

**Uso do Amostrador de Gibbs e
Metropolis-Hastings em
Análise Bayesiana de Modelos AR(p)**

José Roberto Temponi de Oliveira

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

Dissertação apresentada ao Instituto de Ciências Matemáticas de São Carlos-USP,
como parte dos requisitos necessários para obtenção do título de mestre em Ciências - Área:
Ciências de Computação e Matemática Computacional.

USP- São Carlos
Fevereiro - 1998

965905

Há diversidade de dons, mas um mesmo é o Espírito. Há diversidade de ministérios, mas um mesmo é o Senhor. Há diferentes atividades, mas um mesmo Deus que realiza todas as coisas em todos.

A cada um é dada a manifestação do Espírito em vista do bem comum. Como o corpo é um, embora tenha muitos membros, e como todos os membros do corpo, embora sejam muitos, formam um só corpo, assim também acontece com Cristo.

(1Cor 12,4b-7.12)

*À minha esposa Eliana,
e nossas filhas Fernanda e Tatiane*

Agradecimentos

À Deus por mais essa realização.

Ao Prof. Marinho Gomes de Andrade Filho, por sua orientação no decorrer do presente trabalho.

Aos Profs. Josemar Rodrigues e Carlos Alberto Ribeiro Diniz, pelas sugestões no exame de qualificação.

Em especial à Profa. Maria Creusa Bretas Salles, pela sempre boa vontade em ajudar, e ao Prof. Jorge Alberto Achcar.

À CAPES pelo apoio financeiro.

Aos professores, funcionários, colegas de pós-graduação e a todos aqueles que de alguma maneira contribuíram para a realização deste trabalho.

Cap.1 - Conceitos de Processos Estocásticos	1
1.1 Introdução	1
1.2 Conceitos de Processos Estocásticos	2
1.3 Modelos Lineares Estacionários	4
1.4 Objetivo do Trabalho	6
Cap.2 - Modelos Auto-regressivos de Ordem p - AR(p)	8
2.1 Processo AR(p)	8
2.2 Função de Autocorrelação para o AR(p)	10
2.3 Função de Autocorrelação Parcial	11
2.4 Modelos AR(1)	11
2.5 Modelos AR(2)	13
2.6 Comentários Gerais	14
Cap.3 - Método dos Momentos	15
3.1 Estimação da FAC e FACP para o AR(p)	15
Cap.4 - Abordagem Bayesiana para Modelos AR(p) com função de Verossimilhança Aproximada	22
4.1 Função de Verossimilhança Aproximada	22
4.2 Abordagem Bayesiana de Modelos Lineares AR(p)	25
4.2.1 Considerando Priori não Informativa	26
4.2.2 Considerando Priori Conjugada Normal-Gama	28
4.2.3 Considerando Priori t -Student	30
4.3 Algoritmos de MCMC	30
Cap.5 - Previsão	34
5.1 Análise preditiva	34
5.2 Previsão k passos a frente	37
5.3 Previsão Usando MCMC	40
5.4 Considerações sobre Seleção de Modelo	41

Cap.6 - Abordagem Bayesiana para Modelos AR(p) com função de Verossimilhança Exata	42
6.1 A Função de Verossimilhança Exata para o AR(p)	42
6.1.1 Modelos AR(1)	47
6.1.2 Modelos AR(2)	48
6.2 Abordagem Bayesiana e uso do MCMC	49
6.3 Uso do Algoritmo MCMC	52
Cap.7 - Aplicação	54
7.1 Estudo de Casos	54
7.2 Inferência Bayesiana dos Parâmetros dos Modelos	58
7.3 Previsões	66
7.4 Comparação do uso da Verossimilhança Exata e Aproximada	69
Conclusão	70
Referências	71
Apêndice 1	73
Apêndice 2	74

Resumo

Neste trabalho comparamos modelos de séries temporais auto-regressivos de ordem p $AR(p)$, ajustados pela abordagem clássica e Bayesiana. Na análise clássica a identificação do modelo é feita através da função de autocorrelação (FAC) e função de autocorrelação parcial (FACP), a escolha do melhor modelo para um conjunto de dados é feita usando-se o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). Na análise Bayesiana consideramos três alternativas de densidades a priori para os parâmetros, aqui a escolha do melhor modelo é feita pela densidade preditiva. Primeiramente consideramos a priori não informativa de Jeffreys, onde a densidade a posteriori marginal, para os parâmetros do modelo, pode ser calculada analiticamente e mostra-se que o valor esperado dessa posteriori coincide com o estimador de máxima verossimilhança. No segundo caso, adotamos uma função densidade a priori conjugada normal-gama. Aqui, a densidade a posteriori também pode ser calculada analiticamente, resultando em uma densidade t -Student p -dimensional, no entanto em muitas situações reais adotar priori conjugada é pouco realista. Para contornar esse problema, no terceiro caso adotamos uma densidade a priori informativa t -Student, p -dimensional, para os parâmetros e uma densidade a priori gama para o inverso da variância dos resíduos. Isto resulta em uma densidade a posteriori não padronizada. Neste caso a análise a posteriori só pode ser feita usando-se algoritmos de simulação em cadeia de Markov, MCMC.

Abstract

In this work we compare time series models, $AR(p)$, using the Bayesian and classical approach.

The classical approach uses the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) to choose the best fitted models for a time series record data. The estimation of the model parameters are obtained by maximum likelihood estimation.

In the Bayesian approach to time series the likelihood function is combined with the prior density of the parameters, via Bayes theorem, to produce the posterior distribution of the parameters and the predictive distribution of future observation. In this paper we use three different prior density functions. First, the likelihood function is approximated by a normal-gamma density over the parameter space, then combined with either a Jeffreys' vague prior or a normal-gamma prior. The result is that the analysis is similar to that obtained by a multiple linear regression analysis. Thus, the posterior analysis is done with normal-gamma distribution. Thirdly the likelihood function is combined with prior t -Student distribution then the posterior and predictive analysis is made via Markov Chain Monte Carlo.

The results are applied to two simulated series and two stream flow series of Furnas and Itumbiara reservoirs.

Capítulo 1

Conceitos de Processos Estocásticos

1.1 Introdução

Uma série temporal consiste em um conjunto de observações de uma variável aleatória indexada no tempo, que denotamos aqui por $Z_t(\omega)$, onde ω representa a natureza aleatória da série, ou seja ω é um evento no espaço de probabilidade $(\Omega, \mathcal{A}, \mathcal{P})$, (Ω é o espaço amostral, \mathcal{A} uma σ -álgebra de um subconjunto de Ω e \mathcal{P} uma medida de probabilidade definida em \mathcal{A}), e t é um número real positivo, ou seja $t \in \mathbb{R}_+$ (\mathbb{R}_+ conjunto dos números reais positivos). Neste trabalho estaremos considerando séries temporais cujos valores foram observados somente para valores inteiros de t , assim faremos $t \in \mathcal{N}_+$ (\mathcal{N}_+ conjunto dos números inteiros positivos). Desta forma vamos denotar as séries consideradas neste trabalho como $\{Z_1, Z_2, \dots, Z_t, \dots\}$ onde Z_t é a observação da série no instante $t = 1, 2, \dots$.

Analisar uma série temporal $\{Z_t, t \in \mathcal{N}_+\}$ consiste em ajustar um modelo para explicar o valor da série Z_t , usando um conjunto de observações $Z = \{Z_1, Z_2, \dots, Z_n\}$. Em geral essas observações não podem ser vistas como uma amostra aleatória independente da série Z_t , pois $cov(Z_i, Z_j) \neq 0$ para $i \neq j$. Portanto os modelos propostos geralmente expressam o valor da série Z_t em função de seus valores passados Z_{t-1}, Z_{t-2}, \dots , e de variáveis exógenas.

Neste trabalho, vamos considerar, somente, os modelos que expressam os valores de Z_t como uma função linear de seus valores passados Z_{t-1}, Z_{t-2}, \dots , esses modelos recebem o nome de auto-regressivos, [Box, Jenkins & Reisel -1994].

Uma série temporal $\{Z_t, t \in \mathcal{N}_+\}$ pode ser interpretada como uma possível realização de um processo estocástico $\{Z_t(\omega), t \in \mathcal{N}_+, \omega \in (\Omega, \mathcal{A}, \mathcal{P})\}$, por isso apresentamos a seguir algumas propriedades dos processos estocásticos que são úteis na elaboração dos modelos para uma série temporal.

1.2 Conceitos de Processos Estocásticos

Definição 1.1. Seja \mathcal{N}_+ um conjunto de inteiros positivos, o processo estocástico $\{Z_t(\omega), t \in \mathcal{N}_+, \omega \in (\Omega, \mathcal{A}, \mathcal{P})\}$ é uma sequência de variáveis aleatórias, tal que,

$Z_{t^*}(\omega)$, $t^* \in \mathcal{N}_+$ é uma variável aleatória,

$Z_t(\omega^*)$, $\omega^* \in \Omega$ é uma função de tempo.

ou seja para cada $\omega^* \in (\Omega, \mathcal{A}, \mathcal{P})$ fixado $Z_t(\omega^*)$ é uma função do tempo a qual denominamos de realização do processo estocástico, por outro lado, para um $t^* \in \mathcal{N}_+$ fixo, $Z_{t^*}(\omega)$ é uma variável aleatória.

Se $f(z_t)$ é a densidade de probabilidade do processo no instante t e $f(z_{t_1}, z_{t_2})$ a densidade de probabilidade conjunta do processo para dois instantes t_1 e t_2 , então define-se:

$$\mu_t = E(z_t) = \int_{-\infty}^{+\infty} z_t f(z_t) dz_t;$$

$$\sigma_t^2 = Var(z_t) = \int_{-\infty}^{+\infty} (z_t - \mu_t)^2 f(z_t) dz_t;$$

$$cov(z_{t_1}, z_{t_2}) = \int_{-\infty}^{+\infty} (z_{t_1} - \mu_{t_1})(z_{t_2} - \mu_{t_2}) f(z_{t_1}, z_{t_2}) dz_{t_1} dz_{t_2};$$

Um processo estocástico fica completamente definido por sua densidade de probabilidade conjunta, assim se Z_1, \dots, Z_n é um conjunto de n -observações de Z_t , existe uma função densidade conjunta, associada a essa realização do processo, que denotamos por $f(z_1, \dots, z_n)$. Um processo é denominado estacionário quando a densidade conjunta das observações $\{Z_t(\omega), t \in T \subset \mathcal{N}_+\}$, $T = \{t | 1 \leq t \leq n\}$ não se altera quando deslocamos o tempo das observações, por exemplo para $\{Z_t, t \in T_k \subset \mathcal{N}_+\}$, $T_k = \{t | 1 + k \leq t \leq n + k\}$ para qualquer valor de k e n .

Definição 2.1. Estacionariedade no Sentido Estrito

Especificando um processo estocástico por suas distribuições de dimensão finita, consideramos um processo estocástico $Z = \{Z_t; t \in \mathcal{N}_+\}$ estritamente estacionário se todas as distribuições finito-dimensionais deste processo, permanecem inalteradas sob translações no tempo, ou seja:

$$f(z_1, \dots, z_n) = f(z_{1+\tau}, \dots, z_{n+\tau})$$

para quaisquer τ e $n \geq 1$.

Isso significa, em particular, que todas as distribuições unidimensionais são invariantes sob translações do tempo, logo o valor esperado $\mu_t = E(z_t)$ e a variância $\sigma_t^2 = E(z_t - \mu_t)^2$ são invariantes no tempo. Uma subclasse dos processos estacionários são os processos fracamente estacionários, definidos como:

Definição 3.1. Estacionariedade no Sentido Amplo

Um processo $Z = \{Z_t; t \in \mathcal{N}_+\}$ é estacionário no sentido amplo quando:

- i) $E(z_t) = \mu$ (constante)
- ii) $Var(z_t) = \sigma^2$ (constante)
- iii) $Cov(z_t, z_{t+k}) = \gamma(k)$ (função só de k)

Verifica-se facilmente que qualquer processo estacionário no sentido estrito será estacionário no sentido amplo [Box, et al - 1994].

Ergodicidade

Um processo estocástico é denominado ergódico quando todas as propriedades estatísticas do processo, no sentido da sua caracterização completa, está contida em cada realização com probabilidade 1. Ergodicidade implica na seguinte propriedade: Se Z_t é um processo ergódico, então todas as suas médias estatísticas são iguais as correspondentes médias temporais. Assim por exemplo, Z_t é definido como um processo ergódico na média se e somente se:

$$\mu = E(z_t) = \int_{-\infty}^{+\infty} z_t f(z_t) dz_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T z_t dt \quad (1.1)$$

Um processo é ergódico na variância se:

$$\begin{aligned} \sigma^2 = Var(Z_t) &= \int_{-\infty}^{+\infty} (z_t - \mu)^2 f(z_t) dz_t \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (z_t - \mu)^2 dt \end{aligned} \quad (1.2)$$

Analogamente, z_t será ergódico na autocorrelação se e somente se:

$$\begin{aligned} \gamma_k = cov(z_t, z_{t+k}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (z_t - \mu)(z_{t+k} - \mu) f(z_t, z_{t+k}) dz_t dz_{t+k} = \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (z_t - \mu)(z_{t+k} - \mu) dt \end{aligned} \quad (1.3)$$

Um processo ergódico na média na variância e na autocorrelação é estacionário no sentido amplo.

