
Abordagem clássica e bayesiana para os
modelos de séries temporais da
família GARMA com aplicações para
dados de contagem

Adriana Strieder Philippsen

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura:

Abordagem clássica e bayesiana para os
modelos de séries temporais da
família GARMA com aplicações para
dados de contagem

Adriana Strieder Philippsen

Orientador: *Prof. Dr. Marinho Gomes de Andrade Filho*

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA.*

USP – São Carlos
Maio/2011

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

P551a Philippsen, Adriana Strieder
Abordagem clássica e bayesiana para os modelos de séries temporais da família GARMA com aplicações para dados de contagem / Adriana Strieder Philippsen; orientador Marinho Gomes de Andrade Filho -- São Carlos, 2011.
105 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2011.

1. Séries Temporais. 2. Inferência Bayesiana. 3. Inferência Clássica. I. Filho, Marinho Gomes de Andrade, orient. II. Título.

Aos meus pais e amigos, Carlito e Melita.
Às minhas irmãs e sobrinho Gisele, Juliana e Gabriel.
Pelo amor e carinho.

Agradecimentos

Primeiramente agradeço aos meus pais, Carlito e Melita, que me deram não somente a vida, mas principalmente a minha educação e condições de estudo, os mais profundos agradecimentos. As minhas irmãs e ao meu sobrinho, Gisele, Juliana e Gabriel, que sempre estiveram carinhosamente presentes, pelas críticas e força para superar obstáculos todos os dias na difícil tarefa de estudar.

Ao meu namorado, companheiro e amigo Marcos, por aguentar meus momentos de ansiedade e estresse, pelo amor, paciência, total apoio e horas de estudo compartilhadas.

Agradeço ao meu orientador, Marinho Gomes de Andrade Filho, pela paciente tolerância e oportunidade que a mim foi confiada, pelas excelentes aulas ministradas e por estar sempre tão solícito para dirimir toda e qualquer dúvida, pela orientação e incentivo na elaboração e condução deste trabalho.

A todos os meus amigos e amigas que sempre se preocuparam comigo, pela amizade e por todos os momentos divertidos.

Aos professores Maria Sílvia de Assis Moura e Ricardo Sandes Ehlers, membros da banca do exame de qualificação, pelas sugestões feitas.

Aos diversos professores do Departamento de Estatística da Universidade Estadual de Maringá, em especial aos professores Josmar Mazucheli, Rosangela Getirana Santana e Isolde Terezinha S. Previdelli pelo incentivo que me levaram a fazer pós-graduação.

A todas as pessoas que, direta ou indiretamente, contribuíram para a execução dessa dissertação.

Finalmente, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio concedido para este trabalho.

Resumo

Nesta dissertação estudou-se o modelo GARMA para modelar séries temporais de dados de contagem com as distribuições condicionais de Poisson, binomial e binomial negativa. A principal finalidade foi analisar no contexto clássico e bayesiano, o desempenho e a qualidade do ajuste dos modelos de interesse, bem como o desempenho dos percentis de cobertura dos intervalos de confiança dos parâmetros para os modelos adotados. Para atingir tal finalidade considerou-se a análise dos estimadores pontuais bayesianos e foram analisados intervalos de credibilidade. Neste estudo é proposta uma distribuição a priori conjugada para os parâmetros dos modelos e busca-se a distribuição a posteriori, a qual associada a certas funções de perda permite encontrar estimativas bayesianas para os parâmetros. Na abordagem clássica foram calculados estimadores de máxima verossimilhança, usando-se o método de score de Fisher e verificou-se por meio de simulação a consistência dos mesmos. Com os estudos desenvolvidos pode-se observar que, tanto a inferência clássica quanto a inferência bayesiana para os parâmetros dos modelos em questão, apresentou boas propriedades analisadas por meio das propriedades dos estimadores pontuais. A última etapa do trabalho consiste na análise de um conjunto de dados reais, sendo uma série real correspondente ao número de internações por causa da dengue em Campina Grande. Estes resultados mostram que tanto o estudo clássico, quanto o bayesiano, são capazes de descrever bem o comportamento da série.

Palavras-chave: Modelo GARMA, Distribuição de Poisson, Distribuição Binomial, Distribuição Binomial Negativa, Inferência Clássica, Inferência Bayesiana.

Abstract

In this work, it was studied the GARMA model to model time series count data with Poisson, binomial and negative binomial discrete conditional distributions. The main goal is to analyze, in the bayesian and classic context, the performance and the quality of fit of the corresponding models, as well as the coverage percentages performance to these models. To achieve this purpose we considered the analysis of Bayesian estimators and credible intervals were analyzed. To the Bayesian study it was proposed a priori distribution joined to the models parameters and sought a posteriori distribution, which one associate with to certain loss functions allows finding out Bayesian estimates to the parameters. In the classical approach, it was calculated the maximum likelihood estimators using the method of Fisher scoring, whose interest was to verify, by simulation, the consistence. With the studies developed we can notice that, both classical and inference Bayesian inference for the parameters of those models, presented good properties analysed through the properties of the punctual estimators. The last stage of the work consisted of the analysis of one real data set, being a real serie corresponding to the admission number because of dengue in the city of Campina Grande. These results show that both the classic and the Bayesian studies are able to describe well the behavior of the serie.

Keywords: GARMA model, Poisson distribution, Binomial distribution, Negative Binomial distribution, Classic Inference, Bayesian Inference.

Sumário

1	Introdução	6
1.1	Organização dos capítulos	8
2	Modelo Autorregressivo e de Médias Móveis ARMA(p,q)	10
3	Modelos Generalizados ARMA	14
3.1	Modelos Lineares Generalizados	15
3.2	Modelo GARMA	16
3.3	Alguns modelos GARMA	18
3.3.1	Modelo Poisson GARMA	18
3.3.2	Modelo Binomial GARMA	19
3.3.3	Modelo Binomial Negativo GARMA	19
4	Inferência Clássica	21
4.1	O algoritmo de estimação	21
4.2	Seleção de Modelos	25
4.3	Análise de Resíduos	26
4.4	Previsão para o modelo GARMA	27
5	Inferência Bayesiana	29

5.1	Distribuições a priori	31
5.1.1	Priori não informativa	31
5.1.2	Priori conjugada	34
5.2	Abordagem bayesiana para o modelo ARMA	35
5.3	Abordagem bayesiana para o modelo GARMA	38
5.4	Monte Carlo com Cadeias de Markov	40
5.5	Seleção de modelos	43
5.6	Previsão para o modelo GARMA	44
6	Aplicação com dados simulados	45
6.1	Estudo de simulação	47
7	Aplicação com dados reais	58
7.1	Número de internações com dengue em Campina Grande	58
8	Considerações Finais	70
	Referências	72

Capítulo 1

Introdução

A análise de séries temporais é uma área da estatística dedicada ao estudo de observações que apresentam dependência no tempo. Uma série temporal pode ser definida como um conjunto de observações de uma variável disposta ao longo do tempo. Este ramo da estatística surge nas mais variadas áreas de aplicação, como: epidemiologia, finanças, econometria, redes neurais, meteorologia, ciências sociais, física, geofísica, medicina, entre outras.

Na análise do comportamento de uma série temporal sem tendência e sazonalidade podem ser utilizados modelos autorregressivos (AR) ou que incorporem médias móveis (MA). Quando há tendência, utilizam-se os modelos autorregressivos integrados de médias móveis (ARIMA) e, para incorporar o componente de sazonalidade, utilizam-se os modelos SARIMA. Dentre os modelos de séries temporais citados, destacamos o modelo autorregressivo e de médias móveis $ARMA(p, q)$ (Box *et al.*, 2008), pois o mesmo forma uma classe de modelos muito útil e parcimoniosa para descrever dados de séries temporais. Porém este modelo é apropriado para séries temporais estacionárias, mas na prática, muitas delas não são. Assim, para ajustar estes modelos a uma série temporal observada é necessário retirar as fontes de variação não estacionárias. Uma maneira de remover a não estacionariedade consiste em diferenciar a série até que ela se torne estacionária. Outro inconveniente é que muitas séries não apresentam distribuição normal, dificultando o uso dos modelos ARMA. Tendo em vista este fato foram propostos os modelos GARMA (Benjamin *et al.*, 2003) como uma alternativa.

Os modelos GARMA são uma extensão dos Modelos Lineares Generalizados (MLG) de McCullagh & Nelder (1989) para situações de dados dependentes, caracterizando-se pela adição de um termo extra ao preditor linear, no qual este passa a incorporar termos autor-regressivos (AR) e de médias móveis (MA). Ou seja, se considerarmos uma extensão dos modelos ARMA, para uma série não gaussiana, teremos que garantir que a distribuição condicional da variável resposta utilizada pertença à família exponencial e com isso, poderemos tratar o modelo de forma muito semelhante ao tratamento dado aos MLGs, mas com algumas diferenças que serão citadas no decorrer do texto. Estes modelos são indicados quando as variáveis em estudo não têm aderência à distribuição normal, principalmente pelo fato de serem processos de contagem.

Os Modelos Lineares Generalizados foram introduzidos primeiramente por Nelder & Wedderburn (1972), sendo baseados em distribuições que são modelos de dispersão exponencial. Os MLGs podem ser estendidos para séries temporais e vários autores apresentam casos especiais de MLGs para análise de série temporal. Veja, por exemplo, Li (1994); Fokianos & Kedem (2004), nos quais os modelos GARMA Poisson, binomial logística, binomial negativa e gamma são discutidos e aplicados a conjunto de dados de série temporal real.

Esta dissertação aborda o modelo GARMA com distribuições condicionais discretas pertencentes à família exponencial (distribuições Poisson, binomial e binomial negativa) em que o objetivo principal é analisar, nos contextos clássico e bayesiano, o desempenho e a qualidade do ajuste dos modelos de interesse, tais como o desempenho dos percentis de cobertura para os modelos adotados e os intervalos de credibilidade no estudo bayesiano para as três distribuições.

A distribuição binomial conta o número de sucessos em um número fixo de ensaios de Bernoulli independentes, a distribuição Poisson conta o número de ocorrências de um evento durante um intervalo de tempo e a distribuição binomial negativa conta o número de ensaios de Bernoulli necessários para se obter um número de sucessos fixos (Casella & Berger (2002)).

Visto que na literatura ainda não se encontram resultados da teoria assintótica para os modelos GARMA, adicionalmente este trabalho também se propõe a verificar, por meio

do estudo de simulação, se os estimadores de máxima verossimilhança são consistentes, e também é calculado o viés e o erro quadrático médio dos mesmos.

A qualidade do procedimento bayesiano é analisada no contexto frequentista, isto é, amostras foram simuladas para a construção dos intervalos de credibilidade. Esse mesmo procedimento foi considerado para os intervalos de confiança.

1.1 Organização dos capítulos

No Capítulo 2 é explanado rapidamente o modelo $ARMA(p, q)$, citando as condições de estacionariedade e invertibilidade para o modelo.

No Capítulo 3 é definido o modelo GARMA que é uma extensão dos MLGs de McCullagh e Nelder (1989) para situações de dados dependentes, caracterizando-se pela adição de um termo extra ao preditor linear, que passa a incorporar termos autorregressivos e de médias móveis. Também são abordados os casos discretos desse modelo atribuindo distribuições discretas pertencentes à família exponencial: Poisson, Binomial e Binomial Negativa.

O foco principal do Capítulo 4 é mostrar o algoritmo de estimação clássico dos parâmetros dos modelos estudados, bem como apresentar a função desvio para cada um deles. Também são apresentados os critérios de seleção de modelo, a análise de resíduos e a previsão.

No Capítulo 5, é explorada a inferência bayesiana usando uma distribuição a priori conjugada para os modelos GARMA em questão, sendo também desenvolvido um estudo bayesiano para os estimadores do modelo ARMA. Para a estimação dos parâmetros de ambos os modelos, utilizou-se métodos de Monte Carlo com Cadeias de Markov, como Metropolis e Gibbs Sampling.

A verificação do percentil de cobertura dos intervalos de confiança será desenvolvida no Capítulo 6, no qual são apresentados os resultados do estudo de simulação clássico mencionados nos Capítulos 4. Neste capítulo, também são exploradas a consistência, o viés e o erro quadrático médio dos estimadores pontuais gerados no estudo de simulação.

No Capítulo 7 foi feita a última etapa do trabalho. Neste capítulo, são apresentados resultados obtidos pelos métodos clássico e bayesiano para um conjunto de dados reais que representa o número de internações por causa da dengue em Campina Grande.

Capítulo 2

Modelo Autorregressivo e de Médias Móveis ARMA(p,q)

Seja $\{y_t, t = 1, 2, \dots\}$ uma série temporal. Box *et al.* (2008) definem o modelo autorregressivo e de médias móveis de ordem (p, q) , denominado ARMA(p, q), como sendo um modelo constituído de um processo autorregressivo de ordem p , denotado por AR(p) e um processo de médias móveis de ordem q , MA(q). De maneira geral,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

ou

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} + \dots - \phi_p y_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

ou ainda

$$\phi(B)y_t = \theta(B)a_t \tag{2.1}$$

em que a_t é um ruído branco, $\phi(B)$ e $\theta(B)$ são funções polinômiais do operador de defasagem B de ordem p e q , respectivamente e têm a seguinte forma:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

e

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

e o operador B é tal que $B^k y_t = y_{t-k}$.

Uma das suposições mais comuns feitas quando se trabalha com uma série temporal é a de que ela seja estacionária. Porém, grande parte das séries que são encontradas na prática, apresentam alguma forma de não estacionariedade, como por exemplo as séries econômicas e financeiras que apresentam em geral tendências crescentes ou decrescentes ou variância que muda ao longo do tempo. Séries que não apresentam estacionariedade podem apresentar as formas ditas de não estacionariedade explosiva e aquelas que não apresentam comportamento explosivo. Este último tipo de não estacionariedade é chamado homogêneo; a série pode ser estacionária, flutuando ao redor de um nível, por certo tempo, depois mudar de nível e flutuar ao redor de outro, e assim por diante, ou então mudar de inclinação, ou ambas as coisas. Uma série pode ser estacionária durante um período muito longo ou pode ser estacionária apenas em períodos muito curtos, mudando de nível e/ou inclinação. Esta é apenas uma ideia intuitiva de estacionariedade e agora serão apresentadas algumas definições: estacionariedade fraca (ou ampla ou de segunda ordem) e forte (ou estrita).

Definição 2.1 *Um processo estocástico $\{y_t; t \in T\}$, é estacionário no sentido forte (ou estritamente estacionário) se $F_{y_{t_1}, \dots, y_{t_n}}(y_1, \dots, y_n) = F_{y_{t_1+\tau}, \dots, y_{t_n+\tau}}(y_1, \dots, y_n)$, para todo $t_1 < t_2 < \dots \leq t_n \in T$, $n \geq 1$ e para todo $\tau \in T$, ou seja, se todas as distribuições finito-dimensionais $F_{y_{t_1}, \dots, y_{t_n}}(y_1, \dots, y_n) = P(y_{t_1}(w) \leq y_1, \dots, y_{t_n}(w) \leq y_n)$, (função distribuição de probabilidade conjunta), permanecem as mesmas sob translações do tempo.*

Na prática é muito difícil usar esta definição e costuma-se definir estacionariedade de uma forma menos restrita, como segue:

Definição 2.2 *Um processo estocástico $\{y_t; t \in T\}$, é fracamente estacionário (ou estacionário de segunda ordem) se $E(y_t) = \mu$ (média constante), $\text{var}(y_t) = \sigma^2$ (variância constante) e $\text{cov}(y_t, y_{t+k}) = \gamma(k)$ (covariância entre y_t e y_{t+k} é uma função que só depende de k), para todo $t \in T$ e $k \in \mathbb{N}$.*

Definição 2.3 *Um processo estocástico $\{y_t; t \in T\}$, é Gaussiano se, para qualquer conjunto $t_1 < t_2 < \dots \leq t_n \in T$, $n \geq 1$, as variáveis aleatórias y_{t_1}, \dots, y_{t_n} têm uma distribuição Normal n -variada.*

Para definir a estacionariedade para modelos lineares estacionários precisaremos do Teorema de Wald.

Teorema 2.1 (Teorema de Wald) Se y_t é um processo estocástico, então existe um processo estocástico $\{a_t, t = 1, 2, \dots\}$ tal que

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j a_{t-j} = \mu + \Psi(b)a_t$$

em que $\Psi(b)$ é um polinômio definido por $\Psi(B) = \Psi_0 + \Psi_1 B + \Psi_2 B^2 + \dots$, com $\Psi_0 = 1$, $E(a_t) = 0, \forall t$ e $E(a_t a_s) = 0$ para todo $s \neq t$ e $E(a_t^2) = \sigma_a^2$, em que σ_a^2 denota a variância constante de a_t .

Se a_t , com $t = 1, 2, \dots, n$ são variáveis aleatórias, independentes e identicamente distribuídas (iid), normais com média zero e variância constante σ_a^2 , isto é, $a_t \stackrel{iid}{\sim} N(0, \sigma_a^2)$, então neste caso, $\{a_t, t \geq 0\}$ recebe o nome de ruído branco.

O processo $y_t = \psi(B)a_t$ é estacionário no sentido fraco se as raízes do polinômio $\psi(B) = 0$ estão todas dentro ou sobre o círculo de raio unitário. Se a_t é um processo i.i.d. $N(0, \sigma_a^2)$, esta condição das raízes de $\psi(B) = 0$ garantem a estacionariedade no sentido forte para o processo.

Se $\{y_t, t \in T \subseteq Y\}$ é uma série temporal, o modelo linear estacionário para y_t , dado o Teorema de Wald, é estacionário se $|\psi(B)| < \infty$ para $|B| \leq 1$. Também, $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $\psi_0 = 1$.

O modelo ARMA(p, q), dado em (2.1), é estacionário se sua componente AR(p) é estacionário, ou seja, as raízes de $\phi(B) = 0$ estão fora do círculo de raio unitário. De fato: Para verificar a estacionariedade precisamos reescrever (2.1) na forma $y_t = \psi(B)a_t$, em que $\psi(B) = \phi(B)^{-1}\theta(B)$. Mas $\psi(B) = \phi(B)^{-1}\theta(B) = (\prod_{i=1}^p (1 - G_i B)^{-1}\theta(B))$, com $\phi(B) = 0$ tem raízes $G_i^{-1}, i = 1, 2, \dots, p$ e $\phi(B) = \prod_{i=1}^p (1 - G_i B)$, então

$$\psi(B) = \left(\sum_{i=1}^p c_i (1 - G_i B)^{-1} \right) \theta(B) \implies |\psi(B)| \leq \left(\sum_{i=1}^p |c_i| \cdot |(1 - G_i B)^{-1}| \right) |\theta(B)|$$

para $|B| \leq 1$.

