculadas, respectivamente, a partir da fração $0.1n$

inicial e $0.5n$ final da amostra. Considerando os respectivos estimadores das variâncias assintóticas de \bar{x}_1 e \bar{x}_2 , dados por $V(\bar{x}_1)$ e $V(\bar{x}_2)$, tem-se que, quando $n \rightarrow \infty$

$$G_k = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{V(\bar{x}_1)/0.1n + V(\bar{x}_2)/0.5n}} \rightarrow N(0, 1).$$

Assim, valores extremos de G_k indicam falta de convergência. A técnica de Geweke está implementada no pacote CODA (Best et al., 1995) executável no *software* R (R Development Core Team, 2009).

2.4 Critério de seleção de modelos

A escolha de modelos é uma atividade fundamental que vem tornando-se cada vez mais importante na análise estatística, uma vez que, devido aos avanços computacionais, é possível construir modelos cada vez mais complexos. Tal complexidade normalmente aumenta de acordo com a estrutura imposta pelos modelos que requerem especificações em cada um de seus níveis.

O critério DIC (*Deviance Information Criterion*) proposto por Spiegelhalter et al. (2002) favorece modelos com melhor bondade de ajuste e penaliza pelo excesso de parâmetros. Ele é baseado na distribuição a *posteriori* da deviance, que é dada por

$$D(\boldsymbol{\theta}) = -2 \log l(\boldsymbol{\theta}; \mathbf{y}),$$

em que $\boldsymbol{\theta}$ é o vetor paramétrico, \mathbf{y} é o vetor de observações e $l(\cdot; \cdot)$ é a função de verossimilhança. O DIC é definido como

$$DIC = \bar{D} + p_D$$

em que \bar{D} define a esperança a *posteriori* da deviance e p_D é o número efetivo de parâmetros que é dado por

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}),$$

em que $\bar{\boldsymbol{\theta}}$ representa a média a *posteriori* dos parâmetros. A primeira parcela do DIC, avalia a bondade do ajuste, enquanto a segunda parcela penaliza pela complexidade do modelo. Os melhores modelos são os que apresentam o menor valor do DIC. Este

critério é particularmente útil em problemas de seleção de modelos Bayesianos onde as distribuições *a posteriori* foram obtidas por simulações via MCMC. O ponto negativo do DIC é que caso a média *a posteriori* não seja uma boa estimativa, os resultados se tornam não confiáveis. A seguir serão apresentados e discutidos os modelos dinâmicos e como é realizado o seu procedimento de inferência.

2.5 Modelos dinâmicos

Os modelos dinâmicos, também conhecidos como modelos de espaço de estados, são formulados para permitir alterações nos valores dos parâmetros com o passar do tempo. Tal característica torna esta classe de modelos uma classe de grande versatilidade e utilização.

Os modelos dinâmicos são constituídos por dois processos: o processo dos estados não observáveis $\{\boldsymbol{\theta}_t\}$ e o processo observacional $\{\mathbf{y}_t\}$, em que t é um índice que indica tempo. Com a evolução do tempo, toda a informação relevante para prever o futuro é recebida e pode ser usada na revisão do modelo. Suponha que o tempo inicial seja $t = 0$ e que D_0 represente a informação relevante e disponível sobre o modelo até o tempo $t = 0$. Esta informação será usada pelo modelador para fazer as previsões iniciais do futuro. De forma similar, suponha que para qualquer tempo $t > 0$, a informação disponível e relevante seja denotada por D_t . Qualquer afirmação sobre o futuro será condicionada nesta informação. Uma vez que \mathbf{y}_t foi observado no tempo t , define-se $D_t = \{\mathbf{y}_t, D_{t-1}\}$.

O modelo dinâmico geral (MD) é formulado usando-se a terminologia Bayesiana segundo West e Harrison (1997). Tal formulação é particularmente útil na derivação dos resultados em qualquer subclasse dos MD.

Definição 2. Para cada t o MD geral é definido por

$$\mathbf{y}_t \sim p(\mathbf{y}_t | \boldsymbol{\theta}, \Psi) \quad (2.5)$$

$$\boldsymbol{\theta}_t \sim p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \Psi) \quad (2.6)$$

$$\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta}_0 | \Psi) \quad (2.7)$$

em que \mathbf{y}_t é o vetor de observações e $\boldsymbol{\theta}_t$ é o vetor de estados não observáveis. $\boldsymbol{\theta}_0$ tem

distribuição $p(\boldsymbol{\theta}_0 | \Psi)$ a qual pode ser interpretada como a distribuição a priori do estado inicial do sistema e Ψ é o vetor de parâmetros. O modelo é completamente especificado, assumindo que $[\mathbf{y}_{t+i} | \boldsymbol{\theta}_t]$ é independente de $\boldsymbol{\theta}_{t-j}$, para $i = 0, 1, \dots$ e $j = 1, \dots, t$.

2.5.1 Modelos lineares dinâmicos

A classe de modelos dinâmicos é bastante abrangente. Uma subclasse destes modelos bastante conhecida e utilizada na literatura é a subclasse dos modelos lineares dinâmicos (MLD), em que é suposto normalidade para a variável resposta e para a evolução dos parâmetros dinâmicos através do tempo.

Definição 3. Para cada t o MLD é definido por

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{F}'_t \boldsymbol{\theta}_t + \mathbf{v}_t \quad \mathbf{v}_t \sim N_d(\mathbf{0}, \mathbf{V}_t) \quad (2.8)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad \mathbf{w}_t \sim N_p(\mathbf{0}, \mathbf{W}_t) \quad (2.9)$$

$$\boldsymbol{\theta}_0 | D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0) \quad (2.10)$$

em que \mathbf{y}_t é o vetor de observações e $\boldsymbol{\theta}_t$ é o vetor de estados não observáveis. \mathbf{m}_0 e \mathbf{C}_0 são os momentos da distribuição a priori inicial e são supostos conhecidos. Assume-se que os erros observacionais, $\{\mathbf{v}_t\}$, e os erros da evolução, $\{\mathbf{w}_t\}$, são mutuamente independentes e também independentes da distribuição inicial. As componentes de erro introduzem as incertezas e perdas de informações ao longo do tempo. $N_r(\mathbf{a}, \mathbf{B})$ denota distribuição normal multivariada (com dimensão r) em que \mathbf{a} é o vetor de médias e \mathbf{B} é a matriz de covariância.

As matrizes \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t e \mathbf{W}_t e o vetor \mathbf{c}_t podem depender de um vetor de parâmetros $\boldsymbol{\psi}$. A inclusão do termo \mathbf{c}_t na definição 3 foi motivada pela classe de modelos que serão apresentados na seção 3.2 deste trabalho.

A equação (2.8) é denominada equação da observação e relaciona o vetor de observações \mathbf{y}_t ao parâmetro de estado $\boldsymbol{\theta}_t$, enquanto que a equação (2.9) é denominada equação do sistema e é responsável pela evolução dos parâmetros de estado através do

tempo. Estas duas equações podem ser reescritas, para $t = 1, 2, \dots$, na forma

$$\mathbf{y}_t | \boldsymbol{\theta}_t \sim N(\mathbf{c}_t + \mathbf{F}'_t \boldsymbol{\theta}_t, \mathbf{V}_t) \quad (2.11)$$

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1} \sim N(\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t) \quad (2.12)$$

O modelo descrito na definição 3 é completamente especificado através da quádrupla $\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t$ e da distribuição *a priori* inicial assumida para os parâmetros de estado. Devido à própria estrutura markoviana do modelo, $\boldsymbol{\theta}_t$ dado $\boldsymbol{\theta}_{t-1}$, para $t = 1, 2, \dots$, tem uma distribuição normal, conforme a equação (2.9). Esta subclasse de modelos permite diversos tipos de componentes, como uma estrutura de média, sazonalidade e a inclusão de covariáveis. Essas componentes podem ser especificadas de forma a serem independentes no modelo.

A escolha de \mathbf{F}_t e \mathbf{G}_t depende do modelo e da natureza dos dados que estão sendo analisados. Casos particulares dos modelos dinâmicos lineares incluem o modelo de regressão (quando $\mathbf{G}_t = \mathbf{I}$, a matriz identidade, e $\mathbf{W}_t = 0$), e modelos lineares de séries temporais (quando $\mathbf{F}_t = \mathbf{F}$, $\mathbf{G}_t = \mathbf{G}$, $\mathbf{V}_t = \mathbf{V}$ e $\mathbf{W}_t = \mathbf{W}$).

Esta definição do MLD foi introduzida por Harrison e Stevens (1976) no contexto Bayesiano. Um amplo tratamento usando esta abordagem pode ser encontrado em West e Harrison (1997). Quando os valores dos elementos que compõem a quádrupla $\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t$ são conhecidos, o procedimento de inferência sobre os parâmetros de estado nesta subclasse de modelos pode ser feito através de algoritmos sequenciais, como o filtro de Kalman e o filtro de perturbações. A seguir, tais algoritmos serão descritos.

2.5.2 Inferência nos modelos lineares dinâmicos

Os aspectos de inferência dos modelos lineares dinâmicos seguem os passos usuais da estatística Bayesiana, explorando seu aspecto sequencial e combinando duas operações principais: a evolução para construir a distribuição *a priori*, e a atualização para incorporar a nova observação feita no tempo t . Um dos principais aspectos dos modelos dinâmicos é o fato de que se pode fazer inferência sobre a distribuição atualizada de $\boldsymbol{\theta}_t | \mathbf{y}_{1:t}$, em que $\mathbf{y}_{1:t}$ é a informação até o tempo t . Para um modelo fechado, isto é, que só recebe informação dos dados, temos que $\mathbf{y}_{1:t} = (\mathbf{y}_1, \dots, \mathbf{y}_t)'$. Essa inferência sequencial

é baseada em três passos: evolução, previsão e atualização. Serão apresentados a seguir os algoritmos usados para a realização da inferência sequencial.

Filtro de Kalman

O filtro de Kalman é um método para avaliar a distribuição a *posteriori* ($\boldsymbol{\theta}_t|D_t$) baseado na distribuição a *priori* ($\boldsymbol{\theta}_t|D_{t-1}$), em que D_t é a informação disponível até o tempo t . Essencialmente, o filtro de Kalman é um algoritmo que permite atualizar de forma recursiva a distribuição dos estados quando uma nova observação está disponível. O filtro de Kalman é composto pelo seguinte esquema:

$$\begin{array}{ccccc}
 p(\boldsymbol{\theta}_{t-1} | D_{t-1}) & \xrightarrow{\text{Evolução}} & p(\boldsymbol{\theta}_t | D_{t-1}) & \xrightarrow{\text{Atualização}} & p(\boldsymbol{\theta}_t | D_t) \\
 & & \downarrow \text{Previsão} & & \\
 & & p(\mathbf{y}_t | D_{t-1}) & &
 \end{array} \tag{2.13}$$

No MLD da definição 3, a distribuição a *priori*, a previsão um passo a frente e a distribuição a *posteriori*, para cada t , são dadas respectivamente por

$$\boldsymbol{\theta}_t|D_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t) \tag{2.14}$$

$$\mathbf{y}_t|D_{t-1} \sim N(\mathbf{f}_t, \mathbf{Q}_t) \tag{2.15}$$

$$\boldsymbol{\theta}_t|D_t \sim N(\mathbf{m}_t, \mathbf{C}_t) \tag{2.16}$$

em que

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \text{ e } \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \tag{2.17}$$

$$\mathbf{f}_t = \mathbf{c}_t + \mathbf{F}_t' \mathbf{a}_t \text{ e } \mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t \tag{2.18}$$

$$\mathbf{m}_t = \mathbf{a}_t + \underbrace{\mathbf{R}_t \mathbf{F}_t' \mathbf{Q}_t^{-1}}_{\mathbf{A}_t} \underbrace{(\mathbf{y}_t - \mathbf{f}_t)}_{\mathbf{e}_t} \text{ e } \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t' \tag{2.19}$$

A prova de como se chega nestas equações é baseada no princípio de indução. Pela hipótese de indução, $\boldsymbol{\theta}_{t-1} | D_{t-1} \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$, e da equação dos estados (2.9) tem-se (2.14). Usando-a conjuntamente com a equação observacional (2.8) resulta (2.15). (2.8) e (2.15) determinam a distribuição conjunta de \mathbf{y}_t e $\boldsymbol{\theta}_t$ exceto pela covariância, a qual segue de (2.8) e é dada por $Cov(\mathbf{y}_t, \boldsymbol{\theta}_t | D_{t-1}) = \mathbf{F}_t' \mathbf{R}_t$. Logo, da teoria da distribuição Normal

tem-se (2.16) e estabelecendo-se assim a fórmula recursiva. Uma derivação detalhada pode ser vista em West e Harrison (1997).

Nas fórmulas do filtro de Kalman foi introduzida a matriz adaptativa \mathbf{A}_t , (também conhecida como ganho de Kalman), a previsão um passo a frente, \mathbf{f}_t , com sua variância associada, \mathbf{Q}_t , e o erro de previsão um passo a frente, \mathbf{e}_t .

Suavizador de Kalman

A distribuição do vetor de estados, $\boldsymbol{\theta}_t$, utilizando toda a informação disponível, D_T , é chamada de distribuição suavizada. O algoritmo que permite obter estas distribuições suavizadas para todo t é chamado de algoritmo suavizador de Kalman.

No MLD da definição 3, define-se

$$\mathbf{B}_t = \mathbf{C}_t \mathbf{G}_{t+1} \mathbf{R}_{t+1}^{-1}$$

para todo t . Então a distribuição suavizada de $\boldsymbol{\theta}_t$ dado D_T é dada por

$$\boldsymbol{\theta}_t | D_T \sim N(\tilde{\mathbf{m}}_t, \tilde{\mathbf{C}}_t)$$

em que,

$$\tilde{\mathbf{m}}_t = \mathbf{m}_t + \mathbf{B}_t(\tilde{\mathbf{m}}_{t+1} - \mathbf{a}_{t+1})$$

e

$$\tilde{\mathbf{C}}_t = \mathbf{C}_t + \mathbf{B}_t(\tilde{\mathbf{C}}_{t+1} - \mathbf{R}_{t+1})\mathbf{B}_t'$$

em que $\tilde{\mathbf{m}}_T = \mathbf{m}_T$ e $\tilde{\mathbf{C}}_T = \mathbf{C}_T$.

A dedução das fórmulas recursivas do algoritmo de suavização de Kalman é baseada no princípio de indução para trás em t e pode ser vista em West e Harrison (1997). O algoritmo suavizador de Kalman é processado depois do Filtro de Kalman e inicializado por $\boldsymbol{\theta}_T | D_T \sim N(\tilde{\mathbf{m}}_T, \tilde{\mathbf{C}}_T)$.

Uma alternativa ao filtro de Kalman e ao suavizador de Kalman são, respectivamente, o filtro de perturbações e o algoritmo de perturbações suavizadas que serão explicados a seguir.

Filtro de perturbações

O filtro de perturbações é matematicamente equivalente ao filtro de Kalman. A razão para esta terminologia é que a saída do filtro de perturbações fornece os erros de previsão um passo a frente, \mathbf{e}_t , a inversa das variâncias destes erros, \mathbf{Q}_t^{-1} , e as matrizes adaptativas escaladas, \mathbf{K}_t . As equações do filtro de perturbações são obtidas a partir do filtro de Kalman e o algoritmo é inicializado usando \mathbf{m}_0 e \mathbf{C}_0 , os parâmetros de $[\boldsymbol{\theta}_0|D_0]$. No filtro de Kalman, $\mathbf{a}_1 = \mathbf{d}_1 + \mathbf{G}_1\mathbf{m}_0$ e $\mathbf{R}_1 = \mathbf{G}_1\mathbf{C}_0\mathbf{G}'_1 + \mathbf{W}_1$. Então, para $t = 1, \dots, T$, obtem-se a seguinte recursão:

$$\begin{aligned}\mathbf{e}_t &= \mathbf{y}_t - \mathbf{c}_t - \mathbf{F}'_t\mathbf{a}_t \\ \mathbf{Q}_t &= \mathbf{F}'_t\mathbf{R}_t\mathbf{F}_t + \mathbf{V}_t \\ \mathbf{K}_t &= \mathbf{G}_{t+1}\mathbf{F}_t\mathbf{R}_t\mathbf{Q}_t^{-1} \\ \mathbf{a}_{t+1} &= \mathbf{G}_{t+1}\mathbf{a}_t + \mathbf{K}_t\mathbf{e}_t \\ \mathbf{R}_{t+1} &= \mathbf{G}_{t+1}\mathbf{R}_t(\mathbf{G}_{t+1} - \mathbf{K}_t\mathbf{F}'_t)' + \mathbf{W}_{t+1},\end{aligned}$$

em que $\mathbf{K}_t = \mathbf{G}_{t+1}\mathbf{A}_t$ é a conexão com o coeficiente adaptativo definido em (2.19). Deste modo, o filtro de perturbações é equivalente ao filtro de Kalman, exceto pelas saídas. Somente \mathbf{K}_t , \mathbf{Q}_t^{-1} e \mathbf{e}_t são armazenados, enquanto que o filtro de Kalman armazena adicionalmente \mathbf{m}_t e \mathbf{C}_t . Com isso, é possível obter um ganho computacional usando o filtro de perturbações.

Perturbações suavizadas

Como no caso da filtragem, é possível obter um ganho computacional usando as perturbações ao invés dos estados. Isto é, determinam-se $\tilde{\mathbf{v}}_t = E[\mathbf{v}_t|D_T]$ e $\tilde{\mathbf{w}}_t = E[\mathbf{w}_t|D_T]$ em lugar de $\tilde{\mathbf{m}}_t = E[\boldsymbol{\theta}_t|D_T]$. O algoritmo de perturbações suavizadas introduzido por Koopman (1993), fornece o estimador do erro quadrático médio do vetor de perturbações dadas todas as observações, a partir das saídas do filtro de perturbações, \mathbf{K}_t , \mathbf{Q}_t^{-1} e \mathbf{e}_t .

Assim como na derivação do suavizador de Kalman, a obtenção das perturbações suavizadas é uma recursão para trás no tempo. Seja \mathbf{r}_t o vetor p dimensional para as perturbações dos estados e $\boldsymbol{\epsilon}_t$ o vetor d dimensional de perturbações das observações. A

recursão é inicializada com $\mathbf{r}_T = 0$. Então para $t = T, \dots, 1$ tem-se a recursão para trás

$$\boldsymbol{\epsilon}_t = \mathbf{Q}_t^{-1} \mathbf{e}_t + \mathbf{K}_t \mathbf{r}_t \quad (2.20)$$

$$\tilde{\mathbf{v}}_t = \mathbf{V}_t \boldsymbol{\epsilon}_t \quad (2.21)$$

$$\tilde{\mathbf{w}}_t = \mathbf{W}_t \mathbf{r}_{t-1} \quad (2.22)$$

$$\mathbf{r}_{t-1} = \mathbf{F}_t \boldsymbol{\epsilon}_t + \mathbf{G}'_{t+1} \mathbf{r}_t \quad (2.23)$$

Observe-se que comparada ao suavizador de Kalman, não se faz necessária a inversão de matrizes, já que \mathbf{Q}_t tem sido invertida pelo filtro de perturbações. Assim, os valores suavizados $\tilde{\mathbf{m}}_t$ são obtidos pela recursão

$$\tilde{\mathbf{m}}_t = \mathbf{G}_t \tilde{\mathbf{m}}_{t-1} + \tilde{\mathbf{w}}_t,$$

$t = 1, \dots, T$. As fórmulas de $\tilde{\mathbf{C}}_t$ são omitidas já que não tem sido usadas nas aplicações, mais estas podem ser vista em Koopman (1993). A seguir, será discutido o método MCMC para os modelos lineares dinâmicos.

MCMC

Aqui será considerado o problema da inferência Bayesiana para os parâmetros e estados do MLD. Esta não é uma tarefa simples de se fazer, pois geralmente a distribuição conjunta a *posteriori* não tem solução analítica fechada. Com a aparição dos métodos MCMC este problema tem sido resolvido pois, pelo menos a princípio, podem ser implementados com qualquer distribuição a *posteriori*.

A aplicação dos métodos MCMC no MLD, explicitado na definição 3, é baseada na geração de amostras da distribuição a *posteriori*

$$p(\boldsymbol{\theta}_{0:T}, \boldsymbol{\psi} | \mathbf{y}_{1:T})$$

em que $\boldsymbol{\psi}$ é o vetor de parâmetros. Para amostrar desta distribuição procede-se em duas etapas. Inicialmente amostra-se

$$\boldsymbol{\psi} \sim p(\boldsymbol{\psi} | \boldsymbol{\theta}_{0:T}, \mathbf{y}_{1:T})$$

e a seguir os estados são simulados condicionais a este valor a partir de

$$\boldsymbol{\theta}_{0:T} \sim p(\boldsymbol{\theta}_{0:T} | \boldsymbol{\psi}, \mathbf{y}_{1:T}).$$

Em geral, amostrar esta distribuição não é uma tarefa fácil. Existem na literatura duas classes de amostradores para os estados que exploram a natureza do MLD. O primeiro é o amostrador single-move que atualiza um estado por vez, e o segundo é o amostrador multi-move que atualiza blocos de estados simultaneamente, sendo este último mais eficiente do ponto de vista computacional, além de acelerar a convergência à distribuição de equilíbrio. Neste trabalho só será visto o amostrador multi-move.

Algoritmo de simulação de perturbações suavizadas

O algoritmo proposto por de Jong e Shephard (1995) simula as perturbações no lugar dos estados e isto é a principal razão para que o algoritmo seja mais eficiente do que o que simula diretamente dos estados. Isto se deve ao fato de que somente as saídas do filtro de perturbações são necessárias, em lugar das saídas do filtro de Kalman que precisam de maior espaço para armazená-las. Todas as variáveis aleatórias do modelo dinâmico são combinações lineares das perturbações e por isso podem ser construídas através das perturbações simuladas.