1.3 Modelos lineares Estacionários

Os modelos aqui apresentados são casos particulares de um modelo de filtro linear. Este modelo supõe que a série temporal é gerada através de um sistema linear, cuja entrada é um processo $\{a_t; t \in \mathcal{N}_+\}$, e denotamos por $\Psi(B)$ a função de transferência do sistema, onde B é um operador retardo, definido por $B^l a_t = a_{t-l}$. Esse modelo é ilustrado na figura 2.1.

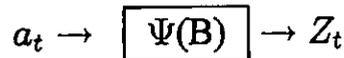


Fig.2.1. - Filtro linear, com entrada a_t , saída Z_t e função de transferência $\Psi(B)$

O processo a_t satisfaz a seguinte propriedade:

$$E(a_t) = 0,$$

$$Var(a_t) = \sigma_a^2 \quad e$$

$$Cov(a_t, a_{t+k}) = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

Desta forma podemos escrever que:

$$\tilde{Z}_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

$$\tilde{Z}_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad \psi_0 = 1 \quad (1.4)$$

denotando por:

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

temos:

$$\tilde{Z}_t = \mu + \psi(B)a_t \quad (1.5)$$

Podemos eliminar o termo constante μ sem perda de generalidade, para isso basta considerar a série $Z_t = \tilde{Z}_t - E(\tilde{Z}_t)$.

A condição de estacionariedade do modelo linear (2.6) está associado ao polinômio $\psi(B)$. Pode ser mostrado [Box, et al -1994] que a série $Z_t = \psi(B)a_t$ é estacionária se o polinômio $\psi(B)$ converge para $|B| \leq 1$.

Uma forma equivalente de representação de uma série temporal é escreve-la como uma combinação linear de seus valores passados, ou seja:

$$Z_t = a_t - \pi_1 Z_{t-1} - \pi_2 Z_{t-2} - \dots$$

$$\sum_{j=0}^{\infty} \pi_j Z_{t-j} = a_t \quad , \pi_0 = 1$$

definindo o polinômio:

$$\pi(B) = 1 + \pi_1 B + \pi_2 B^2 + \dots$$

podemos escrever que:

$$\pi(B) Z_t = a_t \quad (1.6)$$

Com essa representação, temos associado a propriedade de inversibilidade, ou seja o modelo (1.6) é inversível quando o polinômio $\pi(B)$ converge para todo $|B| \leq 1$.

Os modelos (1.5) e (1.6) são equivalentes, se escolhermos $\pi(B)$ e $\psi(B)$ tal que:

$$\psi(B) \pi(B) = 1$$

Os três modelos lineares mais usados para modelar séries temporais estacionárias, são os processos auto-regressivos (AR), médias móveis (MA) e os mixtos auto-regressivos médias móveis (ARMA). Neste trabalho estamos considerando somente os modelos auto-regressivos (AR).

1.4 Objetivo do trabalho

Muitos métodos podem ser utilizados para estimar os parâmetros de um modelo auto-regressivo proposto para uma série temporal, entre esses temos os métodos dos momentos, método de mínimos quadrados e método de máxima verossimilhança, já amplamente estudados a mais de 20 anos [Box et al -1994]. Um método alternativo de análise e inferência para modelos de séries temporais são os métodos Bayesianos, vários trabalhos utilizando abordagem Bayesiana já foram realizadas [Broemeling & Land -1984], [Shaarawy & Broemeling -1984, 85, 88]. A vantagem do uso dos métodos Bayesianos é permitir a consideração de informações a priori sobre os parâmetros do modelo na análise. A limitação desses trabalhos deve-se a dificuldade encontrada para resolver integrais múltiplas, bastante complicadas, representando as densidades a posteriori dos parâmetros e densidades preditivas. Com o surgimento das aplicações dos algoritmos de simulação de Monte Carlo em Cadeia de Markov (MCMC) na análise Bayesiana [Gelfand & Smith -1990], [Gelfand et al -1990] essas dificuldades foram superadas permitindo inclusive escolher densidades a priori sem se prender a propriedade de conjugação.

Neste trabalho vamos comparar os métodos de máxima verossimilhança e os métodos Bayesianos quando ambos são utilizados para estimar os parâmetros dos modelos auto-regressivos.

Nosso objetivo é fazer a análise Bayesiana de modelos de séries temporais quando diferentes densidades a priori são adotadas destacando as características particulares de cada modelo. Na análise Bayesiana estudamos os modelos auto-regressivos considerando três alternativas de densidades a priori para os parâmetros deste modelo. Primeiramente consideramos a priori não informativa de Jeffreys, onde a densidade a posteriori marginal, para os parâmetros do modelo, pode ser calculada analiticamente e mostra-se que o valor esperado dessa posteriori coincide com o estimador de máxima verossimilhança. No segundo caso, adotamos uma função densidade a priori conjugada normal-gama. Aqui, a densidade a posteriori também pode ser calculada analiticamente, resultando em uma densidade t -Student p -dimensional. A análise da expressão do estimador de Bayes, para este caso, mostra que o estimador é uma média ponderada entre o estimador de máxima verossimilhança e o valor esperado a priori.

O uso de densidades a priori conjugadas em geral recebe críticas por serem estas funções pouco realista para tratamento de problemas reais. Para contornar estas críticas adotamos no terceiro caso uma densidade a priori informativa t -Student, p -dimensional, para os parâmetros e uma densidade a priori gama para o inverso da variância dos resíduos. Isto resulta em uma densidade a posteriori não padronizada. Neste caso a análise a posteriori só pode ser feita numericamente. Aqui optamos pelo uso de algoritmos de simulação de Monte Carlo em Cadeia de Markov (MCMC), do tipo Metropolis-Hastings [Chib & Greenberg - 1995] e Gibbs Sampling [Casella & George - 1992] os quais vêm se destacando com muita eficiência neste tipo de abordagem, porém, exigindo para isso considerações práticas.

Um critério de seleção de modelos baseado na densidade preditiva é aplicado. Esses resultados são comparados com a análise clássica onde a identificação do modelo é feita através da função de autocorrelação, FAC e função de autocorrelação parcial, FACP, a escolha do melhor modelo para um conjunto de dados é feita usando-se o Critério de Informação de Akaike AIC e o Critério de Informação Bayesiano BIC.

Este trabalho está organizado como se segue: No Capítulo 2 apresentamos um resumo da teoria dos modelos AR(p). No Capítulo 3 apresentamos a inferência dos parâmetros dos modelos usando o método dos momentos.

A abordagem Bayesiana quando as densidades a posteriori são construídas usando-se a função de verossimilhança aproximada é apresentada no Capítulo 4. No Capítulo 5 apresentamos as previsões de valores futuros utilizando a abordagem Bayesiana do Capítulo 4. A abordagem utilizando a verossimilhança exata é apresentada no Capítulo 6. Por fim reunimos no Capítulo 7 um conjunto de resultados numéricos comparando a inferência clássica e Bayesiana para séries reais de vazões médias mensais, coletadas por empresas do setor elétrico brasileiro nas barragens de Furnas e Itumbiara. Também usamos para fins ilustrativos séries geradas a partir de modelos arbitrariamente construídas.

Capítulo 2

Modelos Auto-regressivos de Ordem p — AR(p)

2.1 Processo AR(p)

Considerando um processo $\{\tilde{Z}_t, t \in \mathcal{N}_+\}$. Um modelo linear auto-regressivo de ordem p é um modelo onde o valor corrente do processo \tilde{Z}_t é expresso como uma combinação linear dos p valores passados $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots, \tilde{Z}_{t-p}$ e de um ruído branco a_t . Aqui estamos considerando $\{a_t, t \in \mathcal{N}\}$ um processo i.i.d. $N(0, \sigma^2)$. Um modelo AR(p) é escrito como:

$$\tilde{Z}_t = \mu + \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad (2.1)$$

onde μ é uma constante.

Usando o operador retardo $B^l \tilde{Z}_t = \tilde{Z}_{t-l}$, reescrevemos (2.1) como:

$$\phi(B) \tilde{Z}_t = \mu + a_t \quad (2.2)$$

onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é chamado polinômio característico do modelo AR(p).

Se $\{\tilde{Z}_t, t \in \mathcal{N}_+\}$ é um processo estacionário no sentido amplo, temos que $E(Z_t)$ é constante então podemos calcular do modelo (2.1) que:

$$E(\tilde{Z}_t) = \frac{\mu}{1 - \sum_{i=1}^p \phi_i} \quad (2.3)$$

Assim podemos eliminar a constante μ de (2.1) sem perda de generalidade, definindo o processo $Z_t = \tilde{Z}_t - E(\tilde{Z}_t)$. Mostra-se facilmente que o processo $\{Z_t, t \in \mathcal{N}_+\}$ satisfaz o seguinte modelo AR(p).

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t \quad (2.4)$$

Neste trabalho, por questão de simplicidade estaremos considerando o modelo (2.4), lembrando sempre que a relação entre os modelos (2.1) e (2.4) é simplesmente a subtração do valor esperado do processo. Na prática esse procedimento sempre pode ser adotado, considerando as observações $(\tilde{Z}_1, \dots, \tilde{Z}_n)$ da série temporal para estimar a constante μ , como:

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^N \tilde{Z}_t \quad (2.5)$$

e fazendo $Z_t = \tilde{Z}_t - \hat{\mu}$

Condições de Estacionariedade e Inversibilidade

Sendo $\phi(B)$ um polinômio de grau finito, não há restrições sobre os parâmetros para assegurar a inversibilidade de Z_t .

A condição de estacionariedade deste modelo pode ser analisada escrevendo o polinômio $\phi(B)$ como:

$$\phi(B) = (1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B) \quad (2.6)$$

onde, G_i^{-1} , $i = 1, \dots, p$, são as raízes da equação característica. Expandindo (2.6) em frações parciais, temos:

$$\phi^{-1}(B) = \sum_{i=1}^p \frac{A_i}{1 - G_i B} \quad (2.7)$$

onde A_i , $i = 1, \dots, p$, são constantes.

Denotando por $\psi(B) = \phi^{-1}(B)$. Desde que a estacionariedade é assegurada se $\psi(B)$ convergir para $|B| \leq 1$, temos de (2.7) que esta condição é satisfeita se $|G_i| \leq 1$, $i = 1, \dots, p$.

Portanto asseguramos que um processo $AR(p)$ é estacionário se as raízes do polinômio $\phi(B) = 0$ estão fora do círculo de raio unitário, $|G_i|^{-1} > 1$, $i = 1, \dots, p$.

2.2 Função de Autocorrelação para o AR(p)

Considerando o modelo (2.4), podemos calcular a função de autocorrelação (FAC) do processo Z_t , por:

$$\rho_k = \gamma_k / \gamma_0$$

onde $\gamma_k = E(Z_t Z_{t-k})$ é a autocovariância entre Z_t e Z_{t-k} e $\gamma_0 = E(Z_t^2)$ é a variância do processo. Assim temos:

$$\begin{aligned} \gamma_k &= E(Z_t Z_{t-k}) = E\{(a_t + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p}) Z_{t-k}\} \\ \gamma_k &= \phi_1 E(Z_{t-1} Z_{t-k}) + \dots + \phi_p E(Z_{t-p} Z_{t-k}) + E(a_t Z_{t-k}) \end{aligned} \quad (2.8)$$

Observe que $E(a_t Z_{t-k}) = 0$ quando $k > 0$, pois Z_{t-k} só é afetado por grandezas que ocorreram até o instante $t - k$. Então:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad (2.9)$$

A variância γ_0 pode ser calculada fazendo $k = 0$ em (2.8), e é dada por:

$$\gamma_0 = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 + \dots + \rho_p \phi_p} \quad (2.10)$$

Dividindo (2.9) por γ_0 temos:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (2.11)$$

A equação (2.11) é conhecida como equação de Yule-Walker. Note que ρ_k satisfaz a equação diferença,

$$\phi(B)\rho_k = 0$$

Cuja solução tem forma geral dada por [Box, et al -1994], como:

$$\rho_k = C_1 G_1^k + C_2 G_2^k + \dots + \phi_p G_p^k \quad (2.12)$$

onde C_i , $i = 1, \dots, p$, são constantes.

Se o modelo é estacionário então $|G_i| < 1$, conseqüentemente ρ_k decai exponencialmente com k . Se G_i^{-1} são raízes complexas conjugadas então G_i^{-1} podem ser escrito como $|G_i| e^{i(2k\pi + f)}$. Neste caso ρ_k decai exponencialmente oscilando.

2.3 Função de Autocorrelação Parcial

Outra ferramenta utilizada na identificação da ordem dos modelos de série temporais é a função de autocorrelação parcial (FACP). Esta função que denotamos por ϕ_{kk} para vários valores de k é definida como:

Definição 1.1. O coeficiente de autocorrelação parcial de ordem k , ϕ_{kk} , é o último coeficiente de um modelo $AR(k)$, ajustado a série temporal Z_t , $t = 1, 2, \dots, n$.

Supondo que se ajuste a Z_t , $t = 1, 2, \dots$, um modelo $AR(1)$, então, $k = 1$ e temos:

$$Z_t = \phi_1 Z_{t-1} + a_t \Rightarrow \phi_{11} = \phi_1$$

Supondo agora que se ajuste a Z_t um modelo $AR(2)$, então, $k = 2$ e temos:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t \Rightarrow \phi_{22} = \phi_2$$

Procedendo desta maneira, temos no caso geral, ao ajustar um $AR(k)$, que:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_k Z_{t-k} + a_t \Rightarrow \phi_{kk} = \phi_k$$

Uma propriedade importante da FACP é: Se a série Z_t é gerada por um modelo $AR(p)$ temos $\phi_{kk} = 0$ para $k > p$. Essa propriedade juntamente com o comportamento da FAC são duas ferramentas muito úteis para identificar a ordem dos modelos AR .