Mas,

$$\begin{aligned} |(1 - G_i B)^{-1}| &\leq 1 + |G_i|B + |G_i|^2 B^2 + \dots \\ &\leq \sum_{j=0}^{\infty} G_i^j < \infty, \text{ se } |G_i| < 1 \end{aligned}$$

Portanto, $|\psi(B)| < \infty$ para $|B| \leq 1$ se $|G_i| < 1$.

Definição 2.4 Um processo linear será invertível se $\pi(B)$ convergir para $|B| \leq 1$, em que $\pi(B) = \phi(B)\theta(B)^{-1}$.

O modelo ARMA(p, q), dado em (2.1), é invertível se sua componente MA(q) é invertível, ou seja, as raízes de $\theta(B) = 0$ estão fora do círculo de raio unitário. De fato: Para verificar a invertibilidade precisamos reescrever (2.1) na forma $\pi(B)y_t = a_t$, em que $\pi(B) = \phi(B)\theta(B)^{-1}$. Mas $\pi(B) = \phi(B)\theta(B)^{-1} = \left(\prod_{i=1}^q (1 - H_i B)^{-1} \phi(B)\right)$, em que H_i^{-1} são as raízes de $\theta(B) = 0$, $i = 1, 2, \dots, q$. Temos

$$\pi(B) = \left(\sum_{i=1}^q c_i (1 - H_i B)^{-1} \right) \phi(B)$$

sendo $|\phi(B)| < \infty$ para $|B| \leq 1$. Tem-se $|\pi(B)| < \infty$ para $|B| \leq 1$ se $|H_i| < 1$.

Neste trabalho considera-se que as condições de estacionariedade e invertibilidade são garantidas, ou seja, as soluções de $\phi(B) = 0$ e $\theta(B) = 0$ estão todas fora do círculo de raio unitário e que os valores de (p, q) são fixos, mas desconhecidos.

Este modelo é muito utilizado na modelagem de série temporal, pois é um modelo com um número não muito grande de parâmetros. Na literatura existe uma grande quantidade de trabalhos a respeito do problema de estimação dos parâmetros deste modelo considerando as abordagens clássica ou bayesiana (ver, por exemplo, Monahan (1983); Marriott *et al.* (1996); Barnett *et al.* (1996); Barnett *et al.* (1997); Box *et al.* (2008); Brockwell & Davis (2009)).

Capítulo 3

Modelos Generalizados ARMA

Séries temporais para dados de contagem são registros da frequência relativa de ocorrência de certos eventos em sucessivos intervalos de tempo, em geral igualmente espaçados. Podem ser citados como exemplos o número mensal de sinistros de carros associados a uma seguradora, o número diário de internações hospitalares em uma cidade em consequência de dengue, entre outros.

Como pode-se observar, dados de contagem são necessariamente números inteiros não negativos, o que elimina a possibilidade do processo gerador de dados ser gaussiano, tornando, à princípio, inadequada a utilização de modelos ARIMA (Box *et al.*, 2008). Os modelos apropriados para dados de contagem devem utilizar distribuições de probabilidade cujo domínio seja formado apenas pelos números inteiros não-negativos como é o caso das distribuições Poisson, binomial e binomial negativa.

Os modelos GARMA são uma extensão dos Modelos Lineares Generalizados (MLG) de McCullagh & Nelder (1989) para situações de dados dependentes, caracterizando-se pela adição de um termo extra ao preditor linear, no qual este passa a incorporar termos autoregressivos (AR) e de médias móveis (MA). Assim sendo, é conveniente introduzir uma pequena descrição dos MLG, antes de abordarmos os modelos GARMA.

3.1 Modelos Lineares Generalizados

Nelder e Wedderburn propuseram, em 1972, os MLGs como uma extensão dos modelos lineares normais. A ideia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta (μ) e o preditor linear (η).

Suponha que Y tem distribuição na família exponencial se

$$f_Y(y|\theta, \varphi) = \exp \left\{ \frac{y\theta - b(\theta)}{\varphi} + c(y, \varphi) \right\},$$

sendo θ e $\varphi > 0$ os parâmetros canônico e de dispersão, respectivamente e $b(\cdot)$ e $c(\cdot)$ funções conhecidas. A média e variância são dadas por

$$E(Y) = \mu = b'(\theta) \quad e \quad Var(Y) = \varphi b''(\theta) = \varphi V(\mu),$$

em que $V(\mu) = b''(\theta)$ é denominada função variância sendo expressa em termos da média μ .

Os MLG podem ser usados quando se tem uma observação do vetor \mathbf{Y} com elementos independentes e um vetor de médias $\boldsymbol{\mu}$, com $E(Y_i) = \mu_i = \sum_{j=1}^p x_{ij}\beta_j$, $i = 1, 2, \dots, n$, sendo que $p < n$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ é o vetor de coeficientes desconhecidos e x_1, x_2, \dots, x_p são as covariáveis.

Em forma matricial, $\boldsymbol{\mu}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}$, em que \mathbf{X} é uma matriz de posto igual a p , chamada de matriz modelo. $\mathbf{X}\boldsymbol{\beta}$ é chamado preditor linear, sendo denotado por $\boldsymbol{\eta}$, isto é, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.

Um MLG envolve três componentes:

- **Componente aleatório:** é representado por um conjunto de variáveis aleatórias independentes pertencentes a uma mesma distribuição, que faça parte da família exponencial, com média $E(Y_i) = \mu_i$, $i = 1, \dots, n$.
- **Componente sistemático:** as variáveis explicativas x_1, x_2, \dots, x_p fornecem o preditor linear $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.

- **Função de Ligação:** é uma função que relaciona o componente aleatório ao componente sistemático, ou seja, a média ao preditor linear

$$g(\mu_i) = \eta_i, \quad i = 1, \dots, n. \quad (3.1)$$

em que $g(\cdot)$ uma função monótona e diferenciável.

Agora, com os conceitos iniciais de MLG definidos, daremos início à explanação dos modelos GARMA.

3.2 Modelo GARMA

Esta seção é baseada no artigo de Benjamin *et al.* (2003). Neste artigo os autores consideram uma extensão dos modelos ARMA com distribuição condicional gaussiana para distribuições condicionais não-gaussianas. Porém, essa nova abordagem faz necessária que uma condição seja satisfeita: a distribuição utilizada deve pertencer à família exponencial e assim, o modelo pode ser tratado de maneira análoga, mas com algumas diferenças, aos Modelos Lineares Generalizados (MLG). A diferença fundamental entre as duas abordagens é que, enquanto em McCullagh e Nelder a distribuição definida é a marginal, nos modelos GARMA a parte aleatória do modelo é especificada pela distribuição condicional, tornando possível o tratamento de dados dependentes.

Em um modelo GARMA a distribuição condicional de cada observação y_t , para cada $t = 1, \dots, n$, dado \mathbf{H}_t que representa o conjunto de informações passadas, denotado por

$$\mathbf{H}_t = \{\mathbf{x}_t, \dots, \mathbf{x}_1, y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}$$

é considerada pertencer à família exponencial se

$$f(y_t|\mathbf{H}_t) = \exp \left\{ \frac{y_t \theta_t - b(\theta_t)}{\varphi} + c(y_t, \varphi) \right\} \quad (3.2)$$

em que θ_t e φ são os parâmetros canônico e de dispersão, respectivamente. As funções $b(\cdot)$ e $c(\cdot)$ definem uma família exponencial específica e \mathbf{x} é o vetor de variáveis explicativas.

Assim como nos modelos lineares generalizados, a média condicional do processo μ_t está relacionada com o preditor η_t pela função de ligação $g(\cdot)$, isto é, $g(\mu_t) = \eta_t$, sendo $g(\cdot)$ uma função monótona e diferenciável.

Como mencionado anteriormente, a diferença dos modelos GARMA para os MLG está no preditor, pois aqui será incluído um componente adicional, chamado τ_t , que fornece a parte relativa aos termos autoregressivos e de médias móveis, além de incluir no preditor valores passados das variáveis explicativas presentes no modelo. Assim, podemos escrever

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \tau_t, \quad t = 1, \dots, n. \quad (3.3)$$

em que

$$\tau_t = \sum_{j=1}^p \phi_j \mathcal{A}(y_{t-j}, \mathbf{x}_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j \mathcal{M}(y_{t-j}, \mu_{t-j})$$

com \mathcal{A} e \mathcal{M} funções que representam os termos autorregressivos e de médias móveis, respectivamente, com $\boldsymbol{\phi}' = (\phi_1, \dots, \phi_p)$ e $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_q)$ os vetores de parâmetros autorregressivos e de médias móveis, respectivamente.

O modelo apresentado em (3.3) é muito geral e pode ser reescrito em uma forma mais simples, ou seja, fazendo as devidas mudanças no termo τ_t , o preditor linear pode ser escrito da seguinte maneira

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\}$$

Então combinando essa equação à equação de $f(y_t | \mathbf{H}_t)$, o modelo GARMA(p,q) fica definido por:

$$f(y_t | \mathbf{H}_t) = \exp \left\{ \frac{y_t \theta_t - b(\theta_t)}{\varphi} + c(y_t, \varphi) \right\} \quad (3.4)$$

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\} \quad (3.5)$$

Assim como nos modelos lineares generalizados, a média e variância condicionais do modelo GARMA(p,q) são dadas por

$$\begin{aligned} E(y_t | \mathbf{H}_t) &= b'(\theta_t) = \mu_t, \\ Var(y_t | \mathbf{H}_t) &= \varphi b''(\theta_t) = \varphi V(\mu_t). \end{aligned}$$

em que V é chamada de função variância dependendo apenas da média μ_t .

Uma das vantagens dos modelos GARMA para séries temporais é devido ao fato de que o componente aleatório do modelo pode assumir várias distribuições, desde que esta pertença à família exponencial, podendo ser utilizado inclusive para dados de contagem.

Na próxima seção serão explanados alguns modelos GARMA para dados de contagem.

3.3 Alguns modelos GARMA

Nesta seção serão explanados três casos particulares do modelo GARMA, atribuindo as distribuições Poisson, Binomial e Binomial Negativa, sendo estas as distribuições discretas mais comuns.

3.3.1 Modelo Poisson GARMA

Supondo que $y_t|\mathbf{H}_t$ possui distribuição de Poisson(μ_t), então

$$f(y_t|\mathbf{H}_t) = \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!}, \quad t = 1, \dots, n.,$$

que pode ser reescrita como

$$f(y_t|\mathbf{H}_t) = \exp \{y_t \log(\mu_t) - \mu_t - \log(y_t!)\} \quad (3.6)$$

Note que $\varphi = 1$ e $\theta_t = \log(\mu_t)$, donde segue que $\mu_t = e^{\theta_t}$. Reescrevendo a equação (3.6), temos

$$f(y_t|\mathbf{H}_t) = \exp \{y_t \theta_t - e^{\theta_t} - \log(y_t!)\} \quad (3.7)$$

Logo, $y_t|\mathbf{H}_t$ pertence à família exponencial com

$$\varphi = 1, \quad \theta_t = \log(\mu_t), \quad b(\theta_t) = e^{\theta_t}, \quad c(y_t, \varphi) = -\log(y_t!), \quad \mu_t = \exp(\theta_t) \text{ e } V(\mu_t) = \mu_t.$$

Para este modelo a função de ligação é a função logarítmica, pois a mesma fornece valores não negativos para $\mu_t = g^{-1}(\eta_t)$ não importando os valores atribuídos a η_t . Assim, o preditor linear fica

$$\log(\mu_t) = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j \{\log(y_{t-j}^*) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^* / \mu_{t-j})\}$$

e, sem perda de generalidade, podemos desconsiderar as variáveis x_t , pois são variáveis exógenas ao processo e escrever $\alpha_0 = \mathbf{x}'_t \boldsymbol{\beta}$ e assim reescrevendo a equação acima como

$$\log(\mu_t) = \alpha_0 + \sum_{j=1}^p \phi_j \log(y_{t-j}^*) + \sum_{j=1}^q \theta_j \{\log(y_{t-j}^*/\mu_{t-j})\} \quad (3.8)$$

com $y_{t-j}^* = \max(y_{t-j}, c)$, $0 < c < 1$.

Portanto, o modelo Poisson GARMA fica definido por (3.7) e (3.8).

3.3.2 Modelo Binomial GARMA

Supondo que $y_t | \mathbf{H}_t$ segue distribuição de binomial, ou seja, $y_t | \mathbf{H}_t \sim B(m, p_t)$ então

$$\begin{aligned} f(y_t | \mathbf{H}_t) &= \frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} p_t^{y_t} (1-p_t)^{m-y_t} \\ &= \exp \left\{ y_t \log \left(\frac{p_t}{1-p_t} \right) + m \log(1-p_t) + \log \left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} \right) \right\} \end{aligned}$$

mas $\mu_t = mp_t$, então $p_t = \frac{\mu_t}{m}$. Logo

$$f(y_t | \mathbf{H}_t) = \exp \left\{ y_t \log \left(\frac{\mu_t}{m-\mu_t} \right) + m \log \left(\frac{m-\mu_t}{m} \right) + \log \left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} \right) \right\} \quad (3.9)$$

que pertence à família exponencial com

$$\theta_t = \log \left(\frac{\mu_t}{m-\mu_t} \right), \quad b(\theta_t) = -m \log \left(\frac{m-\mu_t}{m} \right), \quad \varphi = 1,$$

$$c(y_t, \varphi) = \log \left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} \right), \quad \mu_t = \frac{m \exp(\theta_t)}{1 + \exp(\theta_t)} \quad \text{e} \quad V(\mu_t) = \frac{\mu_t}{m} (m - \mu_t).$$

As equações (3.8) e (3.9) definem o modelo binomial GARMA.

3.3.3 Modelo Binomial Negativo GARMA

No modelo binomial negativo GARMA consideramos a distribuição de probabilidade condicional como sendo a distribuição binomial negativa, ou seja, $y_t | \mathbf{H}_t \sim BN(k, p_t)$ com função de probabilidade

$$f(y_t | \mathbf{H}_t) = \frac{\Gamma(k+y_t)}{\Gamma(y_t+1)\Gamma(k)} p_t^k (1-p_t)^{y_t}$$

com $y_t = 0, 1, 2, \dots$

Mas $\mu_t = \frac{k(1-p_t)}{p_t}$, então $p_t = \frac{k}{\mu_t+k}$ e assim

$$\begin{aligned} f(y_t|\mathbf{H}_t) &= \frac{\Gamma(k+y_t)}{\Gamma(y_t+1)\Gamma(k)} \left(\frac{\mu_t}{\mu_t+k}\right)^{y_t} \left(\frac{k}{\mu_t+k}\right)^k \\ &= \exp\left\{k \log\left(\frac{k}{\mu_t+k}\right) + y_t \log\left(\frac{\mu_t}{\mu_t+k}\right) + \log\left(\frac{\Gamma(k+y_t)}{\Gamma(y_t+1)\Gamma(k)}\right)\right\} \end{aligned} \quad (3.10)$$

que pertence a família exponencial com

$$\theta_t = \log\left(\frac{\mu_t}{k+\mu_t}\right), \quad b(\theta_t) = -k \log\left(\frac{k}{k+\mu_t}\right), \quad c(y_t, \varphi) = \log\left(\frac{\Gamma(k+y_t)}{\Gamma(y_t+1)\Gamma(k)}\right),$$

$$\varphi = 1, \quad \mu_t = \frac{k \exp(\theta_t)}{1 - \exp(\theta_t)} \quad \text{e} \quad V(\mu_t) = \frac{\mu_t}{k}(k + \mu_t).$$

As equações (3.8) e (3.10) definem o modelo binomial negativo GARMA.

As informações descritas acima podem ser resumidas na seguinte tabela:

Tabela 3.1: Características das distribuições discretas na família exponencial.

	Poisson	binomial	binomial negativa
Notação	$P(\mu_t)$	$B(m, p_t)$	$BN(k, p_t)$
Parâmetro de dispersão: φ	1	1	1
Função cumulante: $b(\theta_t)$	$\exp(\theta_t)$	$-m \log\left(\frac{m-\mu_t}{m}\right)$	$-k \log\left(\frac{k}{k+\mu_t}\right)$
$c(y_t, \phi)$	$-\log(y_t!)$	$\log\left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)}\right)$	$\log\left(\frac{\Gamma(k+y_t)}{\Gamma(y_t+1)\Gamma(k)}\right)$
$\mu_t(\theta_t) = E(Y; \theta_t)$	$\exp(\theta_t)$	$\frac{m \exp(\theta_t)}{1 + \exp(\theta_t)}$	$\frac{k \exp(\theta_t)}{1 - \exp(\theta_t)}$
Função de ligação	$\log(\mu_t)$	$\log\left(\frac{\mu_t}{m-\mu_t}\right)$	$\log\left(\frac{\mu_t}{k+\mu_t}\right)$
Função de variância: $V(\mu_t)$	μ_t	$\frac{\mu_t}{m}(m - \mu_t)$	$\frac{\mu_t}{k}(k + \mu_t)$

Capítulo 4

Inferência Clássica

4.1 O algoritmo de estimação

Neste capítulo considera-se o método de máxima verossimilhança (MV) para estimar os parâmetros do modelo. O vetor escore é formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança. O logaritmo da função de verossimilhança é escrito como função apenas do vetor de parâmetros $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\phi}', \boldsymbol{\theta}')$. Considerando o parâmetro de dispersão conhecido e usando a expressão (3.2) tem-se, aproximando nas $r = \max(p, q)$ primeiras observações

$$\ell(\boldsymbol{\gamma}, \varphi) = \frac{1}{\varphi} \sum_{t=r+1}^n \{y_t \theta_t - b(\theta_t)\} + \sum_{t=r+1}^n c(y_t, \varphi). \quad (4.1)$$

em que $\mu_t = g^{-1}(\eta_t)$ e $\theta_t = q(\mu_t)$, q é uma função conhecida.