A recursão é iniciada fazendo $\mathbf{r}_T^* = \mathbf{0}$, $\mathbf{N}_T = \mathbf{0}$. Logo, para $t = T, \dots, 1$,

$$\begin{aligned} \boldsymbol{\Omega}_t &= \mathbf{W}_t - \mathbf{W}_t \mathbf{N}_t \mathbf{W}_t \\ \boldsymbol{\nu}_t &\sim N(\mathbf{0}, \boldsymbol{\Omega}_t) \\ \boldsymbol{\Upsilon}_t &= \mathbf{W}_t \mathbf{N}_t \overbrace{\{\mathbf{G}_{t+1} - \mathbf{K}_t \mathbf{F}'_t\}}^{\boldsymbol{\Lambda}_t} \\ \mathbf{r}_{t-1}^* &= \mathbf{F}_t \mathbf{Q}_t^{-1} \mathbf{e}_t + \boldsymbol{\Lambda}_t \mathbf{r}_t^* - \boldsymbol{\Upsilon}'_t \boldsymbol{\Omega}_t^{-1} \boldsymbol{\nu}_t \\ \mathbf{N}_{t-1} &= \mathbf{F}_t \mathbf{Q}_t^{-1} \mathbf{F}'_t + \boldsymbol{\Lambda}'_t \mathbf{N}_t \boldsymbol{\Lambda}_t + \boldsymbol{\Upsilon}_t \boldsymbol{\Omega}_t^{-1} \boldsymbol{\Upsilon}_t \end{aligned} \tag{2.24}$$

em que \mathbf{e}_t vem do filtro de perturbações. Note que se $\boldsymbol{\nu}_t = \mathbf{0}$ e $\boldsymbol{\Upsilon}_t = \mathbf{0}$, então $\mathbf{r}_t^* = r_t$, o algoritmo de Jong e Shepard se reduz ao algoritmo de perturbações suavizadas. Uma amostra de $\mathbf{w}_t^{(i)}$ de $p(\mathbf{w}_t | \mathbf{y}_{1:T})$ é dada por

$$\mathbf{w}_t^{(i)} = \mathbf{W}_t \mathbf{r}_t^* + \boldsymbol{\nu}_t \tag{2.25}$$

e desta

$$\boldsymbol{\theta}_t^{(i)} = \mathbf{G}_t \boldsymbol{\theta}_{t-1}^{(i)} + \mathbf{w}_t^{(i)}. \quad (2.26)$$

Para obter uma amostra *i.i.d* de $p(\boldsymbol{\theta}_{0:T} \mid \boldsymbol{\psi}, \mathbf{y}_{1:T})$, repete-se o procedimento.

O algoritmo de simulação de perturbações suavizadas é baseado em simular da distribuição *a posteriori* das perturbações do modelo, \mathbf{w}_t , que então permite, como necessário, formar a simulação dos estados $\boldsymbol{\theta}_t$. As vantagens deste algoritmo sobre a simulação diretamente dos estados são dadas por:

- a exigência de armazenamento do algoritmo é menor pois utiliza o filtro de perturbações ao invés do filtro de Kalman, portanto, quando os dados são de grandes dimensões ganha-se muito em tempo computacional;
- a recursão usada no algoritmo opera em dimensão mínima, pois trabalha com as perturbações ao invés dos estados e a recursão não exige inversão da matriz \mathbf{R}_t , que geralmente é de grande dimensão e/ou singular;
- a recursão pode ser operada em forma de raiz quadrada, melhorando a estabilidade numérica;

O procedimento de inferência realizado neste trabalho utilizará somente o algoritmo de simulação de perturbações suavizadas de de Jong e Shephard (1995) para simular os estados do modelo dinâmico. Na próxima seção será feita uma revisão de como é feita a modelagem com dados binários.

2.6 Modelos para dados binários

Dados binários são aqueles que admitem dois resultados possíveis para a variável resposta, seja presença ou ausência de alguma característica em uma unidade observacional. Tais dados são bastante recorrentes em diversas áreas da Ciência. Para exemplificar, seguem alguns casos práticos do uso deste tipo de variável resposta: concessão de crédito de um banco (aprovado ou não); resultado de um diagnóstico (positivo ou negativo);

intenção de voto de um eleitor em relação a um candidato (vota ou não); inspeção de uma peça recém fabricada (defeituosa ou não); teste da publicidade de um novo produto (vendeu ou não), etc. Neste caso, considera-se um problema de sucesso ou fracasso, sendo sucesso o resultado mais importante da pesquisa. Devido à dicotomização da variável resposta, é comum atribuir o código "1" ao "sucesso" e o código "0" ao "fracasso".

Por muitos anos a regressão linear era usada para explicar a maioria dos fenômenos aleatórios. Mesmo quando não era razoável assumir normalidade, utilizava-se algum tipo de transformação para alcançar a normalidade desejada. Com o desenvolvimento computacional alguns modelos que exigiam a utilização de processos iterativos para a estimação dos parâmetros começaram a ser mais utilizados. Uma proposta inovadora no assunto foi apresentada por Nelder e Wedderburn (1972), que propuseram os modelos lineares generalizados que serão discutidos a seguir.

2.6.1 Modelos lineares generalizados

Os modelos lineares generalizados (MLGs) são uma extensão dos modelos lineares clássicos. A principal idéia do MLG é expandir as opções da distribuição da variável resposta, incluindo todas as distribuições que pertençam à família exponencial. Pertencem à família exponencial toda família de distribuições com função de densidade $p(x|\theta)$ que pode ser escrita na forma $p(x|\theta) = a(x) \exp\{b(x)d(\theta) + c(\theta)\}$, em que a , b , c e d são funções conhecidas. Uma das principais fontes de informação sobre os modelos lineares generalizados é McCullagh e Nelder (1989). Estes modelos têm um grande impacto tanto na teoria como na prática da estatística pois a generalização do modelo linear trouxe maior flexibilidade para a relação funcional entre a média da variável resposta e a parte linear do modelo.

Assim, para dados binários, pode-se assumir que a variável resposta segue uma distribuição de Bernoulli. Isto é $Y_i \sim Ber(p_i)$ em que $i = 1, \dots, n$. A relação entre a média da distribuição e as variáveis explicativas é dada por

$$E(Y_i) = p_i = g(\mathbf{x}_i' \boldsymbol{\alpha}),$$

em que \mathbf{x}_i é o vetor de variáveis explicativas e $\boldsymbol{\alpha}$ é o vetor dos coeficientes da regressão

linear. $g^{-1}(\cdot)$ é a função de ligação.

A função de ligação é uma transformação do intervalo da proporção de sucessos $(0; 1)$ para $(-\infty; \infty)$ e pode ser qualquer função monótona diferenciável. No caso da variável resposta ser binária, a função de ligação deve garantir que para quaisquer valores dos parâmetros do modelo a proporção seja estimada em um valor entre 0 e 1. Portanto, é razoável que funções densidade acumulada sejam usadas para gerarem ligações. Uma das funções de ligação mais utilizada na literatura é a ligação probito que é dada por

$$p_i = \Phi(\mathbf{x}'_i \boldsymbol{\alpha}),$$

em que Φ denota a função de distribuição acumulada da normal padrão. Assim, a função de ligação é o inverso da função acumulada da normal padrão, Φ^{-1} .

A seguir será visto uma extensão dos modelos lineares generalizados para uma série temporal de observações binárias.

2.6.2 Modelos dinâmicos binários

Quando os dados consistem em observações dependentes como as séries temporais, a classe dos modelos lineares generalizados não pode ser utilizada diretamente. Uma opção muito utilizada na prática para analisar séries temporais é a classe de modelos lineares dinâmicos visto na Subseção 2.5.1. Geralmente eles assumem observações Gaussianas, mas existe uma generalização para a família exponencial formalizada por West et al. (1985), chamada modelos lineares generalizados dinâmicos. Esta classe também pode ser vista como uma generalização dos modelos lineares generalizados, com os parâmetros mudando no tempo.

A extensão dinâmica dos modelos lineares generalizados é obtida substituindo a equação de observação do modelo linear dinâmico pela equação da média da distribuição que utiliza a função de ligação. No caso da variável resposta ser binária, o modelo é denominado modelo dinâmico binário como em Cox (1981).

Um modelo dinâmico binário consiste em dois processos. No primeiro, o processo observado $\{Y_t, t = 1, \dots, T\}$, cuja distribuição condicional dado um vetor q dimensional de estados $\boldsymbol{\theta}_t$ é $Y_t | \boldsymbol{\theta}_t \sim Ber(p_t)$, a probabilidade condicional de sucesso $p_t = P(Y_t = 1 | \boldsymbol{\theta}_t)$

segue a equação de observação

$$p_t = g(\mathbf{F}'_t \boldsymbol{\theta}_t), \quad (2.27)$$

em que $g^{-1}(\cdot)$ denota uma dada função de ligação. O vetor q dimensional \mathbf{F}_t é constituído de covariáveis que variam no tempo. No segundo processo, assume-se que o vetor de estados $\boldsymbol{\theta}_t$ segue um processo de Markov q dimensional governado pela equação dos estados,

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad (2.28)$$

em que \mathbf{G}_t é uma matriz de transição conhecida $q \times q$ dimensional, e o vetor de erro \mathbf{w}_t tem média zero. Para o caso especial de um processo unidimensional ($q = 1$), dois modelos são comumente usados na literatura. O primeiro é o passeio aleatório em que $G_t = 1$ e portanto o processo é não estacionário. O segundo é o processo AR(1) em que $G_t = a \in (-1, 1)$ sendo então um processo estacionário, a é um parâmetro que representa o coeficiente de autocorrelação da série dos estados. A diferença essencial entre estes dois tipos de processos é em suas variâncias: o AR(1) possui uma variância limitada mas o passeio aleatório tem uma variância ilimitada que aumenta com o tempo.

Em muitos estudos práticos, avaliar o efeito de covariáveis é um dos principais interesses. Para atender tal necessidade, a equação de observação (2.27) pode ser reescrita de uma forma que as covariáveis sejam explicitadas. Componentes representando tendências ou comportamento cíclicos podem ser incorporadas e modeladas como funções lineares de covariáveis. Assim, a probabilidade condicional de sucesso toma a forma:

$$p_t = g(\mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t), \quad (2.29)$$

em que $\mathbf{x}'_t = (x_{t1}, \dots, x_{tp})'$ representa as covariáveis e $\boldsymbol{\alpha}$ é o vetor paramétrico que contém os coeficientes da regressão a serem estimados. O vetor \mathbf{F}_t é um vetor conhecido. A inclusão do termo $\mathbf{x}'_t \boldsymbol{\alpha}$ permite quantificar a relação entre a probabilidade de se obter um sucesso e as covariáveis, que é frequentemente um dos interesses científicos.

Esta classe de modelos dinâmicos binários é largamente usada para a análise de dados binários em diversas áreas de estudo. Na seção a seguir será apresentada a distribuição normal assimétrica que é uma generalização da distribuição normal.

2.7 Distribuição normal assimétrica

Em muitas situações práticas, onde a suposição usual de normalidade não é satisfeita devido à falta de simetria dos dados, propõe-se como alternativa a utilização de uma família mais geral de distribuições, de forma que se consiga modelar a assimetria dos dados e além disso, incluir a distribuição normal como um caso particular. Esta família de distribuições é denominada normal assimétrica.

Esta seção é iniciada com uma descrição da distribuição normal assimétrica padrão e algumas de suas propriedades. Em seguida será feita a relação entre a forma uniparamétrica com a forma normal assimétrica de três parâmetros (locação e escala), usada para construir a classe de distribuições de mistura de escala normal assimétrica.

Distribuição normal assimétrica padrão

A distribuição normal assimétrica uniparamétrica foi introduzida formalmente por Azzalini (1985) e é denominada distribuição normal assimétrica padrão. Esta depende apenas de um parâmetro, o qual caracteriza a assimetria da sua função densidade. A definição desta distribuição é dada por:

Definição 4. *Uma variável aleatória Z tem distribuição normal assimétrica padrão se sua função de densidade de probabilidade é dada por*

$$f(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R} \quad (2.30)$$

em que $\phi(\cdot)$ e $\Phi(\cdot)$ indicam a função de densidade de probabilidade e a função de distribuição acumulada de uma normal padrão, respectivamente.

O parâmetro $\lambda \in \mathbb{R}$ que caracteriza a forma da distribuição é também denominado parâmetro de assimetria, pois valores negativos de λ indicam assimetria negativa e valores positivos de λ indicam assimetria positiva. Se $\lambda = 0$, a densidade em (2.30) coincide com a densidade da distribuição normal padrão e portanto é simétrica. Será usada a seguinte notação $Z \sim NA(\lambda)$. A Figura 2.1 ilustra a função de densidade descrita em (2.30) para diferentes valores do parâmetro de assimetria(λ). Através da figura, verifica-se como o parâmetro de assimetria influencia na densidade da distribuição.

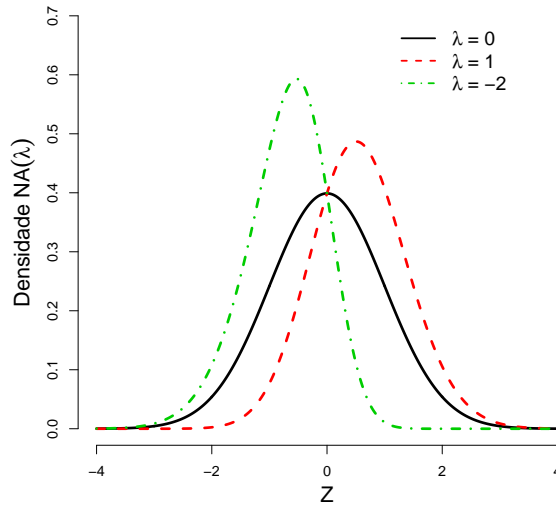


Figura 2.1: Gráfico da função densidade $NA(\lambda)$ para diferentes valores de λ .

A densidade em (2.30) possui algumas propriedades interessantes e cujas provas podem ser obtidas em Azzalini (1985) e Azzalini (2005). Algumas propriedades são:

- ▶ Se $Z \sim NA(\lambda)$, então $|Z| \sim N_+(0, 1)$ em que N_+ denota distribuição normal truncada em \mathbb{R}^+ ;
- ▶ Quando $\lambda \rightarrow \infty$, a densidade em (2.30) converge para a densidade $N_+(0, 1)$;
- ▶ Se $Z \sim NA(\lambda)$, então $-Z \sim NA(-\lambda)$;
- ▶ A densidade (2.30) é log-côncava;
- ▶ Se $Z \sim NA(\lambda)$, então $Z^2 \sim \chi_1^2$. χ_1^2 denota a distribuição qui-quadrado com um grau de liberdade.
- ▶ Representação estocástica: Se $U, V \sim N(0, 1)$, independentes, então

$$\frac{\lambda}{\sqrt{1+\lambda^2}} |U| + \frac{1}{\sqrt{1+\lambda^2}} V \sim NA(\lambda). \quad (2.31)$$

Esta última propriedade é útil para gerar amostras da distribuição normal assimétrica a partir da normal padrão.

Se $Z \sim NA(\lambda)$ então a média e a variância de Z são dadas por

$$E(Z) = \sqrt{\frac{2}{\pi}} \delta \quad \text{e} \quad Var(Z) = 1 - \frac{2}{\pi} \delta^2, \quad (2.32)$$

em que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Da mesma forma que foi definida a distribuição normal assimétrica padrão pode-se generalizar a distribuição agregando parâmetros de escala e de locação.

Distribuição normal assimétrica com três parâmetros

A função de densidade dada em (2.30) é estendida introduzindo parâmetros de locação $\mu \in \mathbb{R}$ e de escala $\sigma > 0$. Neste caso, será utilizada a notação $NA(\mu, \sigma^2, \lambda)$.

Uma variável aleatória Y tem distribuição normal assimétrica com parâmetros de posição μ e de escala σ^2 se sua função de densidade de probabilidade é da forma

$$f(y) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right), \quad y \in \mathbb{R}. \quad (2.33)$$

É fácil verificar que se $Z \sim NA(\lambda)$ e $Y = \mu + \sigma Z$, então $Y \sim NA(\mu, \sigma^2, \lambda)$. A média e a variância da variável aleatória Y são dadas por

$$E(Z) = \mu + \sigma \delta \sqrt{\frac{2}{\pi}} \quad \text{e} \quad Var(Z) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2\right). \quad (2.34)$$

Uma propriedade interessante da distribuição normal assimétrica é se X e Y são duas variáveis aleatórias tais que $X \sim NA(\mu, \sigma^2, \lambda)$ e $Y = a + bX$, a e $b \in \mathbb{R}$, então

$$Y \sim NA(a + b\mu, b^2\sigma^2, \text{sinal}(b)\lambda), \quad (2.35)$$

em que $\text{sinal}(b) = 1$ se $b \geq 0$ e $\text{sinal}(b) = -1$ se $b < 0$.

Embora a distribuição normal assimétrica, tenha sido uma grande contribuição na análise de dados com estruturas assimétricas, ela não se comporta bem na presença de excesso de curtose. Portanto, é necessário buscar uma distribuição assimétrica, que incorpore além da assimetria dos dados a presença de caudas pesadas com intuito de modelar observações extremas. Na próxima seção, serão apresentadas as distribuições que tratam simultaneamente estas duas questões.

2.8 Distribuições de mistura de escala normal assimétrica

Em muitas aplicações de diferentes áreas de pesquisa nota-se que a estrutura dos dados e/ou modelos utilizados não satisfazem a suposição de normalidade devido, entre outras causas, à presença de alta assimetria ou de observações atípicas. Por esta razão, a construção e estudo de famílias paramétricas flexíveis, que são capazes de acomodar valores práticos de assimetria e curtose, tem recebido considerável atenção na literatura recente.

Nesse contexto, misturas de escala da distribuição normal (Andrews e Mallows, 1974) são frequentemente usadas para procedimentos estatísticos de dados simétricos como em Lange e Sinsheimer (1993). No entanto, apesar desta classe representar uma interessante alternativa à distribuição normal, esta pode não ser apropriada quando a distribuição dos dados for assimétrica.

Assim, distribuições apropriadas para ajustar e simular esses dados assimétricos e concentrados nas caudas se fazem necessárias. Uma nova família de distribuições que combine assimetria com caudas pesadas é desenvolvida.

Nesta seção será feita uma apresentação das distribuições de mistura de escala normal assimétrica. A proposta desta família de distribuições generaliza estudos encontrados na maior parte em Lange e Sinsheimer (1993).

Uma variável aleatória Y segue uma distribuição de mistura de escala normal assimétrica (MENA) com parâmetro de locação $\mu \in \mathbb{R}$, parâmetro de escala $\sigma^2 > 0$ e parâmetro de assimetria $\lambda \in \mathbb{R}$, se sua função de densidade de probabilidade é dada por

$$f(y) = 2 \int_0^\infty \phi(y; \mu, \kappa(u)\sigma^2) \Phi\left(\frac{\lambda(y - \mu)}{\sigma}\right) dH(u; \nu), \quad y \in \mathbb{R}, \quad (2.36)$$

em que U é uma variável aleatória positiva com função de distribuição $H(u; \nu)$ indexada pelo parâmetro ν . Este parâmetro está relacionado com a curtose, e assim, pode controlar as caudas da distribuição sob investigação. $\kappa(\cdot)$ é uma função estritamente positiva. Neste caso, denota-se $Y \sim MENA(\mu, \sigma^2, \lambda; H)$. Se $\mu = 0$ e $\sigma^2 = 1$, tem-se uma distribuição MENA padrão, denotada por $MENA(\lambda; H)$.

Embora seja possível lidar com qualquer função positiva para $\kappa(\cdot)$, neste trabalho a atenção será restringida para o caso em que $\kappa(U) = 1/U$, já que sob esta restrição interessantes propriedades matemáticas são obtidas. Esta restrição faz com que a distribuição pertença à classe de distribuições normal/independentes assimétricas discutida em Lachos et al. (2010).

Para uma variável aleatória $Y \sim MENA(\mu, \sigma^2, \lambda; H)$, a representação estocástica (Henze, 1986) é dada por

$$Y = \mu + \sigma\delta\rho U^{-1/2} + \sigma(1 - \delta^2)^{1/2}U^{-1/2}\epsilon \quad (2.37)$$

em que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $U \sim H(\cdot; \nu)$, $\rho \sim N_+(0, 1)$ e $\epsilon \sim N(0, 1)$. ρ e ϵ são mutuamente independentes. Portanto, pode-se gerar de uma distribuição independente normal assimétrica, procedendo em dois passos, gerando primeiro de uma distribuição de U e depois de uma distribuição condicional $Y|U$ usando, por exemplo, a representação estocástica dada em (2.37).

Seja $Y \sim MENA(\mu, \sigma^2, \lambda; H)$, então,

- Se $E[U^{-1/2}] < \infty$, então $E[Y] = \mu + \sqrt{\frac{2}{\pi}} k_1 \sigma \delta$,
- Se $E[U^{-1}] < \infty$, então $Var[Y] = k_2 \sigma^2 - \frac{2}{\pi} k_1^2 \sigma^2 \delta^2$

em que $k_m = E[U^{-m/2}]$.

A classe de distribuições MENA permite a acomodação de observações atípicas e conseqüentemente a obtenção de procedimentos estatísticos robustos, considerando tanto assimetria e caudas mais pesadas que a distribuição normal assimétrica. Com isso, permite a extensão de modelos desenvolvidos sob a hipótese de normalidade. Alguns exemplos de distribuições MENA são:

- **Distribuição t de *Student* assimétrica:** $Y \sim TA(\mu, \sigma^2, \lambda, \nu)$, $\nu > 0$.

A distribuição t de *Student* assimétrica com parâmetros de locação μ , escala σ^2 , assimetria λ e graus de liberdade ν , pode ser expressa de forma hierárquica como:

$$Y|U = u, \mu, \sigma^2, \lambda \sim NA\left(\mu, \frac{\sigma^2}{u}, \lambda\right); U|\nu \sim G(\nu/2, \nu/2).$$

O uso desta distribuição é uma alternativa robusta ao modelo com distribuição normal assimétrica usado na literatura. Um caso particular da distribuição t de *Student* assimétrica é a distribuição Cauchy assimétrica quando $\nu = 1$. Assim, também quando $\nu \rightarrow \infty$, tem-se a distribuição normal assimétrica como caso limite.

- **Distribuição *slash* assimétrica:** $Y \sim SA(\mu, \sigma^2, \lambda, \nu)$, $\nu > 0$.

A distribuição *slash* assimétrica possui parâmetros de locação μ , escala σ^2 e assimetria λ . Pode ser expressa como

$$Y|U = u, \mu, \sigma^2, \lambda \sim NA\left(\mu, \frac{\sigma^2}{u}, \lambda\right); U|\nu \sim Be(\nu, 1).$$

A *slash* assimétrica é uma distribuição de cauda pesada, tendo como limite a distribuição normal assimétrica (quando $\nu \rightarrow \infty$).

- **Distribuição variância gama assimétrica:** $Y \sim VGA(\mu, \sigma^2, \nu)$, $\nu > 0$.

A distribuição simétrica variância gama (VG) foi proposta por Madan e Seneta (1990) para modelar retornos de mercado. A distribuição VG apresenta caudas mais pesadas que a da distribuição normal. Aqui é apresentada a versão assimétrica que além da cauda pesada, esta distribuição leva em conta a assimetria. Assim como nas outras distribuições, a variância gama assimétrica possui parâmetros de locação μ , escala σ^2 e assimetria λ . A representação da distribuição na forma hierárquica é dada por:

$$Y|U = u, \mu, \sigma^2, \lambda \sim NA\left(\mu, \frac{\sigma^2}{u}, \lambda\right); U|\nu \sim GI(\nu/2, \nu/2).$$

Quando o parâmetro de assimetria é igual a zero obtém-se a classe de distribuições de mistura de escala normal (MEN), que é um caso particular da classe MENA.