Como exemplo vamos considerar dois casos particulares de modelos $AR(p)$, quando $p = 1$ e $p = 2$.

2.4 Modelo $AR(1)$

Considere o processo $\{Z_t, t \in \mathcal{N}_+\}$ gerado pelo modelo $AR(1)$, dado por:

$$Z_t = \phi Z_{t-1} + a_t$$

Neste caso temos a equação de Yule-Walker dada por:

$$\rho_k = \phi \rho_{k-1}, \quad k \geq 1$$

Assim temos:

$$\begin{aligned} \rho_1 &= \phi \\ \rho_2 &= \phi\rho = \phi^2 \\ &\vdots \\ \rho_k &= \phi^k \end{aligned}$$

A condição de estacionariedade para o processo AR(p) afirma que a raiz do polinômio $\phi(B) = 1 - \phi B = 0$, deve estar fora do círculo de raio unitário. Denotando essa raiz por:

$$G^{-1} = \frac{1}{\phi}$$

Temos que o processo AR(1) é estacionário se:

$$|G^{-1}| = \frac{1}{|\phi|} > 1 \Rightarrow |\phi| < 1$$

Portanto a região de estacionariedade do modelo AR(1) no espaço do parâmetro ϕ é dado por $-1 < \phi < 1$.

À figura 2.1a e 2.1b mostra o comportamento da FAC e FACP do processo AR(1) no espaço dos parâmetros.

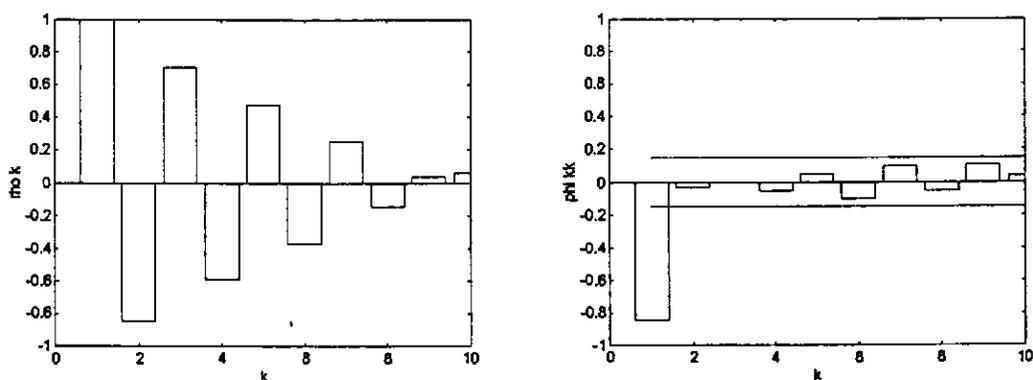


Figura 2.1a- FAC e FACP para um AR(1) quando $-1 < \phi < 0$.

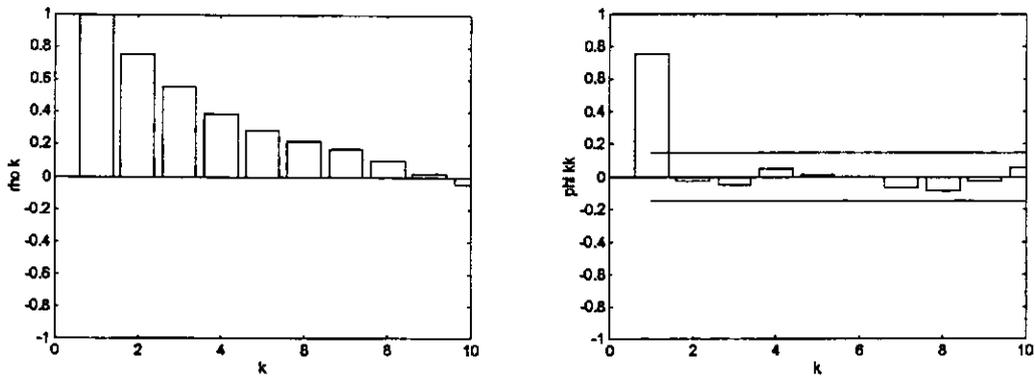


Figura 2.1b- FAC e FACP para um AR(1) quando $0 < \phi < 1$.

2.5 Modelo AR(2)

Considere o processo $\{Z_t, t \in \mathcal{N}_+\}$ gerado pelo modelo AR(2), dado por:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$$

A equação de Yule-Walker para esse modelo é dada por:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 1$$

Neste caso temos:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$$

e os parâmetros ϕ_1 e ϕ_2 podem ser expressos em função das autocorrelações ρ_1 e ρ_2 como:

$$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}$$

$$\phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$$

Pode ser mostrado [Box, et al -1994] que a condição de estacionariedade dos processos AR(2) pode ser expressa em função dos parâmetros. A região no espaço dos parâmetros (ϕ_1, ϕ_2) onde os processos AR(2) são estacionários é dado por:

$$\begin{cases} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ |\phi_2| < 1 \end{cases}$$

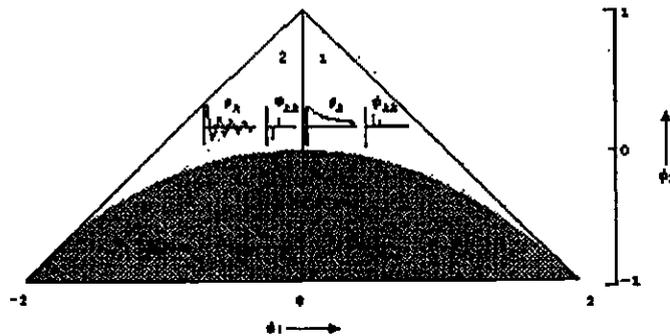


Figura 2.2 - Região de estacionariedade para um AR(2).

2.6 Comentários Gerais

A propriedade principal que precisa ser verificada antes de se ajustar um modelo linear auto-regressivo é a estacionariedade. No entanto essa propriedade dificilmente pode ser verificada na prática, a partir de um conjunto de observações de uma série temporal. Uma forma encontrada de contornar essa dificuldade e assegurar que os modelos ajustados são válidos é supor estacionariedade conjuntamente com a hipótese de normalidade do processo. A estacionariedade ampla pode ser verificada a partir dos dados, estimando-se a média, variância e função de autocorrelação da série. A hipótese de normalidade geralmente é verificada por métodos gráficos ou testes de aderência. Sob essa duas condições podemos assegurar a estacionariedade do processo no sentido estrito. No entanto quando essas propriedades falham geralmente não se tem um bom ajuste dos modelos. Nos próximos capítulos vamos apresentar métodos de estimação de modelos quando essas suposições são válidas.

Capítulo 3

Método dos Momentos

3.1 Estimação da FAC e FACP para o AR(p)

A relação entre a FAC e os parâmetros de um modelo AR(p) é dada pela equação de Yule-Walker (2.11) que reescrevemos aqui:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (3.1)$$

Escrevendo essa equação para $k = 1, 2, \dots, p$ temos um sistema de equações lineares $p \times p$ dado por:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & & \rho_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad (3.2)$$

Denotando por:

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}; \quad \mathbf{P} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}; \quad \rho = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

Podemos escrever (3.2) na forma matricial dada por:

$$\mathbf{P}\phi = \rho \quad \Rightarrow \quad \phi = \mathbf{P}^{-1}\rho \quad (3.3)$$

Notamos de (3.3) que se estimarmos as autocorrelações (ρ_1, \dots, ρ_p) poderemos estimar os parâmetros do modelo AR(p).

Quando as autocorrelações ρ_k são estimadas usando-se o método dos momentos, e esses estimadores são substituídos em (3.3), os estimadores de ϕ obtidos assim são chamados estimadores por método dos momentos. Esse procedimento geralmente não apresenta boa precisão nas estimativas de ϕ quando o modelo está próximo das fronteiras da região de estacionariedade no espaço dos parâmetros [Box, et al - 1994].

Vamos considerar agora a estimativa das autocorrelações pelo método dos momentos.

Seja $Z = (Z_1, \dots, Z_N)$ um conjunto de observações da série temporal $\{Z_t, t \geq 0\}$ supostamente estacionária. Então estimamos ρ_k por:

$$r_k = \frac{c_k}{c_0} \quad (3.4)$$

onde:

$$c_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (Z_t - \hat{\mu})(Z_{t+k} - \hat{\mu}) \quad (3.5)$$

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^N Z_t \quad (3.6)$$

$$c_0 = \frac{1}{N} \sum_{t=1}^n (Z_t - \hat{\mu})^2$$

c_k e c_0 são estimativas da autocovariância e variância do processo respectivamente e $\hat{\mu}$ é uma estimativa do valor esperado do processo.

Devemos observar que r_k é calculado para cada $k = 0, 1, 2, \dots, K$, o número de observações usadas para estimar c_k depende do valor de k , na prática o valor de K não é maior que $N/4$ quando $N \approx 100$. Quando N é grande vamos limitar $K \leq 25$.

Erro padrão do estimador de autocorrelação

Para construir o intervalo de confiança da autocorrelação ρ_k , podemos usar a aproximação de Bartlett [Bartlett -1946, apud, Box et al -1994] para a variância do estimador do coeficiente de autocorrelação, de um processo estacionário normal.

$$Var[r_k] \simeq \frac{1}{N} \sum_{v=-\infty}^{+\infty} \{\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2\} \quad (3.7)$$

Supondo que todas as autocorrelações ρ_v são zero para $v > q$, todos os termos exceto os primeiros que aparecem a direita da expressão (3.7) são zero quando $k > q$. Então a variância do estimador de autocorrelação r_k , para $k > q$, é dada por:

$$Var[r_k] \simeq \frac{1}{N} \left\{ 1 + 2 \sum_{v=1}^q \rho_v^2 \right\}, \quad k > q \quad (3.8)$$

Na prática para usar (3.8), as autocorrelações estimadas r_v ($v = 1, 2, \dots, q$) são substituídas pelas autocorrelações teóricas ρ_v e quando fazemos isto nos referimos a raiz quadrada de (3.8) como o desvio padrão do estimador de r_k .

$$SD[r_k] = \sqrt{Var[r_k]} \quad (3.9)$$

Sobre a suposição que as autocorrelações teóricas ρ_v são todas zeros, podemos construir intervalos de confiança para r_k .

Similarmente a expressão aproximada para a covariância entre as correlações r_k e r_{k+s} de dois passos diferentes k e $k+s$ é dado por Bartlett [Bartlett -1946, apud, Box et al - 1994], como:

$$Cov[r_k, r_{k+s}] \simeq \frac{1}{N} \sum_{v=-\infty}^{+\infty} \rho_v \rho_{v+s} \quad (3.10)$$

A equação (3.10) mostra que é preciso tomar cuidado na interpretação de autocorrelações individuais, pois covariâncias altas podem existir entre valores vizinhos. Este efeito poderá em algum momento distorcer a aparência visual da função de autocorrelação diferindo do comportamento esperado.

Estimação da Função de autocorrelação parcial

A função de autocorrelação parcial pode ser estimada por sucessivos ajustes do processos auto-regressivos de ordens $1, 2, \dots$, usando o método dos momentos. Uma forma alternativa, quando os parâmetros do modelo não estão próximo da fronteira da região de estacionariedade, é usar uma aproximação da equação de Yule-Walker substituindo o estimador r_j no lugar da autocorrelação ρ_j . Assim temos:

$$r_j = \hat{\phi}_{k,1} r_{j-1} + \hat{\phi}_{k,2} r_{j-2} + \dots + \hat{\phi}_{k,(k-1)} r_{j-k+1} + \hat{\phi}_{k,k} r_{j-k} \quad (3.11)$$

para $j = 1, 2, \dots, k$.