Considere as matrizes A e B definidas abaixo cujos elementos são funções dos parâmetros do modelo.

$$\mathbf{A} = \begin{bmatrix} g(y_r) - x_r \boldsymbol{\beta} & \cdots & g(y_{r+1-p}) - x_{r+1-p} \boldsymbol{\beta} \\ g(y_{r+1}) - x_{r+1} \boldsymbol{\beta} & \cdots & g(y_{r+2-p}) - x_{r+2-p} \boldsymbol{\beta} \\ \vdots & \ddots & \vdots \\ g(y_{n-1}) - x_{n-1} \boldsymbol{\beta} & \cdots & g(y_{n-p}) - x_{n-p} \boldsymbol{\beta} \end{bmatrix}_{(n-r) \times p},$$

$$\mathbf{B} = \begin{bmatrix} g(y_r) - \eta_r & \cdots & g(y_{r+1-q}) - \eta_{r+1-q} \\ g(y_{r+1}) - \eta_{r+1} & \cdots & g(y_{r+2-q}) - \eta_{r+2-q} \\ \vdots & \ddots & \vdots \\ g(y_{n-1}) - \eta_{n-1} & \cdots & g(y_{n-q}) - \eta_{n-q} \end{bmatrix}_{(n-r) \times q}.$$

O componente sistemático do modelo GARMA escrito em forma matricial correspondente as observações y_{r+1}, \dots, y_n é

$$\boldsymbol{\eta} = \begin{bmatrix} \mathbf{X} & \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\phi} \\ \boldsymbol{\theta} \end{bmatrix},$$

em que \mathbf{X} é a matriz do modelo formada pelas linhas \mathbf{x}_t , $t = r+1, \dots, n$. Pode-se também denotar o componente sistemático reduzindo-o a apenas $\boldsymbol{\eta} = \mathbf{M}\boldsymbol{\gamma}$, em que $\mathbf{M} = [\mathbf{X} \ \mathbf{A} \ \mathbf{B}]$ é a matriz modelo de ordem $(n-r) \times (m+p+q)$.

Seja $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\eta}} = \widehat{\mathbf{M}}\boldsymbol{\gamma}$ e $\hat{\varphi}$ os estimadores de máxima verossimilhança (EMV) de $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ e φ respectivamente. Uma característica importante no processo de estimação dos parâmetros é que a estimação do vetor de parâmetros $\boldsymbol{\gamma}$ não depende de φ . Os EMV's $\hat{\boldsymbol{\gamma}}$ podem ser obtidos através do ajuste dos modelos (3.2) e (3.3) para y pelo método de mínimos quadrados ponderados reiterados, dado por

$$\hat{\boldsymbol{\gamma}}^{(k+1)} = \left(\widehat{\mathbf{M}}'^{(k)} \widehat{\mathbf{W}}^{(k)} \widehat{\mathbf{M}}^{(k)} \right)^{-1} \widehat{\mathbf{M}}'^{(k)} \widehat{\mathbf{W}}^{(k)} \hat{\mathbf{z}}^{(k)}, \quad (4.2)$$

com k sendo o passo da iteração, em que $\widehat{\mathbf{W}} = \text{diag}\{w_{r+1}, \dots, w_n\}$ é a matriz de pesos, com $w_t = V_t^{-1} \left(\frac{dg(\mu_t)}{d\mu_t} \right)^{-2}$ e a variável z_t é dada por

$$z_t = \eta_t + (y_t - \mu_t) \frac{dg(\mu_t)}{d\mu_t}.$$

O procedimento iterativo usando a equação (4.2) parte de um valor inicial $\hat{\boldsymbol{\gamma}}^{(1)}$ e calcula $\widehat{\mathbf{M}}^{(1)}$, $\widehat{\mathbf{W}}^{(1)}$ e $\hat{\mathbf{z}}^{(1)}$, usados para obter a próxima estimativa $\hat{\boldsymbol{\gamma}}^{(2)}$ em (4.2). Este procedimento é realizado até que a convergência seja observada.

As estimativas $\hat{\eta}_t$, para $t = r+1, \dots, n$ são obtidas de

$$\hat{\eta}_t = \mathbf{x}_t \hat{\boldsymbol{\beta}} + \sum_{j=1}^p \hat{\phi}_j \left\{ g(y_{t-j}) - \mathbf{x}_{t-j} \hat{\boldsymbol{\beta}} \right\} + \sum_{j=1}^q \hat{\theta}_j \left\{ g(y_{t-j}) - \hat{\eta}_{t-j} \right\},$$

e as médias ajustadas são obtidas através da inversa da função g , isto é, $\hat{\mu}_t = g^{-1}(\hat{\eta}_t)$, com $t = r + 1, \dots, n$.

Para estimar φ basta derivar a log-verossimilhança em relação a φ . Fazendo isso obtemos

$$\hat{\varphi}^2 \sum_{t=r+1}^n \frac{dc(y_t, \varphi)}{d\varphi} \Big|_{\phi=\hat{\phi}} = \sum_{t=r+1}^n \left\{ y_t \hat{\theta}_t - b(\hat{\theta}_t) \right\}$$

em que $\hat{\theta}_t = q(g^{-1}(\mathbf{x}_t \hat{\beta}))$ e dada a função de variância $V(x)$, podemos obter $q(x) = \int V^{-1}(x)$.

Intervalo de Confiança

A estimação intervalar consiste na determinação de um intervalo onde, com uma certa confiança (probabilidade), esteja o parâmetro λ desconhecido, levando-se em conta o seu estimador, ou seja, em vez de estimar o parâmetro por um único valor, é dado um intervalo de estimativas prováveis. Quão prováveis são estas estimativas é determinado pelo coeficiente de confiança e quanto maior a probabilidade do intervalo conter o parâmetro, maior será o intervalo.

Assim, $P(L_1 < \lambda < L_2) = 1 - \alpha$ significa que a probabilidade do intervalo aleatório (L_1, L_2) conter o valor exato λ é $1 - \alpha$.

O intervalo (L_1, L_2) é designado por intervalo de confiança para o parâmetro λ , com um nível de confiança $1 - \alpha$. Estes intervalos são usados para indicar a confiabilidade de uma estimativa.

Considerando uma confiabilidade de 95%, um intervalo de confiança para os parâmetros dos modelos GARMA discretos é dado por

$$P(\hat{\beta} - 1.96 \cdot Var(\hat{\beta}) < \beta < \hat{\beta} + 1.96 \cdot Var(\hat{\beta})) = 0,95$$

com $\hat{\beta}$ sendo a estimativa do vetor de parâmetros β do modelo e $Var(\hat{\beta})$, sua variância. Para maiores detalhes ver Cordeiro & Andrade (2009).

Função Desvio

A função desvio D para o modelo GARMA é

$$D = 2 \sum_{t=r+1}^n \{y_t(\hat{\theta}_t^0 - \hat{\theta}_t) + (b(\hat{\theta}_t) - b(\hat{\theta}_t^0))\}$$

sendo que $\hat{\theta}_t^0 = q(y_t)$. Assim a função desvio pode ser reescrita na seguinte forma

$$D = 2 \sum_{t=r+1}^n \{y_t q(y_t) - b(q(y_t))\} - 2 \sum_{t=r+1}^n \{y_t q(\hat{\mu}_t) - b(q(\hat{\mu}_t))\}$$

e a estimativa de máxima verossimilhança parcial de $\hat{\varphi}$ pode ser escrita como

$$\hat{\varphi}^2 \sum_{t=r+1}^n \frac{dc(y_t, \varphi)}{d\varphi} \Big|_{\varphi=\hat{\varphi}} = \sum_{t=r+1}^n e(y_t) - \frac{D}{2}, \quad (4.3)$$

com $e(x) = xq(x) - b(q(x))$. A equação (4.3) é em geral não linear.

Quanto melhor for o ajuste do modelo aos dados tanto menor será o valor do desvio D . Assim, um modelo bem ajustado aos dados, juntamente com uma verossimilhança grande, tem um pequeno desvio.

Modelo Poisson GARMA

No modelo Poisson GARMA temos que $q(\hat{\mu}_t) = \log(\hat{\mu}_t)$ e $b(q(\hat{\mu}_t)) = \exp(\log(\hat{\mu}_t)) = \hat{\mu}_t$, logo a função desvio é dada por

$$D = 2 \sum_{t=r+1}^n \{y_t(\log(y_t) - \log(\hat{\mu}_t)) + \hat{\mu}_t - y_t\} \quad (4.4)$$

$$= 2 \sum_{t=r+1}^n \left\{ y_t \log \left(\frac{y_t}{\hat{\mu}_t} \right) + \hat{\mu}_t - y_t \right\}. \quad (4.5)$$

Se $y_t = 0$, deve-se tomar o limite de D quando $y_t \rightarrow 0$, donde segue que $D = 2\hat{\mu}_t$.

Modelo Binomial GARMA

Para o modelo binomial GARMA tem-se que $q(\hat{\mu}_t) = \log\left(\frac{\hat{\mu}_t}{m-\hat{\mu}_t}\right)$ e $b(q(\hat{\mu}_t)) = m \log\left(\frac{m}{m-\hat{\mu}_t}\right)$, logo a função desvio é dada por

$$\begin{aligned} D &= 2 \sum_{t=r+1}^n \left\{ y_t \left[\log\left(\frac{y_t}{m-y_t}\right) - \log\left(\frac{\hat{\mu}_t}{m-\hat{\mu}_t}\right) \right] + m \log\left(\frac{m}{m-\hat{\mu}_t}\right) - m \log\left(\frac{m}{m-y_t}\right) \right\} \\ &= 2 \sum_{t=r+1}^n \left\{ y_t \log\left(\frac{y_t}{\hat{\mu}_t}\right) + (m-y_t) \log\left(\frac{m-y_t}{m-\hat{\mu}_t}\right) \right\}. \end{aligned} \quad (4.6)$$

Esta expressão é válida para $0 < y_t < m$. Se $y_t = 0$ ou $y_t = m$, o t -ésimo termo de D deve ser substituído por $2m \log\left(\frac{m}{m-\hat{\mu}_t}\right)$ ou $2m \log\left(\frac{m}{\hat{\mu}_t}\right)$, respectivamente, ver Paula (2004).

Modelo Binomial Negativa GARMA

Para o modelo binomial negativa GARMA tem-se que $q(\hat{\mu}_t) = \log\left(\frac{\hat{\mu}_t}{k-\hat{\mu}_t}\right)$ e $b(q(\hat{\mu}_t)) = -k \log\left(\frac{k}{k-\hat{\mu}_t}\right)$, logo a função desvio é

$$\begin{aligned} D &= 2 \sum_{t=r+1}^n \left\{ y_t \left[\log\left(\frac{y_t}{k+y_t}\right) - \log\left(\frac{\hat{\mu}_t}{k+\hat{\mu}_t}\right) \right] - k \log\left(\frac{k}{k+\hat{\mu}_t}\right) + k \log\left(\frac{k}{k+y_t}\right) \right\} \\ &= 2 \sum_{t=r+1}^n \left\{ y_t \log\left(\frac{y_t}{\hat{\mu}_t}\right) + (y_t+k) \log\left(\frac{k+\hat{\mu}_t}{k}\right) \right\}. \end{aligned} \quad (4.7)$$

4.2 Seleção de Modelos

Para decidir qual é o modelo mais apropriado dentre um conjunto de modelos candidatos são utilizados critérios de seleção. A idéia básica dos critérios de seleção de modelos consiste em decidir qual é o modelo mais adequado quando comparamos diversos modelos com quantidades diferentes de parâmetros, isto é, selecionar um modelo que seja parcimonioso, ou em outras palavras, que esteja bem ajustado com um número reduzido ou suficiente de parâmetros. Dentre os critérios mais conhecidos e utilizados estão o AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion).

Akaike (1974) mostrou que o viés é dado assintoticamente por d , em que d é o número de parâmetros a serem estimados no modelo e definiu seu critério de informação como

$$AIC = -2\ell(\gamma) + 2d,$$

em que $\ell(\gamma)$ denota o máximo da função de log-verossimilhança e d é o número de parâmetros a serem estimados do modelo. O critério BIC, proposto por Schwarz (1978), penaliza o modelo a ser escolhido mais fortemente do que o AIC e é definido por

$$BIC = -2\ell(\gamma) + d \log(n),$$

com n sendo o número de observações da série.

O modelo escolhido deve ser aquele que apresenta menor valor de AIC e BIC.

Para verificar se o modelo ajustado é adequado para descrever o comportamento dos dados, foram usados os critérios MSE (do inglês Mean Square Error) e o MAPE (do inglês Mean Absolute Percentile Error) dados respectivamente por

$$MSE = \frac{100\%}{n\hat{\sigma}_y^2} \sum_{t=1}^n (y_t - \hat{\mu}_t)^2 \quad (4.8)$$

e

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{\mu}_t}{y_t} \right|, \quad (4.9)$$

em que $\hat{\sigma}_y^2$ é a variância amostral de y .

Quanto menor for o MSE e o MAPE, mais próximo do real são os valores estimados.

4.3 Análise de Resíduos

Para o modelo selecionado, o diagnóstico para a variável resposta é realizado através de uma análise de resíduos, visto que a normalidade dos resíduos é uma suposição essencial para que os resultados do ajuste do modelo sejam confiáveis. Os resíduos são definidos como

$$r_t = y_t - \hat{\mu}_t$$

com $\hat{\mu}_t$ são os valores da média estimada para cada t e $t = r + 1, \dots, n$.

No modelo de regressão assume-se que os resíduos satisfazem os seguintes pressupostos: devem ser independentes e seguir uma distribuição normal com média zero e variância constante (homocedasticidade). Como era de se esperar nos modelos GARMA discretos em questão, os resíduos podem não ser normalmente distribuídos para dados de contagem.

Dunn & Smyth (1996) apresentam uma definição que produz resíduos que apresentam normalidade: em primeiro lugar, a estimativa da função de densidade acumulada implícita no modelo é usada para transformar as observações em variáveis aleatórias aproximadamente independentes distribuídas uniformemente; por outro lado, o inverso da função de densidade acumulada da distribuição normal padrão é usado para obter as variáveis que são aproximadamente independentes com distribuição normal padrão.

Seja $F(y; \mu, \phi)$ a função de distribuição acumulada de uma distribuição discreta citada anteriormente no texto. De acordo com Dunn & Smyth (1996), para a função de distribuição discreta, o resíduo quantil normalizado é definido por

$$r_t = \Phi^{-1}(u_t)$$

em que Φ^{-1} é a inversa da distribuição acumulada de uma variável normal padrão e u_t é um valor aleatório obtido de uma distribuição uniforme no intervalo $[F(y_{t-1}, \hat{\mu}_t, \hat{\phi}), F(y_t, \hat{\mu}_t, \hat{\phi})]$, com $F(y_t, \hat{\mu}_t, \hat{\phi})$ é a função de distribuição acumulada condicional ajustada.

Desta maneira os resíduos seguem exatamente uma distribuição normal padrão.

Após a seleção de modelos e análise de resíduos, o último passo da análise do modelo GARMA é a previsão.

4.4 Previsão para o modelo GARMA

Muitos estudos envolvendo séries temporais têm como objetivo fazer previsões: prever valores futuros da série com base em valores passados da mesma, ou seja, uma das formas de utilização de um modelo ajustado é para prever um valor y_{t+h} , $h > 1$, supondo que as observações $\dots, y_{t-2}, y_{t-1}, y_t$ são conhecidas.

O método de previsão abordado nesta dissertação foi o método de previsão Naive, uma referência comum encontrada na literatura (Lawrence *et al.*, 2000), que se baseia no conceito

de que a melhor previsão para o instante seguinte será o valor observado no instante atual, isto é, se um valor da série estiver em uma tendência de alta, a previsão é de que ele continuará em alta e vice-versa. A previsão Naive é definida por

$$\hat{y}_{t+h} = y_t$$

em que \hat{y}_{t+h} é a previsão feita h passos à frente e y_t é o valor do retorno observado no instante t.

Como são conhecidas todas as componentes do preditor linear dos modelos GARMA, a previsão é feita de maneira simplificada usando o preditor, ou seja, a previsão h passos à frente é dada por

$$\hat{\mu}_{t+h} = \exp \left(\hat{\alpha}_0 + \sum_{j=1}^p \hat{\phi}_j \log(y_{t+h-j}^*) + \sum_{j=1}^q \hat{\theta}_j \{ \log(y_{t+h-j}^* / \hat{\mu}_{t+h-j}) \} \right) \quad (4.10)$$

Faz-se necessário o uso de algumas técnicas para avaliar a qualidade dos valores previstos, ou seja, é preciso verificar se estes valores previstos estão ou não coerentes com o conjunto de dados. Um método amplamente utilizado na literatura, consiste na retirada dos valores finais da série e calcular os mesmos através da equação (4.10) e, em seguida, compará-los com os valores que foram retirados anteriormente. A quantidade de observações retiradas depende do tamanho do passo da previsão, ou seja, se o passo for de tamanho 3 então deve-se retirar 3 observações da amostra.

Capítulo 5

Inferência Bayesiana

O rápido crescimento do uso do paradigma bayesiano em ciências aplicadas ao longo das últimas décadas, foi facilitado pelo surgimento e disponibilidade de vários algoritmos computacionais usados no cálculo de integrais que são necessárias em uma análise bayesiana.

Na perspectiva bayesiana, a inferência estatística sobre qualquer quantidade de interesse é descrita como a modificação que se processa nas incertezas à medida que são observados novos dados ou novos resultados. É o Teorema de Bayes que permite quantificar esta modificação.

A inferência estatística se compõe de dois elementos: estimação (pontual e por intervalo) de parâmetros desconhecidos e testes de hipóteses sobre estes parâmetros. Existe, nessa abordagem, uma grande quantidade de testes estatísticos especialmente desenvolvidos para as diferentes situações. Estes encontram-se agrupados em extensos manuais de métodos, para maior detalhes ver Zar (1996) e Snedecor & Cochran (1989). Muitas vezes a dificuldade que o usuário encontra é escolher, entre as muitas opções disponíveis, o método adequado para uma situação específica e, se não bastasse isto, em muitos casos os diferentes métodos disponíveis não são diretamente comparáveis entre si, dando a impressão de que a escolha é arbitrária.

Na inferência bayesiana a questão resume-se a construção de uma distribuição de probabilidade a posteriori via o Teorema de Bayes. Esta distribuição resulta da combinação de informações prévias, sumarizadas em uma distribuição denominada priori, com dados

estatísticos descritos por algum modelo probabilístico e resumidos na função de verossimilhança. A distribuição a posteriori representa todo o conhecimento existente sobre o problema investigado. Toda pergunta específica é respondida a partir da análise da distribuição posteriori, pois é ela que contém toda a informação necessária para a inferência. A distribuição a posteriori é o elemento chave de toda análise bayesiana de dados, pois podemos examinar qualquer aspecto do parâmetro de interesse, como a média, mediana, moda, variância, percentis, etc.