A classe MEN possui distribuições assim como a classe MENA com caudas mais pesadas que a distribuição normal mas estas não apresentam assimetria, são todas simétricas. As distribuições descritas anteriormente apresentam suas versões simétricas que pertencem a classe MEN. São elas: a distribuição normal, t de *Student*, *slash* e variância gama. A família de distribuições MEN é usada quando os dados não apresentam assimetria mais apresentam dados com valores atípicos.

Com o objetivo de ilustrar e comparar estas quatro distribuições da família MEN, a Figura 2.2 mostra as densidades da $N(0, 1)$, $T(0, 1, \nu = 3)$, $S(0, 1, \nu = 3)$ e $VG(0, 1, \nu = 10)$, note que a escala das densidades foram alteradas para que possuísem o mesmo valor na origem e o parâmetro ν difere para cada distribuição. Através da Figura 2.2 pode se observar que as três distribuições possuem caudas mais pesadas que a distribuição normal.

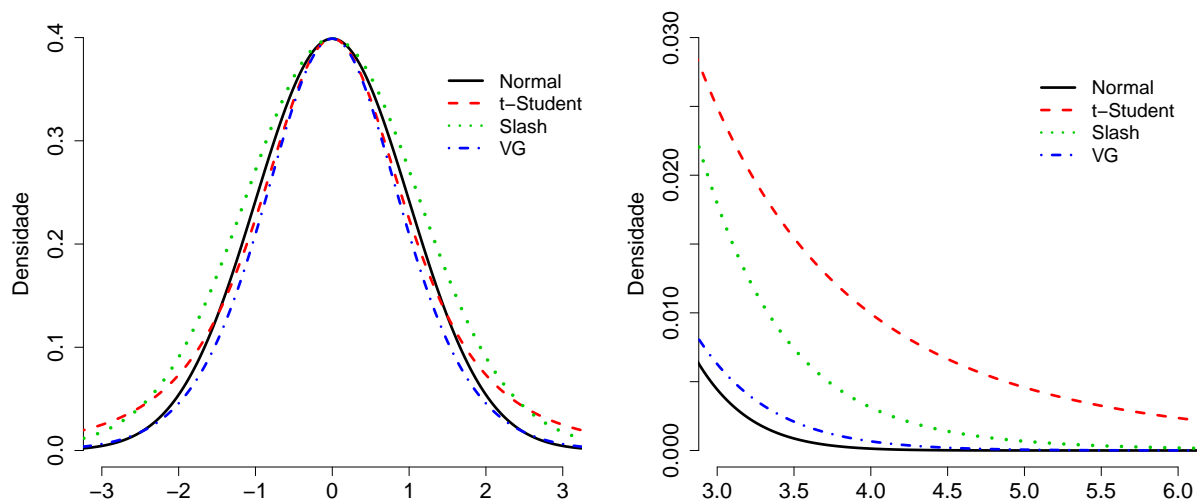


Figura 2.2: Gráfico da função de densidade de quatro diferentes distribuições pertencentes à família MEN. São elas a $N(0,1)$, $T(0,1,3)$, $S(0,1,3)$ e $VG(0,1,10)$.

O próximo capítulo apresenta a descrição detalhada da formulação dos modelos propostos bem como o processo de inferência e um estudo com dados gerados artificialmente.

Capítulo 3

Modelos dinâmicos para observações binárias com função de ligação assimétrica

3.1 Introdução

Neste capítulo é apresentada a classe de modelos dinâmicos para variáveis de respostas binárias com função de ligação, a função de distribuição na classe de mistura de escala normal assimétrica utilizando a idéia de aumento de dados introduzida por Tanner e Wong (1987). Esta classe inclui a ligação proibito como caso particular. A classe proposta é mais flexível pois oferece além de caudas mais pesadas que a da distribuição normal, a modelagem da assimetria dos dados.

O capítulo apresenta detalhes da construção do modelo. A idéia de aumento de dados e toda a modelagem proposta é explicada na Seção 3.2. Na Seção 3.3 são explicitados os quatro modelos propostos que utilizam a abordagem descrita. O procedimento de inferência que será utilizado para a estimação dos parâmetros dos modelos é descrito na Seção 3.4. Finalmente, na Seção 3.5 é realizado um exercício com dados gerados artificialmente cuja finalidade é a verificação da eficiência dos modelos na estimação dos parâmetros.

3.2 Formulação do modelo

Considere uma série temporal de observações binárias $\mathbf{Y}_{1:T} = (Y_1, \dots, Y_T)'$, em que t indica o tempo. Assume-se que

$$Y_t = \begin{cases} 1, & p_t \\ 0, & 1 - p_t \end{cases}, \quad t = 1, \dots, T,$$

em que p_t representa a probabilidade de sucesso e, por definição, deve pertencer ao intervalo $[0, 1]$. Portanto, é razoável modelar p_t por uma função de distribuição acumulada.

Além de usar as covariáveis na modelagem das observações deve-se levar em consideração a dependência temporal na modelagem. Entre os modelos para séries temporais de observações binárias, a classe de modelos dinâmicos binários (Cox, 1981) tem ganhado uma grande popularidade. Como visto na Seção 2.6, o modelo dinâmico binário é representado por dois processos:

- **Processo observado:** a distribuição condicional dado um vetor q dimensional de estados $\boldsymbol{\theta}_t$ é $Y_t | \boldsymbol{\theta}_t \sim Ber(p_t)$, a probabilidade condicional de sucesso $p_t = P(Y_t = 1 | \boldsymbol{\theta}_t)$ segue a equação de observação

$$p_t = H(\mathbf{x}_t' \boldsymbol{\alpha} + \mathbf{F}_t' \boldsymbol{\theta}_t).$$

Nesta equação é incorporada a dependência temporal das observações, representada pelo vetor de estados $\boldsymbol{\theta}_t$ e o vetor q dimensional \mathbf{F}_t é um vetor conhecido. Associam-se às observações um vetor de covariáveis \mathbf{x}_t que variam no tempo e $\boldsymbol{\alpha}$ é o vetor de coeficientes da regressão. H é uma função de distribuição acumulada e H^{-1} é chamada de função de ligação.

- **Processo dos estados:** o vetor de estados $\boldsymbol{\theta}_t$ segue um q dimensional processo de Markov dado por,

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t$$

em que \mathbf{G}_t é uma matriz de transição conhecida $q \times q$ dimensional, e o vetor de erro \mathbf{w}_t tem média zero.

Quando H é uma função de distribuição simétrica, a curva de resposta para p_t possui uma forma simétrica em torno de $p_t = 0.5$, assim como a distribuição normal simétrica e a *t-Student* simétrica na Figura 3.1, ou seja, com isso espera-se obter números parecidos de "0" e "1" na amostra binária. Mas como mencionado por Chen et al. (1999) quando a taxa de "0" e "1" são diferentes, ligações simétricas não são apropriadas pois a inferência é sensível a escolha desta função e portanto podem ser estimados valores incorretos para os parâmetros.

A proposta deste trabalho é utilizar a função de distribuição da classe de distribuições de mistura de escala normal assimétrica para modelar H . Esta modelagem é mais geral e se torna mais apropriada no ajuste de dados pois permite que existam taxas diferentes de "0" e "1" na amostra. A curva de resposta será modelada de forma assimétrica, com a introdução do parâmetro de assimetria, que irá controlar a taxa de crescimento ou decréscimo da probabilidade de sucesso da resposta binária. Com isso é permitida que tal probabilidade possa se aproximar de "0" com uma velocidade diferente da que aproxima-se de "1", como mostra na Figura 3.1 a distribuição normal assimétrica e a *t-Student* assimétrica.

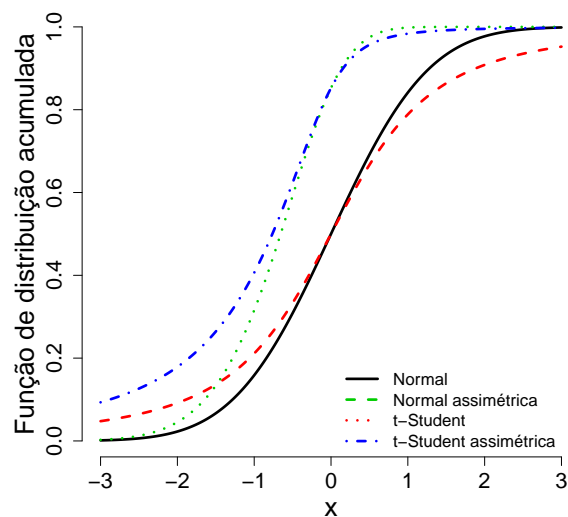


Figura 3.1: Gráfico da função de distribuição acumulada da normal simétrica e assimétrica, e da distribuição *t-Student* simétrica e assimétrica. Sendo o valor da assimetria igual a -2 e os graus de liberdade da *t-Student* igual a 2.

Para facilitar o processo de inferência em conjunto com esta classe de modelos dinâmicos binários será usada a idéia de aumento de dados introduzida por Tanner e Wong (1987) e se refere a um esquema de aumentar os dados observados com a utilização de variáveis latentes, não observáveis. Com isso, a análise se torna mais fácil por simplificar o cálculo da distribuição *a posteriori*, pois utilizando as variáveis latentes evita-se a dificuldade de trabalhar com a função de verossimilhança do tipo Bernoulli.

Portanto, usando a idéia do aumento de dados, foi adotada a abordagem feita em Albert e Chib (1993) em que admite-se que Y_t é gerado por uma dicotomização de um processo contínuo subjacente Z_t que se dá pela relação

$$Y_t = \begin{cases} 1 & \Rightarrow Z_t > 0 \\ 0 & \Rightarrow Z_t \leq 0 \end{cases}, \quad t = 1, \dots, T.$$

Assume-se que Z_t segue uma distribuição na classe MENA.

Com isso, utilizando esta representação, a probabilidade de ocorrer ou não sucesso pode ser escrita na forma

$$P(Y_t = j) = \begin{cases} j = 1, & p_t = P(Z_t > 0) \\ j = 0, & 1 - p_t = P(Z_t \leq 0) \end{cases}, \quad t = 1, \dots, T.$$

Assim, se a distribuição de Z_t for uma distribuição simétrica, como mostra a Figura 3.2(a), a probabilidade de ocorrer "0" e "1" na amostra é a mesma. Por isso, a proposta deste trabalho é considerar a possibilidade de Z_t assumir uma distribuição assimétrica, como nas figuras 3.2(b) e (c), por exemplo. Podendo assim, contemplar os casos em que a taxa de "0" e "1" na amostra não seja a mesma. Se a assimetria for positiva, existirá uma maior probabilidade de ocorrer "0", enquanto se for negativa, a probabilidade maior é de ocorrer "1".

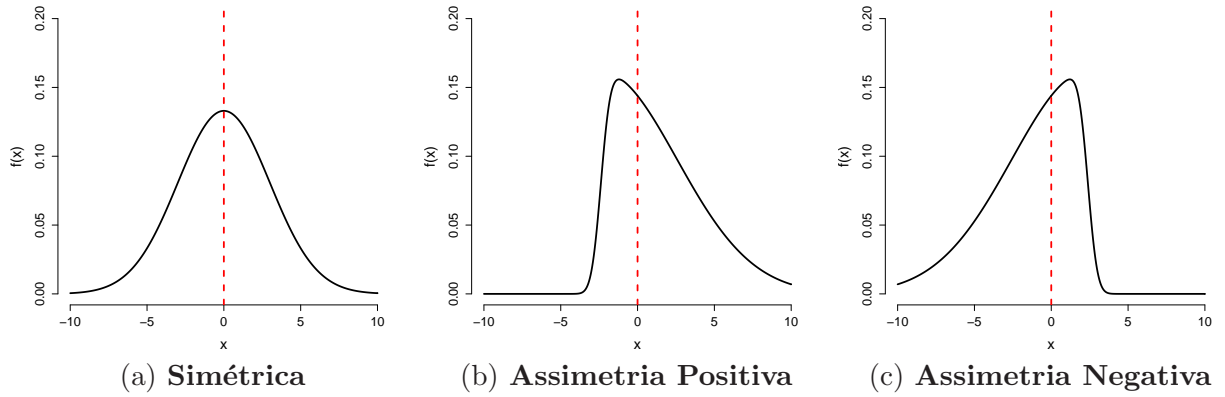


Figura 3.2: Gráficos das funções de densidade da distribuição normal assimétrica, com mesma média e variância, diferindo apenas no parâmetro de assimetria.

O uso da abordagem do aumento de dados permite formular um modelo com uma estrutura mais geral para analisar dados binários e facilitar o procedimento de inferência. Assume-se que o vetor da variável aleatória não observável $\mathbf{Z}_{1:T} = (Z_1, \dots, Z_T)'$ segue um modelo dinâmico binário, cujo o sistema de equações é dado por,

$$Z_t = \mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t + e_t \quad (3.1)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t \quad (3.2)$$

em que o erro da equação de observação, e_t , segue uma distribuição de mistura de escala normal assimétrica e $\mathbf{w}_t \stackrel{iid}{\sim} N(0, W)$. Assumindo que Z_t segue um modelo dinâmico a dependência temporal é incorporada no modelo. Já a assimetria e a cauda pesada são incorporadas em Z_t através de e_t .

A Equação (3.1) pode ser escrita considerando para o erro e_t a representação estocástica da distribuição de mistura apresentada na seção (2.8):

$$Z_t = \mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t + \underbrace{\mu + \sigma \delta \rho_t U_t^{-1/2} + \sigma (1 - \delta^2)^{1/2} U_t^{-1/2} \epsilon_t}_{e_t}.$$

Portanto o modelo é descrito exatamente na forma:

$$Z_t = \mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t + \mu + \sigma \delta \rho_t U_t^{-1/2} + \sigma(1 - \delta^2)^{1/2} U_t^{-1/2} \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} N(0, 1) \quad (3.3)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \stackrel{iid}{\sim} N(0, W) \quad (3.4)$$

$$U_t \sim h(U_t | \nu) \quad (3.5)$$

$$\rho_t \sim N_+(0, 1)$$

Onde \mathbf{x}_t é o vetor de covariáveis referentes ao tempo t , $\boldsymbol{\alpha}$ o vetor dos coeficientes do efeito fixo (regressão) e \mathbf{F}_t é um vetor que pode representar efeitos de tendência. Os parâmetros μ e σ são, respectivamente, os parâmetros de locação e escala do erro da equação de observação (e_t) e são fixados em $\mu = -\sigma \delta \sqrt{\frac{2}{\pi}} k_1$ em que $k_1 = E[U_t^{-1/2}]$ e $\sigma = 1$. Tal fato faz com que a média de e_t seja zero e que esta generalização inclua o caso em que a função de ligação é o probito. λ é o parâmetro de assimetria e $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ tal que $\delta \in (-1, 1)$. U_t é a variável de mistura e segue uma distribuição $h(U|\nu)$ em que ν é um parâmetro de forma e \mathbf{G}_t é a matriz de transição dos estados. O parâmetro ν é responsável pela forma da distribuição da variável de mistura U_t e portanto dependendo da escolha desta distribuição o modelo apresentará caudas mais pesadas acomodando melhor valores atípicos. O parâmetro λ é responsável pela assimetria da distribuição do erro, e_t , e com isso a distribuição se torna uma alternativa mais flexível em relação a distribuições simétricas.

Com a formulação do modelo proposta neste trabalho é introduzida uma nova classe de distribuições para a modelagem dinâmica de variáveis respostas binárias. Esta modelagem é mais geral e se torna mais flexível pois são propostas diferentes funções de ligação.

Fixando o parâmetro de locação, μ , e de escala, σ como foi mencionado anteriormente esta formulação do modelo inclui o caso em que a função de ligação é o probito. Isto ocorre quando a assimetria é nula, ou seja, $\lambda = 0$ e a variável de mistura $U_t = 1$ para todo $t = 1, \dots, T$. Da relação (3.1), a distribuição marginal de Y_t dado $\boldsymbol{\theta}_t$ e $\boldsymbol{\alpha}$ segue o modelo probito da forma

$$p_t = P(Y_t = 1 | \boldsymbol{\theta}_t, \boldsymbol{\alpha}) = \Phi(\mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t), \quad (3.6)$$

em que $\Phi(\cdot)$ representa a função de distribuição acumulada da normal padrão.

O intuito desta proposta é desenvolver uma família paramétrica de ligações assimétricas, proporcionando assim maior flexibilidade na hora do ajuste dos dados. Como esta modelagem é mais geral inclui casos particulares para dados que não apresentem assimetria ou caudas mais pesadas. Se $\lambda = 0$, os modelos passam a ser simétricos e utilizam somente a mistura de distribuição de escala normal, sendo estes com caudas mais pesadas. Se a variável de mistura $U_t = 1$ para todo $t = 1, \dots, T$ cai no caso da função de ligação normal assimétrica.

As equações (3.3) e (3.4), condicionadas a variável de mistura U_t , a variável latente ρ_t e ao vetor α , representam juntas um modelo linear dinâmico descrito na Subseção 2.5.1. O que torna o procedimento de inferência mais simples se caso não fosse usada a idéia do aumento de dados. A inferência nesta classe de modelos foi discutida na Subseção 2.5.2.

Utilizando a modelagem descrita a seção a seguir apresentará os modelos propostos.

3.3 Modelos propostos

Seja o modelo descrito pelas equações (3.3) e (3.4) descritas na Seção 3.2, serão propostas diferentes distribuições $h(U_t|\nu)$ para a variável de mistura U_t que é responsável pela cauda da distribuição do erro da equação de observação. Estas propostas são referentes as distribuições de mistura de escala normal assimétrica descritas e discutidas na Seção 2.8. Tais escolhas apresentam distribuições assimétricas com caudas mais pesadas que a distribuição normal que é frequentemente usada na literatura.

Portanto seguindo esta formulação, são propostos quatro modelos nos quais são assumidas diferentes abordagens para a variável de mistura U_t . São eles:

- **M1:** $U_t = 1$ para todo $t = 1, \dots, T$, com função de ligação normal assimétrica.
- **M2:** $U_t|\nu \sim G(\nu/2, \nu/2)$, com função de ligação *t-Student* assimétrica.
- **M3:** $U_t|\nu \sim Be(\nu, 1)$, com função de ligação *slash* assimétrica.
- **M4:** $U_t|\nu \sim GI(\nu/2, \nu/2)$, com função de ligação variância gama assimétrica.

O processo de estimação dos parâmetros dos modelos propostos será abordado na próxima seção.

3.4 Procedimento de inferência

O procedimento de inferência será feito seguindo a abordagem Bayesiana, assim deve-se atribuir uma distribuição a *priori* ao vetor paramétrico a fim de obter a distribuição a *posteriori*. Para realizar a estimação dos parâmetros do modelo serão utilizadas técnicas de MCMC.

Suponha que o modelo dependa de um vetor paramétrico Ψ . A formulação do modelo usa o princípio de aumento de dados, que considera $\mathbf{Z}_{1:T}$, $\boldsymbol{\theta}_{0:T}$, $U_{1:T}$ e $\rho_{1:T}$ como variáveis latentes. O núcleo da densidade da distribuição a *posteriori* conjunta dos parâmetros e das variáveis latentes pode ser escrita como

$$\begin{aligned}
p(\mathbf{Z}_{1:T}, \boldsymbol{\theta}_{0:T}, U_{1:T}, \rho_{1:T}, \Psi \mid \mathbf{y}_{1:T}) &\propto p(\mathbf{Z}_{1:T} \mid \boldsymbol{\theta}_{0:T}, U_{1:T}, \rho_{1:T}, \Psi, \mathbf{y}_{1:T})p(\boldsymbol{\theta}_{0:T} \mid U_{1:T}, \rho_{1:T}, \Psi) \\
&\times p(U_{1:T} \mid \Psi)p(\rho_{1:T} \mid \Psi)p(\Psi) \\
&= \prod_{t=1}^T \{1(Z_t \geq 0)1(y_t = 1) + 1(Z_t < 0)1(y_t = 0)\} \\
&\times \phi\left(Z_t; \mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t + \mu + \delta \rho_t U_t^{-1/2}, \frac{1 - \delta^2}{U_t}\right) \\
&\times p(\boldsymbol{\theta}_{0:T} \mid U_{1:T}, \rho_{1:T}, \Psi)p(U_{1:T} \mid \Psi)p(\rho_{1:T} \mid \Psi)p(\Psi).
\end{aligned}$$

$\phi(\cdot, \cdot)$ denota a função de densidade de probabilidade da distribuição normal e $1(x \in A)$ é a função indicadora que é igual a 1 se x está contido em A .

A distribuição a *posteriori* não possui forma fechada e com isso seria muito difícil amostrar diretamente dela. Mas a distribuição a *posteriori* marginal pode ser obtida usando o algoritmo de amostragem de Gibbs. O esquema de amostragem é dado por:

Algoritmo 3. Algoritmo de simulação da distribuição a posteriori:

1. Fazer $i = 1$ e inicializar $\boldsymbol{\theta}_{0:T}^{(i)}$, $U_{1:T}^{(i)}$, $\rho_{1:T}^{(i)}$ e $\Psi^{(i)}$;
2. Amostrar $\mathbf{Z}_{1:T}^{(i+1)} \sim p(\mathbf{Z}_{1:T} \mid \boldsymbol{\theta}_{0:T}^{(i)}, U_{1:T}^{(i)}, \rho_{1:T}^{(i)}, \Psi^{(i)}, \mathbf{y}_{1:T})$;

3. Amostrar $\boldsymbol{\theta}_{0:T}^{(i+1)} \sim p(\boldsymbol{\theta}_{0:T} \mid U_{1:T}^{(i)}, \rho_{1:T}^{(i)}, \Psi^{(i)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T});$

4. Amostrar $U_{1:T}^{(i+1)} \sim p(U_{1:T} \mid \boldsymbol{\theta}_{1:T}^{(i+1)}, \rho_{1:T}^{(i)}, \Psi^{(i)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T});$

5. Amostrar $\rho_{1:T}^{(i+1)} \sim p(\rho_{1:T} \mid \boldsymbol{\theta}_{1:T}^{(i+1)}, U_{1:T}^{(i+1)}, \Psi^{(i)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T});$

6. Amostrar $\Psi^{(i+1)} \sim p(\Psi \mid \boldsymbol{\theta}_{1:T}^{(i+1)}, U_{1:T}^{(i+1)}, \rho_{1:T}^{(i+1)}, \mathbf{Z}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$

7. Fazer $i = i + 1$ e voltar para o item 2 até que a convergência seja atingida.

No passo 2, para realizar a amostragem de $\mathbf{Z}_{1:T}$, tem-se que $p(\mathbf{Z}_{1:T} \mid \cdot) = \prod_{t=1}^T p(Z_t \mid \cdot)$ pois condicional as variáveis latentes e ao vetor paramétrico os $Z_{t's}$ são independentes. Portanto para cada $t = 1, \dots, T$ basta amostrar:

- se $y_t = 1$ então $Z_t \sim N_+(\mu_{z_t}, \sigma_{z_t}^2)$;
- se $y_t = 0$ então $Z_t \sim N_-(\mu_{z_t}, \sigma_{z_t}^2)$ em que N_- é uma normal truncada nos \mathbb{R}^- ;

Os parâmetros da normal truncada são definidos por:

$$\begin{aligned} \mu_{z_t} &= \mathbf{x}'_t \boldsymbol{\alpha} + \mathbf{F}'_t \boldsymbol{\theta}_t + \mu + \delta \rho_t U_t^{-1/2}; \\ \sigma_{z_t}^2 &= \frac{1 - \delta^2}{U_t}. \end{aligned}$$

Para obter uma amostra dos estados, $\boldsymbol{\theta}_{0:T}$, no passo 3 será usado o algoritmo de simulação de perturbações suavizadas introduzido por de Jong e Shephard (1995) descrito na Subseção 2.5.2, neste caso o termo \mathbf{c}_t é representado por $\mathbf{c}_t = \mathbf{x}'_t \boldsymbol{\alpha} + \mu + \sigma \delta \rho_t U_t^{-1/2}$.