Um método recursivo [Box, et al -1994] pode ser usado para calcular $\widehat{\phi}_{k,k}$. Para ilustrar o método considere a equação (3.11) para $k = 2$ e $k = 3$. Assim temos:

$$r_2 = \widehat{\phi}_{2,1}r_1 + \widehat{\phi}_{2,2} \quad (3.12)$$

$$r_1 = \widehat{\phi}_{2,1} + \widehat{\phi}_{2,2}r_1 \quad (3.13)$$

$$r_3 = \widehat{\phi}_{3,1}r_2 + \widehat{\phi}_{3,2}r_1 + r_{3,3} \quad (3.14)$$

$$r_2 = \widehat{\phi}_{3,1}r_1 + \widehat{\phi}_{3,3} + \widehat{\phi}_{3,3}r_1 \quad (3.15)$$

$$r_1 = \widehat{\phi}_{3,1} + \widehat{\phi}_{3,2}r_1 + \widehat{\phi}_{3,3}r_2 \quad (3.16)$$

Os coeficientes $\widehat{\phi}_{3,1}$ e $\widehat{\phi}_{3,2}$ podem ser expressos em termos de $\widehat{\phi}_{3,3}$ usando as equações (3.15) e (3.16). A solução pode ser escrito na forma matricial como se segue:

$$\begin{bmatrix} \widehat{\phi}_{3,1} \\ \widehat{\phi}_{3,2} \end{bmatrix} = \mathbf{R}_2^{-1} \begin{bmatrix} r_2 - \widehat{\phi}_{3,3}r_1 \\ r_1 - \widehat{\phi}_{3,3}r_2 \end{bmatrix} \quad (3.17)$$

onde $\mathbf{R}_2 = \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}$

Agora (3.17) pode ser escrito como:

$$\begin{bmatrix} \widehat{\phi}_{3,1} \\ \widehat{\phi}_{3,2} \end{bmatrix} = \mathbf{R}_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix} - \widehat{\phi}_{3,3} \mathbf{R}_2^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad (3.18)$$

Usando o fato de que (3.12) e (3.13) podem ser escritos como:

$$\begin{bmatrix} \widehat{\phi}_{2,1} \\ \widehat{\phi}_{2,2} \end{bmatrix} = \mathbf{R}_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix} \quad (3.19)$$

Substituindo (3.19) em (3.18), temos:

$$\begin{bmatrix} \hat{\phi}_{3,1} \\ \hat{\phi}_{3,2} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{2,1} \\ \hat{\phi}_{2,2} \end{bmatrix} - \hat{\phi}_{3,3} \begin{bmatrix} \hat{\phi}_{2,2} \\ \hat{\phi}_{2,1} \end{bmatrix} \quad (3.20)$$

Agora substituindo as expressões de $\hat{\phi}_{3,1}$ e $\hat{\phi}_{3,2}$ de (3.20) em (3.14), e resolvendo essa equação para $\hat{\phi}_{3,3}$, temos:

$$\hat{\phi}_{3,3} = \frac{r_3 - \hat{\phi}_{2,1}r_2 - \hat{\phi}_{2,2}r_1}{1 - \hat{\phi}_{2,1}r_1 - \hat{\phi}_{2,2}r_2} \quad (3.21)$$

Assim de (3.12) e (3.13) estimamos $\hat{\phi}_{2,1}$ e $\hat{\phi}_{2,2}$ e em seguida usando (3.21) calcula-se $\hat{\phi}_{3,3}$. Usando-se (3.20) calcula-se $\hat{\phi}_{3,1}$ e $\hat{\phi}_{3,2}$. No caso geral, temos:

$$\hat{\phi}_{p+1,j} = \hat{\phi}_{p,j} - \hat{\phi}_{p+1,p+1} \hat{\phi}_{p,p-j+1}, \quad j = 1, 2, \dots, p \quad (3.22)$$

$$\hat{\phi}_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^p \hat{\phi}_{p,j} r_{p+1-j}}{1 - \sum_{j=1}^p \hat{\phi}_{p,j} r_j} \quad (3.23)$$

Para ilustrar a aplicação desse método considere uma série para a qual estimamos $r_1 = 0.806$, $r_2 = 0.428$ e $r_3 = 0.070$. Então:

$$\hat{\phi}_{2,1} = \frac{r_1(1-r_2)}{1-r_1^2} = 1.316$$

$$\hat{\phi}_{2,2} = \frac{r_2 - r_1^2}{1 - r_1^2} = -0.632$$

Usando (3.21) temos:

$$\hat{\phi}_{3,3} = \frac{0.070 - (1.316)(0.428) + (0.632)(0.806)}{1 - (1.316)(0.806) + (0.632)(0.428)} = 0.077$$

Substituindo os valores de $\hat{\phi}_{2,1}$ e $\hat{\phi}_{2,2}$ em (3.20) temos:

$$\hat{\phi}_{3,1} = 1.316 + (0.077)(0.632) = 1.365$$

$$\hat{\phi}_{3,2} = -0.632 - (0.077)(1.316) = -0.733$$

Em seguida podemos usar esses valores de $\hat{\phi}_{3,1}$ e $\hat{\phi}_{3,2}$ em (3.23) com $p = 3$ após estimarmos r_4 e calcular:

$$\hat{\phi}_{4,4} = \frac{r_4 - \hat{\phi}_{3,1}r_3 - \hat{\phi}_{3,2}r_2 - \hat{\phi}_{3,3}r_1}{1 - \hat{\phi}_{3,1}r_1 - \hat{\phi}_{3,2}r_2 - \hat{\phi}_{3,3}r_3}$$

Prosseguindo assim sucessivamente estima-se a FACP de um processo de forma iterativa.

Erro padrão dos estimadores da função de autocorrelação parcial

É mostrado por Quenouilli [Quenouilli M.H. -1949, apud, Box et al -1994] que sobre as hipóteses de que o processo auto-regressivo de ordem p , as funções de autocorrelações parciais estimadas de ordem $p+1$, e superiores, são aproximadamente independente distribuído. Também se o número de observações N usadas é apropriado,

$$Var[\hat{\phi}_{kk}] \simeq \frac{1}{N} \quad k \geq p+1$$

Então o desvio padrão (SE) do estimador de autocorrelação parcial $\hat{\phi}_{kk}$ é,

$$SE[\hat{\phi}_{kk}] = \sqrt{Var[\hat{\phi}_{kk}]} \simeq \frac{1}{\sqrt{N}} \quad k \geq p+1$$

Assim, supondo que N é grande podemos dizer que ϕ_{kk} tem distribuição aproximadamente normal, o que permite a construção de intervalo de confiança para ϕ_{kk} , usando-se o desvio padrão $SE(\hat{\phi}_{kk})$.

Estimativa Inicial para Processos AR(p)

Uma estimativa inicial para os parâmetros de ϕ_1, \dots, ϕ_p de um processo AR(p) muitas vezes é necessário para servir como ponto de partida para métodos mais precisos de estimativa.

Uma estimativa inicial desses parâmetros podem ser obtidos substituindo-se o estimador r_j para a autocorrelação teórica ρ_j na equações de Yule-Walker e resolvendo-se o sistema linear para obter esses parâmetros. Assim, temos para processo AR(p).

$$r_j = \hat{\phi}_1 r_{j-1} + \hat{\phi}_2 r_{j-2} + \dots + \hat{\phi}_p r_{j-p}$$

Fazendo $j = 1, 2, \dots, p$, temos o sistema linear $p \times p$, dado por:

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & r_2 & \dots & r_{p-1} \\ r_1 & 1 & r_1 & \dots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \dots & 1 \end{pmatrix}_{p \times p} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} \quad (3.24)$$

Vamos denotar por \mathbb{R} a matriz $p \times p$ da equação (3.24), \mathbb{R} é a matriz de correlação do processo, esta matriz é simétrica devido ao fato que $\rho_{-j} = \rho_j$. Se o processo AR(p) é estacionário \mathbb{R} é definida positiva [Box, et al -1994].

A estimativa inicial $(\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)$ para os parâmetros do processo podem ser obtidos, usando-se a notação matricial, $\mathbf{r} = (r_1, \dots, r_p)'$ e $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \dots, \hat{\phi}_p)'$, como:

$$\hat{\boldsymbol{\phi}} = \mathbb{R}^{-1} \mathbf{r} \quad (3.25)$$

Usando-se as estimativas r_1, \dots, r_p temos uma estimativa, pelo método dos momentos para a variância do resíduo a_t , substituindo ρ_k por r_k , na equação (2.10), assim temos:

$$\hat{\sigma}_a^2 = c_0(1 - \phi_1 r_1 - \phi_2 r_2 - \dots - \phi_p r_p) \quad (3.26)$$

onde c_0 é uma estimativa da variância do processo γ_0 . Muitas vezes é conveniente usar o inverso da variância $\hat{\tau} = 1/\sigma_a^2$, nos métodos de inferência, isto será feito em capítulos futuros.

Capítulo 4

Abordagem Bayesiana para Modelos AR(p) com Função de Verossimilhança Aproximada

4.1 Função de Verossimilhança Aproximada

Considere a série temporal Z_t , $t = 1, \dots, N$ gerada por um processo auto-regressivo de ordem p , AR(p):

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t \quad (4.1)$$

onde $\phi \in R^p$ é um vetor $p \times 1$, $\{a_t, t = 1, 2, \dots\}$ é uma sequência de variáveis aleatória i.i.d. $N(0, \tau^{-1})$ e $\tau = 1/\sigma_a^2 > 0$. Vamos supor agora um conjunto de N observações para essa série que denotamos por $\mathbf{Z}_N = (Z_1, \dots, Z_N)'$, e vamos denotar por $\mathbf{Z}_p = (Z_1, \dots, Z_p)'$ as p -primeiras observações do processo e denotando por $\mathbf{Z} = (Z_{p+1}, \dots, Z_N)$ o restante das observações. Escrevendo o modelo (4.1) como:

$$a_t = Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} \quad (4.2)$$

Podemos escrever que a densidade de probabilidade conjunta para $\{a_t, t = p+1, \dots, N\}$ como:

$$f(a_{p+1}, \dots, a_N | \phi, \tau) \propto \tau^{\frac{N-p}{2}} \exp\left\{-\frac{\tau}{2} \sum_{t=p+1}^N a_t^2\right\} \quad (4.3)$$

A equação (4.2) relaciona os ruídos (a_{p+1}, \dots, a_N) com as observações \mathbf{Z} . Levando em consideração que a relação (4.1) tem Jacobiano unitário, podemos escrever a densidade de probabilidade conjunta dos dados \mathbf{Z} , assumindo fixas as p -primeiras observações \mathbf{Z}_p , como:

$$f(\mathbf{Z} | \phi, \tau) \propto \tau^{\frac{N-p}{2}} \exp\left\{-\frac{\tau}{2} \sum_{t=p+1}^N (Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p})^2\right\} \quad (4.4)$$

Considerando o conjunto de dados \mathbf{Z} e definindo-se a matriz \mathbf{X} :

$$\mathbf{Z} = \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_N \end{pmatrix}; \mathbf{X} = \begin{pmatrix} Z_1 & \cdots & Z_p \\ Z_2 & \cdots & Z_{p+1} \\ \cdots & \cdots & \cdots \\ Z_{N-p} & \cdots & Z_N \end{pmatrix}_{N \times p}; \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}_{p \times 1}$$

A função de verossimilhança, associada com os parâmetros $(\boldsymbol{\phi}, \tau)$, considerando fixas as p -observações iniciais \mathbf{Z}_p , pode ser escrita a partir de (4.4) como:

$$\mathcal{L}(\boldsymbol{\phi}, \tau | \mathbf{Z}) \propto \tau^{\frac{N-p}{2}} \exp\left\{-\frac{\tau}{2} (\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})' (\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})\right\} \quad (4.5)$$

Podemos verificar que:

$$(\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})' (\mathbf{Z} - \mathbf{X}\boldsymbol{\phi}) = (\mathbf{Z} - \hat{\mathbf{Z}})' (\mathbf{Z} - \hat{\mathbf{Z}}) + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}) \quad (4.6)$$

Onde $\hat{\mathbf{Z}}$ é dado por:

$$\hat{\mathbf{Z}} = \mathbf{X}\hat{\boldsymbol{\phi}}$$

Substituindo (4.6) em (4.5) essa função pode ser escrita como:

$$\mathcal{L}(\boldsymbol{\phi}, \tau | \mathbf{Z}) \propto \tau^{\frac{N-p}{2}} \exp\left\{-\frac{\tau}{2} [(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}) + (\mathbf{Z} - \hat{\mathbf{Z}})' (\mathbf{Z} - \hat{\mathbf{Z}})]\right\} \quad (4.7)$$

A função de verossimilhança (4.5) é chamada de função de verossimilhança aproximada por ter sido construída a partir da densidade conjunta dos ruídos a_t condicionada as p -primeiras observações \mathbf{Z}_p , esta função representa somente a densidade de probabilidade conjunta das observações $\mathbf{Z} = (Z_{p+1}, \dots, Z_N)$, não considerando portanto a densidade conjunta das p -primeiras observações. Por isso denotamos essa função por $\mathcal{L}(\boldsymbol{\phi}, \tau | \mathbf{Z})$, condicionada somente a parte dos dados dada por $\mathbf{Z} = (Z_{p+1}, \dots, Z_N)$. Tomando-se o logarítmo de (4.5) podemos calcular os estimadores de máxima verossimilhança (EMV) para $\boldsymbol{\phi}$ e τ , seja:

$$l(\boldsymbol{\phi}, \tau | \mathbf{Z}) = \frac{N-p}{2} \ln(\tau) - \frac{\tau}{2} (\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})' (\mathbf{Z} - \mathbf{X}\boldsymbol{\phi}) \quad (4.8)$$

Derivando (4.8) com relação a ϕ e τ , e igualando a zero, temos:

$$\frac{\partial l(\phi, \tau | \mathbf{Z})}{\partial \phi} = -\tau \mathbf{X}'(\mathbf{Z} - \mathbf{X}\phi) = 0$$

$$\frac{\partial l(\phi, \tau | \mathbf{Z})}{\partial \tau} = \frac{N-p}{2} \tau^{-1} - \frac{1}{2}(\mathbf{Z} - \mathbf{X}\phi)'(\mathbf{Z} - \mathbf{X}\phi)$$

Os EMV são dados por:

$$\hat{\phi} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z})$$

$$\hat{\tau}^{-1} = \frac{1}{N-p} (\mathbf{Z} - \mathbf{X}\hat{\phi})'(\mathbf{Z} - \mathbf{X}\hat{\phi})$$

A matriz de informação de Fisher, é dada por:

$$I(\phi, \tau) = \begin{bmatrix} -E(\hat{\tau}\mathbf{X}'\mathbf{X}) & E(\mathbf{X}'\mathbf{Z} - \mathbf{X}'\mathbf{X}\phi) \\ E(\mathbf{X}'\mathbf{Z} - \mathbf{X}'\mathbf{X}\phi) & -E\left(\frac{N-p}{2}\hat{\tau}^{-2}\right) \end{bmatrix} \quad (4.9)$$