A análise bayesiana de decisão pode ser utilizada como critério na escolha de estimadores e, dessa forma, compor a inferência bayesiana. No enfoque bayesiano, inferência nada mais é que uma forma particular de análise de decisão em presença de incerteza.

Definição 5.1 *O estimador de Bayes do parâmetro θ , denotado por $\delta_B(x)$, é o valor que minimiza a perda esperada com relação à distribuição a posteriori*

$$\delta_B(x) = \min_{\delta(x)} \int L(\theta, \delta(\theta)) \pi(\theta|x) d\theta = E^{\pi(\theta|x)}[L(\theta, \delta(x))]$$

em que $L(\theta, \delta(\theta))$ é a função que determina o cálculo de perda e $\pi(\theta|x)$ é a distribuição a posteriori.

Esta definição indica que o estimador de Bayes dependerá das incertezas (resumidas na posteriori) e da função de perda que quantifica as consequências. As funções mais utilizadas são:

a. **Perda quadrática:** $L(\theta, \delta(x)) = [\theta - \delta(x)]^2$

Neste caso, pode-se mostrar que ao resolver a equação apresentada na definição 5.1, o estimador de Bayes é $\delta_B(x) = E^{\pi(\theta|x)}[\theta]$, ou seja, a *média* da distribuição a posteriori.

b. **Perda absoluta:** $L(\theta, \delta(x)) = |\theta - \delta(x)|$

Aqui a solução para o estimador de Bayes é a *mediana* da distribuição a posteriori.

c. **Perda 0 - 1:**

$$L(\theta, \delta(\theta)) = \begin{cases} 0, & \text{se } \theta = \delta(x) \\ 1, & \text{caso contrário} \end{cases}$$

Neste caso o estimador de Bayes é a maior *moda* da distribuição a posteriori.

Na análise bayesiana, a definição da variância a posteriori do estimador é

$$V_{\delta}(x) = E^{\pi(\theta|x)}[(\theta - \delta(x))^2] = V(\theta|x) + [E^{\pi}[\theta|x] - \delta(x)]^2.$$

Uma alternativa bayesiana ao intervalo de confiança (IC) empregado na estatística convencional, é o *intervalo de credibilidade* (ICr).

Definição 5.2 *O intervalo de credibilidade de 95% ($ICr_{95\%}$) para θ é o intervalo delimitado pelos percentis 2,5% ($\theta_{[2,5\%]}$) e 97,5% ($\theta_{[97,5\%]}$) da distribuição a posteriori $\pi(\theta|x)$ para θ .*

O $ICr_{95\%}$ é então o intervalo de valores mais prováveis de θ , que soma probabilidade de 0,95.

Como resultado final das análises, costuma-se apresentar um resumo descritivo das distribuições marginais a posteriori dos parâmetros de interesse, incluindo valores da média, desvio padrão, mediana, moda e intervalos de credibilidade dessas distribuições, bem como seus gráficos.

5.1 Distribuições a priori

A distribuição a priori constitui-se no único elemento novo na análise bayesiana em relação à análise clássica. A premissa desta distribuição é que ela reflita a incerteza sobre os possíveis valores dos parâmetros antes da coleta dos dados. Nesta seção, serão discutidas algumas formas de especificação da distribuição a priori.

5.1.1 Priori não informativa

Existem situações em que o conhecimento sobre determinado fenômeno é vago ou inexistente. Nestes casos, a distribuição a priori é dita vaga, difusa ou não informativa, não precisa ou de variância alta, significando que a densidade a priori reflete ignorância (Jeffreys, 1998), sendo caracterizada pela ausência total ou quantidade mínima de informação, pois não privilegia nenhum valor do parâmetro sobre qualquer outro. Também utiliza-se

esse tipo de priori quando se espera que os dados transmitam toda a informação necessária, isto é, a informação dos dados seja dominante.

Método de Bayes-Laplace

A primeira ideia intuitiva para gerar distribuições não informativas foi o Princípio da Razão Insuficiente devido a Bayes e Laplace. De acordo com este princípio, na ausência de razão suficiente para privilegiar umas possibilidades em detrimento a outras, decorrente da escassez informativa a priori, deve-se optar pela equiprobabilidade.

No caso em que Θ é finito, digamos $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$, a distribuição não informativa gerada por este argumento é a distribuição Uniforme Discreta

$$\pi(\theta) \propto c, \quad \theta \in \Theta$$

No entanto, se Θ for um conjunto infinito enumerável, então a distribuição a priori é imprópria, ou seja,

$$\int_{\Theta} \pi(\theta) d\theta = \infty$$

pois não fica garantida a equiprobabilidade de todos os valores de θ .

No caso em que Θ é infinito e não enumerável, o Princípio da Razão Insuficiente é interpretado de modo a conduzir à distribuição Uniforme Contínua, que é, novamente imprópria se Θ não é limitado.

Sendo assim, faz-se necessário verificar se a distribuição a posteriori obtida com densidades a priori impróprias é própria antes de se fazer qualquer inferência. Uma outra objeção que veio a se levantar quanto à ideia de representar a falta de informação por distribuições uniformes é a sua inconsistência no sentido em que se $\psi = \psi(\theta)$ é uma transformação injetiva de um parâmetro θ , que assume uma gama contínua de valores possíveis, as distribuições uniformes para θ e ψ não são compatíveis. De fato, seja $h(\theta)$ uma distribuição a priori para θ , então

$$h(\psi) = h[\theta(\psi)] \left| \frac{d\theta}{d\psi} \right| \quad (5.1)$$

deve ser a correspondente distribuição para a reparametrização injetiva ψ , que não é necessariamente uniforme quando $h(\theta)$ o é (considere transformações não lineares).

Método de Jeffreys

A crítica da inconsistência da distribuição uniforme na representação formal da ignorância implica que esta deve ser invariante sob transformações injetivas. Entre os procedimentos que asseguram esta invariância está aquele proposto por Jeffreys, sendo baseado no uso da informação esperada de Fisher sobre $\theta \in \mathfrak{R}$

$$I(\theta) = E \left[\left(\frac{\partial \log L(x|\theta)}{\partial \theta} \right)^2 \middle| \theta \right] = -E \left[\frac{\partial^2 \log L(x|\theta)}{\partial \theta^2} \middle| \theta \right]$$

Seja $\psi = g(\theta)$ e $\theta = g^{-1}(\psi)$, então

$$\begin{aligned} I(\psi) &= -E \left[\frac{\partial^2 \log L(x|\psi)}{\partial \psi^2} \right] \\ &= -E \left[\frac{\partial^2 \log L(x|\theta = g^{-1}(\psi))}{\partial \theta^2} \right] \left| \frac{dg^{-1}(\psi)}{d\psi} \right|^2 \\ &= I(\theta) \left| \frac{d\theta}{d\psi} \right|^2 \end{aligned}$$

Extraindo raiz quadrada nesta última igualdade e, em seguida comparando com 5.1, tem-se

$$h(\theta) \propto |I(\theta)|^{\frac{1}{2}}$$

e assim pode-se enunciar a priori de Jeffreys para o caso uniparamétrico.

Definição 5.3 *Seja uma observação X com função densidade de probabilidade $h(x|\theta)$. A priori de Jeffreys tem função densidade de probabilidade dada por*

$$h(\theta) \propto |I(\theta)|^{1/2}.$$

Se Θ for um vetor com múltiplos parâmetros, digamos $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$, então a matriz

de informação de Fisher é $I(\Theta) = \begin{pmatrix} I_{11}(\Theta) & I_{12}(\Theta) & \cdots & I_{1k}(\Theta) \\ \cdots & \cdots & \cdots & \cdots \\ I_{k1}(\Theta) & I_{k2}(\Theta) & \cdots & I_{kk}(\Theta) \end{pmatrix}$ e então a priori não

informativa de Jeffreys é $\pi(\Theta) \propto [|\det \mathbf{I}(\Theta)|]^{1/2}$.

Uma simplificação geralmente adotada para modelos multiparamétricos é supor a independência a priori entre os parâmetros θ_i , $i = 1, 2, \dots, k$. Neste caso, a densidade a priori não informativa de Jeffreys é

$$h(\theta) = [\prod_{i=1}^k |diag I(\Theta)|]^{1/2}.$$

5.1.2 Priori conjugada

Existem situações práticas em que há informação a priori mais ou menos substancial sobre os parâmetros do modelo, quer por parte do pesquisador ou por parte de outros indivíduos que o pesquisador pode acorrer. A questão de como obter e quantificar essa informação de modo a transformá-la em uma distribuição a priori que possa ser utilizada para prosseguir com a metodologia bayesiana é um tema que vem sendo amplamente tratado na literatura científica. A escolha da densidade a priori de um parâmetro $\theta \in \Theta$ deve ser baseada em questões relacionadas diretamente a ele e isso depende do tipo de parâmetro, dos conhecimentos do pesquisador, entre outros.

Uma densidade a priori é dita conjugada com a função de verossimilhança se a distribuição a posteriori resultante é da mesma família de distribuições da priori.

Uma vantagem do uso de priori conjugadas é que, em geral, facilita a análise visto que as distribuições a priori e a posteriori estão em uma mesma família de distribuições e uma desvantagem, é que nem sempre a distribuição utilizada representa bem a incerteza a priori.

Definição 5.4 *Se $F = \{p(x|\theta), \theta \in \Theta\}$ é uma classe de distribuições amostrais, então uma classe de distribuições P é conjugada a F se*

$$\forall p(x|\theta) \in F \text{ e } p(\theta) \in P \Rightarrow p(\theta|x) \in P.$$

Uma família de distribuições conjugadas pode ser obtida considerando:

- a. Identificar a classe P de distribuições para θ tal que $\ell(\theta; x)$ seja proporcional a um membro desta classe,
- b. Verificar se P é fechada por amostragem, isto é, se $\forall p_1, p_2 \in P \exists k$ tal que $kp_1p_2 \in P$.

Se além disso existir uma constante k tal que $k^{-1} = \int \ell(\theta; x)d\theta < \infty$ e todo $p \in P$ é definido como $p(\theta) = k\ell(\theta; x)$ então P é a *família conjugada natural* ao modelo amostral gerador de $\ell(\theta; x)$.

Existe também uma outra forma de conjugação de priori, que são as priori conjugadas na família exponencial, bastante utilizadas por incluir grande parte das distribuições contínuas e discretas. Para maiores detalhes desta forma de conjugação, ver Ehlers (2007).

5.2 Abordagem bayesiana para o modelo ARMA

Para o desenvolvimento desta seção é conveniente referir-se ao parâmetro $\tau = 1/\sigma^2$, definido como sendo o inverso da variância, chamado de precisão. Observe que quanto maior a variância, menor será a precisão e vice-versa. Portanto no restante desta seção uma distribuição normal será identificada pela sua precisão e não pelo desvio padrão: a densidade de uma variável aleatória Y com distribuição normal com média μ e precisão τ , $N \sim (\mu, \tau)$, será

$$f(y|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp \left[-\frac{\tau(y - \mu)^2}{2} \right].$$

Considere n observações correspondentes ao processo dado pela equação (2.1) e seja $\mathbf{S} = (y_1, y_2, \dots, y_n)'$ o vetor de observações. Podemos reescrever (2.1) condicionando nas $r = \max(p, q)$ primeiras observações e obter os últimos $n - r$ erros, ou seja

$$a_t = y_t - \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j}$$

com $t = r + 1, r + 2, \dots, n$. Assuma que $a_p = a_{p-1} = a_{p-2} = \dots = a_{1-q} = 0$.

Assumindo que a_t são independentes e identicamente distribuídos com $N(0, \tau)$, temos uma amostra aleatória de tamanho $n - r$ de uma distribuição normal, $a_t \stackrel{iid}{\sim} N(0, \tau)$ e, sob essa suposição, constrói-se a função de verossimilhança aproximada

$$\begin{aligned} L(\phi, \theta, \tau | \mathbf{S}) &= \prod_{t=r+1}^n \sqrt{\frac{\tau}{2\pi}} \exp \left\{ -\frac{\tau}{2} a_t^2 \right\} \\ &\propto \tau^{\frac{n-r}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n a_t^2 \right\} \\ &\propto \tau^{\frac{n-r}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n \left[y_t - \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} \right]^2 \right\} \end{aligned}$$

O modelo escrito na forma vetorial é

$$y_t = \mathbf{W}_t' \phi + a_t + \mathbf{b}_t' \theta$$

em que $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ e $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ são os vetores de parâmetros do modelo, $\mathbf{W}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$ e $\mathbf{b}_t = (a_{t-1}, a_{t-2}, \dots, a_{t-q})'$, para $t = r + 1, \dots, n$.

A função de verossimilhança aproximada fica

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau | \mathbf{W}_t, \mathbf{b}_t) \propto \tau^{\frac{n-r}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n \left[y_t - \mathbf{W}'_t \boldsymbol{\phi} - \mathbf{b}'_t \boldsymbol{\theta} \right]^2 \right\} \quad (5.2)$$

em que

$$\begin{aligned} (y_t - \mathbf{W}'_t \boldsymbol{\phi} - \mathbf{b}'_t \boldsymbol{\theta})^2 &= (y_t - \mathbf{W}'_t \boldsymbol{\phi} - \mathbf{b}'_t \boldsymbol{\theta})' (y_t - \mathbf{W}'_t \boldsymbol{\phi} - \mathbf{b}'_t \boldsymbol{\theta}) \\ &= [(y_t - \mathbf{W}'_t \boldsymbol{\phi})' - (\mathbf{b}'_t \boldsymbol{\theta})'] (y_t - \mathbf{W}'_t \boldsymbol{\phi} - \mathbf{b}'_t \boldsymbol{\theta}) \\ &= (y_t - \boldsymbol{\phi}' \mathbf{W}_t - \boldsymbol{\theta}' \mathbf{b}_t) (y_t - \mathbf{W}'_t \boldsymbol{\phi} - \mathbf{b}'_t \boldsymbol{\theta}) \\ &= y_t^2 - y_t \mathbf{W}'_t \boldsymbol{\phi} - y_t \mathbf{b}'_t \boldsymbol{\theta} - \boldsymbol{\phi}' \mathbf{W}_t y_t + \boldsymbol{\phi}' \mathbf{W}_t \mathbf{W}'_t \boldsymbol{\phi} + \boldsymbol{\phi}' \mathbf{W}_t \mathbf{b}'_t \boldsymbol{\theta} - \\ &\quad - \boldsymbol{\theta}' \mathbf{b}_t y_t + \boldsymbol{\theta}' \mathbf{b}_t \mathbf{W}'_t \boldsymbol{\phi} + \boldsymbol{\theta}' \mathbf{b}_t \mathbf{b}'_t \boldsymbol{\theta} \end{aligned}$$

Para fazer um estudo bayesiano precisamos de uma distribuição a priori para os vetores de parâmetros $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ e τ do modelo ARMA(p, q) descrito na equação (2.1). Poderíamos utilizar densidades a priori informativas, da classe conjugada. No entanto, iremos utilizar priori não-informativas assumindo que $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ e τ são independentes de modo que $\pi_0(\boldsymbol{\phi}) \propto \text{constante}$, $\pi_0(\boldsymbol{\theta}) \propto \text{constante}$ e $\pi_0(\tau) \propto \tau^{-1}$ (priori de Jeffrey's). Assim, a priori conjunta pra $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ e τ é

$$\begin{aligned} \pi_0(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau) &= \pi_0(\boldsymbol{\phi}) \pi_0(\boldsymbol{\theta}) \pi_0(\tau) \\ &\propto \tau^{-1} \end{aligned} \quad (5.3)$$

A distribuição a posteriori é obtida combinando-se a função de verossimilhança do modelo, dada em (5.2) e a priori conjunta, dada por (5.3), ou seja

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau) \propto \tau^{\frac{n-r}{2}-1} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n \left[y_t - \mathbf{W}'_t \boldsymbol{\phi} + \mathbf{a}'_t \boldsymbol{\theta} \right]^2 \right\} \quad (5.4)$$

Encontrada a distribuição a posteriori, o passo seguinte é encontrar as densidades condicionais dos parâmetros do modelo, para que assim possamos gerar amostras a posteriori de cada parâmetro usando algum método iterativo, como por exemplo o método de simulação de Monte Carlo com Cadeia de Markov (MCMC). Comumente usa-se este método, pois, em geral, calcular estimativas a posteriori com as densidades a posteriori desconhecidas, só pode ser feito numericamente. As distribuições condicionais para cada parâmetro são:

- Densidade condicional para τ

$$\tau|\phi, \theta, \mathbf{W}_t, \mathbf{b}_t \sim \text{Gama} \left(\frac{n-r}{2}, 0.5 \sum_{t=r+1}^n [y_t - \mathbf{W}'_t \phi + \mathbf{a}'_t \theta]^2 \right)$$

- Densidade condicional para ϕ

Considere a forma

$$(\phi - \mu_\phi)' \Sigma (\phi - \mu_\phi) = \phi' \Sigma \phi - \phi' \Sigma \mu_\phi - \mu'_\phi \Sigma \phi + \mu'_\phi \mu_\phi \quad (5.5)$$

e seja

$$\begin{aligned} S^2(\phi|\theta) &\propto -y_t \mathbf{W}'_t \phi - \phi' \mathbf{W}_t y_t + \phi' \mathbf{W}_t \mathbf{W}'_t \phi + \phi' \mathbf{W}_t \mathbf{b}'_t \theta + \theta' \mathbf{b}_t \mathbf{W}'_t \phi \\ &\propto \phi' \mathbf{W}_t \mathbf{W}'_t \phi - \phi' (\mathbf{W}_t y_t - \mathbf{W}_t \mathbf{b}'_t \theta) - (y_t \mathbf{W}'_t - \theta' \mathbf{b}_t \mathbf{W}'_t) \phi \end{aligned}$$

Comparando $S^2(\phi|\theta)$ com (5.5) tem-se que

- $\Sigma_\phi = \sum_{t=r+1}^n (\mathbf{W}_t \mathbf{W}'_t)$, de ordem $p \times p$,
- $\mu_\phi = (\sum_{t=r+1}^n (\mathbf{W}_t \mathbf{W}'_t))^{-1} (\sum_{t=r+1}^n (\mathbf{W}_t y_t - \mathbf{W}_t \mathbf{b}'_t \theta))$, de ordem $p \times 1$.

Portanto,

$$\phi|\theta, \tau, \mathbf{W}_t, \mathbf{b}_t \propto N_p(\mu_\phi, \Sigma_\phi^{-1})$$

em que N_p denota a densidade Normal p-variada.