No passo 4 é feita a amostragem de $U_{1:T}$. Note que os $U_{t's}$ são independentes quando condicionado as variáveis latentes e ao vetor paramétrico, portanto $p(U_{1:T} \mid \cdot) = \prod_{t=1}^T p(U_t \mid \cdot)$. As distribuições das condicionais completas da variável latente U_t variam de acordo com o modelo. No caso do modelo **M1** esta variável é considerada igual a 1 para todo tempo. Nos outros modelos as distribuições condicionais completas não são conhecidas e portanto será necessário fazer um passo de Metropolis para poder amostrar delas. Os cálculos das distribuições condicionais completas de U_t estão no Apêndice A.

O passo 5 é para amostrar de $\rho_{1:T}$. Seja $p(\rho_{1:T} | \cdot) = \prod_{t=1}^T p(\rho_t | \cdot)$ pois os ρ_{t_s} são independentes quando condicionado as variáveis latentes e ao vetor paramétrico. Assim, amostra-se ρ_t para cada $t = 1, \dots, T$ de uma $N_+(\mu_{\rho_t}, \sigma_{\rho_t}^2)$ em que

$$\begin{aligned}\mu_{\rho_t} &= (Z_t - \mathbf{x}'_t \boldsymbol{\alpha} - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta U_t^{1/2} \\ \sigma_{\rho_t}^2 &= 1 - \delta^2.\end{aligned}$$

A conta desta distribuição condicional completa pode ser vista no Apêndice A.

No passo 6 realiza-se a amostragem do vetor paramétrico Ψ . Como este vetor varia para cada modelo, sua distribuição condicional completa não possui uma forma geral para os quatro modelos propostos. Na próxima seção, as distribuições das condicionais completas dos hiperparâmetros de cada modelo serão apresentadas.

Na próxima seção, será feita uma análise do processo de estimação dos modelos através de dados gerados artificialmente.

3.5 Estudo com dados artificiais

O estudo com dados gerados artificialmente é realizado a fim de ilustrar os modelos propostos na Seção 3.3 e verificar se o algoritmo de simulação da distribuição a *posteriori* está obtendo resultados satisfatórios, ou seja, se o algoritmo consegue obter boas estimativas para os parâmetros. Seguindo a modelagem descrita na Seção 3.2, em que $Y_{1:T} = (Y_1, \dots, Y_T)$ é uma série temporal binária e segue uma relação com Z_t , uma variável não observável, tal que se $Z_t > 0$ então $Y_t = 1$, caso contrário $Y_t = 0$. O modelo considerado para a realização deste estudo é descrito da seguinte forma:

$$\begin{aligned}Z_t &= \theta_t + \mu + \sigma \delta \rho_t U_t^{-1/2} + \sigma (1 - \delta^2)^{1/2} U_t^{-1/2} \epsilon_t, & \epsilon_t &\sim N(0, 1) \\ \theta_t &= \phi \theta_{t-1} + w_t, & w_t &\sim N(0, W) \\ U_t &\sim h(U_t | \nu) \\ \rho_t &\sim N_+(0, 1)\end{aligned}$$

em que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $\sigma = 1$ e $\mu = -\sigma \delta \sqrt{\frac{2}{\pi}} k_1$ onde $k_1 = E[U_t^{-1/2}]$.

De acordo com os quatro modelos propostos na Seção 3.3, a abordagem para a variável latente U_t , a constante k_1 e o vetor paramétrico para cada modelo é dado por:

M1:

- $U_t = 1$ para todo $t = 1, \dots, T$;
- $k_1 = 1$;
- O vetor paramétrico é igual $\Psi = (\phi, W, \lambda)$. Em que $\phi \in (-1, 1)$, $W > 0$ e $\lambda \in \mathbb{R}$.

M2:

- $U_t|\nu \sim G(\nu/2, \nu/2)$;
- $k_1 = \frac{(\nu/2)^{1/2}}{\Gamma(\nu/2)}\Gamma(\frac{\nu-1}{2})$, em que $\Gamma(\cdot)$ representa a função gama.

Para que k_1 exista, o parâmetro ν deve ser maior que 1;

- O vetor paramétrico é igual $\Psi = (\phi, W, \lambda, \nu)$. Em que $\phi \in (-1, 1)$, $W > 0$, $\lambda \in \mathbb{R}$ e $\nu > 2$.

M3:

- $U_t|\nu \sim Be(\nu, 1)$;
- $k_1 = \frac{2\nu}{2\nu-1}$, condição para sua existência é $\nu \neq 0.5$;
- O vetor paramétrico é igual $\Psi = (\phi, W, \lambda, \nu)$. Em que $\phi \in (-1, 1)$, $W > 0$, $\lambda \in \mathbb{R}$ e

$\nu > 1$.

M4:

- $U_t|\nu \sim GI(\nu/2, \nu/2)$;
- $k_1 = \left(\frac{\nu}{\nu+1}\right)^{\nu/2} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)}$;
- O vetor paramétrico é igual $\Psi = (\phi, W, \lambda, \nu)$. Em que $\phi \in (-1, 1)$, $W > 0$, $\lambda \in \mathbb{R}$ e

$\nu > 0$.

O parâmetro ϕ está definido em $(-1, 1)$ pois será considerado um processo estacionário para os estados, θ , do modelo dinâmico.

Foram gerados dados artificiais a partir destes modelos e os parâmetros foram fixados conforme a Tabela 3.1. O tamanho da série binária foi fixada em $T = 300$. Os dados

foram gerados usando a livreria estatística *Scythe* do programa C++ (Pemstein et al., 2007).

Tabela 3.1: Valores dos parâmetros fixados para a geração dos dados por modelo.

Modelo	ϕ	W	ν	λ
M1	0.95	0.1	-	-5
M2	0.95	0.1	10	-5
M3	0.95	0.1	5	-5
M4	0.95	0.1	10	-5

O parâmetro de assimetria, λ , foi fixado em -5 para todos os modelos. Isto indica uma assimetria negativa nos dados, o que resulta em uma maior quantidade de 1 na amostra binária. Na Tabela 3.2 é apresentada a porcentagem de zeros na amostra binária gerada através de cada modelo proposto.

Tabela 3.2: Porcentagem de zeros na amostra da série binária gerada artificialmente de tamanho $T = 300$.

Modelo	Porcentagem de zeros
M1	32 %
M2	33 %
M3	35 %
M4	36 %

No contexto da análise Bayesiana, é preciso atribuir uma distribuição a *priori* para o vetor paramétrico, Θ , de cada modelo para que sua especificação esteja completa. Será considerada a independência dos parâmetros e portanto a distribuição a *priori* do vetor pode ser dada pela multiplicação das distribuições marginais. Sendo assim, foi atribuída para o parâmetro $\phi \sim N_{(-1,1)}(\mu_\phi, \sigma_\phi^2)$, uma distribuição a *priori* normal truncada em $(-1, 1)$ em que $\mu_\phi = 0.95, \sigma_\phi^2 = 1000$. Para o parâmetro W foi atribuída a *priori* uma

$GI(\alpha_w, \beta_w)$ em que $\alpha_w = 5$ e $\beta_w = 0.3$. Foi atribuída uma normal centrada no valor verdadeiro com uma variância 50 para a distribuição a *priori* do parâmetro de assimetria, ou seja, $\lambda \sim N(\mu_\lambda = -5, \sigma_\lambda^2 = 50)$. Para o parâmetro ν foi atribuída $\nu \sim G(\alpha_\nu, \beta_\nu)$ onde no Modelo 3, $\alpha_\nu = 5$ e $\beta_\nu = 1$ e nos modelos 2 e 4, $\alpha_\nu = 10$ e $\beta_\nu = 1$. Foi feito também um exercício utilizando uma distribuição a *priori* não informativa vista em Fonseca et al. (2008) para o parâmetro ν do Modelo 2, este parâmetro representa os graus de liberdade da distribuição t-*Student* assimétrica. Este resultado pode ser visto no apêndice D. O cálculo da distribuição condicional completa de cada parâmetro é apresentado no Apêndice C e a distribuição proposta é discutida quando preciso.

Após a atribuição da distribuição a *priori* para o vetor paramétrico, foram geradas amostras da distribuição a *posteriori* através do algoritmo de Gibbs descrito na seção 3.4. Foram simuladas ao todo 700 mil iterações. Foram descartadas as 200 mil primeiras considerado o período de aquecimento. E depois foram retiradas as amostras usando um *thin* de 250. Finalmente, a amostra na qual foi feita a inferência possui tamanho 2000. O número alto de iterações se deve ao fato do parâmetro de assimetria, λ , apresentar uma demora em sua convergência. Todo o algoritmo de simulação da distribuição a *posteriori* foi realizado utilizando a livreria estatística *Scythe* do programa C++ (Pemstein et al., 2007).

Na Figura 3.3 estão os gráficos dos estados, θ , para cada modelo considerado. A linha cheia representa o valor verdadeiro e as linhas tracejadas representam o intervalo de credibilidade de 95 %. Note que o intervalo estimado, em geral, contém e acompanha bem o valor verdadeiro. Esta Figura indica que o algoritmo consegue estimar bem os estados em cada modelo proposto.

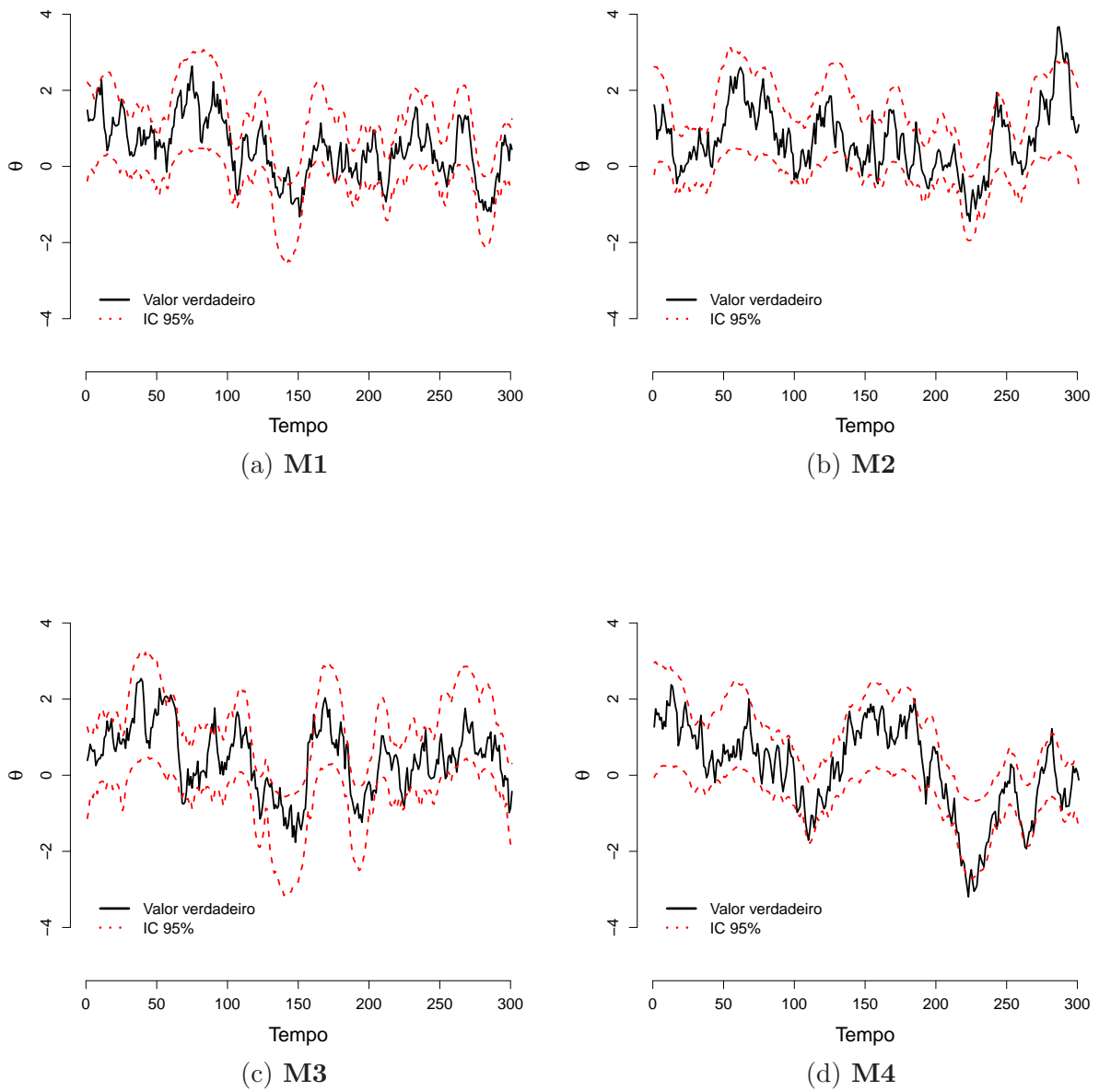
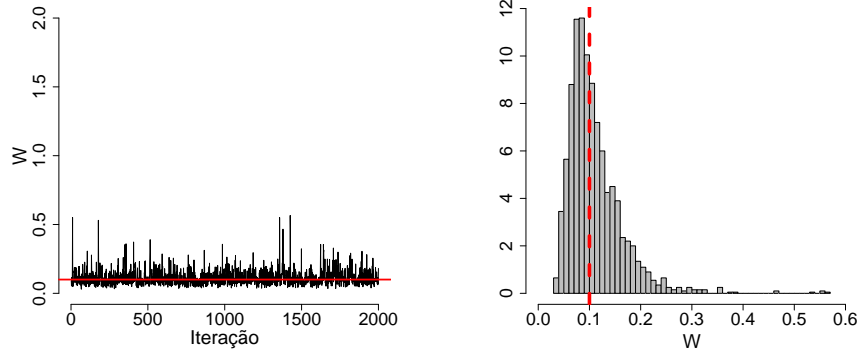


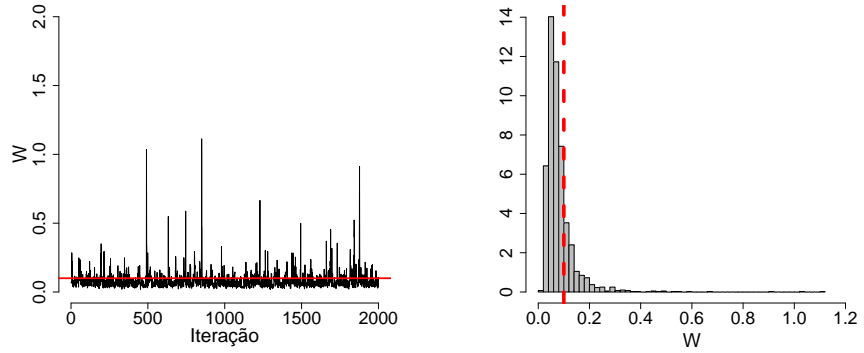
Figura 3.3: Gráficos da série temporal de θ para cada modelo, sendo que a linha cheia representa o valor verdadeiro e a linha tracejada o intervalo de 95% de credibilidade a *posteriori*.

Na Figura 3.4, estão os sumários a *posteriori* do parâmetro W . Os modelos, estimaram bem o valor do parâmetro, sendo que nos modelos 1 e 3 o valor verdadeiro está mais próximo do ponto com maior concentração.

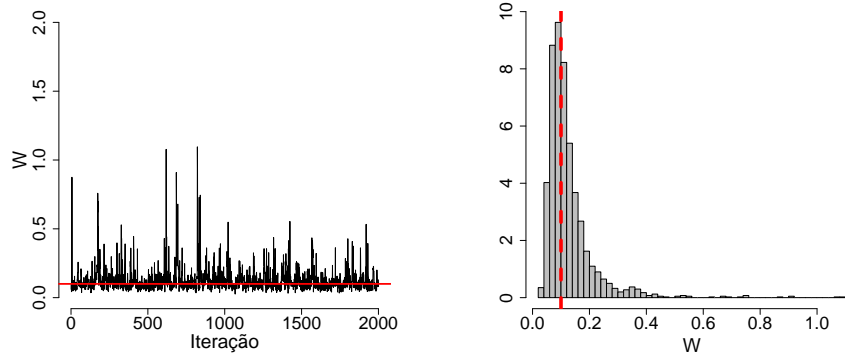
(a) M1



(b) M2



(c) M3



(d) M4

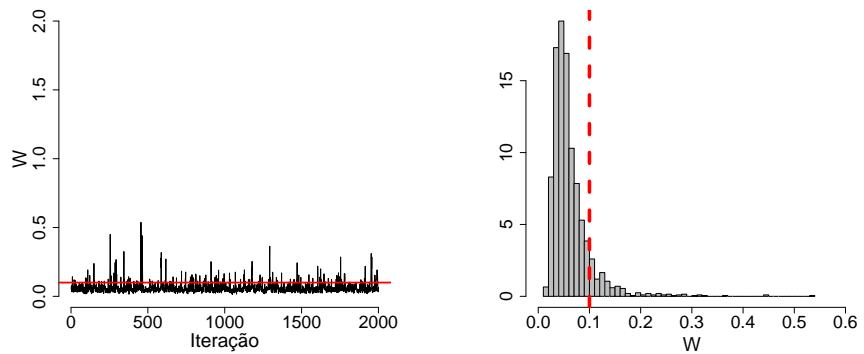
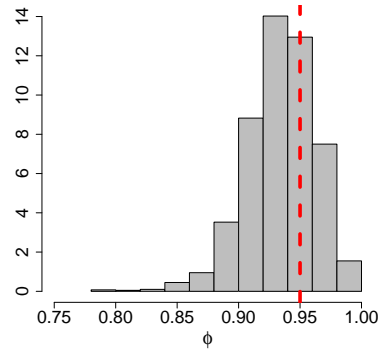
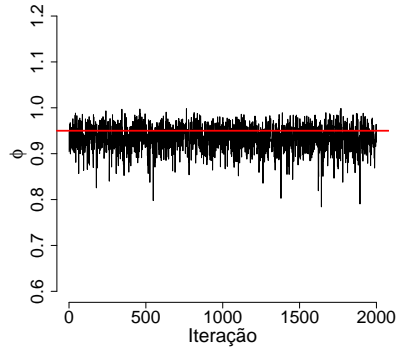


Figura 3.4: Gráficos da cadeia do MCMC do parâmetro W para cada modelo, a linha cheia é o valor verdadeiro. Ao lado, o respectivo histograma em que a linha tracejada representa o valor real.

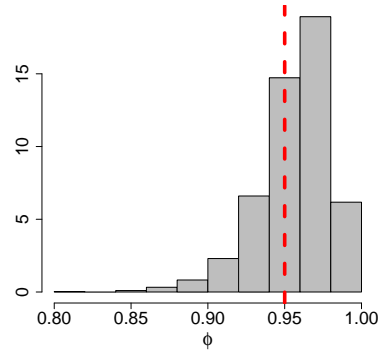
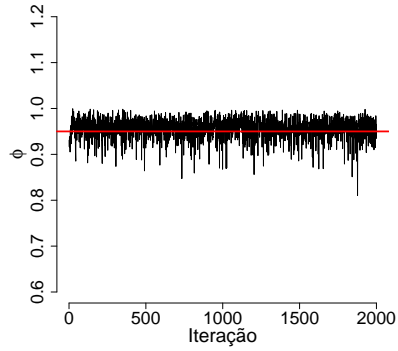
A Figura 3.5, apresenta os sumários a *posteriori* do parâmetro ϕ . Os modelos, estimaram bem o valor de ϕ . O Modelo 1 foi o que obteve maior dispersão para os valores estimados. O Modelo 4 apresenta uma maior concentração dos valores estimados a *posteriori* acima do valor verdadeiro, através de seu histograma nota-se uma assimetria negativa.

Na Figura 3.6, estão os sumários a *posteriori* do parâmetro de assimetria λ . Em geral, a estimação deste parâmetro foi bem. O modelo 4 foi o que apresentou o valor verdadeiro mais próximo do ponto de maior concentração a *posteriori*. O modelo 1 foi o modelo com menor dispersão dos valores na cadeia do MCMC. Os histogramas apresentaram certa simetria para os valores estimados.