A inversa de (4.9) é dada por:

$$I^{-1} = \begin{bmatrix} \hat{\tau}^{-1}(-E(\mathbf{X}'\mathbf{X})^{-1}) & 0 \\ 0 & -E\left(\frac{2}{N-p}\hat{\tau}^2\right) \end{bmatrix} \quad (4.10)$$

Os intervalos de confiança para os parâmetros ϕ podem ser construídos determinando-se as variâncias $Var[\phi] \cong (Var[\phi_1], \dots, Var[\phi_p])'$ através do inverso da matriz de informação de Fisher ou seja:

$$Var[\phi, \tau] = \begin{bmatrix} -E\left(\frac{\partial^2 l}{\partial \phi^2}\right) & -E\left(\frac{\partial^2 l}{\partial \phi \partial \tau}\right) \\ -E\left(\frac{\partial^2 l}{\partial \phi \partial \tau}\right) & -E\left(\frac{\partial^2 l}{\partial \tau^2}\right) \end{bmatrix}^{-1}$$

Assim temos:

$$Var[\phi] = -\tau[E(X'X)]^{-1}$$

Sendo $E(X'X) = N\mathbb{P}$, onde \mathbb{P} é a matriz de autocovariância do processo, temos:

$$Var[\phi] = -N^{-1}\tau\mathbb{P}^{-1}$$

Substituindo $\tau = \gamma_0[1 - \phi\rho]$ temos:

$$Var[\hat{\phi}] = -N^{-1}\gamma_0[1 - \phi\rho]\mathbb{P}^{-1}$$

Estimando γ_0 por c_0 e $\mathbb{P} = \gamma_0\mathbb{R}$, onde \mathbb{R} é a matriz de autocorrelação do processo temos:

$$Var[\hat{\phi}] \approx -N^{-1}[1 - \hat{\phi}\mathbf{r}]\mathbb{R}$$

Onde \mathbf{r} é o vetor com as p -autocorrelações estimadas e $\hat{\phi}$ os estimadores dos parâmetros. Assim temos, o intervalo de confiança para ϕ , com nível de significância α , quando N é muito grande, dado por:

$$\phi = \hat{\phi} \pm Z_{\alpha/2} \sqrt{Var[\hat{\phi}]}$$

Note que a abordagem de máxima verossimilhança para os modelos $AR(p)$ é semelhante a dos modelos de regressão linear multivariado exceto pelo fato das variáveis Z_t não serem independentes.

4.2 Abordagem Bayesiana de Modelos Lineares $AR(p)$

Sob o enfoque Bayesiano a análise de séries temporais consiste em determinar a densidade a posteriori do modelo paramétrico e a densidade preditiva de observações futuras.

Neste capítulo serão descritos os principais conceitos envolvidos no modo de fazer inferência sob o ponto de vista Bayesiano. O interesse principal é mostrar como para um dado problema são formulados argumentos necessários para a construção de um modelo, aqui nós introduzimos análises a priori, posteriori e preditivas de modelos auto-regressivos (AR).

A análise Bayesiana de modelos AR(p) começa pela escolha de uma função densidade a priori para os parâmetros ϕ e τ .

4.2.1 Considerando Priori Não Informativa

Quando temos "pouca" informação sobre os parâmetros podemos adotar a priori não informativa de Jeffreys; [Box & Tiao -1973]. Supondo ϕ e τ , independentes, temos a densidade conjunta a priori dada por:

$$\Pi_0(\phi, \tau) \propto 1/\tau, \tau > 0 \quad (4.11)$$

Combinando-se a função de verossimilhança (4.7) com a priori (4.11), temos a densidade a posteriori:

$$\Pi(\phi, \tau) \propto \mathcal{L}(\phi, \tau | Z) \Pi_0(\phi, \tau) \quad (4.12)$$

A densidade conjunta a posteriori (4.12) é uma densidade Normal-Gama, dada por:

$$\Pi(\phi, \tau) \propto \tau^{\frac{N-p}{2}-1} \exp\left\{-\frac{\tau}{2}[(\phi - \hat{\phi})' X' X (\phi - \hat{\phi}) + (Z - \hat{Z})'(Z - \hat{Z})]\right\} \quad (4.13)$$

Neste caso em particular, onde estamos usando uma priori não informativa, podemos calcular as densidades marginais para os parâmetros ϕ e τ , integrando diretamente (4.13). Assim temos:

$$\begin{aligned} \Pi(\tau) \propto \tau^{\frac{N-p}{2}-1} \exp\left\{-\frac{\tau}{2}(Z - \hat{Z})'(Z - \hat{Z})\right\} \\ \int_{\phi} \exp\left\{-\frac{\tau}{2}[(\phi - \hat{\phi})' X' X (\phi - \hat{\phi})]\right\} d\phi \end{aligned} \quad (4.14)$$

A integral (4.14) representa uma integral múltipla para os parâmetros ϕ . Resolvendo essa integral temos:

$$\Pi(\tau) \propto \tau^{\frac{1}{2}(N-2p)-1} \exp\left\{-\frac{\tau}{2}(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})\right\} \quad (4.15)$$

Portanto τ tem uma densidade a posteriori Gama $(\frac{N-2p}{2}, (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})/2)$.

Temos então:

$$E(\tau|\mathbf{Z}) = (N - 2p)[(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]^{-1}$$

$$Var(\tau|\mathbf{Z}) = 2(N - 2p)[(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]^{-2}$$

Para encontrar agora a densidade a posteriori para os parâmetros ϕ , integramos (4.13) com relação a τ , temos:

$$\Pi(\phi) \propto \int_0^\infty \tau^{\frac{N-2p}{2}-1} \exp\left\{-\frac{\tau}{2}[(\phi - \hat{\phi})' \mathbf{X}' \mathbf{X} (\phi - \hat{\phi}) + (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]\right\} d\tau$$

$$\Pi(\phi) \propto [(\phi - \hat{\phi})' \mathbf{X}' \mathbf{X} (\phi - \hat{\phi}) + (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]^{-\frac{N-2p}{2}}$$

Após algumas simplificações algébricas, podemos escrever a densidade a posteriori dos parâmetros ϕ , como:

$$\Pi(\phi) \propto \left[1 + \frac{(\phi - \hat{\phi})' \mathbf{V} (\phi - \hat{\phi})}{N-2p}\right]^{-\frac{N-2p}{2}} = \left[1 + \frac{(\phi - \hat{\phi})' \mathbf{V} (\phi - \hat{\phi})}{v}\right]^{-\frac{1}{2}(v+p)}$$

onde $\mathbf{V} = (N - 2p)(\mathbf{X}' \mathbf{X})[(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]^{-1}$ e $v = N - 2p$

Portanto ϕ tem densidade a posteriori t -Student, p dimensional, com $v = (N - 2p)$ graus de liberdade, vetor de localização $\hat{\phi}$ e matriz de precisão \mathbf{V} . O estimador bayesiano dos parâmetros, é o vetor esperado a posteriori de ϕ ou seja, $\phi_b = E(\phi|\mathbf{Z})$. sendo:

$$E(\phi|\mathbf{Z}) = \phi_b = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}$$

$$Var(\phi|\mathbf{Z}) = \frac{v}{v-2} \mathbf{V}^{-1} = \frac{(\mathbf{X}' \mathbf{X})^{-1} [(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]}{N-2p-2}$$

Notamos que neste caso, quando usa-se densidade a priori não-informativa, o estimador bayesiano coincide com o EMV, $E(\phi|\mathbf{Z}) = \hat{\phi}$.

4.2.2 Considerando Priori Conjugada Normal-Gama

A função de verossimilhança (4.7) sugere o uso de uma densidade a priori conjugada Normal-Gama. Assim, temos:

$$\Pi_0(\phi, \tau) = \Pi_1(\phi|\tau) \Pi_2(\tau) \quad (4.16)$$

$$\Pi_1(\phi|\tau) \propto \tau^{\frac{p}{2}} \exp\left\{-\frac{\tau}{2}(\phi - \mu)'P(\phi - \mu)\right\}$$

$$\Pi_2(\tau) \propto \tau^{\alpha-1} \exp\{-\beta\tau\}$$

onde $\phi \in R^p$, $P \in R^{p \times p}$ e α, β escalares são hiperparâmetros.

Combinando a função de verossimilhança (4.7) com a densidade a priori (4.16) temos a densidade a posteriori dada por:

$$\Pi(\phi, \tau) \propto \tau^{\frac{(N-p+2\alpha)}{2}-1} \exp\{-\tau D\} \tau^{\frac{p}{2}} \exp\left\{-\frac{\tau}{2}(\phi - \phi_b)'V(\phi - \phi_b)\right\} \quad (4.17)$$

onde,

$$\phi_b = (X'X + P)^{-1}(X'Z + P\mu)$$

$$V = X'X + P$$

$$D = \beta + \frac{1}{2}[(Z'Z + \mu'P\mu) - (X'Z + P\mu)'(X'X + P)^{-1}(X'Z + P\mu)]$$

Devido a conjugação, temos que a densidade conjunta a posteriori (4.17) é uma Normal-Gama. Assim, podemos determinar a densidade a posteriori de τ , integrando (4.17) com relação aos parâmetros ϕ . Procedendo desta forma, temos:

$$\Pi(\tau) \propto \tau^{\frac{(N-p+2\alpha)}{2}-1} \exp\{-\tau D\} \int_{\phi} \tau^{\frac{p}{2}} \exp\left\{-\frac{\tau}{2}[(\phi - \phi_b)'V(\phi - \phi_b)]\right\} d\phi$$

$$\Pi(\tau) \propto \tau^{\frac{(N-p+2\alpha)}{2}-1} \exp\{-\tau D\} \quad (4.18)$$

Temos portanto que o valor esperado e a variância a posteriori de τ são dados por:

$$E(\tau|\mathbf{Z}) = (N - p + 2\alpha)(2\mathbf{D})^{-1} \quad (4.19)$$

$$Var(\tau|\mathbf{Z}) = E(\tau|\mathbf{Z})(\mathbf{D})^{-1} \quad (4.20)$$

Da mesma forma, integrando (4.17) com relação a τ , temos a densidade marginal a posteriori para os parâmetros ϕ , dada por:

$$\Pi(\phi) \propto \int_0^\infty \tau^{\frac{(N+2\alpha)}{2}-1} \exp\left\{-\frac{\tau}{2}[(\phi - \phi_b)'V(\phi - \phi_b) + 2\mathbf{D}]\right\} d\tau$$

$$\Pi(\phi) \propto [2\mathbf{D} + (\phi - \phi_b)'V(\phi - \phi_b)]^{-\frac{N+2\alpha}{2}} \quad (4.21)$$

Reagrupando os termos, podemos escrever (4.21) como:

$$\Pi(\phi) \propto \left[1 + \frac{(\phi - \phi_b)'W(\phi - \phi_b)}{N - p + 2\alpha}\right]^{-\frac{N+2\alpha}{2}} \quad (4.22)$$

onde,

$$W = (N - p + 2\alpha)V(2\mathbf{D})^{-1}$$

Portanto, vemos que ϕ tem densidade a posteriori t -Student, p -dimensional, com $v = (N - p + 2\alpha)$ graus de liberdade, vetor de localização ϕ_b e matriz de precisão W . Então podemos estimar os parâmetros ϕ , pelo valor esperado da densidade (4.22) e a matriz de covariância dos parâmetros pode ser obtida diretamente da matriz de precisão W . Assim temos:

$$E(\phi|\mathbf{Z}) = (\mathbf{X}'\mathbf{X} + P)^{-1}(\mathbf{X}'\mathbf{Z} + \mu P) \quad (4.23)$$

$$Var(\phi|\mathbf{Z}) = \frac{v}{v-2} W^{-1} = \frac{(\mathbf{X}'\mathbf{X} + P)^{-1}(2\mathbf{D})}{(N - p + 2\alpha - 2)} \quad (4.24)$$

Note de (4.23) que o valor esperado a posteriori, ϕ_b , é uma média ponderada do estimador de máxima verossimilhança $\hat{\phi}$ e da média da densidade a priori, μ , ou seja:

$$\phi_b = (\mathbf{X}'\mathbf{X} + P)^{-1}(\mathbf{X}'\mathbf{Z} + P\mu) = (\mathbf{X}'\mathbf{X} + P)^{-1}[(\mathbf{X}'\mathbf{X})\hat{\phi} + P\mu] \quad (4.25)$$

4.2.3 Considerando Priori t -Student

Supondo agora uma densidade a priori para os parâmetros ϕ , como sendo uma densidade t -Student p -dimensional, com graus de liberdade κ , e vamos considerar uma densidade a priori gama com parâmetros α e β para τ , assim temos:

$$\Pi_1(\phi) \propto \left(1 + \frac{(\phi - \mu)' P (\phi - \mu)}{\kappa}\right)^{-\frac{(\kappa+p)}{2}} \quad (4.26)$$

$$\Pi_2(\tau) \propto \tau^{\alpha-1} \exp\{-\beta\tau\} \quad (4.27)$$

com grau de liberdade κ , vetor de localização μ e matriz de precisão P . A análise a posteriori é obtida combinando-se a (4.7) com (4.26) e (4.27), desta forma temos:

$$\Pi(\phi, \tau) \propto \tau^{\frac{N-p+2\alpha}{2}-1} \exp\{-\tau[\beta + B(\phi)/2]\} \Pi_1(\phi) \quad (4.28)$$

onde:

$$B(\phi) = (\phi - \hat{\phi})'(X'X)(\phi - \hat{\phi}) + (Z - \hat{Z})'(Z - \hat{Z}).$$

Notamos aqui que a densidade conjunta a posteriori não é uma densidade padrão, portanto só podemos avaliar as densidades a posteriori marginais por meio de métodos de aproximação, tais como o método de Laplace [Tierney, Kass & Kadane -1986] ou usando métodos de simulação de Monte Carlo em Cadeia de Markov, tais como, Gibbs Sampling [Casela & George -1992] e Metropolis-Hastings [Chib & Greenberg -1995]. Neste trabalho vamos optar pelos métodos de simulação por serem precisos e de fácil implementação computacional. Com a finalidade de fazer uma ilustração do uso de algoritmos de MCMC vamos fazer uma estimação dos parâmetros ϕ e τ para os casos em que temos priori conjugada.