- Densidade condicional para θ

Procedendo de maneira análoga a densidade condicional para ϕ , obtém-se

$$\theta|\phi, \tau, \mathbf{W}_t, \mathbf{b}_t \propto N_q(\mu_\theta, \Sigma_\theta^{-1})$$

em que N_q denota a densidade Normal q-variada e

- $\Sigma_\theta = \sum_{t=r+1}^n (\mathbf{b}_t \mathbf{b}'_t)$, de ordem $q \times q$,
- $\mu_\theta = (\sum_{t=r+1}^n (\mathbf{b}_t \mathbf{b}'_t))^{-1} (\sum_{t=r+1}^n (\mathbf{b}_t y_t - \mathbf{b}_t \mathbf{W}'_t \phi))$, de ordem $q \times 1$.

5.3 Abordagem bayesiana para o modelo GARMA

A função de verossimilhança para o modelo é dada por

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | y_t, \mathbf{H}_t) \propto \exp \left\{ \sum_{t=r+1}^n \frac{y_t \theta_t - b(\theta_t)}{\varphi} + \sum_{t=r+1}^n c(y_t, \varphi) \right\}.$$

Nesta abordagem, serão consideradas distribuições a priori normais com média zero e variância σ^2 para ϕ_i , θ_j e α_0 , $i = 1, \dots, p$ e $j = 1, \dots, q$, isto é, $\phi_i \sim N(0, \sigma^2)$, $\theta_j \sim N(0, \sigma^2)$ e $\alpha_0 \sim N(0, \sigma^2)$. Sendo assim, supondo independência entre os parâmetros, a distribuição a priori conjunta é

$$\pi_0(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0) \propto -\frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=1}^p \phi_i^2 + \sum_{j=1}^q \theta_j^2 \right) \quad (5.6)$$

Para não confundir a notação, observe que $\boldsymbol{\theta}_j$ é o vetor de parâmetros e θ_t é o parâmetro canônico do modelo.

Portanto, a distribuição *a posteriori* para o modelo

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | y_t, \mathbf{H}_t) \propto \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | y_t, \mathbf{H}_t) \pi_0(\boldsymbol{\phi}, \boldsymbol{\theta}, \alpha_0 | y_t, \mathbf{H}_t)$$

A seguir são dados com maiores detalhes as distribuições a posteriori para as distribuições condicionais Poisson, Binomial e Binomial Negativa.

Modelo Poisson GARMA

A função de verossimilhança aproximada para o modelo Poisson GARMA é dada por

$$\begin{aligned} \mathcal{L}(\mu_t | y_t, \mathbf{H}_t) &= \prod_{t=r+1}^n \exp\{y_t \log(\mu_t) - \mu_t - \log(y_t!)\} \\ &= \exp \left\{ \sum_{t=r+1}^n [y_t \log(\mu_t)] - \sum_{t=r+1}^n \mu_t - \sum_{t=r+1}^n \log(y_t!) \right\}. \end{aligned} \quad (5.7)$$

De (3.8) tem-se que

$$\mu_t = \exp \left[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j} / \mu_{t-j}) \right] \quad (5.8)$$

e substituindo este resultado em (5.7) temos que a função de verossimilhança aproximada é

$$\begin{aligned} \mathcal{L}(\mu_t|y_t, \mathbf{H}_t) &\propto \exp \left\{ \sum_{t=r+1}^n \left[y_t \left(\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j}) \right) \right] \right. \\ &\quad \left. - \sum_{t=r+1}^n \left(\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j}) \right) \right\}. \end{aligned} \quad (5.9)$$

Então, a densidade a posteriori para o modelo Poisson GARMA é dada pela combinação de (5.6) e (5.9), isto é

$$\begin{aligned} \pi(\phi_i, \theta_j, \alpha_0|y_t, \mathbf{H}_t) &\propto \exp \left\{ \sum_{t=r+1}^n \left[y_t \left(\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j}) \right) \right] - \sum_{t=r+1}^n \right. \\ &\quad \left. \left(\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j}) \right) - \frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=p}^n \phi_i^2 + \sum_{j=q}^n \theta_j^2 \right) \right\}. \end{aligned} \quad (5.10)$$

Modelo Binomial GARMA

A função de verossimilhança aproximada para o modelo binomial GARMA é dada por

$$\begin{aligned} \mathcal{L}(\mu_t|y_t, \mathbf{H}_t) &= \prod_{t=r+1}^n \exp \left\{ y_t \log \left(\frac{\mu_t}{m - \mu_t} \right) + m \log \left(\frac{m - \mu_t}{m} \right) + \log \left(\frac{\Gamma(m+1)}{\Gamma(y_t+1)\Gamma(m-y_t+1)} \right) \right\} \\ &\propto \exp \left\{ \sum_{t=r+1}^n \left[y_t \log \left(\frac{\mu_t}{m - \mu_t} \right) \right] + \sum_{t=r+1}^n \left[m \log \left(\frac{m - \mu_t}{m} \right) \right] \right\}. \end{aligned}$$

Usando (5.8), temos a função de verossimilhança aproximada escrita na forma

$$\begin{aligned} \mathcal{L}(\mu_t|y_t, \mathbf{H}_t) &\propto \exp \left\{ \sum_{t=r+1}^n \left[y_t \log \left(\frac{\exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]}{m - \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]} \right) \right] \right. \\ &\quad \left. + \sum_{t=r+1}^n \left[m \log \left(\frac{m - \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]}{m} \right) \right] \right\}. \end{aligned} \quad (5.11)$$

Então, a densidade a posteriori para o modelo binomial GARMA é dada pela combinação de (5.6) e (5.11), ou seja

$$\begin{aligned} \pi(\phi_i, \theta_j, \alpha_0|y_t, \mathbf{H}_t) &\propto \exp \left\{ \sum_{t=r+1}^n \left[y_t \log \left(\frac{\exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]}{m - \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]} \right) \right] \right. \\ &\quad \left. + \sum_{t=r+1}^n \left[m \log \left(\frac{m - \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]}{m} \right) \right] \right. \\ &\quad \left. - \frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=p}^n \phi_i^2 + \sum_{j=q}^n \theta_j^2 \right) \right\}. \end{aligned} \quad (5.12)$$

Modelo Binomial Negativa GARMA

A função de verossimilhança aproximada para o modelo binomial negativa GARMA é dada por

$$\begin{aligned} f(y_t|\mathbf{H}_t) &= \prod_{t=r+1}^n \exp \left\{ k \log \left(\frac{k}{\mu_t + k} \right) + y_t \log \left(\frac{\mu_t}{\mu_t + k} \right) + \log \left(\frac{\Gamma(k + y_t)}{\Gamma(y_t + 1)\Gamma(k)} \right) \right\} \\ &\propto \exp \left\{ k \sum_{t=r+1}^n \log \left(\frac{k}{\mu_t + k} \right) + \sum_{t=r+1}^n \left(y_t \log \left(\frac{\mu_t}{\mu_t + k} \right) \right) \right\} \end{aligned}$$

e substituindo (5.8) na equação acima, a função de verossimilhança aproximada fica escrita na forma

$$\begin{aligned} \mathcal{L}(\mu_t|y_t, \mathbf{H}_t) &\propto \exp \left\{ k \sum_{t=r+1}^n \log \left(\frac{k}{k + \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]} \right) + \right. \\ &\quad \left. \sum_{t=r+1}^n \left(y_t \log \left(\frac{\exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]}{k + \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]} \right) \right) \right\}. \quad (5.13) \end{aligned}$$

Então, a densidade a posteriori para o modelo binomial negativa GARMA é dada pela combinação de (5.6) e (5.13), ou seja

$$\begin{aligned} \pi(\mu_t|y_t, \mathbf{H}_t) &\propto \exp \left\{ k \sum_{t=r+1}^n \log \left(\frac{k}{k + \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]} \right) + \sum_{t=r+1}^n (y_t \right. \\ &\quad \left. \log \left(\frac{\exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]}{k + \exp[\alpha_0 + \sum_{i=1}^p \phi_i \log(y_{t-i}) + \sum_{j=1}^q \theta_j \log(y_{t-j}/\mu_{t-j})]} \right) \right) - \frac{1}{2\sigma^2} \left(\alpha_0^2 + \sum_{i=p}^n \phi_i^2 + \sum_{j=q}^n \theta_j^2 \right) \right\}. \quad (5.14) \end{aligned}$$

Em geral calcular estimativas a posteriori com as densidades apresentadas, só pode ser feito numericamente. Para isso comumente usa-se algum método de simulação de Monte Carlo com Cadeia de Markov (MCMC).

5.4 Monte Carlo com Cadeias de Markov

Metropolis-Hastings

O algoritmo de Metropolis-Hastings (Chib & Greenberg (1995)) é um método iterativo usado para gerar amostras de uma densidade de forma indireta, quando não se sabe gerar

diretamente dessa densidade. O algoritmo consiste em escolher um núcleo $q(\cdot|\theta)$, e gerar deste núcleo, usando um critério de aceitação do valor gerado para garantir que a amostra obtida é representativa da amostra gerada.

Seja $\pi(\theta)$ a densidade de probabilidade conjunta de interesse com $\theta = (\theta_1, \dots, \theta_k)$ e suponha que as densidades condicionais completas a posteriori $\pi_i(\theta_i) = \pi_i(\theta_i|\theta_{-i})$, $i = 1, 2, \dots, k$, estejam disponíveis.

O algoritmo é dado por:

Passo 1: Inicialize o contador de iterações da cadeia em $j = 1$ e escolha os valores iniciais

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)});$$

Passo 2: Gere um novo valor θ^* da distribuição $q(\cdot|\theta)$;

Passo 3: Calcule a probabilidade de aceitação $\alpha(\theta, \theta^*)$ e gere $u \sim U(0, 1)$;

Passo 4: Se $u \leq \alpha$ aceite o novo valor e faça $\theta^{(j+1)} = \theta^*$, caso contrário rejeite e faça $\theta^{(j+1)} = \theta$;

Passo 5: Atualizar o contador de j para $j + 1$ e repetir o passo 2 até a convergência,

no qual o novo valor θ^* é aceito com probabilidade

$$\alpha(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)} \right\}. \quad (5.15)$$

O conhecimento da densidade de probabilidade $\pi(\theta)$ pode ser parcial, isto é, a menos de uma constante a probabilidade em (5.15) não se altera. A cadeia pode permanecer no mesmo estado por muitas iterações e na prática é comum monitorar a porcentagem das iterações para as quais novos valores são aceitos.

Para verificar a convergência da cadeia, pode-se utilizar técnicas gráficas como: gráficos de séries temporais, em que a série deve estar em uma faixa, sem apresentar tendências nem oscilações bruscas; gráficos de autocorrelação dos parâmetros, sendo esperado que a primeira correlação seja alta e as demais próximas de zero, indicando que a amostra simulada é não correlacionada. Uma outra avaliação de convergência é feita usando-se técnicas numéricas. Os diagnósticos de convergência mais utilizados são descritos por Gelman & Rubin (1992),

em que consideram a análise de variância para verificar a convergência de cadeias paralelas (com valores iniciais super dispersos) e o critério Geweke (1992), que é baseado em métodos de séries temporais. Neste trabalho, os métodos de MCMC apresentados são utilizados na obtenção de estimativas para quantidades de interesse do modelo.

Para o caso em que as matrizes de transição (matriz de transição é um processo estocástico utilizado para estudar fenômenos que passam, a partir de um estado inicial, por uma sequência de ensaios, onde a transição de um determinado estado ocorre segundo uma certa probabilidade) são simétricas pode-se simplificar o algoritmo de Metropolis-Hastings para o algoritmo de Metropolis, em que a única diferença é que a probabilidade de aceitação é dada por

$$\alpha(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*)}{\pi(\theta)} \right\}. \quad (5.16)$$

A utilização da distribuição uniforme para servir como matriz de transição é apenas uma entre muitas opções. A distribuição normal centrada em θ_i é uma outra alternativa que pode ser usada.

Amostrador de Gibbs

Em muitas situações a geração de uma amostra pode ser complicada ou simplesmente impossível, no entanto, se as condicionais apresentam formas conhecidas (normal, gama, beta, etc.), então o método de Gibbs Sampling (Casella & George, 1992) é uma alternativa de geração de amostras. O Gibbs Sampling é um esquema iterativo de amostragem de uma cadeia de Markov cuja matriz de transição é constituído das distribuições condicionais completas, fornecendo, então, uma alternativa de geração baseada em sucessivas gerações das distribuições condicionais completas.

Seja $\pi(\theta)$ a densidade de probabilidade conjunta de interesse com $\theta = (\theta_1, \dots, \theta_k)$ e suponha que as densidades condicionais completas a posteriori $\pi_i(\theta_i) = \pi_i(\theta_i | \theta_{-i})$, $i = 1, 2, \dots, k$, sejam conhecidas.

O algoritmo é dado por:

Passo 1: Inicialize o contador de iterações da cadeia em $j = 1$ e escolha os valores iniciais

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)});$$

Passo 2: Obtenha um novo valor $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$ a partir de $\boldsymbol{\theta}^{(j-1)}$ através de sucessivas gerações de valores

$$\begin{aligned}\theta_1^{(j)} &\sim \left(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)}\right) \\ \theta_2^{(j)} &\sim \left(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)}\right) \\ &\vdots \\ \theta_k^{(j)} &\sim \left(\theta_k | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{k-1}^{(j)}\right);\end{aligned}$$

Passo 3: Atualizar o contador de j para $j + 1$ e repetir o passo 2 até a convergência.

À medida que o número de iterações aumenta, a cadeia se aproxima de sua condição de equilíbrio. Podemos assumir que a convergência é atingida em uma iteração quando a Cadeia de Markov possui propriedades de regularidade, tais como, reversibilidade e pode-se garantir que o processo iterativo converge para a distribuição de equilíbrio $\pi(\boldsymbol{\theta})$.

5.5 Seleção de modelos

Assim como na metodologia clássica, a metodologia bayesiana também possui ferramentas para selecionar modelos ou ainda, a sua ordem. O critério BIC, definido na Seção 4.2, pode ser usado, usando algumas modificações, para tal feito, porém há critérios mais específicos e que se enquadram melhor nessa abordagem.

Raftery *et al.* (2006) propuseram o BICM (ou BIC-Monte Carlo), que é uma forma alternativa ao critério proposto por Schwarz. Essa proposta tem como objetivo incrementar as estimativas de Monte Carlo de cada parâmetro no critério BIC ((Schwarz, 1978)).

O BICM, definido por (Raftery *et al.*, 2006), é dado por

$$BICM = \sum_{i=1}^M \{2\ell(\hat{\boldsymbol{\gamma}}^{(i)}) + k \log(n)\}, \quad (5.17)$$

em que M é o tamanho do vetor de amostras selecionadas pelo processo de aceitação/rejeição do método MCMC em uso, $\ell(\hat{\boldsymbol{\gamma}}^{(i)})$ é a log-verossimilhança avaliada em cada estimativa dos parâmetros, k é o número de parâmetros do modelo e n é o número de observações da série.

O modelo que deve ser escolhido de acordo com o menor valor de (5.17), analogamente ao BIC usual.

5.6 Previsão para o modelo GARMA

Analogamente a previsão para a inferência clássica (Seção 4.4), no estudo bayesiano, a previsão bayesiana também é feita utilizando o preditor linear, porém com algumas mudanças.

Ao longo do processo de simulação MCMC são geradas amostras para os parâmetros do modelos, as quais são rejeitadas ou não (Metropolis ou Metropolis-Hastings), ou então nunca são rejeitadas (Gibbs sampling). Neste trabalho o método utilizado é o de Metropolis e, portanto, são guardadas amostras dos valores não rejeitados dos parâmetros, é dado um salto para poder retirar a correlação e, com os valores finais guardados, são calculadas a média, moda ou mediana para obter as estimativas dos parâmetros.

Dessa forma, não faz muito sentido calcular a previsão como na forma clássica, onde são utilizadas as estimativas dos parâmetros, neste caso a média, moda ou mediana. Na amostra final guardada, cada estimativa do parâmetro produzirá uma previsão diferente, sendo necessário, então, utilizar a média da previsão calculada para cada estimativa como a previsão obtida. Ou seja, as previsões h-passos a frente é dada por

$$\hat{\mu}_{t+h} = \frac{1}{M} \sum_{i=1}^M \exp \left(\hat{\alpha}_{0,i} + \sum_{j=1}^p \hat{\phi}_{j,i} \log(y_{t+h-j}) + \sum_{j=1}^q \hat{\theta}_{j,i} \log(y_{t+h-j} - \mu_{t+h-j}) \right)$$

Capítulo 6

Aplicação com dados simulados

Neste capítulo é apresentado o método de geração de dados e para facilitar o entendimento, foi conduzido um estudo de simulação para o modelo GARMA(p,q), com as distribuições Poisson, Binomial e Binomial Negativa.

O uso de dados simulados, em geral, tem como objetivo verificar a viabilidade do modelo proposto, a qualidade do ajuste, a qualidade das estimativas calculadas (visto que é possível comparar os verdadeiros valores dos parâmetros com os valores estimados) e estudar propriedades de estimadores, como consistência e vício.

A teoria de MLG diz que para cada observação da variável resposta tem-se uma distribuição com uma determinada média e variância, sendo que a média é referente a cada observação e a variância é dada pela função de variância. Para exemplificar, suponha que temos um conjunto de observações y_t , $t = 1, \dots, n$, e que esses dados são provenientes de uma distribuição de Poisson com média μ_t . Então $y_t \sim P(\mu_t)$, $t = 1, \dots, n$.

De maneira análoga, nos modelos GARMA tem-se que cada observação terá média μ_t variando no tempo e como visto no Capítulo 3, μ_t está associada a um preditor linear, definido por (3.3), e através deste são geradas médias μ_t , tornando possível gerar observações y_t de acordo com a distribuição desejada.