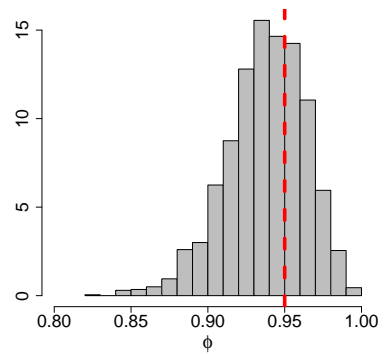
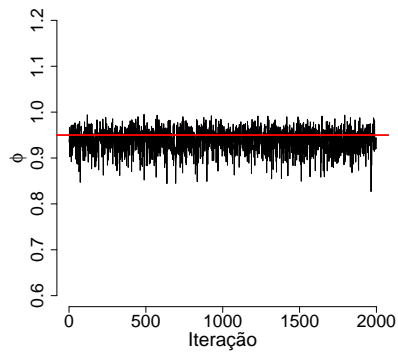
(a) M1



(b) M2



(c) M3



(d) M4

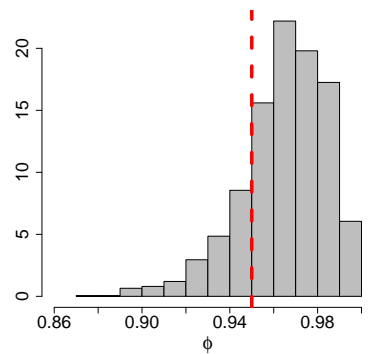
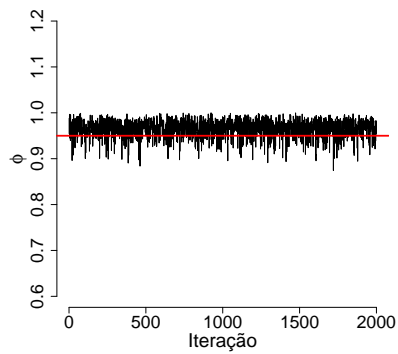
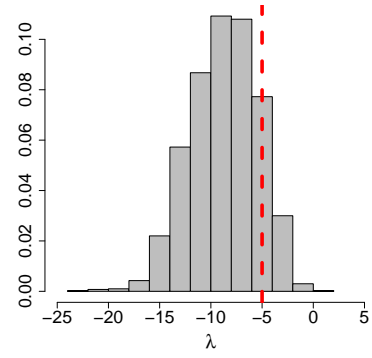
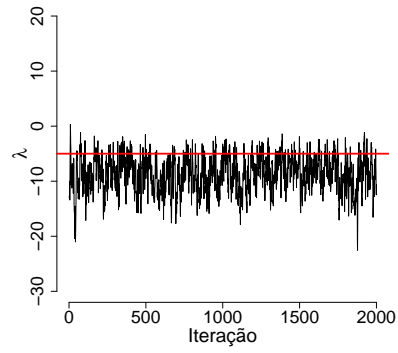
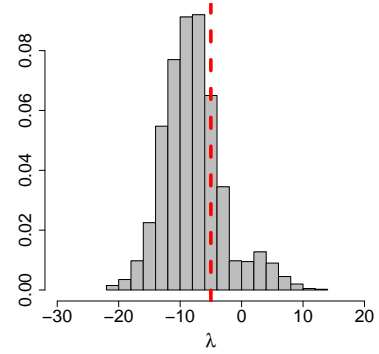
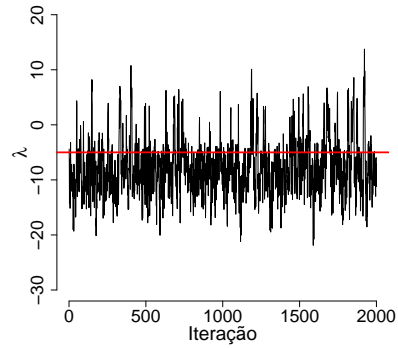


Figura 3.5: Gráficos da cadeia do MCMC do parâmetro ϕ para cada modelo, a linha cheia é o valor verdadeiro. Ao lado, o respectivo histograma em que a linha tracejada representa o valor real.

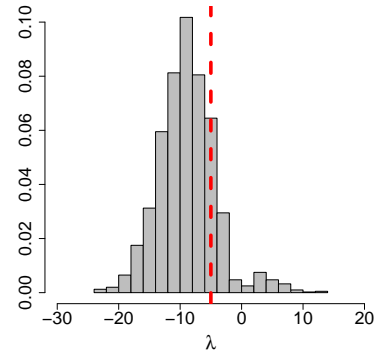
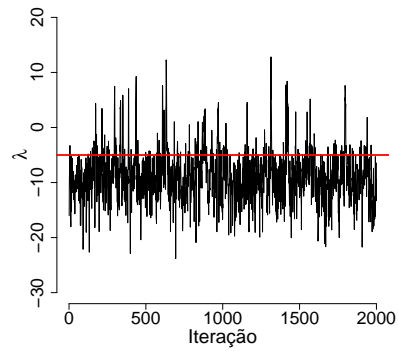
(a) M1



(b) M2



(c) M3



(d) M4

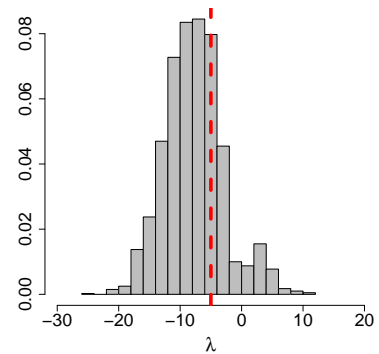
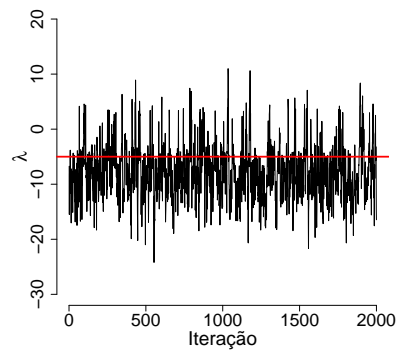


Figura 3.6: Gráficos da cadeia do MCMC do parâmetro λ para cada modelo, a linha cheia é o valor verdadeiro. Ao lado, o respectivo histograma em que a linha tracejada representa o valor real.

Na Figura 3.7, estão os sumários *a posteriori* do parâmetro ν . Este parâmetro está relacionado com a cauda da distribuição do erro da equação de observação. Os modelos conseguem uma boa estimação para ν . O Modelo 3 apresenta uma assimetria positiva e o Modelo 2 apresenta a amostra concentrada no valor verdadeiro.

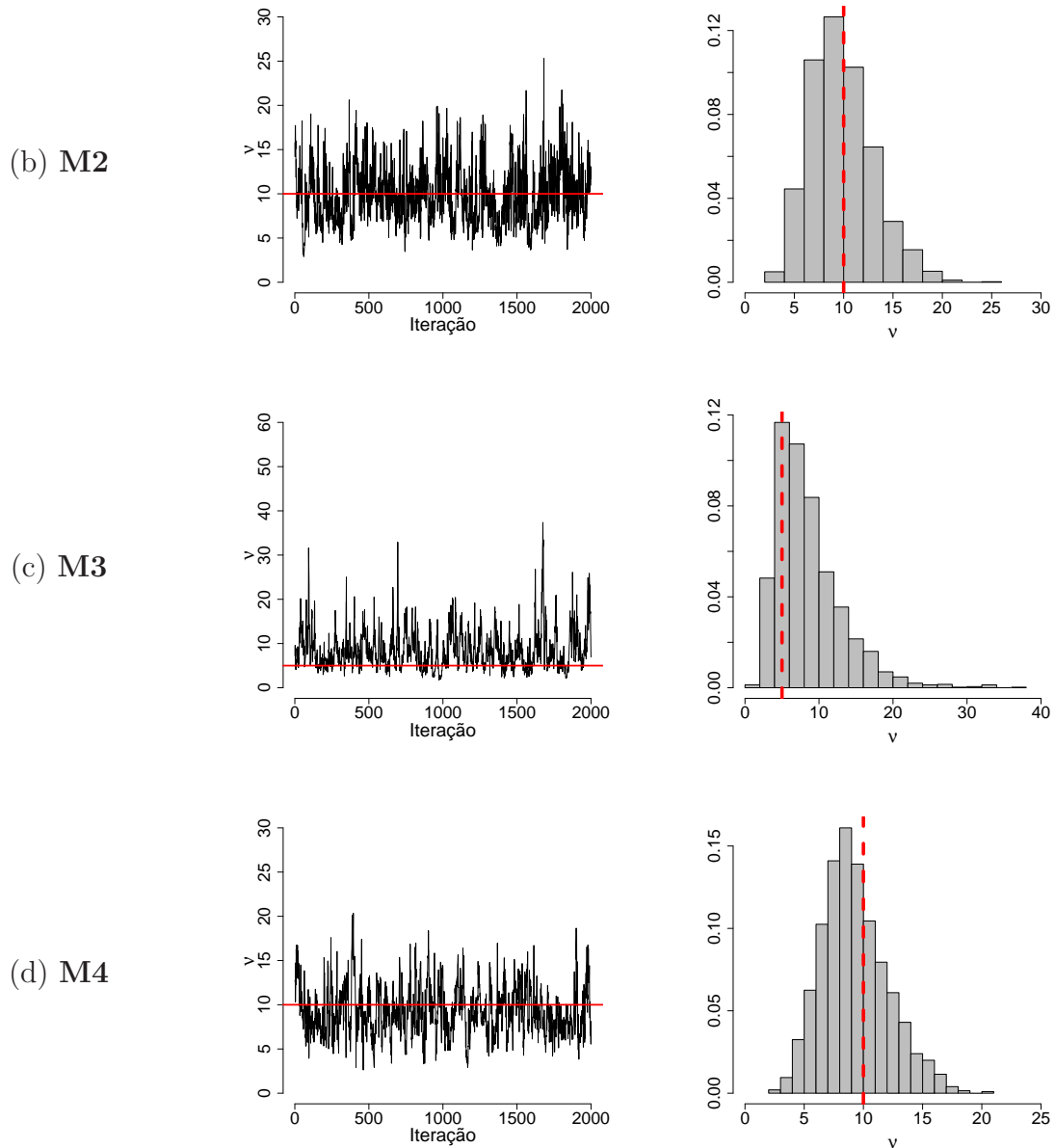


Figura 3.7: Gráficos da cadeia do MCMC do parâmetro ν para cada modelo, a linha cheia é o valor verdadeiro. Ao lado, o respectivo histograma em que a linha tracejada representa o valor real.

Note que no modelo 2 este resultado é referente a utilização da distribuição a *priori* Gama para o parâmetro ν . O resultado usando a distribuição não informativa está no apêndice D e este apresenta resultados mais dispersos com relação ao valor verdadeiro, mas também apresenta um maior ganho de informação a *posteriori*.

A Tabela 3.3 apresenta o sumário a *posteriori* de todos os parâmetros de cada modelo. Na Tabela são exibidas a média, o intervalo de credibilidade de 95%, o desvio padrão, a estatística de Geweke e o fator de ineficiência. A estatística de Geweke (Geweke, 1992) indica se há indícios de convergência da cadeia do MCMC e isto ocorre quando o valor da estatística pertence ao intervalo de $[-1.96, 1.96]$. O fator de ineficiência avalia o esquema de amostragem pelo aumento da variância da amostra a *posteriori* quando comparado com a de amostras independentes, quanto mais longe de 1, menos eficiente é a amostragem.

Note que todos os intervalos de credibilidade de 95% contém o valor verdadeiro do parâmetro em todos os modelos. A estimação dos parâmetros ν e λ são as que apresentam maior dispersão devido a dificuldade na estimação destes parâmetros. Para o parâmetro de assimetria, λ , os resultados além de dispersos, apresentam o valor zero no intervalo de credibilidade, exceto para o modelo 1. Perceba que mesmo o valor verdadeiro sendo negativo o intervalo contém zero. A estimação do parâmetro ϕ é a que apresenta valores estimados da média a *posteriori* mais próximo do valor verdadeiro. Para o parâmetro W os valores estimados também são bem próximos do valor real.

A estatística de Geweke mostra que há indícios de convergência para todos os parâmetros pois os valores estão no intervalo sugerido. O parâmetro ν obteve a amostragem menos eficiente segundo o fator de ineficiência que possui valor maior que três em todos os modelos. E a amostragem mais eficiente foi a do parâmetro ϕ .

Em geral, os modelos estimaram bem os parâmetros e o procedimento de simulação da distribuição a *posteriori* mostrou-se satisfatório.

A seguir no capítulo 4, será apresentada uma aplicação a dados reais.

Tabela 3.3: Sumário *a posteriori* dos parâmetros de cada modelo proposto.

Parâmetro	Estatísticas	Modelos			
		M1	M2	M3	M4
W	Valor real	0.1	0.1	0.1	0.1
	Média	0.111	0.082	0.128	0.063
	IC (95%)	(0.05 ; 0.24)	(0.03 ; 0.23)	(0.05 ; 0.36)	(0.02 ; 0.16)
	Desvio padrão	0.054	0.067	0.090	0.041
	Geweke	0.261	0.727	0.654	-1.003
	Ineficiência	1.508	1.721	2.651	1.601
	ϕ	Valor real	0.95	0.95	0.95
Média		0.934	0.956	0.938	0.965
IC (95%)		(0.87 ; 0.98)	(0.90 ; 0.99)	(0.88 ; 0.98)	(0.92 ; 0.99)
Desvio padrão		0.028	0.023	0.026	0.019
Geweke		0.303	0.836	-1.222	-0.356
Ineficiência		1.135	1.162	1.055	1.048
ν		Valor real	-	10	5
	Média	-	9.851	8.554	9.262
	IC (95%)	-	(4.74 ; 17.03)	(2.85 ; 19.78)	(4.53 ; 15.49)
	Desvio padrão	-	3.179	4.526	2.802
	Geweke	-	-0.287	0.173	0.487
	Ineficiência	-	3.637	3.762	3.775
	λ	Valor real	-5	-5	-5
Média		-8.737	-7.810	-8.878	-7.855
IC (95%)		(-15.33 ; -2.96)	(-16.39 ; 4.70)	(-17.62 ; 3.34)	(-16.64 ; 3.70)
Desvio padrão		3.290	4.941	4.697	4.860
Geweke		-1.104	-1.448	-1.171	-0.103
Ineficiência		3.038	2.835	2.767	2.366

Capítulo 4

Aplicação empírica

4.1 Descrição dos dados

A série binária utilizada neste trabalho é referente à chuva na cidade de Emerald, na Austrália. Os dados são diários no período de 1º de janeiro de 2001 até 15 de setembro de 2002, e foram disponibilizados pelo Departamento de Indústrias Primárias de Queensland (*Queensland Department of Primary Industries*). Ao todo a série possui 623 observações. Existem cinco covariáveis disponíveis para esta série temporal. São elas: o mínimo e máximo da temperatura ($^{\circ}\text{C}$), radiação (MJ/m^2), evaporação (mm) e déficit máximo de pressão de vapor (hPa). O déficit de pressão de vapor é a diferença entre a pressão parcial de vapor d'água na condição de saturação e a pressão parcial de vapor d'água em ar úmido.

A princípio foram usadas estas cinco covariáveis mas duas delas não se apresentaram significativas em nenhum modelo. Portanto a análise será feita com somente três covariáveis, sendo elas, temperatura máxima, radiação e déficit máximo de pressão de vapor.

A Tabela 4.1 apresenta as estatísticas descritivas das três covariáveis selecionadas. A média da temperatura máxima na cidade de Emerald é de 30°C . o máximo de radiação é de $30 \text{ MJ}/\text{m}^2$ e o mínimo do déficit de pressão de vapor é de 6 hPa .

Tabela 4.1: Estatísticas descritivas das covariáveis.

Covariável	Mínimo	Média	Máximo	Desvio Padrão
Temperatura máxima	15.50	30.09	43	5.236
Radiação	6	19.84	30	4.793
Déficit máximo de pressão de vapor	6	17.4	30	6.023

Na série binária que será analisada, o valor 0 representa a ocorrência de chuva no dia e o valor 1 representa a não ocorrência de chuva. A taxa de 0 na amostra é de 18%, o que indica que é possível que funções de ligação simétrica não se ajustem bem a esta série de dados.

4.2 Modelos propostos e Procedimento de inferência

Seguindo a modelagem descrita no capítulo 3, em que $Y_{1:T} = (Y_1, \dots, Y_T)$ é uma série temporal binária e segue uma relação com Z_t , uma variável não observável, tal que se $Z_t > 0$ então $Y_t = 1$, caso contrário $Y_t = 0$. Neste caso a série binária $Y_{1:T}$ é a ocorrência de chuva em Emerald com $Y_t = 0$ quando ocorre chuva no dia t e $Y_t = 1$ caso contrário. O tamanho da série é $T = 623$.

O modelo que será ajustado nesta série é descrito da seguinte forma:

$$\begin{aligned}
 Z_t &= \mathbf{x}'_t \boldsymbol{\alpha} + \theta_t + \mu + \delta \rho_t U_t^{-1/2} + (1 - \delta^2)^{1/2} U_t^{-1/2} \epsilon_t, & \epsilon_t &\sim N(0, 1) \\
 \theta_t &= \phi \theta_{t-1} + w_t, & w_t &\sim N(0, W) \\
 U_t &\sim h(U_t | \nu) \\
 \rho_t &\sim N_+(0, 1)
 \end{aligned}$$

em que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $\mu = -\sigma \delta \sqrt{\frac{2}{\pi}} k_1$ onde $k_1 = E[U_t^{-1/2}]$ e \mathbf{x}'_t é a matriz de covariáveis sendo a primeira coluna somente com o número 1 e as outras três são as covariáveis descritas na seção 4.1. $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ é o vetor dos coeficientes da regressão, sendo α_0 considerado o intercepto e os outros três referentes a cada covariável.

Seguindo esta formulação, serão ajustados quatro modelos nos quais são assumidas diferentes abordagens. São eles:

- **Probito:** $U_t = 1$ para todo $t = 1, \dots, T$, e $\lambda = 0$, no qual representa a função de ligação simétrica probito.
- **M1:** $U_t = 1$ para todo $t = 1, \dots, T$, com função de ligação normal assimétrica.
- **M2:** $U_t|\nu \sim G(\nu/2, \nu/2)$, com função de ligação *t-Student* assimétrica.
- **M3:** $U_t|\nu \sim Be(\nu, 1)$, com função de ligação *slash* assimétrica.

Para cada modelo já foi explicitado a constante k_1 e o vetor paramétrico na seção 3.5, sendo que para o modelo **Probito** não existe a constante k_1 e o vetor paramétrico é dado por $\Psi = (W, \phi, \alpha)$. Note que para todos os modelos existe mais um parâmetro a ser estimado que não tinha no exercício com dados artificiais, este é o vetor de coeficientes da regressão, α . Assim como na Seção 3.5, será considerado um processo estacionário para os estados do modelo dinâmico, com isso o parâmetro $\phi \in (-1, 1)$.

Portanto, são propostos quatro modelos para esta série de dados referente à chuva em Emerald. O modelo probito é um modelo cuja a ligação é simétrica e bastante utilizada na literatura. No modelo 1 é inserido o parâmetro de assimetria, permitindo assim os dados indicarem se existe assimetria ou não. Os modelos 2 e 3 além de serem assimétricos trazem uma distribuição dos erros com cauda mais pesada que os do modelo 1 e probito. Vale ressaltar que o modelo VG assimétrico não foi considerado por problemas numéricos decorrentes da rotina computacional usada.

Será considerada a independência dos parâmetros *a priori*. Sendo assim, para os parâmetros em comum nos quatro modelos foram atribuídas *a priori*: $\phi \sim N_{(-1,1)}(0.95, 100)$, $W \sim GI(5, 0.3)$ e $\alpha \sim N_4(\mathbf{0}, 50I_4)$ em que I_4 denota uma matriz identidade de tamanho 4. Para o parâmetro de assimetria nos modelos 1, 2 e 3, foi atribuída $\lambda \sim N(0, 50)$. Para o parâmetro ν , no modelo 3 foi atribuída $\nu \sim G(0.4, 0.1)$ e no modelo 2 foi usada uma distribuição *a priori* não informativa vista em Fonseca et al. (2008). O cálculo da distribuição condicional completa de α pode ser visto no Apêndice C. A seguir os resultados serão mostrados.

4.3 Resultados a *posteriori*

Após a atribuição da distribuição a *priori* para o vetor paramétrico, os quatro modelos foram ajustados aos dados. Foram geradas amostras da distribuição a *posteriori* através do algoritmo de Gibbs descrito na Seção 3.4. Ao todo 700 mil iterações foram simuladas, sendo as 200 mil primeiras descartadas e depois foram retiradas as amostras usando um *thin* de 250. Finalmente, a amostra na qual foi feita a inferência possui tamanho 2000. Assim como nos dados artificiais, todo o algoritmo de simulação da distribuição a *posteriori* foi realizado utilizando a livreria estatística *Scythe* do programa C++ (Pemstein et al., 2007).

A Tabela 4.2 apresenta o sumário a *posteriori* dos parâmetros dos modelos sem considerar os coeficientes da regressão. O parâmetro W , que representa a variância do erro da equação dos estados, foi estimado por todos os modelos próximo do valor 0.2, o modelo com menor desvio em relação a estimativa foi o modelo 1. Podem ser vistos na Figura 4.2 os seus respectivos histogramas em que todos os modelos apresentaram uma maior concentração nos valores menores que 0.5. O parâmetro ϕ está diretamente relacionado com a dependência temporal dos dados. Nos modelos probito e 1 os intervalos de credibilidade não contém o zero e apesar dos modelos 2 e 3 conterem o zero pode ser visto na Figura 4.4 que a concentração de ambos são nos valores positivos, sendo que o modelo probito e o 1 estão com menor dispersão na estimação. No modelo probito a média a *posteriori* de ϕ foi estimada em 0.77, isto indica mais persistência nos dados que o modelo 1 que estimou a média em 0.65. O intervalo de credibilidade do parâmetro de assimetria λ no modelo 1 não conteve o zero o que indica uma assimetria negativa nos dados. Isto também é evidente nos modelos 2 e 3 pois possuem uma grande concentração de valores negativos que pode ser vista na Figura 4.3. O modelo probito não possui este parâmetro pois é um modelo simétrico. O parâmetro ν foi estimado em 2.956 no modelo 2 e 1.648 no modelo 3, indicando presença de caudas mais pesadas que a da distribuição normal, pois quanto menor o valor de ν mais pesada é a cauda. Na Figura 4.5 observa-se que os modelos concentram os valores estimados em valores menores que 5.

Segundo a estatística de Geweke houve indícios de convergência das cadeias dos parâmetros dos quatro modelos. Os parâmetros λ e ν foram os menos eficiente na amostragem, pois obtiveram os maiores valores do fator de ineficiência em todos os modelos.

Tabela 4.2: Sumário a *posteriori* dos parâmetros de cada modelo ajustado.

Parâmetro	Estatísticas	Modelos			
		Probit	M1	M2	M3
W	Média	0.205	0.145	0.133	0.306
	IC (95%)	(0.03 ; 1.03)	(0.03 ; 0.66)	(0.03 ; 0.56)	(0.02 ; 2.91)
	Desvio padrão	0.310	0.163	0.186	0.749
	Geweke	-0.872	-0.755	1.098	-0.267
	Ineficiência	1.884	3.852	3.190	5.241
ϕ	Média	0.773	0.655	0.498	0.4678
	IC (95%)	(0.36 ; 0.96)	(0.11 ; 0.92)	(-0.68 ; 0.96)	(-0.72 ; 0.97)
	Desvio padrão	0.163	0.204	0.450	0.464
	Geweke	0.865	1.787	0.603	0.4195
	Ineficiência	1.343	2.247	2.857	2.889
λ	Média	-	-5.674	-4.141	-4.402
	IC (95%)	-	(-13.04 ; -0.50)	(-10.10 ; 0.09)	(-11.08 ; -0.15)
	Desvio padrão	-	3.100	3.026	2.971
	Geweke	-	-1.339	1.615	-0.216
	Ineficiência	-	5.185	5.281	5.035
ν	Média	-	-	2.956	1.648
	IC (95%)	-	-	(2.02 ; 6.22)	(1.01 ; 4.92)
	Desvio padrão	-	-	1.192	1.096
	Geweke	-	-	-0.656	-0.8388
	Ineficiência	-	-	5.485	6.669

Na Figura 4.1 estão os gráficos dos estados, θ , para cada modelo ajustado. A linha cheia representa o valor da mediana e o intervalo de credibilidade de 95 % a *posteriori* é representado através do intervalo mais claro. Em geral, os gráficos para cada modelo são parecidos, sendo que o modelo 1 e o modelo 2 apresentam uma menor dispersão e o modelo 4 apresenta a maior.