4.3 Algoritmos de MCMC

Algoritmo Gibbs Sampling:

A técnica do Amostrador de Gibbs é usada para gerar variáveis aleatórias de uma distribuição sem utilizar sua densidade. O Amostrador de Gibbs é essencialmente um esquema

iterativo de amostragem de uma cadeia de Markov cujo núcleo de transição é formado pelas distribuições condicionais.

Modelo com Priori não Informativa

Neste caso, temos uma densidade conjunta a posteriori, Normal-Gama, (4.13), dada por:

$$\Pi(\phi, \tau) \propto \tau^{\frac{N-p}{2}-1} \exp\left\{-\frac{\tau}{2}[(\phi - \hat{\phi})'(\mathbf{X}'\mathbf{X})(\phi - \hat{\phi}) + (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]\right\}$$

As densidades a posteriori condicionais, são:

$$\Pi(\phi|\tau) \sim N(\hat{\phi}, (\tau\mathbf{V})^{-1}) \quad (4.29)$$

$$\Pi(\tau|\phi) \sim \Gamma\left(\frac{N-p}{2}, \mathbf{B}\right) \quad (4.30)$$

onde, $\mathbf{B} = \frac{1}{2}[(\phi - \hat{\phi})'(\mathbf{X}'\mathbf{X})(\phi - \hat{\phi}) + (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})]$, $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}$

Modelo com Priori Informativa Normal-Gama

A posteriori conjugada para este caso dada por (4.17) é:

$$\Pi(\phi, \tau) \propto \tau^{\frac{(N-p+2\alpha)}{2}-1} \exp\{-\tau\mathbf{D}\} \tau^{\frac{p}{2}} \exp\left\{-\frac{\tau}{2}(\phi - \phi_b)' \mathbf{V}(\phi - \phi_b)\right\}$$

As densidades a posteriori condicionais, são:

$$\Pi(\phi|\tau) \sim N(\phi_b, (\tau\mathbf{V})^{-1}) \quad (4.31)$$

$$\Pi(\tau|\phi) \sim \Gamma\left(\frac{N+2\alpha}{2}, \mathbf{D} + \frac{1}{2}(\phi - \phi_b)' \mathbf{V}(\phi - \phi_b)\right) \quad (4.32)$$

onde,

$$\mathbf{V} = (\mathbf{X}'\mathbf{X} + \mathbf{P})$$

$$\mathbf{D} = \beta + \frac{1}{2}[(\mathbf{Z}'\mathbf{Z} + \mu'\mathbf{P}\mu) - (\mathbf{X}'\mathbf{Z} + \mathbf{P}\mu)'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Z} + \mathbf{P}\mu)]$$

e usamos as condicionais para gerar os valores dos parâmetros como segue:

passo 1: Inicialize o contador de iterações $j = 1$ e arbitre valores iniciais:

$$\phi^{(0)} = (\phi_1^{(0)}, \dots, \phi_p^{(0)}), \tau^{(0)}$$

passo 2: Obtenha os novos valores gerando-se a partir das densidades condicionais como se segue:

$$\phi^{(j+1)} \sim \Pi(\phi | \tau^{(j)})$$

$$\tau^{(j+1)} \sim \Pi(\tau | \phi^{(j+1)})$$

passo 3: Atualize o contador de iterações para $j \leftarrow j + 1$ e volte ao passo 2, repetir este procedimento até a convergência.

Após um número grande de iterações as sequências geradas (ϕ^j e τ^j) convergirão para uma amostra aleatória das variáveis $\phi \sim \Pi(\phi)$ e $\tau \sim \Pi(\tau)$. Para verificar se a convergência já foi atingida existem vários métodos [Cowles & Carlin -1996]. Neste trabalho vamos usar o critério de Gelman & Rubin (1992); ver apêndice 2.

Algoritmo Metropolis-Hastings & Gibbs Sampling:

Quando as distribuições condicionais não são facilmente identificadas e não podemos gerar amostras dessas distribuições diretamente, nós usamos o algoritmo de Metropolis-Hastings.

Modelo com Priori Informativa t -Student

Neste caso a posteriori conjugada é dada em (4.28).

$$\Pi(\phi, \tau) \propto \tau^{\frac{N-p+2\alpha}{2}-1} \exp\{-\tau[\beta + B(\phi)/2]\} \Pi_1(\phi)$$

As posteriori condicionais são:

$$\Pi(\phi|\tau) \propto \tau^{\frac{N-p+2\alpha}{2}-1} \exp\left\{-\tau\left[\beta + \frac{1}{2}(\phi - \hat{\phi})'(X'X)(\phi - \hat{\phi})\right]\right\} \Pi_1(\phi) \quad (4.33)$$

$$\Pi(\tau|\phi) \sim \Gamma\left(\frac{N-p-2\alpha}{2}, \beta + \frac{1}{2}B(\phi)\right) \quad (4.34)$$

Aqui é conveniente escrever:

$$\Pi(\phi|\tau) \propto \Psi(\phi) \Pi_1(\phi)$$

onde:
$$\Psi(\phi) = \tau^{\frac{N-p+2\alpha}{2}-1} \exp\left\{-\tau[\beta + B(\phi)/2]\right\}$$

e usamos $\Pi_1(\phi)$ como núcleo para gerar os valores do parâmetro ϕ como se segue:

1) Inicialize o contador de iterações com $j = 1$ e arbitre valores iniciais com

$$\phi^{(0)} = (\phi_1^{(0)}, \dots, \phi_p^{(0)}), \tau^{(0)}$$

2) Obtenha um novo valor a partir da função transição $\Pi_1(\phi, \phi^*)$

3) Calcule a probabilidade de aceitação do novo valor na iteração j por:

$$\alpha(\phi^j, \phi^*) = \begin{cases} \min\left\{1, \frac{\psi(\phi^*)}{\psi(\phi^j)}\right\} & \text{se } \psi(\phi^j) > 0 \\ 1 & \text{caso contrário.} \end{cases}$$

4) Gerar u de Uniforme[0, 1] e faça:

$$\phi^{j+1} = \begin{cases} \phi^* & \text{se } u \leq \alpha(\phi, \phi^*) \\ \phi^j & \text{caso contrário} \end{cases}$$

O valor gerado ϕ^* é aceito com probabilidade α ou rejeitado com probabilidade $(1 - \alpha)$.

5) Gerar $\tau^{(j)} \sim \text{Gama}(\tau|\phi^{j+1})$

6) Atualize o contador de iteração j para $j + 1$ e volte para o passo 2 até a convergência.

Capítulo 5

Previsão

5.1 Análise Preditiva

Um dos objetivos do uso de modelos de séries temporais é fazer previsão de valores futuros da série, Z_{N+1}, Z_{N+2}, \dots .

Previsão um passo a frente

Considerando uma observação futura Z_{N+1} , dos dados $\mathbf{Z} = (Z_1, Z_2, \dots, Z_N)'$.

Seja,

$$\Pi(Z_{N+1} | \mathbf{Z}, \phi, \tau) \propto \tau^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} \left(Z_{N+1} - \sum_{i=1}^p \phi_i Z_{N-(i-1)}\right)^2\right\} \quad (5.1)$$

sendo:

$$\left(Z_{N+1} - \sum_{i=1}^p \phi_i Z_{N-(i-1)}\right)^2 = Z_{N+1}^2 - 2 Z_{N+1} \sum_{i=1}^p \phi_i Z_{N-(i-1)} + \left(\sum_{i=1}^p \phi_i Z_{N-(i-1)}\right)^2$$

fazendo:

$$Z_{N+1} \sum_{i=1}^p \phi_i Z_{N-(i-1)} = \underbrace{[\phi_1, \dots, \phi_p]}_{\phi'} Z_{N+1} \underbrace{\begin{bmatrix} Z_N \\ \vdots \\ Z_{N-(p-1)} \end{bmatrix}}_F = \phi' F$$

$F = Z_{N+1} \circ L$

e

$$\left(\sum_{i=1}^p \phi_i Z_{N-(i-1)}\right)^2 = \phi' \underbrace{E}_{(LL')} \phi$$

onde E é uma matriz $p \times p$ (onde o ij^{th} elemento é $Z_{t-(i-1)} Z_{t-(j-1)}$) e F é uma matriz $p \times 1$ (onde o j^{th} elemento é $Z_{N+1} Z_{N-(j-1)}$). Assim temos a densidade preditiva condicional,

$$\Pi(Z_{N+1} | \mathbf{Z}, \phi, \tau) \propto \tau^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} (Z_{N+1}^2 + \phi' E \phi - 2 \phi' F)\right\} \quad (5.2)$$

A preditiva posteriori condicional é dada por:

$$\Pi(Z_{N+1}, \phi, \tau | \mathbf{Z}) \propto \Pi(Z_{N+1} | \mathbf{Z}, \phi, \tau) \Pi(\phi, \tau | \mathbf{Z}) \quad (5.3)$$

onde $\Pi(\phi, \tau | \mathbf{Z})$ é a posteriori condicional dos parâmetros dado as observações.

Modelo com Priori não Informativa

Considerando uma priori não informativa, a posteriori é dada por (4.13), fazendo o produto temos:

$$\begin{aligned} \Pi(Z_{N+1}, \phi, \tau | \mathbf{Z}) \propto & \tau^{\frac{N-2p-1}{2}} \exp\left\{-\frac{\tau}{2}[(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}}) + Z_{N+1}^2]\right\} \\ & \exp\left\{-\frac{\tau}{2}[(\phi - \phi_b)' \mathbf{V}(\phi - \phi_b) + \phi' \mathbf{E} \phi - 2 \phi' \mathbf{F}]\right\} \end{aligned} \quad (5.4)$$

Integrando (5.4) em ϕ e τ temos:

$$\Pi(Z_{N+1} | \mathbf{Z}) \propto \left[1 + \frac{\mathbf{P}(Z_{N+1} - \eta)^2}{N - 2p}\right]^{-\frac{N-2p+1}{2}} \quad (5.5)$$

com

$$\eta = (1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1} \mathbf{F}^*)^{-1} \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1} \mathbf{C}$$

$$\mathbf{V} = \mathbf{X}' \mathbf{X}$$

$$\mathbf{C} = \mathbf{X}' \mathbf{Z}$$

$$\mathbf{D} = \frac{1}{2}[\mathbf{Z}' \mathbf{Z} - \mathbf{X}' \mathbf{Z}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}]$$

$$\mathbf{P} = (N - 2p)(1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1} \mathbf{F}^*)^{-1}$$

$$[2\mathbf{D} + \mathbf{C}' \mathbf{V}^{-1} \mathbf{C} - \mathbf{C}'(\mathbf{V} + \mathbf{E})^{-1}(1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1} \mathbf{F}^*)^{-1} \mathbf{C}]^{-1}$$

onde \mathbf{F}^* é um vetor $p \times 1$ com o j^{th} componente dado por $Z_{N-(j-1)}$ e $j = 1, \dots, p$.

Portanto (5.5) é a função densidade preditiva um passo a frente, e é uma t -Student com $v = (N - 2p)$ graus de liberdade, localização η e precisão \mathbf{P} .

Então o valor esperado e a variância da densidade (5.5) pode ser obtida diretamente como:

$$E(Z_{N+1}|\mathbf{Z}) = (1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{F}^{*})^{-1}\mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{C}$$

$$Var(Z_{N+1}|\mathbf{Z}) = \frac{\nu}{\nu-2} \mathbb{P}^{-1}$$

Modelo com Priori Informativa Normal-Gama

Considerando uma priori conjugada Normal-Gama, a posteriori é dada por (4.17), fazendo o produto temos:

$$\Pi(Z_{N+1}, \phi, \tau | \mathbf{Z}) \propto \tau^{\frac{(N-p+2\alpha)-1}{2}} \exp\left\{-\frac{\tau}{2}[(2\mathbf{D} + Z_{N+1}^2)]\right\}$$

$$\tau^{\frac{p}{2}} \exp\left\{-\frac{\tau}{2}[(\phi - \phi_b)' \mathbf{V}(\phi - \phi_b) + \phi' \mathbf{E} \phi - 2\phi' \mathbf{F}]\right\} \quad (5.6)$$

Integrando (5.6) em ϕ e τ temos:

$$\Pi(Z_{N+1}|\mathbf{Z}) \propto \left[1 + \frac{\mathbf{P}(Z_{N+1}-\eta)^2}{N-p+2\alpha}\right]^{-\frac{N-p+2\alpha+1}{2}} \quad (5.7)$$

com

$$\eta = (1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{F}^{*})^{-1}\mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{C}$$

$$\mathbf{V} = \mathbf{X}'\mathbf{X} + \mathbb{P}$$

$$\mathbf{C} = (\mathbf{X}'\mathbf{Z} + \mathbb{P}\boldsymbol{\mu})$$

$$\mathbf{D} = \beta + \frac{1}{2}[(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\mu}'\mathbb{P}\boldsymbol{\mu}) - (\mathbf{X}'\mathbf{Z} + \mathbb{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbb{P})^{-1}(\mathbf{X}'\mathbf{Z} + \mathbb{P}\boldsymbol{\mu})]$$

$$\mathbb{P} = (N - p + 2\alpha)(1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{F}^{*})^{-1}$$

$$[2\mathbf{D} + \mathbf{C}'\mathbf{V}^{-1}\mathbf{C} - \mathbf{C}'(\mathbf{V} + \mathbf{E})^{-1}(1 - \mathbf{F}^{*'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{F}^{*})^{-1}\mathbf{C}]^{-1}$$

Portanto (5.7) é a função densidade preditiva um passo a frente, e é uma t -Student, p dimensional, com $\nu = (N - p + 2\alpha)$ graus de liberdade, localização η e precisão \mathbb{P} .