Suponha por exemplo, o modelo Poisson GARMA com $p = 2$ e $q = 2$. O preditor linear é dado por $\mu_t = \alpha_0 + \phi_1 \log(y_{t-1}) + \phi_2 \log(y_{t-2}) + \theta_1 \log(y_{t-1}/\mu_{t-1}) + \theta_2 \log(y_{t-2}/\mu_{t-2})$, ou generalizando, $g(\mu_t) = \eta_t = \alpha_0 + \sum_{j=1}^2 \phi_j g(y_{t-j}) + \sum_{j=1}^2 \theta_j [g(y_{t-j}) - g(\mu_{t-j})]$, em que $g(\cdot)$

é a função de ligação conhecida. As médias são gerados atribuindo valores iniciais, neste caso, para $\mu_1, \mu_2, y_2, y_1, \alpha_0, \phi_1, \phi_2, \theta_1$ e θ_2 . Com o valor das médias geradas, uma nova observação é gerada de uma distribuição Poisson de média μ_t calculada anteriormente, ou seja, $y_3 \sim P(\mu_3)$, com $\mu_3 = \alpha_0 + \phi_1 \log(y_2) + \phi_2 \log(y_1) + \theta_1 \log(y_2/\mu_2) + \theta_2 \log(y_1/\mu_1)$, $y_4 \sim P(\mu_4)$, com $\mu_4 = \alpha_0 + \phi_1 \log(y_3) + \phi_2 \log(y_2) + \theta_1 \log(y_3/\mu_3) + \theta_2 \log(y_2/\mu_2)$ e assim sucessivamente até y_n .

O algoritmo abaixo exemplifica o processo de simulação de dados que foi descrito para um modelo GARMA(p,q) mais geral. Todo o processo de simulação e aplicação foi realizado no software livre R Development Core Team (2010).

1. Atribua valores para os parâmetros $\varphi, \alpha_0, \boldsymbol{\phi}$ e $\boldsymbol{\theta}$, e também para observações iniciais y_i e $\mu_i, i = 1, \dots, \max(p, q)$;
2. Calcular a média usando $g(\mu_t) = \alpha_0 + \sum_{i=1}^p \phi_i g(y_{t-i}) + \sum_{j=1}^q \theta_j (g(y_{t-j}) - g(\mu_{t-j}))$;
3. Gerar $y_t \sim \Psi(\mu_t, \varphi)$, em que Ψ é uma distribuição pertencente à família exponencial dada em (3.2);
4. Repetir os passos acima até que o número desejado de observações seja atingido.

Uma dificuldade encontrada, neste trabalho foi que, no estudo clássico, não foi possível implementar a parte MA do modelo, pois a matriz $\widehat{\mathbf{M}}' \widehat{\mathbf{W}} \widehat{\mathbf{M}}$ (ver a equação 4.2) mostrou-se singular, além de terem sido encontrados problemas com a convergência dos parâmetros, que não foi alcançada usando o método Escore de Fisher. Para os conjuntos de dados reais o problema persistiu e não foi possível ajustar aos dados um modelo que incorporasse os termos de médias móveis. Em virtude desses problemas encontrados, as aplicações nesse trabalho contam apenas com os parâmetros autorregressivos do modelo. Os problemas aqui encontrados, precisam ser estudados com maiores detalhes.

6.1 Estudo de simulação

O interesse desta seção é analisar o desempenho do percentil de cobertura gerado por intervalos de confiança e confirmar, por meio de simulação, a consistência e calcular o viés e o erro quadrático médio dos estimadores pontuais clássicos.

O percentil de cobertura descreve a proporção de intervalos gerados que contém o verdadeiro valor do parâmetro, o qual é fixado ao realizar o estudo de simulação. O ideal é que esta taxa de cobertura dos intervalos de confiança estejam muito próxima do nível estipulado, $100(1 - \alpha)\%$, independente do valor do parâmetro.

Para atingir este objetivo foram simuladas 10.000 réplicas dos processos $\text{GAR}(p)$, para valores de p variando entre 1, 2 e 3, com as distribuições Poisson, binomial e binomial negativa com tamanho amostral igual a 30, 300 e 1000. No caso da distribuição binomial, a simulação foi repetida para valores de $m = 10, 30$ e 100; e para a distribuição binomial negativa, $k = 15, 20$ e 30. O nível de confiança fixado para a construção dos intervalos foi de 95%.

Para cada amostra gerada, ajustou-se o modelo e calculou-se os intervalos de confiança estimados para os parâmetros do modelo e então, contamos o número de tais intervalos que continham o verdadeiro valor do parâmetros para obter o percentil de cobertura.

As Tabelas 6.15 a 6.35 apresentam os percentis de cobertura para intervalos de confiança de 95% com base no método de máxima verossimilhança proposto no Capítulo 4, Seção 4.1. Pode-se dizer que com o estudo de simulação desenvolvido, há evidências de que os intervalos de confiança possuem boas propriedades. Também observa-se que os valores dos parâmetros estimados tornam-se próximos do valor real a medida em que se aumenta o tamanho da amostra. Ainda, como o vício e o MSE dos estimadores tendem a zero, há indícios de que os EMV são assintoticamente não viesados e consistentes para os modelos em questão.

Usualmente quando o tamanho da amostra aumenta, o percentil de cobertura para os parâmetros tende a ser próximo dos intervalos de confiança de 95%, de acordo com a teoria assintótica. Nota-se que o método utilizado é robusto, pois quando a ordem dos modelos é pequena, os percentis variam pouco com o tamanho da amostra, e para ordem maiores,

como por exemplo nas tabelas 6.17 e 6.23, verifica-se que estes percentuais chegam a valores próximos de 95% somente para $n=300$.

Tabela 6.1: Estudo de simulação para o modelo Poisson GAR(1), com $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9785	0.1733	0.0214	0.0304	95.72
	ϕ_1	-0.3896	0.1268	-0.0103	0.0158	95.56
300	α_0	1.9975	0.0467	0.0024	0.0021	95.43
	ϕ_1	-0.3975	0.0341	-0.0024	0.0011	95.60
1000	α_0	1.9988	0.0252	0.0011	0.0006	94.57
	ϕ_1	-0.3981	0.0185	-0.0018	0.0003	94.95

Tabela 6.2: Estudo de simulação para o modelo Poisson GAR(2), com $\alpha_0 = 1.5$, $\phi_1 = 0.7$ e $\phi_2 = -0.2$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.7612	0.5520	-0.2612	0.4463	94.57
	ϕ_1	0.6523	0.1936	0.0476	0.0808	94.01
	ϕ_2	-0.2412	0.1886	0.0412	0.0475	95.60
300	α_0	15.293	0.1508	-0.0293	0.0235	94.78
	ϕ_1	0.6945	0.0562	0.0054	0.0032	94.34
	ϕ_2	-0.2045	0.0543	0.0045	0.0029	95.09
1000	α_0	1.5070	0.0816	-0.0070	0.0068	94.89
	ϕ_1	0.6990	0.0306	0.0009	0.0009	94.73
	ϕ_2	-0.2014	0.0296	0.0014	0.0008	94.77

Tabela 6.3: Estudo de simulação para o modelo Poisson GAR(3), com $\alpha_0 = 2.0$, $\phi_1 = 0.4$, $\phi_2 = -0.2$ e $\phi_3 = -0.6$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.7124	0.1917	0.2875	1.8649	87.47
	ϕ_1	0.5568	0.1150	-0.1568	0.6537	85.98
	ϕ_2	-0.2944	0.1147	0.6944	0.4086	85.88
	ϕ_3	-0.6023	0.1042	1.0023	0.2129	87.66
300	α_0	2.0128	0.0469	-0.0128	0.0023	94.28
	ϕ_1	0.3940	0.0281	0.0059	0.0008	94.74
	ϕ_2	-0.1968	0.0280	0.5968	0.0008	94.96
	ϕ_3	-0.5937	0.0257	0.9937	0.0006	95.3
1000	α_0	2.0141	0.0253	-0.0141	0.0008	91.82
	ϕ_1	0.3939	0.0151	0.0060	0.0002	93.40
	ϕ_2	-0.1962	0.0150	0.5962	0.0002	94.69
	ϕ_3	-0.5943	0.0138	0.9943	0.0002	93.79

Tabela 6.4: Estudo de simulação para o modelo binomial GAR(1), com $m = 10$, $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9855	0.1852	0.0144	0.0258	97.94
	ϕ_1	-0.3933	0.1340	-0.0066	0.0149	96.95
300	α_0	1.9972	0.0514	0.0027	0.0016	98.40
	ϕ_1	-0.3974	0.0373	-0.0025	0.0009	97.44
1000	α_0	1.9986	0.0278	0.0013	0.0004	98.56
	ϕ_1	-0.3984	0.0202	-0.0015	0.0002	97.77

Tabela 6.5: Estudo de simulação para o modelo binomial GAR(1), com $m = 30$, $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9815	0.1830	0.0184	0.0302	97.00
	ϕ_1	-0.3914	0.1343	-0.0085	0.0167	96.45
300	α_0	1.9968	0.0517	0.0031	0.0023	96.19
	ϕ_1	-0.3966	0.0380	-0.0033	0.0013	96.04
1000	α_0	1.9985	0.0280	0.0014	0.0007	96.25
	ϕ_1	-0.3976	0.0206	-0.0023	0.0003	96.04

Tabela 6.6: Estudo de simulação para o modelo binomial GAR(1), com $m = 100$, $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9804	0.1830	0.0195	0.0327	96.04
	ϕ_1	-0.3895	0.1349	-0.0104	0.0173	95.98
300	α_0	1.9985	0.0517	0.0014	0.0025	95.73
	ϕ_1	-0.3975	0.0382	-0.0024	0.0013	95.70
1000	α_0	1.9984	0.0281	0.0015	0.0007	95.19
	ϕ_1	-0.3973	0.0207	-0.0026	0.0004	95.32

Tabela 6.7: Estudo de simulação para o modelo binomial GAR(2), com $m = 10$, $\alpha_0 = 1.0$, $\phi_1 = 0.5$ e $\phi_2 = -0.3$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.0619	0.2306	-0.0619	0.0547	94.91
	ϕ_1	0.4543	0.1733	0.0456	0.0304	94.54
	ϕ_2	-0.3118	0.1547	0.0118	0.0187	97.33
300	α_0	1.0192	0.0652	-0.0192	0.0046	94.24
	ϕ_1	0.4828	0.0502	0.0171	0.0027	93.87
	ϕ_2	-0.2954	0.0652	-0.0045	0.0016	96.12
1000	α_0	1.0153	0.0354	-0.0153	0.0014	92.33
	ϕ_1	0.4857	0.0273	0.0142	0.0009	91.66
	ϕ_2	-0.2945	0.0237	-0.0054	0.0005	96.00

Tabela 6.8: Estudo de simulação para o modelo binomial GAR(2), com $m = 30$, $\alpha_0 = 1.0$, $\phi_1 = 0.5$ e $\phi_2 = -0.3$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.0644	0.1574	-0.0644	0.0568	94.88
	ϕ_1	0.4527	0.1757	0.0472	0.0309	94.61
	ϕ_2	-0.3125	0.2340	0.0125	0.0211	96.31
300	α_0	1.0204	0.0674	-0.0204	0.0046	94.83
	ϕ_1	0.4821	0.0514	0.0178	0.0027	94.57
	ϕ_2	-0.2944	0.0452	-0.0055	0.0019	95.87
1000	α_0	1.0194	0.0367	-0.0194	0.0016	92.34
	ϕ_1	0.4829	0.0280	0.0170	0.0010	91.30
	ϕ_2	-0.2931	0.0245	-0.0068	0.0005	95.06

Tabela 6.9: Estudo de simulação para o modelo binomial GAR(2), com $m = 100$, $\alpha_0 = 1.0$, $\phi_1 = 0.5$ e $\phi_2 = -0.3$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.0611	0.1588	-0.0611	0.0545	95.39
	ϕ_1	0.4538	0.1769	0.0461	0.0298	94.89
	ϕ_2	-0.3137	0.2352	0.0137	0.0208	96.64
300	α_0	1.0232	0.0682	-0.0232	0.0047	95.07
	ϕ_1	0.4807	0.0518	0.0192	0.0028	94.16
	ϕ_2	-0.2946	0.0456	-0.0053	0.0019	95.74
1000	α_0	1.0202	0.0371	-0.0202	0.0016	92.28
	ϕ_1	0.4821	0.0282	0.0178	0.0010	91.43
	ϕ_2	-0.2921	0.0248	-0.0078	0.0006	94.69

Tabela 6.10: Estudo de simulação para o modelo binomial GAR(3), com $m = 10$, $\alpha_0 = 1.0$, $\phi_1 = 0.4$, $\phi_2 = -0.2$ e $\phi_3 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.0304	0.2024	-0.0304	0.0339	96.37
	ϕ_1	0.3593	0.1862	0.0406	0.0287	95.79
	ϕ_2	-0.2039	0.1837	0.0039	0.0243	97.77
	ϕ_3	-0.3753	0.1706	-0.0246	0.0216	97.48
300	α_0	1.0302	0.0559	-0.0302	0.0036	92.39
	ϕ_1	0.3700	0.0512	0.0299	0.0030	92.22
	ϕ_2	-0.1864	0.0506	-0.0135	0.0022	95.93
	ϕ_3	-0.3806	0.0464	-0.0193	0.0020	95.21
1000	α_0	1.0311	0.0304	-0.0311	0.0018	93.58
	ϕ_1	0.3704	0.0278	0.0295	0.0015	94.20
	ϕ_2	-0.1855	0.0275	-0.0144	0.0008	93.84
	ϕ_3	-0.3813	0.0252	-0.0186	0.0008	93.70

Tabela 6.11: Estudo de simulação para o modelo binomial GAR(3), com $m = 30$, $\alpha_0 = 1.0$, $\phi_1 = 0.4$, $\phi_2 = -0.2$ e $\phi_3 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.0353	0.1785	-0.0353	0.0389	96.17
	ϕ_1	0.3538	0.1900	0.0461	0.0310	96.04
	ϕ_2	-0.2017	0.1906	0.0017	0.0275	97.69
	ϕ_3	-0.3715	0.2116	-0.0284	0.0252	97.35
300	α_0	1.0331	0.0590	-0.0331	0.0041	92.68
	ϕ_1	0.3694	0.0526	0.0305	0.0032	92.13
	ϕ_2	-0.1859	0.0527	-0.0140	0.0025	95.88
	ϕ_3	-0.3776	0.0490	-0.0223	0.0025	94.65
1000	α_0	1.0339	0.0320	-0.0339	0.0020	92.56
	ϕ_1	0.3702	0.0285	0.0297	0.0015	93.39
	ϕ_2	-0.1858	0.0286	-0.0141	0.0009	93.80
	ϕ_3	-0.3779	0.0266	-0.0220	0.0011	94.79

Tabela 6.12: Estudo de simulação para o modelo binomial GAR(3), com $m = 100$, $\alpha_0 = 1.0$, $\phi_1 = 0.4$, $\phi_2 = -0.2$ e $\phi_3 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.0310	0.1817	-0.0310	0.0404	95.90
	ϕ_1	0.3573	0.1921	0.0426	0.0318	95.91
	ϕ_2	-0.2066	0.1918	0.0066	0.0286	97.47
	ϕ_3	-0.3704	0.2133	-0.0295	0.0264	97.33
300	α_0	1.0338	0.0599	-0.0338	0.0043	92.45
	ϕ_1	0.3695	0.0530	0.0304	0.0033	92.34
	ϕ_2	-0.1864	0.0533	-0.0135	0.0026	95.76
	ϕ_3	-0.3758	0.0499	-0.0241	0.0026	94.33
1000	α_0	1.0345	0.0325	-0.0345	0.0021	93.77
	ϕ_1	0.3700	0.0287	0.0299	0.0015	96.34
	ϕ_2	-0.1853	0.0289	-0.0146	0.0009	95.72
	ϕ_3	-0.3767	0.0270	-0.0232	0.0011	94.88

Tabela 6.13: Estudo de simulação para o modelo binomial negativa GAR(1), com $k = 15$, $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9761	0.1835	0.0238	0.0378	94.16
	ϕ_1	-0.3872	0.1369	-0.0127	0.0194	94.93
300	α_0	1.9971	0.0526	0.0028	0.0030	93.90
	ϕ_1	-0.3958	0.0393	-0.0041	0.0016	94.47
1000	α_0	1.9987	0.0286	0.0012	0.0009	93.12
	ϕ_1	-0.3962	0.0213	-0.0037	0.0004	94.17

Tabela 6.14: Estudo de simulação para o modelo binomial negativa GAR(1), com $k = 20$, $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9787	0.1829	0.0212	0.0353	95.08
	ϕ_1	-0.3883	0.1360	-0.0116	0.0185	95.81
300	α_0	1.9976	0.0524	0.0023	0.0030	93.53
	ϕ_1	-0.3956	0.0390	-0.0043	0.0016	94.37
1000	α_0	1.9986	0.0284	0.0013	0.0009	93.55
	ϕ_1	-0.3966	0.0212	-0.0033	0.0004	93.97

Tabela 6.15: Estudo de simulação para o modelo binomial negativa GAR(1), com $k = 30$, $\alpha_0 = 2.0$ e $\phi_1 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	1.9789	0.1828	0.0210	0.0353	95.02
	ϕ_1	-0.3892	0.1358	-0.0107	0.0182	95.78
300	α_0	1.9964	0.0522	0.0035	0.0029	94.38
	ϕ_1	-0.3957	0.0388	-0.0042	0.0015	94.79
1000	α_0	1.9988	0.0283	0.0011	0.0008	93.70
	ϕ_1	-0.3968	0.0211	-0.0031	0.0004	94.48

Tabela 6.16: Estudo de simulação para o modelo binomial negativa GAR(2), com $k = 15$, $\alpha_0 = 2.0$, $\phi_1 = 0.6$ e $\phi_2 = -0.2$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	2.2321	0.6516	-0.2321	2.8450	93.87
	ϕ_1	0.5834	0.1930	0.0165	0.5379	93.85
	ϕ_2	-0.2598	0.1874	0.0598	0.1881	94.98
300	α_0	2.0250	0.1816	-0.0250	0.0333	94.68
	ϕ_1	0.5958	0.0563	0.0041	0.0032	93.94
	ϕ_2	-0.2036	0.0543	0.0036	0.0029	94.68
1000	α_0	2.0082	0.0984	-0.0082	0.0098	94.67
	ϕ_1	0.5983	0.0306	0.0016	0.0009	94.65
	ϕ_2	-0.2008	0.0294	0.0008	0.0009	93.78

Tabela 6.17: Estudo de simulação para o modelo binomial negativa GAR(2), com $k = 20$, $\alpha_0 = 2.0$, $\phi_1 = 0.6$ e $\phi_2 = -0.2$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	2.2702	0.6551	-0.2702	1.6245	94.52
	ϕ_1	0.5577	0.1931	0.0422	0.1303	94.11
	ϕ_2	-0.2429	0.1881	0.0429	0.0663	95.63
300	α_0	2.0302	0.1827	-0.0302	0.0338	94.85
	ϕ_1	0.5952	0.0564	0.0047	0.0032	94.84
	ϕ_2	-0.2046	0.0545	0.0046	0.0029	94.45
1000	α_0	2.0068	0.0990	-0.0068	0.0094	95.22
	ϕ_1	0.5989	0.0306	0.0010	0.0009	94.94
	ϕ_2	-0.2009	0.0296	0.0009	0.0008	94.60