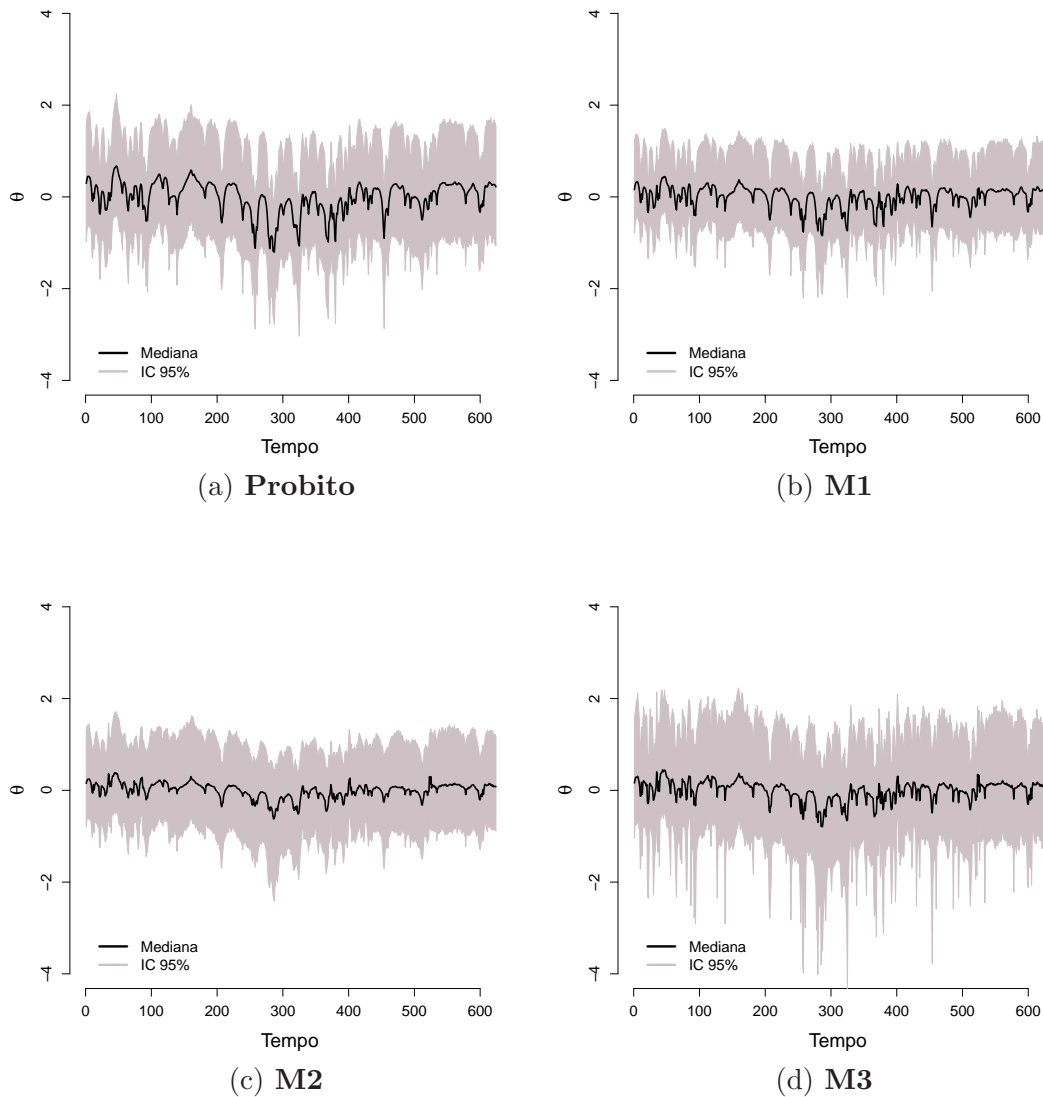


Figura 4.1: Gráficos da série temporal de θ para cada modelo, sendo que a linha cheia representa a mediana e a superfície cinza o intervalo de 95% de credibilidade a *posteriori*.

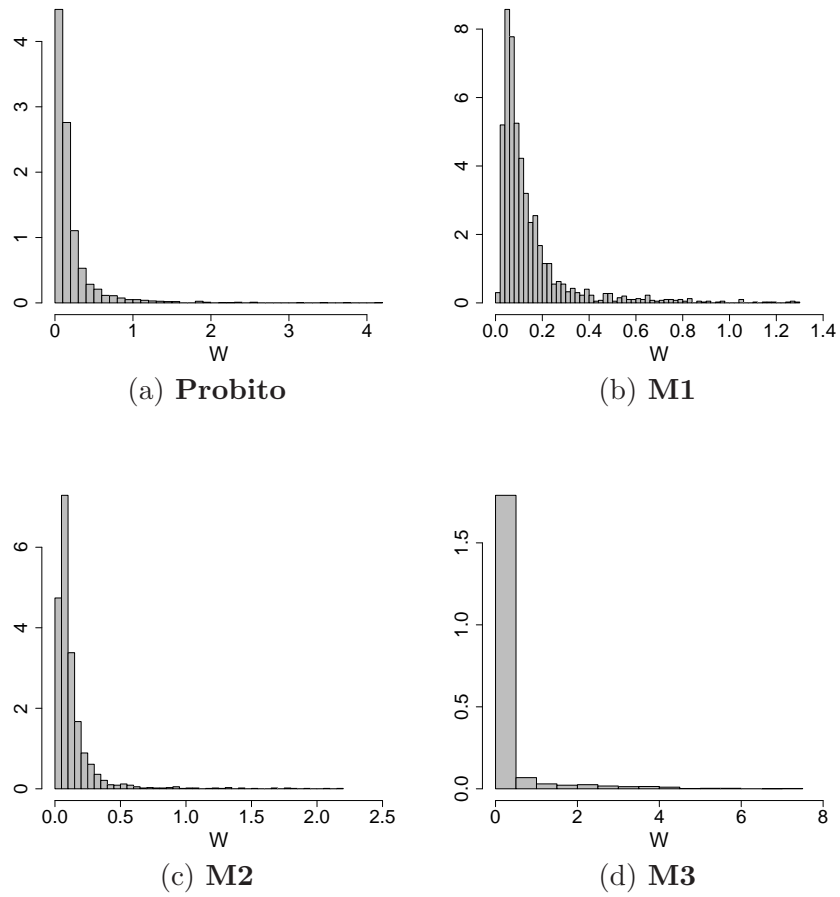


Figura 4.2: Histograma da cadeia do MCMC do parâmetro W para cada modelo.

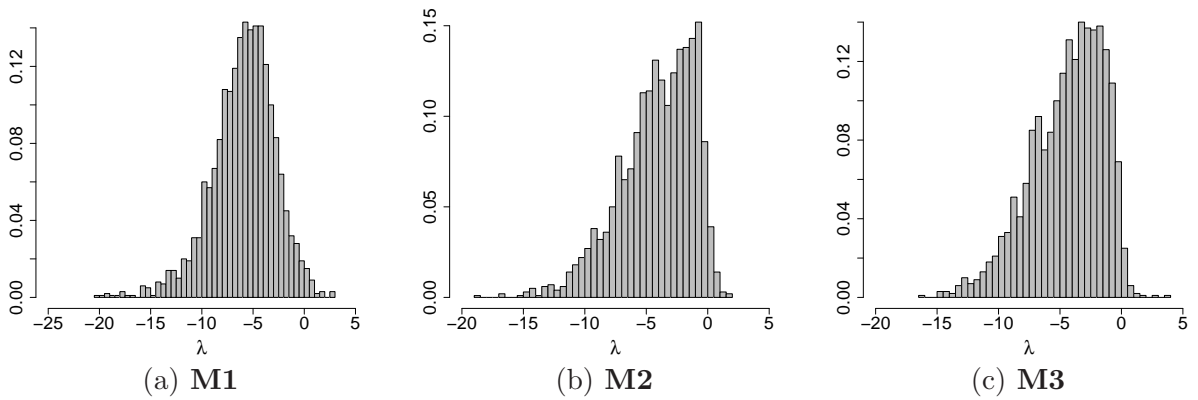


Figura 4.3: Histograma da cadeia do MCMC do parâmetro λ para cada modelo.

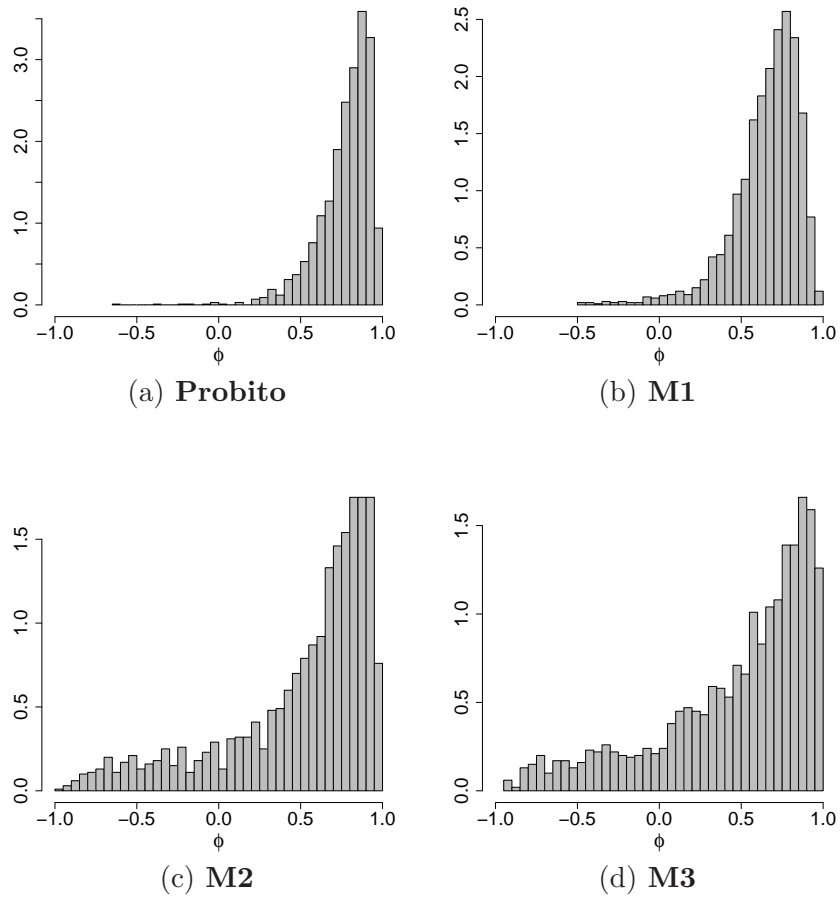


Figura 4.4: Histograma da cadeia do MCMC do parâmetro ϕ para cada modelo.

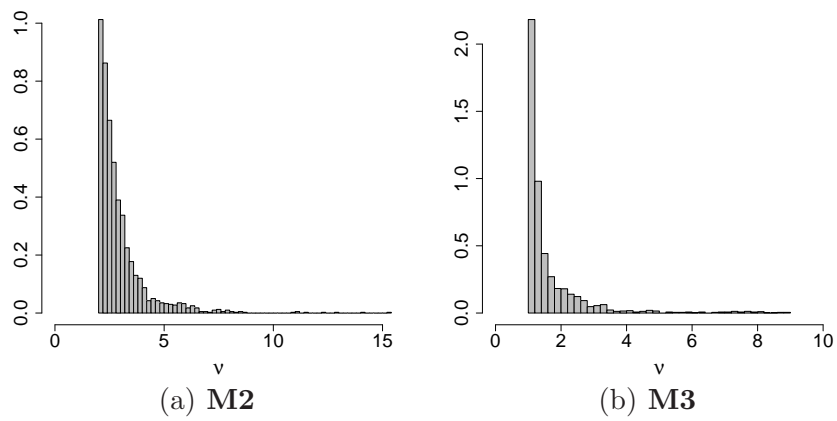


Figura 4.5: Histograma da cadeia do MCMC do parâmetro ν para cada modelo.

A Tabela 4.3 apresenta o sumário a *posteriori* dos coeficientes da regressão para cada modelo ajustado. O coeficiente α_1 referente à temperatura máxima foi o único estimado com um valor positivo, ou seja, altas temperaturas indicam maior probabilidade de não chover. Os outros coeficientes tem um efeito oposto.

Tabela 4.3: Sumário a *posteriori* dos coeficientes da regressão de cada modelo ajustado.

Parâmetro	Estatísticas	Modelos			
		Probita	M1	M2	M3
α_0	Média	-0.7572	-0.611	-1.682	-1.625
	IC (95%)	(-2.324;0.600)	(-1.601;0.308)	(-3.45 ; -0.14)	(-3.70 ; 0.06)
	Desvio padrão	0.731	0.488	0.847	0.953
	Geweke	-0.263	-0.930	-1.907	-0.458
	Ineficiência	1.043	1.325	2.280	3.359
α_1	Média	0.278	0.217	0.421	0.425
	IC (95%)	(0.188;0.395)	(0.151;0.310)	(0.25 ; 0.62)	(0.22 ; 0.65)
	Desvio padrão	0.053	0.040	0.097	0.115
	Geweke	-0.554	0.485	1.495	1.628
	Ineficiência	1.326	2.382	3.771	4.772
α_2	Média	-0.071	-0.060	-0.125	-0.120
	IC (95%)	(-0.135;-0.013)	(-0.109;-0.018)	(-0.22 ; -0.04)	(-0.22 ; -0.03)
	Desvio padrão	0.031	0.023	0.045	0.049
	Geweke	0.463	-0.819	-1.191	-1.942
	Ineficiência	1.219	1.825	2.840	3.688
α_3	Média	-0.276	-0.214	-0.3967	-0.407
	IC (95%)	(-0.377;-0.209)	(-0.295;-0.163)	(-0.57 ; -0.25)	(-0.61 ; -0.22)
	Desvio padrão	0.043	0.033	0.081	0.101
	Geweke	0.801	0.109	-0.770	-1.661
	Ineficiência	1.607	3.162	4.247	5.016

Na Figura 4.6 estão os histogramas *a posteriori* do parâmetro α_0 . Este parâmetro representa o intercepto da regressão. Os modelos apresentaram uma maior concentração nos valores negativos.

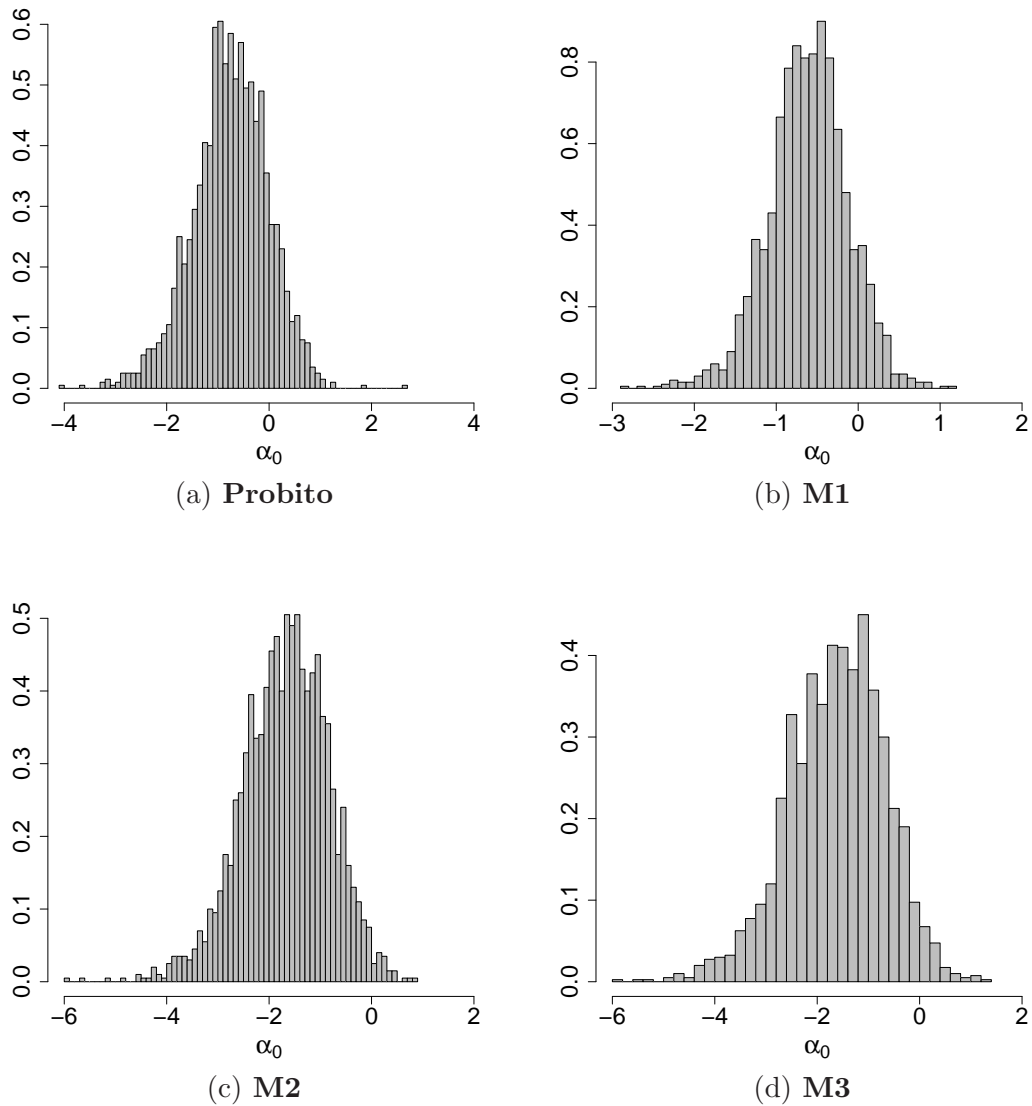


Figura 4.6: Histograma da cadeia do MCMC do parâmetro α_0 para cada modelo.

A Figura 4.7 apresenta os histogramas *a posteriori* do parâmetro α_1 . Este parâmetro representa a influência da covariável temperatura máxima na probabilidade de não chover. Através da Figura 4.7 nota-se que a influência é positiva, ou seja quanto maior a temperatura maior é a probabilidade de não chover.

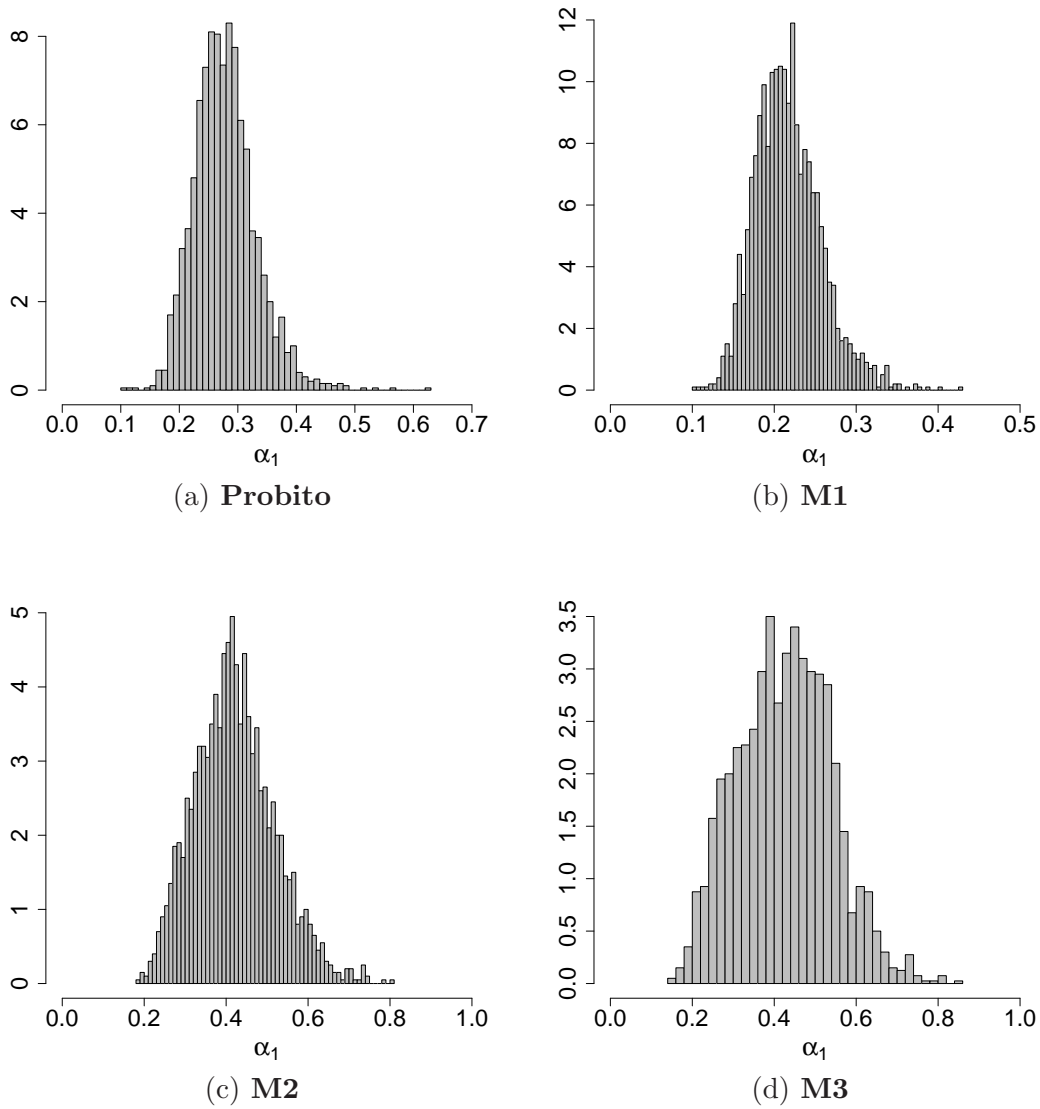


Figura 4.7: Histograma da cadeia do MCMC do parâmetro α_1 para cada modelo.

A Figura 4.8 mostra os histogramas *a posteriori* do parâmetro α_2 . Este parâmetro indica a influência da covariável de radiação na probabilidade de não chover. A Figura 4.8 indica uma influência negativa, ou seja quanto maior a radiação maior é a probabilidade de chover.