Então podemos estimar o valor esperado da densidade (5.7) e a variância pode ser obtida diretamente da matriz de precisão \mathbb{P} . Assim temos:

$$E(Z_{N+1}|\mathbf{Z}) = (1 - \mathbf{F}^{**'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{F}^{*})^{-1}\mathbf{F}^{**'}(\mathbf{V} + \mathbf{E})^{-1}\mathbf{C}$$

$$Var(Z_{N+1}|\mathbf{Z}) = \frac{v}{v-2} \mathbb{P}^{-1}$$

Modelo com Priori Informativa t -Student

Considerando uma priori t -Student, a posteriori é dada por (4.28), fazendo o produto temos:

$$\begin{aligned} \Pi(Z_{N+1}, \phi, \tau | \mathbf{Z}) &\propto \tau^{\frac{(N-p+2\alpha)-1}{2}} \exp\left\{-\frac{\tau}{2}[\beta + Z_{N+1}^2]\right\} \\ &\exp\left\{-\frac{\tau}{2}[\mathbf{B}(\phi) + \phi' \mathbf{E} \phi - 2 \phi' \mathbf{F}]\right\} \Pi_1(\phi) \end{aligned} \quad (5.8)$$

Neste caso não é possível integrar a equação (5.8), a previsão de valores futuros só podem ser calculados por simulação ou aproximação numérica.

5.2 Previsão k passos a frente

Denotando por \mathbf{Z}_f os k valores da série a serem previstos e por \mathbf{Z}_p as últimas p observações, ou seja:

$$\mathbf{Z}_f = (Z_{(N+1)}, Z_{(N+2)}, \dots, Z_{(N+k)})'$$

$$\mathbf{Z}_p = (Z_{N-(p-1)}, Z_{N-(p-2)}, \dots, Z_N)'$$

Seja $\Pi(\phi, \tau | \mathbf{Z})$ a densidade a posteriori para os parâmetros ϕ e τ . A densidade preditiva pode ser calculada por:

$$\Pi(\mathbf{Z}_f|\mathbf{Z}) = \int_{\phi} \int_{\tau} \Pi(\mathbf{Z}_f|\phi, \tau, \mathbf{Z}_p) \Pi(\phi, \tau|\mathbf{Z}_p) d\phi d\tau$$

onde \int_{ϕ} denota as p -integrais $\int \dots \int d\phi_1 \dots d\phi_p$.

A previsão dos valores futuros \mathbf{Z}_f pode ser obtida resolvendo-se a integral múltipla,

$$E(\mathbf{Z}_f|\mathbf{Z}) = \int \mathbf{Z}_f \Pi(\mathbf{Z}_f|\mathbf{Z}) d\mathbf{Z}_f$$

$$E(\mathbf{Z}_f|\mathbf{Z}) = \int \mathbf{Z}_f \left[\int_{\phi} \int_{\tau} \Pi(\mathbf{Z}_f|\phi, \tau, \mathbf{Z}_p) \Pi(\phi, \tau|\mathbf{Z}_p) d\phi d\tau \right] d\mathbf{Z}_f$$

Alterando a ordem das integrações temos,

$$E(\mathbf{Z}_f|\mathbf{Z}) = \int_{\phi} \int_{\tau} E(\mathbf{Z}_f|\phi, \tau, \mathbf{Z}_p) \Pi(\phi, \tau|\mathbf{Z}_p) d\phi d\tau \quad (5.9)$$

Usando o modelo AR(p), temos:

$$Z_{N+k} = \phi_1 Z_{N+k-1} + \dots + \phi_p Z_{N-(p-k)} + a_{N+k}$$

Colocando na forma matricial

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\phi_1 & 1 & 0 & \dots & 0 \\ -\phi_2 & -\phi_1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -\phi_{k-1} & -\phi_{k-2} & -\phi_{k-3} & \dots & 1 \end{pmatrix}_{k \times k}$$

$$\mathbf{B} = \begin{pmatrix} -\phi_p & -\phi_{p-1} & -\phi_{p-2} & \dots & -\phi_1 \\ 0 & -\phi_p & -\phi_{p-1} & \dots & -\phi_2 \\ 0 & 0 & -\phi_p & \dots & -\phi_3 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\phi_k \end{pmatrix}_{k \times p}$$

Podemos escrever:

$$\mathbf{AZ}_f + \mathbf{BZ}_p = \mathbf{a}$$

$$\mathbf{Z}_f = -\mathbf{A}^{-1}\mathbf{BZ}_p + \mathbf{A}^{-1}\mathbf{a}$$

Onde \mathbf{a} é um vetor de variáveis aleatória i.i.d. $N(0, \tau^{-1})$. Assim temos que \mathbf{Z}_f também é um vetor de variáveis aleatória normalmente distribuído $N[E(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z}), \text{Var}(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z})]$ onde:

$$E(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z}) = -\mathbf{A}^{-1}\mathbf{BZ}_p \quad (5.10)$$

$$\text{Var}(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z}) = \tau^{-1}(\mathbf{A}^{-1})(\mathbf{A}^{-1})' \quad (5.11)$$

Substituindo (5.10) em (5.9) podemos calcular a previsão dos valores futuros \mathbf{Z}_f , no entanto, para isso é necessário resolver a integral

$$E(\mathbf{Z}_f|\mathbf{Z}) = \int_{\boldsymbol{\phi}} \int_{\tau} (-\mathbf{A}^{-1}\mathbf{BZ}_p) \Pi(\boldsymbol{\phi}, \tau|\mathbf{Z}) d\boldsymbol{\phi} d\tau \quad (5.12)$$

A variância é calculada como:

$$\text{Var}(\mathbf{Z}_f|\mathbf{Z}) = E_{\boldsymbol{\phi}, \tau}[\text{Var}(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z})] + \text{Var}_{\boldsymbol{\phi}, \tau}[E(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z})] \quad (5.13)$$

onde:

$$\begin{aligned} E_{\boldsymbol{\phi}, \tau}[\text{Var}(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z})] &= \int_{\boldsymbol{\phi}} \int_{\tau} \text{Var}(\mathbf{Z}_f|\boldsymbol{\phi}, \tau, \mathbf{Z}) \Pi(\boldsymbol{\phi}, \tau|\mathbf{Z}) d\boldsymbol{\phi} d\tau \\ &= \int_{\boldsymbol{\phi}} \int_{\tau} \tau^{-1}(\mathbf{A}^{-1})(\mathbf{A}^{-1})' \Pi(\boldsymbol{\phi}, \tau|\mathbf{Z}) d\boldsymbol{\phi} d\tau \end{aligned} \quad (5.14)$$

Devido a complexidade das expressões (5.12) e (5.13) quando $p \geq 2$, geralmente essa integral só pode ser calculada de forma aproximada. No entanto, mostraremos que $E(\mathbf{Z}_f|\mathbf{Z})$ e $\text{Var}(\mathbf{Z}_f|\mathbf{Z})$ podem ser calculadas facilmente usando métodos de simulação de Monte-Carlo.

5.3 Previsão usando MCMC

Retomando a equação (5.12) dada por:

$$E(\mathbf{Z}_f|\mathbf{Z}) = \int_{\phi} \int_{\tau} (-\mathbf{A}^{-1}\mathbf{B}\mathbf{Z}_p) \Pi(\phi, \tau|\mathbf{Z}) d\phi d\tau$$

O valor esperado pode ser calculado usando algoritmos de simulação de Monte-Carlo. Para isso seja a amostra $\phi^{(j)} = (\phi_1^{(j)}, \phi_2^{(j)}, \dots, \phi_p^{(j)})$; $j = 1, \dots, m$ gerados por simulação (Amostrador de Gibbs ou Metropolis-Hastings). Podemos calcular $E(\mathbf{Z}_f|\mathbf{Z})$ usando essa amostra, por:

$$E(\mathbf{Z}_f|\mathbf{Z}) = \frac{1}{m} \sum_{j=1}^m -\mathbf{A}^{-1}(\phi^{(j)})\mathbf{B}(\phi^{(j)})\mathbf{Z}_p \quad (5.15)$$

Retomando as equações (5.13) e (5.14) temos:

$$Var(\mathbf{Z}_f|\mathbf{Z}) = E_{\phi, \tau}[Var(\mathbf{Z}_f|\phi, \tau, \mathbf{Z})] + Var_{\phi, \tau}[E(\mathbf{Z}_f|\phi, \tau, \mathbf{Z})] \quad (5.16)$$

Usando simulação de Monte Carlo podemos calcular cada termo desta equação como:

$$E_{\phi, \tau}[Var(\mathbf{Z}_f|\phi, \tau, \mathbf{Z})] = \frac{1}{m} \sum_{j=1}^m (\tau^{(j)})^{-1} [\mathbf{A}^{-1}(\phi^{(j)})][\mathbf{A}^{-1}(\phi^{(j)})]' \quad (5.17)$$

sendo:

$$E(\mathbf{Z}_f|\phi^{(j)}, \tau^{(j)}, \mathbf{Z}) = -\mathbf{A}^{-1}(\phi^{(j)})\mathbf{B}(\phi^{(j)})\mathbf{Z}_p, \quad j = 1, \dots, m$$

Podemos calcular a $Var_{\phi, \tau}$ como:

$$\begin{aligned} Var_{\phi, \tau}[E(\mathbf{Z}_f|\phi, \tau, \mathbf{Z})] &= \frac{1}{m} \sum_{j=1}^m (\mathbf{A}^{-1}(\phi^{(j)})\mathbf{B}(\phi^{(j)})\mathbf{Z}_p)^2 \\ &\quad - \left(\frac{1}{m} \sum_{j=1}^m \mathbf{A}^{-1}(\phi^{(j)})\mathbf{B}(\phi^{(j)})\mathbf{Z}_p \right)^2 \end{aligned} \quad (5.18)$$

Substituindo as equações (5.17) e (5.18) na equação (5.16) podemos calcular $Var(\mathbf{Z}_f|\mathbf{Z})$ por simulação MCMC.

5.4 Considerações sobre Seleção de Modelo

Na análise Bayesiana existem vários critérios para seleção de modelo [Chang -1995]. Aqui vamos considerar o critério baseado na densidade preditiva. A densidade preditiva para Z_{N+k} dado $\mathbf{Z}_k = (Z_1, Z_2, \dots, Z_N, \dots, Z_{N+k-1})'$ e os parâmetros ϕ e τ é dada por:

$$\Pi(Z_{N+k} | \mathbf{Z}_{N+k-1}) \propto \int_{\phi} \int_{\tau} \tau^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} \left(Z_{N+k} - \sum_{i=1}^p \phi_i Z_{N-(i-1)}\right)^2\right\} \Pi(\phi, \tau | \mathbf{Z}) d\phi d\tau \quad (5.19)$$

A equação (5.19) pode também ser calculada usando simulação de Monte Carlo. Neste caso a densidade preditiva é estimada por $c_k = \hat{\Pi}(Z_{N+k} | \mathbf{Z}_{N+k-1})$. Então fazemos o gráfico de c_k versus k ($k = 1, 2, \dots, n$) para diferentes modelos, o maior valor de c_k indica o melhor modelo. Escolhemos o modelo tal que $c(l) = \prod_{i=1}^n c_i(l)$ é máximo (l é o índice do modelo).

O critério de seleção de modelo utilizando simulação de Monte Carlo é feito estimando $\Pi(Z_{N+k} | \mathbf{Z}_{N+k-1})$, usando (5.19), ou seja:

$$\hat{c}_k = \frac{1}{m} \sum_{j=1}^m (\tau^{(j)})^{-1} \exp\left\{-\frac{\tau^{(j)}}{2} \left[Z_{N+k} - \sum_{i=1}^p \phi_i^{(j)} Z_{N-(i-1)}\right]^2\right\}$$

Fazendo o gráfico de c_k versus k ($k = 1, 2, \dots$) e calculando $\hat{c}(l) = \prod_{k=1}^n \hat{c}_k(l)$ para cada modelo l , escolhemos o modelo que apresenta o maior $\hat{c}(l)$.