Tabela 6.18: Estudo de simulação para o modelo binomial negativa GAR(2), com $k = 30$, $\alpha_0 = 2.0$, $\phi_1 = 0.6$ e $\phi_2 = -0.2$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	2.2952	0.6630	-0.2952	0.5663	94.16
	ϕ_1	0.5502	0.1942	0.0497	0.0758	94.57
	ϕ_2	-0.2414	0.1899	0.0414	0.0479	95.33
300	α_0	2.0258	0.1836	-0.0258	0.0337	95.22
	ϕ_1	0.5952	0.0564	0.0047	0.0031	95.13
	ϕ_2	-0.2031	0.0548	0.0031	0.0030	94.77
1000	α_0	2.0089	0.0298	-0.0089	0.0097	95.14
	ϕ_1	0.5983	0.0306	0.0016	0.0009	94.92
	ϕ_2	-0.2011	0.0995	0.0011	0.0008	94.62

Tabela 6.19: Estudo de simulação para o modelo binomial negativa GAR(3), com $k = 15$, $\alpha_0 = 2.0$, $\phi_1 = 0.5$, $\phi_2 = -0.2$ e $\phi_3 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	2.0047	0.3427	-0.0047	0.1160	94.97
	ϕ_1	0.4747	0.1740	0.0252	0.0300	94.52
	ϕ_2	-0.2015	0.1779	0.0015	0.03017	95.19
	ϕ_3	-0.3830	0.1512	-0.0169	0.0241	93.90
300	α_0	2.0073	0.0884	-0.0073	0.0084	93.59
	ϕ_1	0.4929	0.0466	0.0070	0.0023	93.50
	ϕ_2	-0.1979	0.0475	-0.0020	0.0025	93.04
	ϕ_3	-0.3979	0.0389	-0.0020	0.0018	92.35
1000	α_0	2.0056	0.0475	-0.0056	0.0025	93.01
	ϕ_1	0.4952	0.0251	0.0047	0.0007	92.59
	ϕ_2	-0.1979	0.0256	-0.0020	0.0007	92.54
	ϕ_3	-0.3987	0.0209	-0.0012	0.0005	91.50

Tabela 6.20: Estudo de simulação para o modelo binomial negativa GAR(3), com $k = 20$, $\alpha_0 = 2.0$, $\phi_1 = 0.5$, $\phi_2 = -0.2$ e $\phi_3 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	2.0047	0.3429	-0.0047	0.1156	95.06
	ϕ_1	0.4782	0.1737	0.0217	0.0291	95.41
	ϕ_2	-0.2057	0.1773	0.0057	0.0292	95.65
	ϕ_3	-0.3812	0.1508	-0.0187	0.0236	94.66
300	α_0	2.0026	0.0875	-0.0026	0.0081	94.00
	ϕ_1	0.4945	0.0464	0.0054	0.0023	93.95
	ϕ_2	-0.1978	0.0472	-0.0021	0.0024	93.44
	ϕ_3	-0.3971	0.0385	-0.0028	0.0017	92.74
1000	α_0	2.0047	0.0472	-0.0047	0.0023	94.07
	ϕ_1	0.4957	0.0251	0.0042	0.0006	93.81
	ϕ_2	-0.1979	0.0255	-0.0020	0.0007	92.83
	ϕ_3	-0.3990	0.0208	-0.0009	0.0005	92.69

Tabela 6.21: Estudo de simulação para o modelo binomial negativa GAR(3), com $k = 30$, $\alpha_0 = 2.0$, $\phi_1 = 0.5$, $\phi_2 = -0.2$ e $\phi_3 = -0.4$.

n	Parâmetros	Média	D.P. Médio	Vício	MSE	% Cobertura
30	α_0	2.0061	0.3415	-0.0061	0.1107	95.63
	ϕ_1	0.4769	0.1730	0.0230	0.0277	95.46
	ϕ_2	-0.2036	0.1767	0.0036	0.0277	96.34
	ϕ_3	-0.3829	0.1503	-0.0170	0.0216	95.27
300	α_0	2.0051	0.0869	-0.0051	0.0078	94.44
	ϕ_1	0.4937	0.0461	0.0062	0.0022	94.41
	ϕ_2	-0.1973	0.0470	-0.0026	0.0023	94.08
	ϕ_3	-0.3984	0.0382	-0.0015	0.0016	93.70
1000	α_0	2.0049	0.0469	-0.0049	0.0024	93.69
	ϕ_1	0.4954	0.0250	-0.0045	0.0006	93.93
	ϕ_2	-0.1977	0.0254	-0.0022	0.0006	93.91
	ϕ_3	-0.3990	0.0206	0.0090	0.0004	93.58

Capítulo 7

Aplicação com dados reais

Neste capítulo segue uma aplicação com um conjunto de dados reais para ilustrar o uso dos modelos abordados durante este texto. O propósito geral deste capítulo é mostrar a viabilidade e a precisão dos modelos GARMA na modelagem de séries temporais reais usando as abordagens clássica e bayesiana.

7.1 Número de internações com dengue em Campina Grande

Como aplicação dos modelos GARMA é proposto analisar a série temporal do número de internações com dengue em Campina Grande, Paraíba, Brasil, no período de janeiro de 1998 a outubro de 2003. Esta série tem 70 observações.

Dengue é uma doença infecciosa aguda de curta duração, de gravidade variável, causada por um arbovírus, do gênero *Flavivírus* (sorotipos: 1,2,3 e 4), transmitida principalmente pelo mosquito *Aedes aegypti* infectado, mas também pelo *Aedes albopictus*. No Brasil circulam os tipos 1, 2 e 3. O vírus 3 está presente desde dezembro de 2000 e foi isolado em janeiro de 2001, no Rio de Janeiro.

O *Aedes aegypti* é encontrado principalmente em áreas tropicais e subtropicais do mundo, inclusive no Brasil, pois as condições do meio ambiente favorecem seu desenvol-

vimento e proliferação. Esses mosquitos picam durante o dia, ao contrário do mosquito comum (*Culex*) que pica durante a noite.

As epidemias geralmente ocorrem no verão, durante ou imediatamente após períodos chuvosos. A dengue está se expandindo rapidamente e a grande preocupação é que nos próximos anos a transmissão aumente por todas as áreas tropicais do mundo, se medidas eficientes não forem tomadas para a contenção das mesmas.

A transmissão se dá pela picada do mosquito *Aedes aegypti* que ficou infectado porque picou uma pessoa doente. Esse mosquito infectado picando uma pessoa sadia, passa o vírus da dengue e a mesma adoece. A doença só acomete a população humana. Não há contágio direto entre pessoas sadias e doentes, e também não há transmissão pela água, por alimentos ou por quaisquer objetos.

Geralmente a dengue inicia-se de uma hora para outra e dura entre 5 a 7 dias. A pessoa infectada tem febre alta (39° a 40°C), dores de cabeça, cansaço, dor muscular e nas articulações, indisposição, enjoos, vômitos, manchas vermelhas na pele, dor abdominal (principalmente em crianças), entre outros sintomas.

Em alguns casos (a minoria), nos três primeiros dias depois que a febre começa a ceder, pode ocorrer diminuição acentuada da pressão sanguínea. Esta queda da pressão caracteriza a forma mais grave da doença, chamada de dengue “hemorrágica”. Este nome pode fazer com que se pense que sempre ocorrem sangramentos, o que não é verdadeiro. A gravidade está relacionada, principalmente, à diminuição da pressão sanguínea, que deve ser tratada rapidamente, uma vez que pode levar ao óbito.

O melhor método para se combater a dengue é evitando a procriação do mosquito *Aedes aegypti*, que é feita em ambientes úmidos em água parada, seja ela limpa ou suja. O combate ao mosquito, então, deve ser feito de duas maneiras: eliminando os mosquitos adultos e, principalmente, acabando com os criadouros de larvas. Para eliminação dos criadouros é importante que não deixar objetos que possam acumular água expostos à chuva. Os recipientes de água devem ser cuidadosamente limpos e tampados. Não adianta apenas trocar a água, pois os ovos do mosquito ficam aderidos às paredes dos recipientes. Para isso, deve-se tampar cuidadosamente caixas d’água, filtros, barris, tambores, cisternas etc,

e não deixar expostos à chuva pneus velhos ou objetos (latas, garrafas, cacos de vidro) que possam acumular água.

Para maiores informações, visitar o site:

<http://www.cetesb.sp.gov.br/institucional/dengue/dengue.asp>

Abordagem Clássica

Os gráficos na Figura 7.1 são a série, o histograma, a função de autocorrelação e o normal plot. Pode-se observar desses gráficos que o número de internações com dengue em Campina Grande não segue a distribuição normal e há evidência de que existe uma tendência crescente na série.

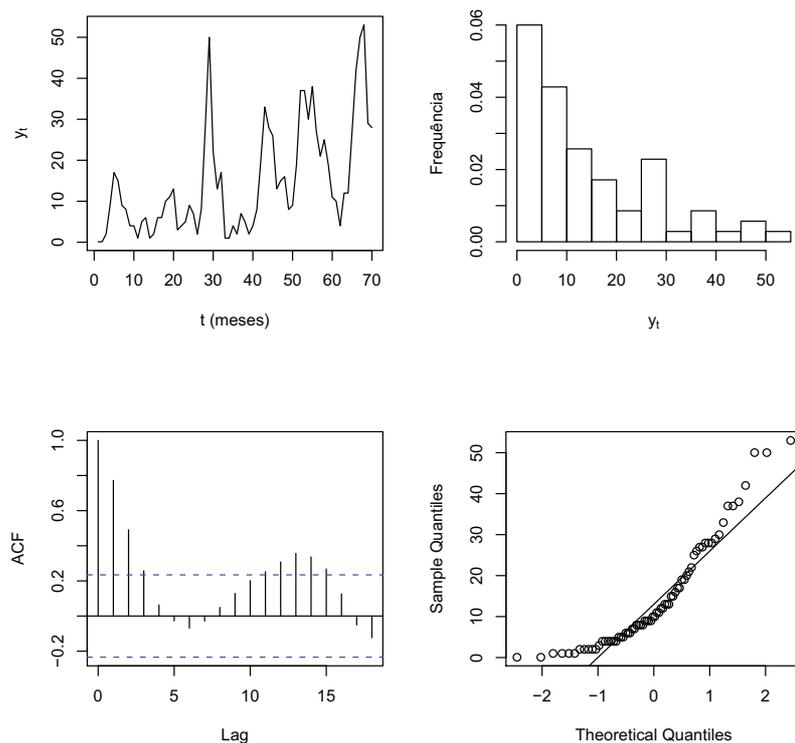


Figura 7.1: Série temporal do número de internações com dengue em Campina Grande de janeiro de 1998 a outubro de 2003.

Para este conjunto de dados foi ajustado o processo Poisson GARMA e binomial negativa GARMA, de ordem $q=0$ e $p=1,2$ e 3 . Para a escolha da ordem do modelo foram utilizados

os critérios AIC e BIC e comparamos os modelos usando a estatística MSE, todos definidos no Capítulo 4, seção 4.2 e a função desvio, definida no Capítulo 4, seção 4.1. Seguem na tabela seguinte os valores para AIC, BIC e MSE dos modelos adotados. O modelo binomial GARMA não apresentou ajuste ao conjunto de dados.

Tabela 7.1: Resultados dos critérios AIC, BIC e MSE para a série do número de internações por causa da dengue em Campina Grande.

(p,q)	Poisson			Binomial Negativa		
	AIC	BIC	MSE(%)	AIC	BIC	MSE(%)
(1,0)	558.3165	562.7847	3.2e-25	515.0517	519.5199	0.0811
(2,0)	534.3672	541.0258	5.1e-25	498.9999	505.6585	0.2558
(3,0)	530.6130	539.4318	1.1e-25	495.4838	504.3026	0.1681

Então de acordo com a Tabela 7.1, fica evidente que pelos dois critérios utilizados, que a estrutura do modelo que deveria ser escolhido é o modelo binomial negativa GARMA(3,0), pois este foi o modelo que apresentou menor AIC e BIC. Na Tabela 7.2 seguem as estimativas de máxima verossimilhança (EMV), desvio padrão (D.P.) e o intervalo de confiança de 95% ($IC_{95\%}$) para os parâmetros deste modelo.

Tabela 7.2: Resumo descritivo para o modelo binomial negativa GARMA(3,0) para a série do número de internações por causa da dengue em Campina Grande.

	EMV	D.P.	$IC_{95\%}$
α_0	0.8611	0.1304	[0.6054, 1.1168]
ϕ_1	0.9188	0.0733	[0.7750, 1.0625]
ϕ_2	-0.1967	0.0788	[-0.3512, -0.0422]
ϕ_3	-0.0234	0.0491	[-0.1198, 0.0729]

Observa-se que a estimativa do parâmetro ϕ_3 é não significativo, visto seu intervalo de confiança. Logo, sugere-se o uso de outra ordem do modelo, como o modelo binomial negativa GARMA(2,0). A Tabela 7.2 traz as estimativas de máxima verossimilhança (EMV), desvio padrão (D.P.) e o intervalo de confiança de 95% ($IC_{95\%}$) para os parâmetros do modelo escolhido: o modelo binomial negativa GARMA(2,0).

Tabela 7.3: Resumo descritivo para o modelo binomial negativa GARMA(2,0) para a série do número de internações por causa da dengue em Campina Grande.

	EMV	D.P.	$IC_{95\%}$
α_0	0.8897	0.1250	[0.6446, 1.1349]
ϕ_1	0.9135	0.0695	[0.7771, 1.0499]
ϕ_2	-0.2226	0.0543	[-0.3292, -0.1161]

Apesar deste modelo não ter o menor para o MSE, o mesmo apresentou um baixo valor (0.2558%), visto que quanto menor for seu valor mais próximo do real está a estimativa dos parâmetros do modelo selecionado. Abaixo estão alguns gráficos referentes à estimativa da série pelo modelo escolhido. A Figura 7.2 mostra o gráfico de dispersão da série e a série real *versus* série ajustada. A Figura 7.3 apresenta o gráfico dos resíduos, o histograma, a função de autocorrelação e autocorrelação parcial e o normal plot.

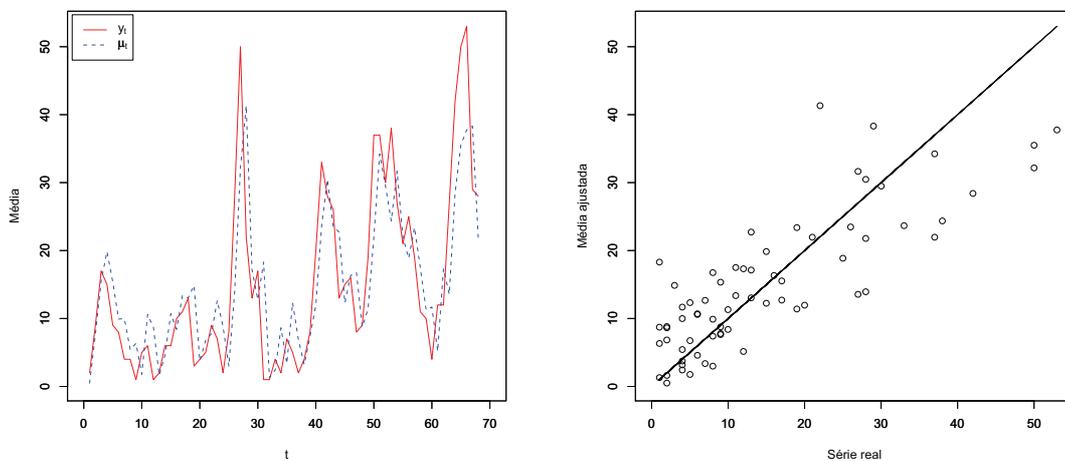


Figura 7.2: Série de casos de dengue em Campina Grande *versus* média estimada e dispersão da série.

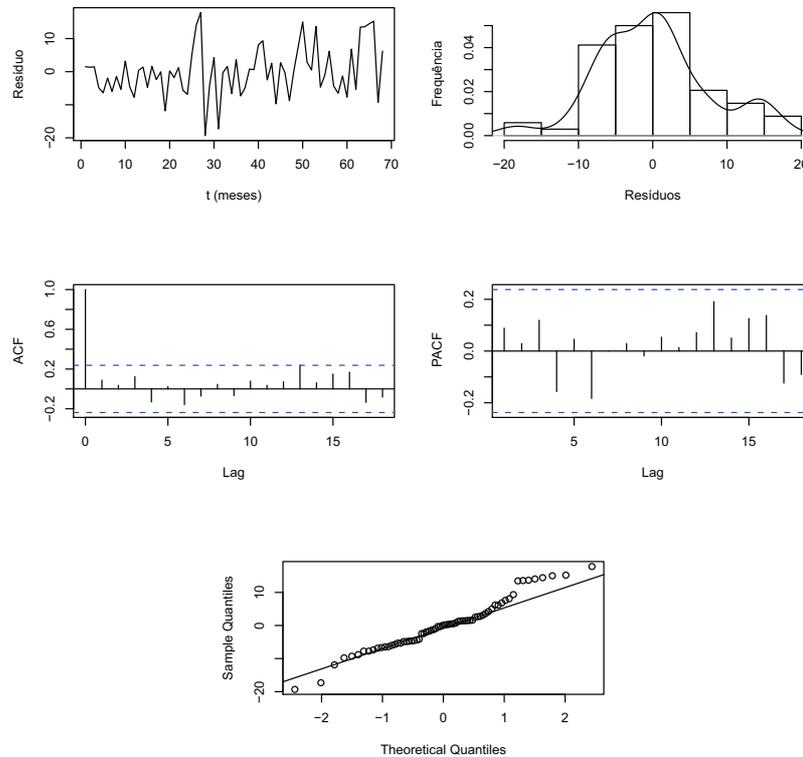


Figura 7.3: Resíduo, histograma, função de autocorrelação e autocorrelação parcial, normal plot do modelo binomial negativa GARCH(2,0) ajustado a série do número de internações por causa da dengue em Campina Grande.