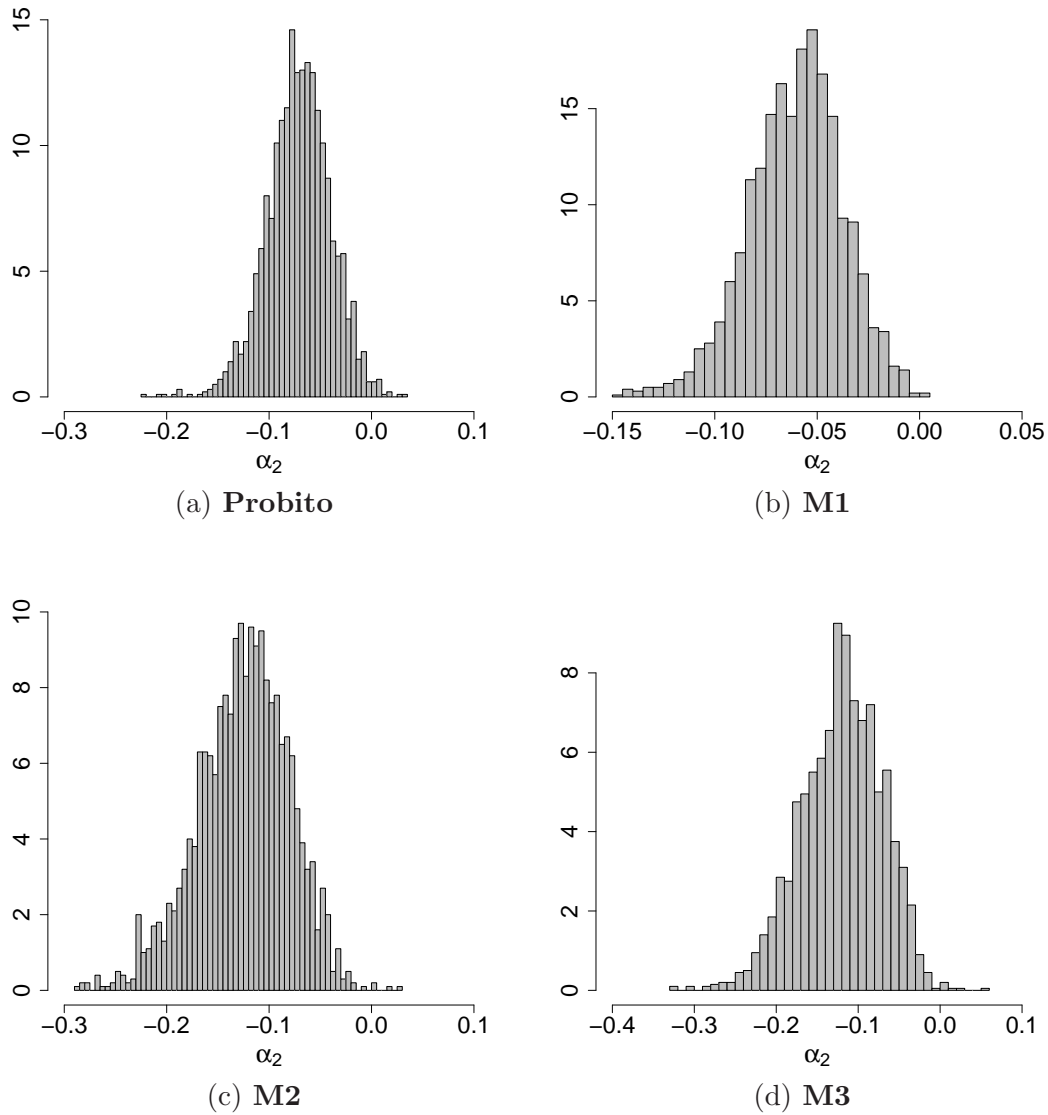


Figura 4.8: Histograma da cadeia do MCMC do parâmetro α_2 para cada modelo.

Na Figura 4.9 os histogramas *a posteriori* do parâmetro α_3 são apresentados. Este parâmetro indica a influência da covariável déficit de pressão de vapor na probabilidade de não chover. A Figura 4.9 indica uma influência negativa, ou seja quanto maior o déficit de pressão maior é a probabilidade de chover.

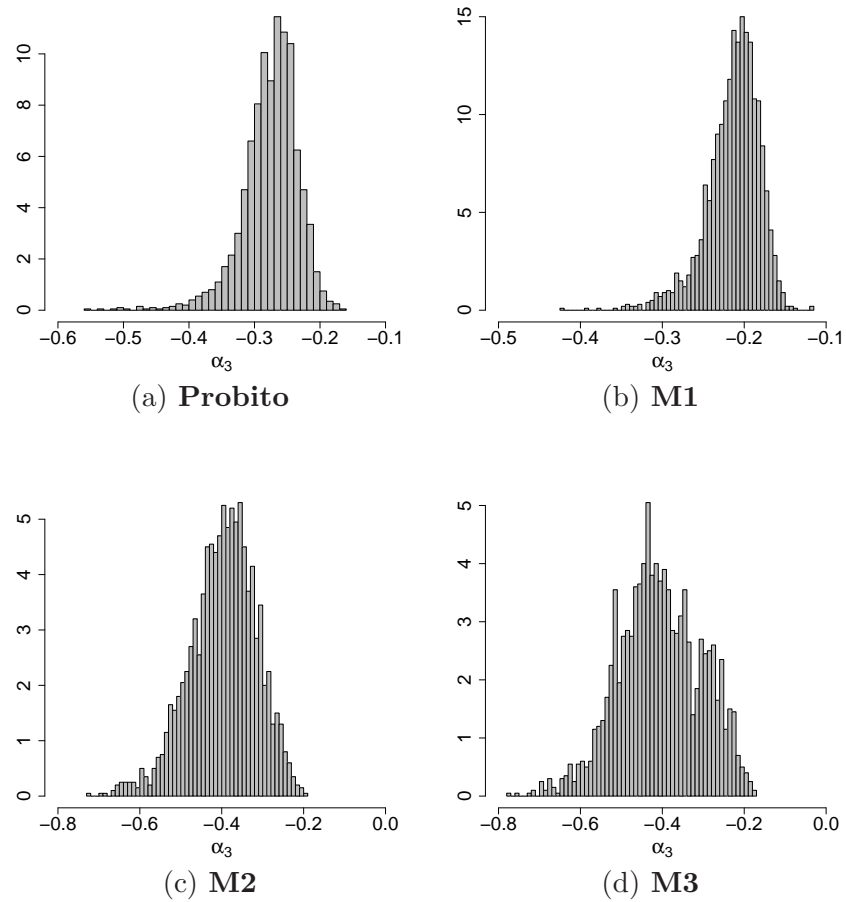


Figura 4.9: Histograma da cadeia do MCMC do parâmetro α_3 para cada modelo.

4.4 Comparação dos modelos

A Figura 4.10, mostra a probabilidade a *posteriori* condicional em todos os parâmetros de ocorrer chuva. Esta probabilidade é calculada para todos os modelos em todas as iterações do MCMC. Na Figura 4.10 estão as probabilidades médias de chover para cada tempo e em cada modelo. Note que é conhecido se choveu ou não mas esta probabilidade é calculada através da variável latente, Z_t . Ou seja, $P(Y_t = 0) \Leftrightarrow P(Z_t \leq 0)$. No modelo probito a probabilidade se comporta mais suave em torno de 0.5, já nos outros modelos saltos ocorrem, indicando que a probabilidade de chover é estimada em maiores valores do que no modelo probito.

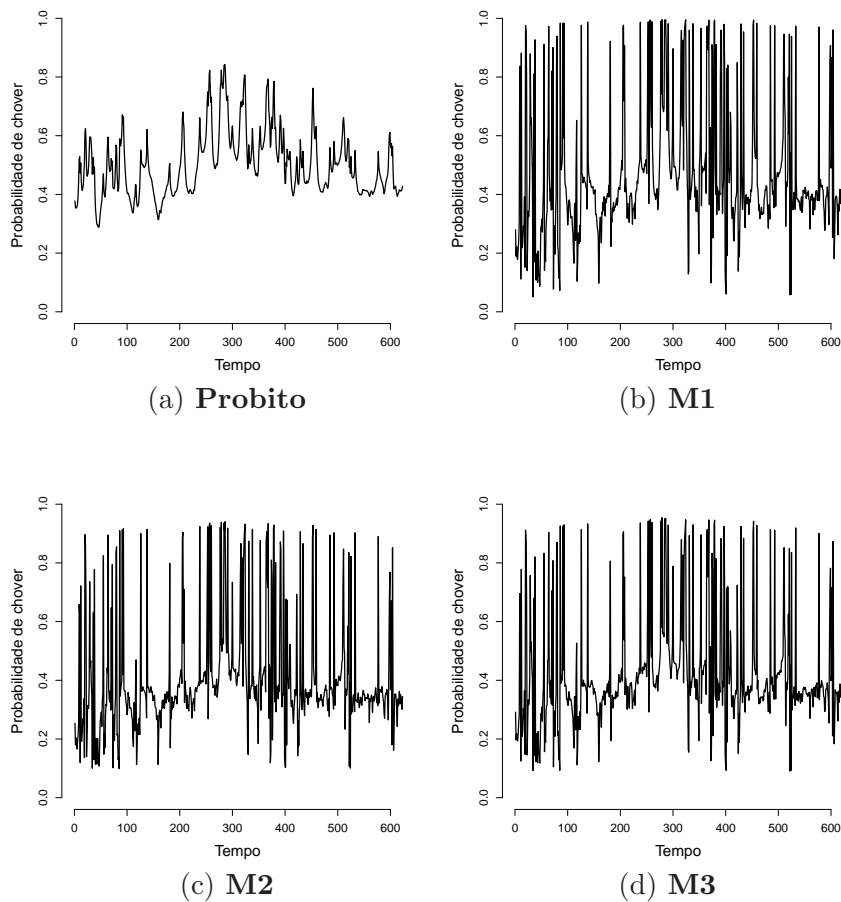


Figura 4.10: Gráficos da probabilidade condicional a *posteriori* de chover para cada modelo.

Para a melhor visualização da diferença na estimação da probabilidade de chover em cada modelo, a Figura 4.11 apresenta a probabilidade de chover nos primeiros vinte dias. Note que se for considerado que choveu caso a probabilidade de chover for maior que 0.5 (a linha na horizontal) todos os modelos acertaram os vinte dias mas os modelos 1, 2 e 3 apresentam tanto a maior probabilidade de chover quando realmente choveu quanto a menor probabilidade de chover quando não choveu que o modelo probito. Isto mostra que os modelos propostos ajustam melhor uma série de dados binários que possui taxas diferentes de 0 e 1.

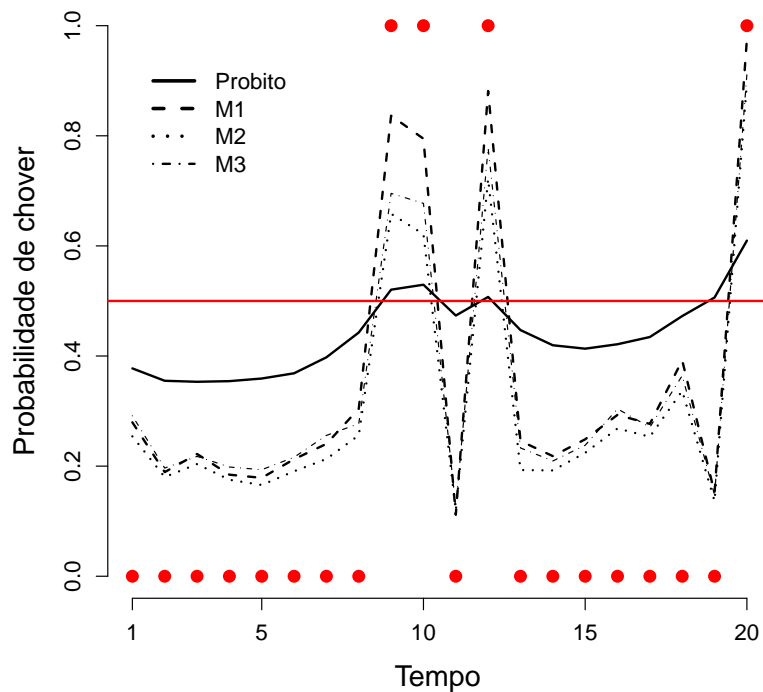


Figura 4.11: Gráfico da probabilidade condicional *a posteriori* de chover para cada modelo somente nos primeiros 20 dias. Os círculos representam 1 se choveu e 0 se não choveu.

Para quantificar quantas vezes os modelos acertaram se choveu ou não a Tabela 4.4 mostra a porcentagem de acertos admitindo que se a probabilidade condicional a *posteriori* de chover for maior que 0.5 indica que o modelo determina que choveu e se for menor ou igual a 0.5 determina que não choveu. O modelo probito foi o modelo que apresentou a menor porcentagem e os outros apresentaram uma porcentagem superior a 90%.

Tabela 4.4: Porcentagem de acertos de chover na amostra.

Modelo	Porcentagem de acertos
Probito	71 %
M1	91 %
M2	95 %
M3	96 %

A Tabela 4.5 apresenta o valor do DIC para cada modelo, este é o critério de comparação dos modelos que foi adotado neste trabalho. Através do DIC, o modelo que melhor ajustou os dados foi o modelo 3, pois obteve o menor valor. Este modelo leva em consideração tanto assimetria como caudas mais pesadas, usando a ligação *Slash* assimétrica. Depois deste, o modelo 2 foi considerado o que melhor ajustou os dados. O modelo probito é o menos indicado para ajustar esta amostra de dados.

Tabela 4.5: O valor do DIC para cada modelo.

Modelo	DIC
Probito	2231.08
M1	799.647
M2	-7381.38
M3	-106442

Capítulo 5

Conclusões e trabalhos futuros

Este trabalho teve como objetivo propor modelos que melhor ajustassem uma série temporal de dados binários. Foi discutido que quando uma série binária apresenta diferentes taxas de 0 e 1 em sua amostra funções de ligação simétricas não são indicadas, pois o procedimento de inferência é sensível a esta escolha e valores errados podem ser estimados para os parâmetros. Diante desta afirmação o interesse deste trabalho foi desenvolver modelos mais flexíveis que pudessem acomodar desvios na função de ligação.

Foi usada uma abordagem que considera a idéia de aumento de dados, ou seja, modela uma série binária através de uma variável contínua não observável com distribuição normal. Esta abordagem em conjunto com a modelagem dinâmica possibilitou a incorporação tanto da assimetria como caudas mais pesadas através do erro na equação de observação no modelo dinâmico binário. Portanto, utilizando esta modelagem, este trabalho apresenta uma classe de modelos dinâmicos binários com função de ligação assimétrica. Entre as opções apresentadas estão a distribuição assimétrica normal, *t-Student*, *Slash* e variância Gama.

Os modelos propostos são modelos mais flexíveis na modelagem de dados temporais binários. O aumento de dados permitiu a construção de um algoritmo de estimação que não trabalha com a função de verossimilhança do tipo Bernoulli, o que torna o procedimento de inferência menos complicado. O enfoque do trabalho foi todo Bayesiano e o algoritmo de estimação foi desenvolvido usando métodos de simulação estocásticas via cadeias de Markov e em particular para amostrar dos estados do modelo dinâmico foi

usado o algoritmo de perturbações suavizadas. Este eficiente algoritmo faz a atualização dos estados através das perturbações do modelo.

Foi feito um exemplo com dados gerados artificialmente, que serviram para testar o procedimento de inferência. Foram geradas séries binárias a partir de cada modelo e então depois os modelos foram ajustados. O procedimento de inferência se mostrou satisfatório em todos os modelos, pois as cadeias de MCMC convergiram e os intervalos de credibilidade de 95 % contiveram os valores verdadeiros dos parâmetros com que foram gerados os dados.

Uma aplicação à dados reais foi realizada com a finalidade de comparar quatro modelos. Tais dados se referem a uma série de ocorrência de chuva na cidade de Emerald na Austrália nos anos de 2001 a 2002. Tal série apresenta uma taxa bem diferente de 0 e 1 na amostra. Os quatro modelos ajustados foram o modelo probito, o modelo normal assimétrico, o modelo *t-Student* assimétrico e o modelo *Slash* assimétrico. O intuito foi verificar se os modelos com funções de ligação assimétricas e caudas mais pesadas teriam um melhor ajuste que o modelo simétrico, probito, muito usado na literatura. Os modelos foram ajustados e segundo o critério de comparação de modelo, DIC, o modelo probito foi o menos indicado para esta série de dados binários. A probabilidade condicional *a posteriori* de ocorrer chuva quando realmente choveu no modelo probito apresentou-se sempre abaixo das outras. Isto indica que os modelos assimétricos estimam com maior probabilidade a ocorrência de chuva quando realmente choveu, ou seja, além de acertar mais a ocorrência de chuva os modelos ainda acertam com uma probabilidade maior.

Propostas futuras de trabalho incluem a extensão destes modelos para dados binomiais, a inclusão de outros critérios para a comparação dos modelos e a estimação sequencial das variáveis latentes e parâmetros dos modelos propostos.

Apêndice A

Cálculo das distribuições condicionais completas das variáveis latentes

As contas das condicionais completas para as variáveis latentes U_t e ρ_t são apresentadas a seguir. Para a variável U_t é feito o cálculo para cada modelo, pois a *priori* são atribuídas diferentes distribuições. Portanto, para cada modelo obtém-se:

- Modelo **M2** em que $p(U_t|\nu) \sim G(\nu/2, \nu/2)$:

$$\begin{aligned} p(U_t | \cdot) &\propto p(Z_t | \cdot)p(U_t) \\ &\propto U_t^{1/2} \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} U_t^{\frac{\nu}{2}-1} \exp \left\{ -\frac{\nu}{2} U_t \right\} \\ &= U_t^{\frac{\nu+1}{2}-1} \exp \left\{ -U_t \left[\frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu)^2}{2(1-\delta^2)} + \frac{\nu}{2} \right] \right\} \exp \left\{ \frac{U_t^{1/2} (Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta \rho_t}{1-\delta^2} \right\}. \end{aligned}$$

O núcleo desta distribuição não é de uma distribuição conhecida, portanto é usado um passo de Metropolis para poder amostrar dela. A distribuição proposta usada é

$$G \left(\frac{\nu+1}{2}, \frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu)^2}{2(1-\delta^2)} + \frac{\nu}{2} \right).$$

- Modelo **M3** em que $p(U_t|\nu) \sim Be(\nu, 1)$:

$$\begin{aligned}
p(U_t | \cdot) &\propto p(Z_t | \cdot)p(U_t) \\
&\propto U_t^{1/2} \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} U_t^{\nu-1} 1(U_t \in [0, 1]) \\
&= U_t^{\nu+\frac{1}{2}-1} \exp \left\{ -U_t \left[\frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu)^2}{2(1-\delta^2)} \right] \right\} \exp \left\{ \frac{U_t^{1/2} (Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta \rho_t}{1-\delta^2} \right\} \\
&\times 1(U_t \in [0, 1]).
\end{aligned}$$

Este núcleo não é de uma distribuição conhecida, portanto um passo de Metropolis para poder amostrar dela é usado. A distribuição proposta utilizada é uma Gama truncada em $[0, 1]$, com parâmetros:

$$G \left(\nu + \frac{1}{2}, \frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu)^2}{2(1-\delta^2)} \right).$$

Para gerar da Gama truncada foi usado um algoritmo visto em Philippe (1997).

- Modelo **M4** em que $p(U_t|\nu) \sim GI(\nu/2, \nu/2)$:

$$\begin{aligned}
p(U_t | \cdot) &\propto p(Z_t | \cdot)p(U_t) \\
&\propto U_t^{1/2} \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} U_t^{-(\frac{\nu}{2}+1)} \exp \left\{ -\frac{\nu}{2} \frac{1}{U_t} \right\} \\
&= U_t^{-\frac{\nu}{2}+\frac{1}{2}-1} \exp \left\{ -\frac{1}{2} \left[\frac{\nu}{U_t} + \frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu)^2}{1-\delta^2} U_t \right] \right\} \exp \left\{ \frac{U_t^{1/2} (Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta \rho_t}{1-\delta^2} \right\}.
\end{aligned}$$

Para poder amostrar desta distribuição também é realizado um passo de Metropolis pois a distribuição não é conhecida. A distribuição proposta utilizada é a distribuição gaussiana inversa generalizada (GIG), dada por

$$GIG \left(\frac{-\nu + 1}{2}, \sqrt{\nu}, \frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu)}{\sqrt{(1-\delta^2)}} \right).$$

Para gerar desta distribuição foi adaptada a rotina do livro de Dagpunar (1988) que estava na linguagem do Fortran 77 para o código do programa C++. Informações sobre a distribuição GIG estão no apêndice B.

O cálculo da condicional completa da variável ρ_t é similar para todos os modelos, note que para o modelo **M1** a variável $U_t = 1$ para todo t . A conta é dada por:

$$\begin{aligned}
p(\rho_t | \cdot) &\propto p(Z_t | \cdot) p(\rho_t) \\
&\propto \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} \exp \left\{ -\frac{\rho_t^2}{2} \right\} 1(\rho_t > 0) \\
&\propto \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left[-2(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta U_t^{-1/2} \rho_t + \frac{\delta^2 \rho_t^2}{U_t} \right] - \frac{\rho_t^2}{2} \right\} 1(\rho_t > 0) \\
&= \exp \left\{ -\frac{1}{2} \left[\frac{-2(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta U_t^{1/2}}{(1-\delta^2)} \rho_t + \left(\frac{\delta^2}{1-\delta^2} + 1 \right) \rho_t^2 \right] \right\} 1(\rho_t > 0) \\
&= \exp \left\{ -\frac{1}{2} \left[\frac{-2(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta U_t^{1/2}}{(1-\delta^2)} \rho_t + \left(\frac{1}{1-\delta^2} \right) \rho_t^2 \right] \right\} 1(\rho_t > 0) \\
&= \exp \left\{ -\frac{1}{2} \left(\frac{1}{1-\delta^2} \right) \left[-2\rho_t \left(\frac{(Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta U_t^{1/2}}{1-\delta^2} \right) \left(\frac{1}{1-\delta^2} \right)^{-1} + \rho_t^2 \right] \right\} \\
&\times 1(\rho_t > 0),
\end{aligned}$$

que é o núcleo de uma distribuição normal truncada

$$N_+ \left((Z_t - \mathbf{F}'_t \boldsymbol{\theta}_t - \mu) \delta U_t^{1/2}, 1 - \delta^2 \right).$$

Apêndice B

Definição da distribuição gaussiana inversa generalizada

Se uma variável aleatória X possui uma distribuição gaussiana inversa generalizada (GIG), com parâmetros λ , δ e γ , denotada por $X \sim GIG(\lambda, \delta, \gamma)$, então a sua função de densidade de probabilidade, para todo $x > 0$, é dada por:

$$f(x) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\gamma\delta)} x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \gamma^2 x\right)\right\}, \quad (\text{B.1})$$

em que $-\infty < \lambda < \infty$, $\delta > 0$, $\gamma > 0$ e K_λ é função de Bessel modificada do terceiro tipo com índice λ . A dependência dos parâmetros é a seguinte:

$$\delta \geq 0, \gamma > 0, \text{ se } \lambda > 0;$$

$$\delta > 0, \gamma > 0, \text{ se } \lambda = 0;$$

$$\delta > 0, \gamma \geq 0, \text{ se } \lambda < 0.$$

Quando $\lambda > 0$ e $\delta = 0$, a função de densidade em (B.1) se reduz a função de densidade de uma distribuição gama, enquanto que a função de densidade de uma distribuição gama inversa é obtida quando $\lambda < 0$ e $\gamma = 0$.