Observa-se da Figura 7.3 que há evidências de que os resíduos não seguem a distribuição normal, o que era de se esperar, assim como a não evidência de correlação dos mesmos. Então, procedendo da mesma forma que na Seção 4.3, Capítulo 4, foi realizada a normalização dos resíduos. A Figura 7.4 apresenta o histograma dos resíduos normalizados, a função de autocorrelação e autocorrelação parcial, o normal plot e o gráfico dos resíduos normalizados versus o tempo.

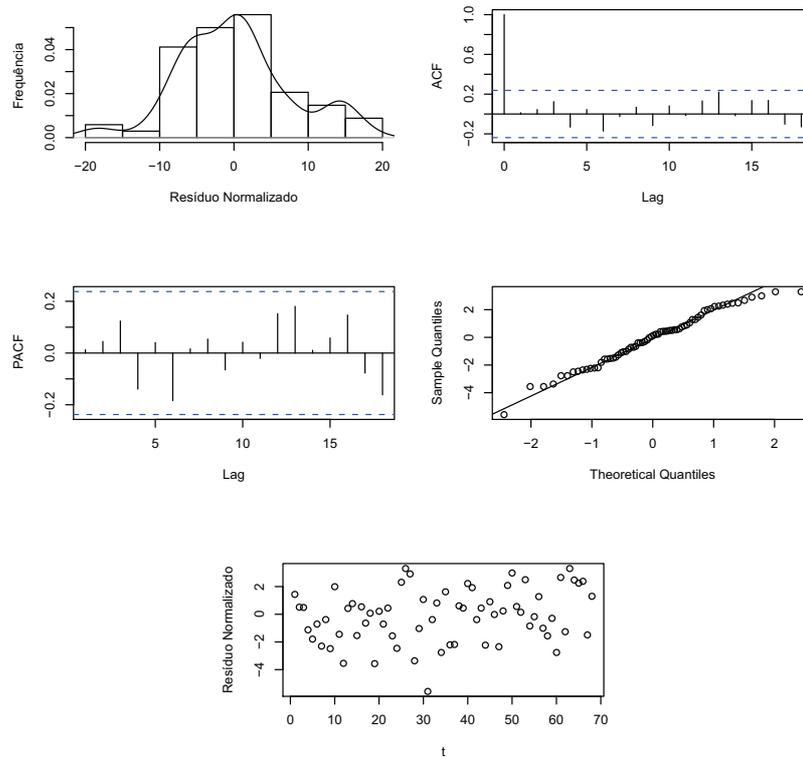


Figura 7.4: Resíduo normalizado, histograma, função de autocorrelação e normal plot do modelo GARMA Binomial Negativa(2,0) aplicado a série.

Da Figura 7.4 não há nenhuma evidência de qualquer correlação entre os resíduos, o que pode ser confirmado pela estatística de Box-Ljung, (Morettin & Tolo, 2004), de 11.92 com base em 15 lags (porque $\chi_{13,0.05}^2 = 22.36$). Também traz o normal plot, onde há evidências de que os resíduos parecem ter distribuição aproximadamente normal conforme o esperado. Ainda, o gráfico dos resíduos normalizados versus o tempo indicando uma dispersão aleatória dos mesmos.

Finalmente será apresentada a previsão da série ajustada, conforme apresentado na Seção 4.4, Capítulo 4. Assim sendo, utilizando o modelo binomial negativa GARMA(2,0), que foi o modelo que melhor se ajustou à série e, tomando o passo da previsão igual a 3, ou seja, $h = 3$, serão previstos os 3 últimos valores da série, os quais foram retirados do conjunto de dados. Foi calculado o MAPE, dado em (4.9), para verificar a qualidade da previsão.

Tabela 7.4: Previsão para $h=3$ meses.

Número da observação	Valor original	Valor ajustado	Valor previsto
68	53	37,75	30,32
69	29	38,30	30,31
70	28	21,79	24,40

A Tabela 7.4 traz os três últimos valores da série original, juntamente com seus valores ajustados e os previstos. Nota-se que os valores previstos tendem a manter uma média, sem a variância que os valores retirados possuem, o que é de se esperar dado que os valores calculados são os $\hat{\mu}_t$ dados no preditor linear. O MAPE calculado para avaliar a previsão ficou em torno de 35%, que é um valor relativamente alto. Este fato pode ser justificado devido a grande variação dos dados da série. A figura 7.5 mostra os valores observados e previstos.

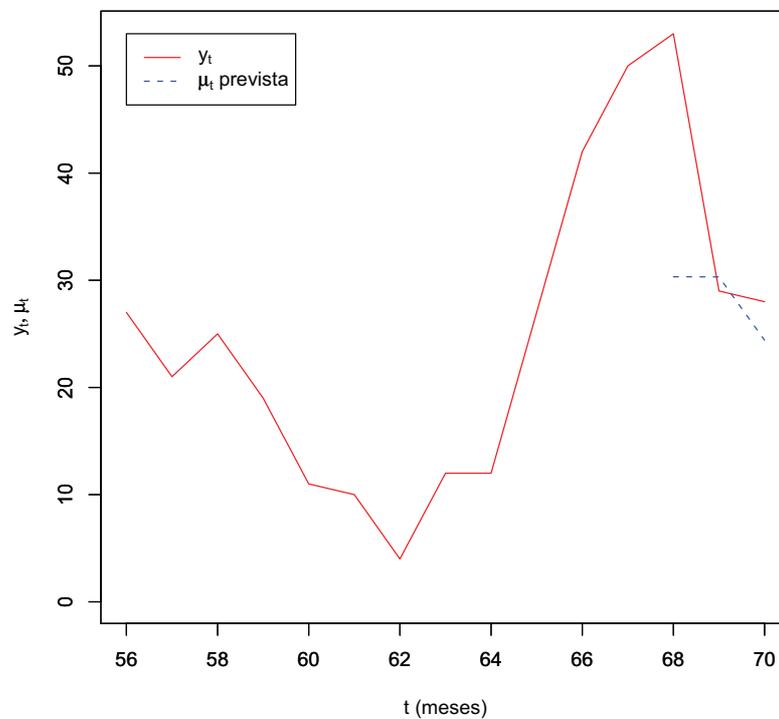


Figura 7.5: Previsão 3 meses à frente.

Diante do exposto, concluí-se que o modelo binomial negativa GARMA(2,0) descreve bem a série temporal do número de internações com dengue em Campina Grande, no período de janeiro de 1998 a outubro de 2003.

Abordagem Bayesiana

Na abordagem bayesiana para este conjunto de dados, foram analisados os modelos Poisson e binomial negativa GARMA, de ordem com $q=0,1,2$ e $p=0,1,2$. Para a escolha da ordem do modelo foi utilizado o critério BICM, citado no Capítulo 5, Seção 5.5. Segue na tabela seguinte os valores da estatística BICM para os modelos Poisson e binomial negativa GARMA(p,q). O modelo binomial GARMA não apresentou ajuste ao conjunto de dados.

Tabela 7.5: Valores do critério BICM para a série do número de internações por causa da dengue em Campina Grande.

Poisson		Binomial Negativa	
(p,q)	BICM	(p,q)	BICM
(0,1)	*	(0,1)	18,47
(1,0)	562,79	(1,0)	15,65
(0,2)	1068,60	(0,2)	12,75
(2,0)	541,04	(2,0)	12,65
(1,1)	566,39	(1,1)	12,70
(1,2)	572,56	(1,2)	524,73
(2,1)	545,21	(2,1)	16,87
(2,2)	550,98	(2,2)	514,44

Na Tabela 7.5 o símbolo * indica que não houve convergência no parâmetro do modelo. Ainda, de acordo com a mesma tabela, há evidências de que a estrutura do modelo a ser escolhido é o modelo binomial negativa GARMA(2,0), pois este foi o modelo que apresentou menor valor de BICM. Na Tabela 7.6 segue um resumo descritivo das estimativas bayesianas para os parâmetros do modelo escolhido como: média, desvio padrão (D.P.), moda, mediana, intervalo de credibilidade de 95% ($ICr_{95\%}$) e a estatística de Geweke (z_g). Obser-

vando os resultados na Tabela 7.6, nota-se que o modelo produziu resultados condizentes, pois nenhum dos valores estimados para os parâmetros são não significativos.

Tabela 7.6: Resumo descritivo para o modelo binomial negativa GARMA(2,0).

	Média	Mediana	Moda	D.P.	$ICr_{95\%}$	z_g
α_0	0.8780	0.8793	0.8322	0.1255	[0,6237; 1,1139]	-0.6726
ϕ_1	0.9130	0.9120	0.9114	0.0690	[0.7807; 1,0487]	0.4439
ϕ_2	-0.2178	-0.2184	-0.2261	0.0547	[-0,3255; -0,1042]	-0.4532

A escolha do modelo binomial negativa GARMA(2,0) parece ser apropriada visto que o mesmo apresenta um ajuste adequado da série. A Figura 7.6 mostra o gráfico de dispersão da série e a série real *versus* série ajustada.

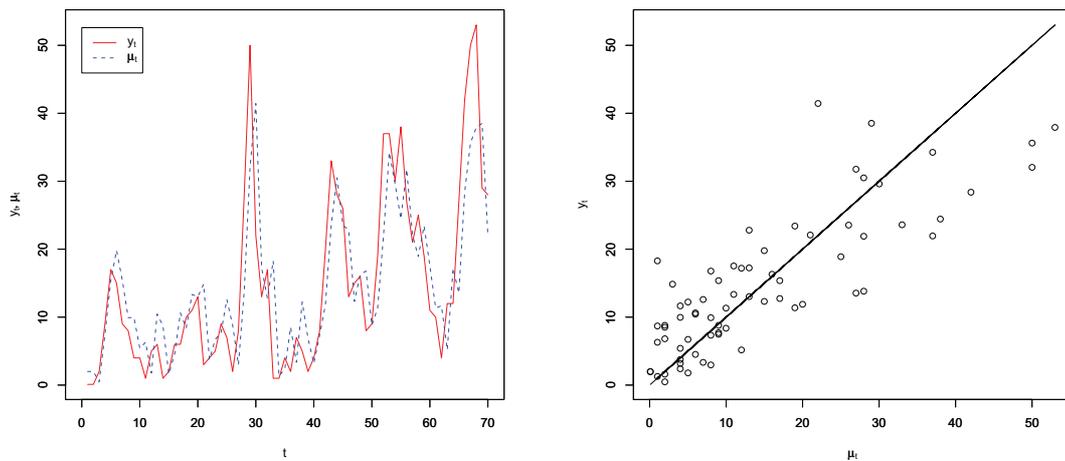


Figura 7.6: Série de casos de dengue em Campina Grande *versus* média estimada e dispersão da série.

Finalmente será apresentada a previsão da série ajustada, conforme apresentado na Seção 5.6, Capítulo 5. Assim sendo, utilizando o modelo binomial negativa GARMA(2,0), que foi o modelo que melhor se ajustou à série e, tomando o passo da previsão igual a 3, ou seja, $h = 3$, serão previstos os 3 últimos valores da série, os quais foram retirados do conjunto de dados. Foi calculado o MAPE, dado em (4.9), para verificar a qualidade da previsão.

Tabela 7.7: Previsão para $h=3$ meses.

Número da observação	Valor original	Valor ajustado	Valor previsto
68	53	13,5	35,66
69	29	28,46	37,94
70	28	35,66	38,50

A Tabela 7.7 traz os três últimos valores da série original, juntamente com seus valores ajustados e os previstos. Nota-se que os valores previstos tendem a manter uma média, sem a variância que os valores retirados possuem, o que é de se esperar dado que os valores calculados são os $\hat{\mu}_t$ dados no preditor linear. O MAPE calculado para avaliar a previsão ficou em torno de 33%, que é um valor relativamente alto. Este fato pode ser justificado devido a grande variação dos dados na série. A figura 7.7 mostra os valores observados e previstos.

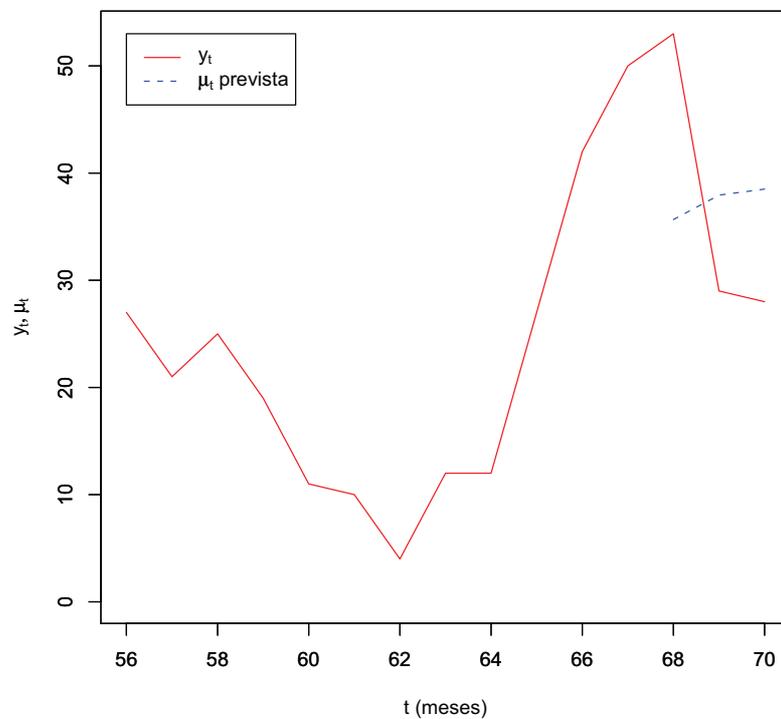


Figura 7.7: Previsão 3 meses à frente.

Diante do exposto, concluí-se que o modelo binomial negativa GARMA(2,0) descreve bem a série temporal do número de internações com dengue em Campina Grande, no período de janeiro de 1998 a outubro de 2003.

Capítulo 8

Considerações Finais

As ideias propostas no início desse texto foram exploradas e os resultados encontrados foram bastante interessantes. Com os estudos desenvolvidos pode-se observar que a Inferência Clássica para os modelos GARMA discretos apresentaram boas propriedades, analisadas a partir do desempenho do percentil de cobertura dos intervalos de confiança, dos quais ainda foram explorados a consistência, o viés e o erro quadrático médio dos estimadores pontuais.

Quanto ao estudo bayesiano, usando método de simulação de Monte Carlo via Cadeia de Markov, os resultados evidenciam que as inferências obtidas são bastante precisas e apresentou-se como uma excelente alternativa para estimação dos parâmetros, pois permite a obtenção de estimativas apuradas e com maior flexibilidade, devido ao fato das distribuições marginais a posterior gerarem inferências mais precisas.

As estimativas pontuais fornecidas por ambas as abordagens são próximas para a parte AR dos modelos, mas o modelo Bayesiano apresentou a vantagem de trazer também a parte MA dos modelos e os intervalos de credibilidade como medidas da variabilidade amostral dos parâmetros estimados. Pode-se dizer então que o modelo GARMA consegue, em ambos os ajustes, descrever bem o comportamento da série.

Tendo em vista os resultados encontrados, algumas questões podem vir a serem respondidas e despertam o interesse mais profundo no estudo dos modelos GARMA. Como proposta futura pretende-se comparar o desempenho dos modelos GARMA discretos com o

desempenho de outros modelos não lineares e não gaussianos e implementar outros modelos GARMA com diferentes distribuições condicionais para a média.

Na etapa final do trabalho observou-se que as séries reais não estacionárias de dengue ficam bem ajustadas utilizando os modelos GARMA, o que mostra a viabilidade e a precisão dos mesmos na modelagem, usando as abordagens clássica e bayesiana.

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.
- Barnett, G., Kohn, R. & Sheather, S. (1996). Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics*, **74**(2), 237–254.
- Barnett, G., Kohn, R. & Sheather, S. (1997). Robust bayesian estimation of autoregressive–moving-average models. *Journal of Time Series Analysis*, **18**(1), 11–28.
- Benjamin, M. A., Rigby, R. A. & Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, **98**, 214–223.
- Box, G., Jenkins, G. & Reinsel, G. (2008). *Time series analysis: forecasting and control*. John Wiley, fourth edition.
- Brockwell, P. & Davis, R. (2009). *Time series: theory and methods*. Springer Verlag. ISBN 1441903194.
- Casella, G. & Berger, R. (2002). *Statistical inference*. Duxbury, second edition.
- Casella, G. & George, E. (1992). Explaining the gibbs sampler. *The American Statistician*, **46**(3), 167–174.
- Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *American Statistician*, pages 327–335.
- Cordeiro, G. M. & Andrade, M. G. (2009). Transformed generalized linear models. *Journal of Statistical Planning and Inference*, **138**(9), 2970–2987.

- Dunn, P. & Smyth, G. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244.
- Ehlers, R. (2007). Introdução à Inferência bayesiana. *Notas de aula, Disponível em <http://www.icmc.usp.br/bayes>. Acesso em Março, 2009.*
- Fokianos, K. & Kedem, B. (2004). Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, **25**(2), 173–197.
- Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**(4), 457–472.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, (4), 625–631.
- Jeffreys, H. (1998). *Theory of probability*. Oxford University Press, USA, third edition.
- Lawrence, M., O'Connor, M. & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, **122**(1), 151–160.
- Li, W. K. (1994). Times series model based on generalized linear models: Some further results. *Biometrics*, **50**, 506–511.
- Marriott, J., Ravishanker, N., Gelfand, A. & Pai, J. (1996). Bayesian analysis of ARMA processes: Complete sampling-based inference under exact likelihoods. *Bayesian analysis in statistics and econometrics: essays in honor of Arnold Zellner*, pages 243–256.
- McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. London: Chapman and Hall, second edition.
- Monahan, J. (1983). Full Bayesian analysis of ARMA time series models. *Journal of Econometrics*, **21**(3), 307–331.
- Morettin, P. & Toloi, C. (2004). *Análise de séries temporais*. São Paulo: Edgard Blücher.
- Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, **135**, 370–384.

- Paula, G. (2004). Modelos de regressão: com apoio computacional. *Disponível em <http://www.ime.usp.br/giapaula/>. Acesso em Agosto, 2009.*
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Raftery, A., Newton, M., Satagopan, J. & Krivitsky, P. (2006). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Memorial Sloan-Kettering Cancer Center Department of Epidemiology and Biostatistics Working Paper Series*, page 6.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.
- Snedecor, G. W. & Cochran, W. G. (1989). *Statistical Methods*. Iowa State University Press, Ame, **8**.
- Zar, J. (1996). *Biostatistical analysis*. New Jersey, Prentice Hall, **3**.