O momento de ordem r da variável aleatória X é dado por

$$E(X^r) = \left(\frac{\delta}{\gamma}\right)^{r/2} \frac{K_{\lambda+r}(2\sqrt{\delta\gamma})}{K_\lambda(2\sqrt{\delta\gamma})}$$

Mais detalhes sobre a distribuição GIG podem ser visto em Jørgensen (1982).

Apêndice C

Cálculo das distribuições condicionais completas dos parâmetros

As contas das condicionais completas para as parâmetros são apresentadas a seguir. Admitindo uma distribuição a *priori* para $\theta_0 \sim N(0, \frac{W}{1-\phi^2})$. Então para o parâmetro ϕ , o cálculo da condicional completa que é similar para todos os modelos é dado por:

$$\begin{aligned} p(\phi | \cdot) &\propto p(\theta_{0:T} | \cdot) p(\phi) \\ &\propto (1 - \phi^2)^{1/2} \exp \left\{ -\frac{1}{2W} (1 - \phi^2) \theta_0^2 \right\} \exp \left\{ -\frac{1}{2W} \sum_{t=1}^T (\theta_t - \phi \theta_{t-1})^2 \right\} \\ &\times \exp \left\{ -\frac{1}{2\sigma_\phi^2} (\phi - \mu_\phi)^2 \right\} \\ &\propto (1 - \phi^2)^{1/2} \exp \left\{ \frac{1}{2W} \phi^2 \theta_0^2 \right\} \\ &\times \exp \left\{ -\frac{1}{2} \left[-2\phi \left[\frac{\sum_{t=1}^T \theta_t \theta_{t-1}}{W} + \frac{\mu_\phi}{\sigma_\phi^2} \right] + \phi^2 \left(\frac{\sum_{t=1}^T \theta_{t-1}^2}{W} + \frac{1}{\sigma_\phi^2} \right) \right] \right\}. \end{aligned}$$

O núcleo desta distribuição não é de uma distribuição conhecida, portanto será necessário um passo de Metropolis no algoritmo de Gibbs para amostrar desta distribuição. A distribuição proposta usada para a amostragem deste parâmetro foi a distribuição normal.

O cálculo da condicional completa para o parâmetro W também é similar para todos os modelos e é dado por:

$$\begin{aligned}
p(W | \cdot) &\propto p(\theta_{0:T} | \cdot)p(W) \\
&\propto \sqrt{\frac{(1-\phi^2)}{W}} \exp\left\{-\frac{1}{2W}(1-\phi^2)\theta_0^2\right\} \prod_{t=1}^T \exp\left\{\frac{1}{\sqrt{W}}(\theta_t - \phi\theta_{t-1})^2\right\} \\
&\times \left(\frac{1}{W}\right)^{\alpha_w+1} \exp\left\{-\frac{1}{W}\beta_w\right\} \\
&\propto \left(\frac{1}{W}\right)^{\frac{T+1}{2}+\alpha_w+1} \exp\left\{-\frac{1}{W}\left[\frac{\theta_0^2(1-\phi^2)}{2} + \frac{1}{2}\sum_{t=1}^T(\theta_t - \phi\theta_{t-1})^2 + \beta_w\right]\right\}
\end{aligned}$$

que é o núcleo de uma distribuição gama inversa

$$GI\left(\frac{T+1}{2} + \alpha_w, \frac{\theta_0^2(1-\phi^2) + \sum_{t=1}^T(\theta_t - \phi\theta_{t-1})^2}{2} + \beta_w\right).$$

O cálculo da condicional completa para o parâmetro de assimetria λ é similar para todos os modelos e é dado por:

$$\begin{aligned}
p(\lambda | \cdot) &\propto \prod_{t=1}^T p(Z_t | \cdot)p(\lambda) \\
&\propto \prod_{t=1}^T \frac{1}{(1-\delta^2)^{1/2}} \exp\left\{-\frac{U_t}{2(1-\delta^2)}\left(Z_t - \theta_t - \mu - \delta\rho_t U_t^{-1/2}\right)^2\right\} \exp\left\{-\frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2\right\} \\
&\propto \left(\frac{1}{1-\delta^2}\right)^{T/2} \exp\left\{-\frac{1}{2(1-\delta^2)}\sum_{t=1}^T U_t\left(Z_t - \theta_t - \mu - \delta\rho_t U_t^{-1/2}\right)^2 - \frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2\right\}.
\end{aligned}$$

Como este núcleo não é de uma distribuição conhecida será necessário realizar um passo de Metropolis para amostrar desta distribuição. Em especial, para este parâmetro foi utilizado um passo de Langevin-Hastings, que é uma opção para o passeio aleatório

do algoritmo Metropolis-Hastings, cuja proposta é acelerar a velocidade de convergência do método de Metropolis-Hastings. A distribuição proposta pelo algoritmo de Langevin-Hastings é dada por $\lambda^{(i)} \sim N(\lambda^{(i-1)} + h \frac{\log'[p(\lambda; \cdot)]}{2}, h)$, em que h é um valor fixo e $\log'[p(\lambda; \cdot)]$ é a derivada do logaritmo da distribuição condicional completa. Mais detalhes sobre esse algoritmo podem ser vistos em Christensen et al. (2001) e Christensen e Waagepetersen (2002).

Para o parâmetro ν , o cálculo da condicional completa é feito para cada modelo, pois são abordagens diferentes. Portanto, por modelo obtém-se:

- Modelo **M2**:

$$\begin{aligned}
p(\nu | \cdot) &\propto \prod_{t=1}^T p(Z_t | \cdot) p(U_t | \nu) p(\nu) \\
&\propto \prod_{t=1}^T \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \theta_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} \\
&\times \prod_{t=1}^T \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} U_t^{\frac{\nu}{2}-1} \exp \left\{ -\frac{\nu}{2} U_t \right\} \nu^{\alpha_\nu-1} \exp\{\beta_\nu \nu\} \\
&\propto \frac{(\nu/2)^{T\nu/2}}{[\Gamma(\nu/2)]^T} \nu^{\alpha_\nu-1} \\
&\times \exp \left\{ -\frac{\sum_{t=1}^T U_t (-2(Z_t - \theta_t - \delta \rho_t U_t^{-1/2})\mu + \mu^2)}{2(1-\delta^2)} + \frac{\nu}{2} \sum_{t=1}^T \log(U_t) - \frac{\nu}{2} \sum_{t=1}^T U_t - \beta_\nu \nu \right\}.
\end{aligned}$$

Será necessário um passo de Metropolis no algoritmo de Gibbs para amostrar desta distribuição, pois seu núcleo não pertence a uma distribuição conhecida. A distribuição proposta usada para a amostragem deste parâmetro foi a distribuição normal truncada.

- Modelo **M3**:

$$\begin{aligned}
p(\nu | \cdot) &\propto \prod_{t=1}^T p(Z_t | \cdot) p(U_t | \nu) p(\nu) \\
&\propto \prod_{t=1}^T \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \theta_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} \\
&\times \prod_{t=1}^T \nu U_t^{\nu-1} \nu^{\alpha_\nu-1} \exp\{-\beta_\nu \nu\} \\
&\propto \exp \left\{ -\frac{U_t[-2(Z_t - \theta_t - \delta \rho_t U_t^{-1/2})\mu + \mu^2]}{2\sigma^2(1-\delta^2)} \right\} \nu^{T+\alpha_\nu-1} \left[\prod_{t=1}^T U_t \right]^{\nu-1} \exp\{-\beta_\nu \nu\}.
\end{aligned}$$

Neste caso também será necessário para amostrar desta distribuição, realizar um passo de Metropolis no algoritmo de Gibbs, pois seu núcleo não pertence a uma distribuição conhecida. A distribuição proposta usada para a amostragem deste parâmetro foi a distribuição normal truncada.

- Modelo **M4**:

$$\begin{aligned}
p(\nu | \cdot) &\propto \prod_{t=1}^T p(Z_t | \cdot) p(U_t | \nu) p(\nu) \\
&\propto \prod_{t=1}^T \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \theta_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} \\
&\times \prod_{t=1}^T \frac{(\nu/2)^{\nu/2}}{\Gamma(\frac{\nu}{2})} U_t^{-(\frac{\nu}{2}-1)} \exp \left\{ -\frac{\nu}{2U_t} \right\} \nu^{\alpha_\nu-1} \exp\{-\beta_\nu \nu\} \\
&\propto \exp \left\{ -\frac{\sum_{t=1}^T U_t[-2(Z_t - \theta_t - \delta \rho_t U_t^{-1/2})\mu + \mu^2]}{2(1-\delta^2)} - \frac{\nu}{2} \sum_{t=1}^T \log(U_t) - \frac{\nu}{2} \sum_{t=1}^T \frac{1}{U_t} - \beta_\nu \nu \right\} \\
&\times \frac{(\nu/2)^{T\nu/2}}{[\Gamma(\frac{\nu}{2})]^T} \nu^{\alpha_\nu-1}.
\end{aligned}$$

O núcleo da distribuição condicional completa neste caso também não é de uma distribuição conhecida. Assim, será necessário realizar um passo de Metropolis no algoritmo de Gibbs para amostrar desta distribuição. A distribuição proposta usada foi a distribuição normal truncada.

O cálculo da condicional completa para o parâmetro $\boldsymbol{\alpha}$ é similar para todos os modelos. Seja a distribuição a priori, $p(\boldsymbol{\alpha}) \sim N_p(\boldsymbol{\mu}_0, \Sigma_0)$ em que p é a quantidade de covariáveis no modelo. A distribuição condicional completa é dada por:

$$\begin{aligned}
p(\boldsymbol{\alpha} | \cdot) &\propto \prod_{t=1}^T p(Z_t | \cdot) p(\boldsymbol{\alpha}) \\
&\propto \prod_{t=1}^T \exp \left\{ -\frac{U_t}{2(1-\delta^2)} \left(Z_t - \mathbf{x}'_t \boldsymbol{\alpha} - \theta_t - \mu - \delta \rho_t U_t^{-1/2} \right)^2 \right\} \\
&\times \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\mu}_0)' \Sigma_0^{-1} (\boldsymbol{\alpha} - \boldsymbol{\mu}_0) \right\}, \text{ seja } S_t = (Z_t - \theta_t - \mu - \delta \rho_t U_t^{-1/2}) \\
&\propto \exp \left\{ -\frac{1}{2(1-\delta^2)} \sum_{t=1}^T U_t (S_t - \mathbf{x}'_t \boldsymbol{\alpha})^2 - \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\mu}_0)' \Sigma_0^{-1} (\boldsymbol{\alpha} - \boldsymbol{\mu}_0) \right\} \\
&\propto \exp \left\{ -\frac{1}{2(1-\delta^2)} \sum_{t=1}^T U_t (\boldsymbol{\alpha}' \mathbf{x}_t \mathbf{x}'_t \boldsymbol{\alpha} - 2 S_t \mathbf{x}'_t \boldsymbol{\alpha}) - \frac{1}{2} (\boldsymbol{\alpha}' \Sigma_0^{-1} \boldsymbol{\alpha} - 2 \boldsymbol{\mu}'_0 \Sigma_0^{-1} \boldsymbol{\alpha}) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\alpha}' \left(\frac{\sum_{t=1}^T U_t \mathbf{x}_t \mathbf{x}'_t}{1-\delta^2} \right) \boldsymbol{\alpha} - 2 \left(\frac{\sum_{t=1}^T U_t S_t \mathbf{x}'_t}{1-\delta^2} \right) \boldsymbol{\alpha} \right] - \frac{1}{2} (\boldsymbol{\alpha}' \Sigma_0^{-1} \boldsymbol{\alpha} - 2 \boldsymbol{\mu}'_0 \Sigma_0^{-1} \boldsymbol{\alpha}) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\alpha}' \left(\Sigma_0^{-1} + \frac{\sum_{t=1}^T U_t \mathbf{x}_t \mathbf{x}'_t}{1-\delta^2} \right) \boldsymbol{\alpha} - 2 \left(\boldsymbol{\mu}'_0 \Sigma_0^{-1} + \frac{\sum_{t=1}^T U_t S_t \mathbf{x}'_t}{1-\delta^2} \right) \boldsymbol{\alpha} \right] \right\}.
\end{aligned}$$

Este é o núcleo de uma distribuição normal multivariada $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ em que,

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \left(\boldsymbol{\mu}'_0 \Sigma_0^{-1} + \frac{\sum_{t=1}^T U_t S_t \mathbf{x}'_t}{1-\delta^2} \right) \Sigma_1 \\
\Sigma_1 &= \left(\Sigma_0^{-1} + \frac{\sum_{t=1}^T U_t \mathbf{x}_t \mathbf{x}'_t}{1-\delta^2} \right)^{-1}.
\end{aligned}$$

Apêndice D

Resultados *a posteriori* do modelo 2 para dados artificiais utilizando uma distribuição *a priori* não informativa para o parâmetro ν

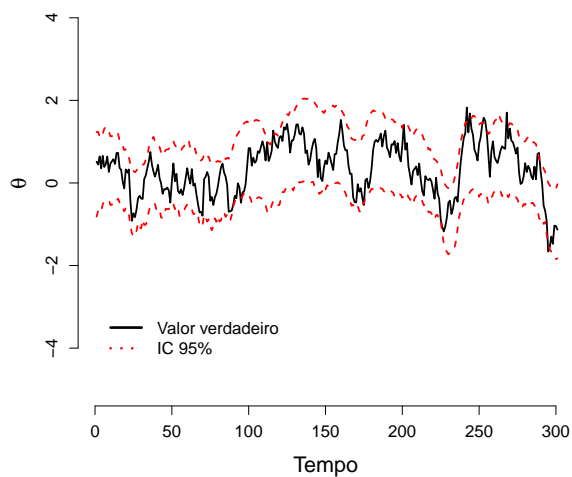


Figura D.1: Gráfico da série temporal de θ , sendo que a linha cheia representa o valor verdadeiro e a linha tracejada o intervalo de 95% de credibilidade *a posteriori*.

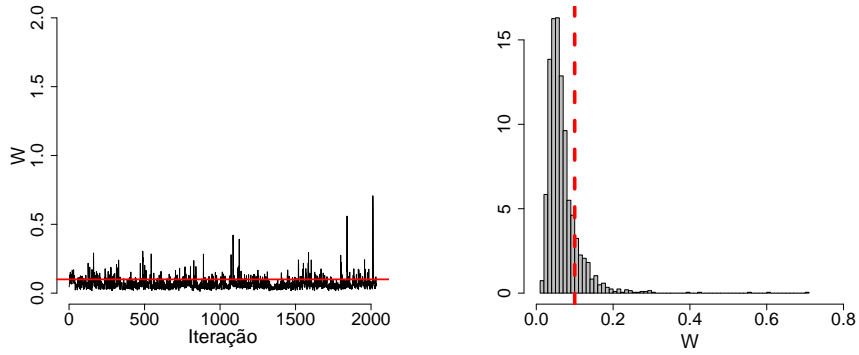


Figura D.2: Gráfico da cadeia do MCMC do parâmetro W , a linha cheia é o valor verdadeiro. Ao lado, o histograma em que a linha tracejada representa o valor real.

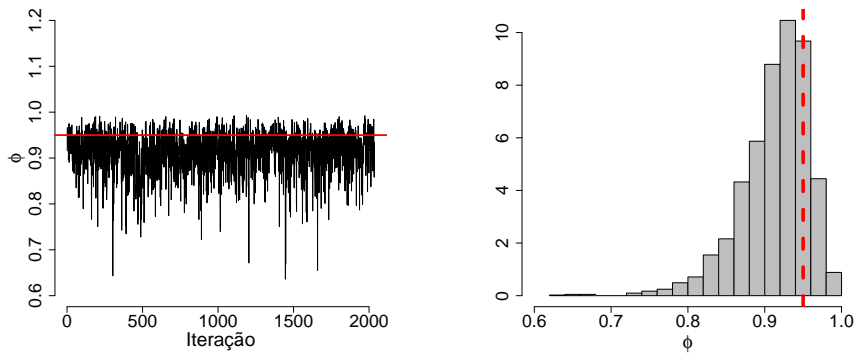


Figura D.3: Gráfico da cadeia do MCMC do parâmetro ϕ , a linha cheia é o valor verdadeiro. Ao lado, o histograma em que a linha tracejada representa o valor real.

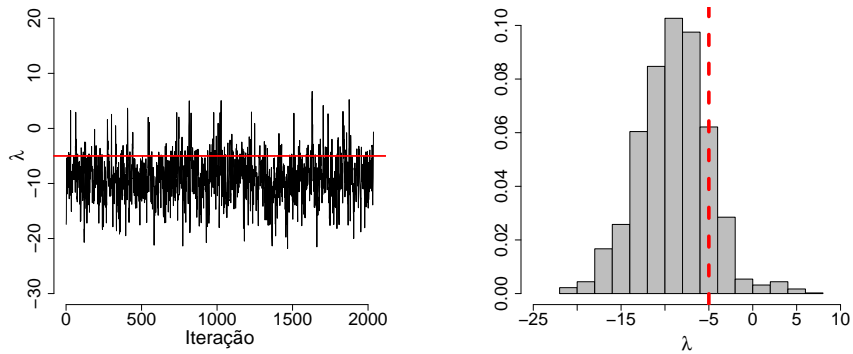


Figura D.4: Gráfico da cadeia do MCMC do parâmetro λ , a linha cheia é o valor verdadeiro. Ao lado, o histograma em que a linha tracejada representa o valor real.

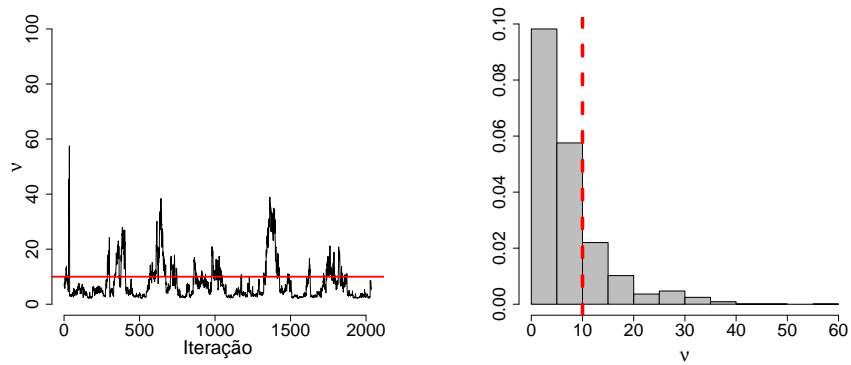


Figura D.5: Gráfico da cadeia do MCMC do parâmetro ν , a linha cheia é o valor verdadeiro. Ao lado, o histograma em que a linha tracejada representa o valor real.

Referências Bibliográficas

- Albert, J. H. e Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.
- Andrews, D. R. e Mallows, C. L. (1974) Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, B*, **36**, 99–102.
- Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- (2005) The skew-normal distribution and related multivariate families(with discussion). *Scandinavian Journal of Statistics*, **32**, 159–188.
- Best, N. G., Cowles, M. K. e Vines, K. (1995) Coda : Convergence diagnosis and output analysis software for gibbs sampling output. version 0.4. *Relatório técnico*, Biostatistics Unit MRC, Cambridge, Inglaterra.
- Chen, M. H., Dipak, K. D. e Shao, Q. M. (1999) A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, **94**, 1172–1186.
- Christensen, O. F., Møller, J. e Waagepetersen, R. (2001) Geometric ergodicity of metropolis-hastings algorithms for conditional simulation in generalised linear mixed models. *Methodology and Computing in Applied Probability*, **3**, 309–327.
- Christensen, O. F. e Waagepetersen, R. (2002) Bayesian prediction of spatial count data using generalised linear mixed models. *Biometrics*, **58**, 280–286.
- Cox, D. R. (1981) Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, **8**, 93–115.

- Czado, C. e Song, P. X. K. (2008) State space mixed models for longitudinal observations with binary and binomial responses. *Statistical Papers*, **49**, 691–714.
- Dagpunar, J. (1988) *Principles of Random Variate Generation*. Clarendon Oxford Science Publications, Oxford, U.K.
- Damien, P. e Walker, S. G. (2001) Sampling truncated normal, beta and gamma densities. *Journal of Computational and Graphical Statistics*, **10**, 206–215.
- Fonseca, T., Ferreira, M. A. R. e Migon, H. (2008) Objective bayesian analysis for the student-t regression model. *Biometrika*, **95**, 325–333.
- Gamerman, D. e Lopes, H. F. (2006) *Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London.
- Gelfand, S. e Smith, A. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.
- Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). Em *Bayesian Statistics*, vol. 4. Bernardo J. M. et al. (eds). Oxford: Oxford University Press, 169-193.
- Harrison, P. J. e Stevens, C. (1976) Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society*, **38**, 205–247.
- Hastings, W. K. (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.
- de Jong, P. d. e Shephard, N. (1995) The simulation smoother for time series models. *Biometrika*, **82**, 339–350.
- Jørgensen, B. (1982) *Statistical properties of the generalized inverse Gaussian distribution*. New York: Springer-Verlag.

- Koopman, S. (1993) Disturbance smoothers for state space models. *Biometrika*, **80**, 117–126.
- Lachos, V. H., Ghosh, P. e Arellano-Valle, R. B. (2010) Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**, SS–08–045.
- Lange, K. L. e Sinsheimer, J. S. (1993) Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.
- Madan, D. B. e Seneta, E. (1990) The variance gamma (v.g) model for share market return. *Journal Business*, **63**, 511–524.
- McCullagh, S. e Nelder, J. A. (1989) *Generalized linear models*. London: Chapman & Hall, second edn.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.
- Migon, H. e Gamerman, D. (1999) *Statistical Inference: an Integrated Approach*. London: Arnold.
- Nelder, J. A. e Wedderburn, R. (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series A*, **135**, 370–384.
- Pemstein, D., V., Q. K. e D., M. A. (2007) The scythe statistical library: an open source c++ library for statistical computation. *Journal of Statistical Software*, **V**, 1–29.
- Philippe, A. (1997) Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing*, **7**, 173–181.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. ISBN 3-900051-07-0.

- Spiegelhalter, D., Best, N., Carlin, B. e Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, **64**, 583–639.
- Tanner, T. A. e Wong, W. H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528–549.
- West, M., Harrison, J. e Migon, H. (1985) Dynamic generalized linear models and bayesian forecasting (with discussion). *Journal of the American Statistical Association*, **80**, 73–97.
- West, M. e Harrison, P. (1997) *Bayesian Forecasting and Dynamic Models*. Springer Verlag, New York.