Capítulo 6

Abordagem Bayesiana para modelos AR(p) com Função de Verossimilhança Exata

6.1 A Função de Verossimilhança Exata para AR(p)

Supondo que uma série $\mathbf{Z} = (Z_1, \dots, Z_n)'$ é gerada por um processo AR(p) dado por:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t$$

Denotando agora por $\mathbf{Z} = (Z_{p+1}, \dots, Z_n)'$ e $\mathbf{Z}_p = (Z_1, Z_2, \dots, Z_p)'$ de forma que $\tilde{\mathbf{Z}} = (\mathbf{Z}_p, \mathbf{Z})$. Além disso seja o vetor dos parâmetros $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$. Então podemos escrever:

$$f(\tilde{\mathbf{Z}} | \boldsymbol{\phi}, \tau) = f(\mathbf{Z} | \mathbf{Z}_p, \boldsymbol{\phi}, \tau) f(\mathbf{Z}_p | \boldsymbol{\phi}, \tau) \quad (6.1)$$

Assumindo normalidade para o ruído $\{a_t, t = 1, 2, \dots\}$, ou seja $a_t \sim N(0, \tau^{-1})$, $\tau > 0$, além disso, assumimos que:

$$E(a_t a_{t+k}) = \begin{cases} \tau^{-1} & k = 0 \\ 0 & k \neq 0 \end{cases}$$

$f(\mathbf{Z} | \mathbf{Z}_p, \boldsymbol{\phi}, \tau)$ pode ser calculado usando a distribuição conjunta dos ruídos $f(a_{p+1}, \dots, a_n)$ dada por:

$$f(a_{p+1}, \dots, a_n) \propto \tau^{\frac{(n-p)}{2}} \exp\left\{-\frac{\tau}{2} \sum_{t=p+1}^n a_t^2\right\} \quad (6.2)$$

e considerando a relação:

$$\begin{aligned} a_{p+1} &= Z_{p+1} - \phi_1 Z_p - \dots - \phi_p Z_1 \\ &\vdots \\ a_n &= Z_n - \phi_1 Z_{n-1} - \dots - \phi_p Z_{n-p} \end{aligned} \quad (6.3)$$

a qual tem jacobiano unitário, temos:

$$f(\mathbf{Z}|\mathbf{Z}_p, \boldsymbol{\phi}, \tau) = \tau^{\frac{(n-p)}{2}} \exp\left\{-\frac{\tau}{2} \sum_{t=p+1}^n (Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p})^2\right\} \quad (6.4)$$

Usando a notação matricial temos:

$$\sum_{t=p+1}^n (Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p})^2 = (\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})'(\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})$$

onde:

$$\mathbf{X} = \begin{pmatrix} Z_p & \dots & Z_1 \\ Z_{p+1} & \dots & Z_2 \\ \vdots & \vdots & \vdots \\ Z_{N-1} & \dots & Z_{N-p} \end{pmatrix}_{n \times p}, \quad \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}_{p \times 1}, \quad \mathbf{Z} = \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_N \end{pmatrix}$$

Usando a relação

$$(\mathbf{Z} - \mathbf{X}\boldsymbol{\phi})'(\mathbf{Z} - \mathbf{X}\boldsymbol{\phi}) = (\mathbf{Z} - \widehat{\mathbf{Z}})'(\mathbf{Z} - \widehat{\mathbf{Z}}) + (\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}}) \quad (6.5)$$

onde:

$$\widehat{\boldsymbol{\phi}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}) \quad (6.6)$$

$$\widehat{\mathbf{Z}} = \mathbf{X}\widehat{\boldsymbol{\phi}} \quad (6.7)$$

Podemos escrever (6.4) como:

$$\begin{aligned} f(\mathbf{Z}|\mathbf{Z}_p, \boldsymbol{\phi}, \tau) &= \tau^{\frac{(n-p)}{2}} \exp\left\{-\frac{\tau}{2} [(\mathbf{Z} - \widehat{\mathbf{Z}})'(\mathbf{Z} - \widehat{\mathbf{Z}}) \right. \\ &\quad \left. + (\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\phi} - \widehat{\boldsymbol{\phi}})]\right\} \end{aligned} \quad (6.8)$$

Usando o fato de que $\mathbf{Z}_p = (Z_1, Z_2, \dots, Z_p)'$ tem distribuição conjunta normal, temos:

$$f(\mathbf{Z} | \mathbf{Z}_p, \boldsymbol{\phi}, \tau) = \tau^{\frac{p}{2}} |\mathbf{M}_p^{(p)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p]\right\} \quad (6.9)$$

onde:

$$\mathbf{M}_p^{(p)} = \{m_{ij}^{(p)}\} = \{\gamma_{|i-j|}\}^{-1} \tau^{-1} = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \dots & \gamma_0 \end{bmatrix}^{-1} \tau^{-1} \quad (6.10)$$

onde $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ são as autocovariâncias teóricas do processo, assim $\gamma_{|i-j|} = \text{cov}(Z_{t+i}, Z_{t+j})$.

Substituindo-se (6.8) e (6.9) em (6.1) temos a verossimilhança exata dada por:

$$f(\tilde{\mathbf{Z}} | \boldsymbol{\phi}, \tau) = \tau^{\frac{(n-p)}{2}} \exp\left\{-\frac{\tau}{2} [(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}}) + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})]\right\} \tau^{\frac{p}{2}} |\mathbf{M}_p^{(p)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p]\right\}$$

Ou seja:

$$f(\tilde{\mathbf{Z}} | \boldsymbol{\phi}, \tau) = \tau^{\frac{N}{2}} |\mathbf{M}_p^{(p)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p]\right\} \exp\left\{-\frac{\tau}{2} [\mathcal{R} + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})]\right\} \quad (6.11)$$

onde,

$$\mathcal{R} = (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}}) \text{ e}$$

$|\mathbf{M}_p^{(p)}|$ denota o determinante da matriz $\mathbf{M}_p^{(p)}$.

A maior dificuldade no uso da verossimilhança exata está nos cálculos para determinar a matriz $\mathbf{M}_p^{(p)}$, Um procedimento iterativo apresentado em [Box, et al-1994] pode ser usado para facilitar os cálculos. Para aplicar esse procedimento a um modelo AR(p) vamos

considerar um modelo AR(p) quando temos $N = p + 1$ observações; Neste caso a verossimilhança exata é dada por:

$$f(\mathbf{Z}|\boldsymbol{\phi}, \tau) \propto \tau^{\frac{p+1}{2}} |\mathbf{M}_p^{(p)}|^{-\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p + (Z_{p+1} - \phi_1 Z_p - \dots - \phi_p Z_1)^2]\right\} \quad (6.12)$$

então temos que:

$$\mathbf{Z}'_{p+1} \mathbf{M}_{p+1}^{(p)} \mathbf{Z}_{p+1} = \mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p + (Z_{p+1} - \phi_1 Z_p - \dots - \phi_p Z_1)^2$$

Onde $\mathbf{M}_{p+1}^{(p)}$ é dada por:

$$\mathbf{M}_{p+1}^{(p)} = \begin{bmatrix} & & & \vdots & 0 \\ & & & \vdots & 0 \\ & \mathbf{M}_p^{(p)} & & \vdots & \\ \dots & \dots & \dots & \vdots & \\ 0 & 0 & \dots & \vdots & 0 \end{bmatrix} + \begin{bmatrix} \phi_p^2 & \phi_p \phi_{p-1} & \dots & \vdots & -\phi_p \\ \phi_p \phi_{p-1} & \phi_{p-1}^2 & \dots & \vdots & -\phi_{p-1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \dots & \dots & \dots & \dots & \dots \\ -\phi_p & -\phi_{p-1} & \dots & \vdots & 1 \end{bmatrix} \quad (6.13)$$

A matriz $\mathbf{M}_{p+1}^{(p)}$ é um caso particular da matriz de autocovariância para N -observações $\mathbf{M}_N^{(p)}$. A matriz $\mathbf{M}_N^{(p)}$ é simétrica em torno de ambas diagonais principais. Se diz que essa matrizes são duplamente simétricas. Assim os elementos de $\mathbf{M}_{p+1}^{(p)}$ e $\mathbf{M}_p^{(p)}$ podem ser deduzidos da consideração de que ambas são duplamente simétricas. Assim, por exemplo, fazendo $p = 1$ e $N = p + 1 = 2$, temos:

$$\mathbf{M}_2^{(1)} = m_{11}^{(1)} Z_1^2 + (Z_2 - \phi_1 Z_1)^2$$

$$\mathbf{M}_2^{(1)} = (Z_1 \ Z_2) \begin{bmatrix} m_{11}^{(1)} + \phi_1^2 & -\phi_1 \\ -\phi_1 & 1 \end{bmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

Usando a propriedade de duplamente simétrica, temos:

$$\begin{bmatrix} m_{11}^{(1)} + \phi_1^2 & -\phi_1 \\ -\phi_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\phi_1 \\ -\phi_1 & m_{11}^{(1)} + \phi_1^2 \end{bmatrix}$$

portanto quando $p = 1$ temos:

$$M_1^{(1)} = m_{11}^{(1)} = 1 - \phi_1^2$$

$$|M_1^{(1)}| = 1 - \phi_1^2 \tag{6.14}$$

Outro exemplo é quando $p = 2$ e $N = p + 1 = 3$, neste caso temos:

$$M_3^{(2)} = \begin{bmatrix} M_2^{(2)} & \vdots & 0 \\ \dots & \vdots & 0 \\ 0 & 0 & \vdots & 0 \end{bmatrix} + \begin{bmatrix} \phi_2^2 & \phi_2\phi_1 & \vdots & -\phi_2 \\ \phi_2\phi_1 & \phi_1^2 & \vdots & -\phi_1 \\ \dots & \dots & \dots & \dots \\ -\phi_2 & -\phi_1 & \vdots & 1 \end{bmatrix}$$

Fazendo:

$$M_2^{(2)} = \begin{bmatrix} m_{11}^{(2)} & m_{12}^{(2)} \\ m_{21}^{(2)} & m_{22}^{(2)} \end{bmatrix}$$

Usando a propriedade de dupla simetria, temos:

$$\begin{bmatrix} m_{11}^{(2)} + \phi_2^2 & m_{12}^{(2)} + \phi_2\phi_1 & -\phi_2 \\ m_{21}^{(2)} + \phi_2\phi_1 & m_{22}^{(2)} + \phi_1^2 & -\phi_1 \\ -\phi_2 & -\phi_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 \\ -\phi_1 & m_{22}^{(2)} + \phi_1^2 & m_{12}^{(2)} + \phi_2\phi_1 \\ -\phi_2 & m_{21}^{(2)} + \phi_2\phi_1 & m_{11}^{(2)} + \phi_2^2 \end{bmatrix}$$

Assim temos:

$$m_{11}^{(2)} = 1 - \phi_2^2$$

$$m_{12}^{(2)} = m_{21}^{(2)} = -\phi_1(1 + \phi_2)$$

$$m_{22}^{(2)} = m_{11}^{(2)} = 1 - \phi_2^2$$

Portanto temos:

$$\mathbf{M}_2^{(2)} = \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{bmatrix} \quad (6.15)$$

$$|\mathbf{M}_2^{(2)}| = (1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2] \quad (6.16)$$

Com esses dois exemplos podemos construir as funções de verossimilhança exata para o AR(p) quando $p = 1$ ou $p = 2$. Como mostraremos a seguir.

6.1.1 Modelos AR(1)

A função de verossimilhança exata do AR(1) pode ser construída fazendo $p = 1$ na equação (6.11).

$$f(\mathbf{Z}|\phi, \tau) \propto \tau^{\frac{N}{2}} |\mathbf{M}_1^{(1)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2}[\mathbf{Z}'_1 \mathbf{M}_1^{(1)} \mathbf{Z}_1]\right\} \\ \exp\left\{-\frac{\tau}{2}[\mathcal{R} + (\phi - \hat{\phi})'(\mathbf{X}'\mathbf{X})(\phi - \hat{\phi})]\right\}$$

Neste caso, temos:

$$|\mathbf{M}_1^{(1)}| = (1 - \phi^2)$$

Então temos:

$$f(\mathbf{Z}|\phi, \tau) \propto \tau^{\frac{N}{2}} (1 - \phi)^2 \exp\left\{-\frac{\tau}{2} Z_1^2 (1 - \phi^2)\right\} \\ \exp\left\{-\frac{\tau}{2} [\mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\} \quad (6.17)$$

onde: $\mathbf{X} = (Z_1, Z_2, \dots, Z_{N-1})'$, $\mathbf{Z} = (Z_2, \dots, Z_n)'$, $\hat{\phi} = (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \tilde{\mathbf{Z}})$ e $\hat{\mathbf{Z}} = \mathbf{X} \hat{\phi}$.

6.1.2 Modelos AR(2)

A função de verossimilhança do processo AR(2) pode ser construída fazendo $p = 2$ em (6.12), assim temos:

$$f(\mathbf{Z}|\phi, \tau) \propto \tau^{\frac{N}{2}} |\mathbf{M}_2^{(2)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}_2' \mathbf{M}_2^{(2)} \mathbf{Z}_2]\right\} \\ \exp\left\{-\frac{\tau}{2} [\mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\}$$

Neste caso temos:

$$\mathbf{M}_2^{(2)} = \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{bmatrix}$$

$$|\mathbf{M}_2^{(2)}| = (1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2]$$

Então:

$$f(\mathbf{Z}|\phi, \tau) \propto \tau^{\frac{N}{2}} (1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2] \\ \exp\{Z_1(1 - \phi_2^2) - 2Z_1Z_2\phi_1(1 + \phi_2) + Z_2^2(1 - \phi_2^2)\} \\ \exp\left\{-\frac{\tau}{2} [\mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\} \quad (6.18)$$

Notamos que a função de verossimilhança exata torna-se mais complexa devido a presença do determinante da matriz $\mathbf{M}_p^{(p)}$, quando p cresce ($p > 2$).

A determinação dos estimadores de máxima verossimilhança para os parâmetros ϕ_i , $i = 1, \dots, p$ só podem ser calculados de forma aproximada e usando-se métodos numéricos, três métodos são apresentados em Box, et al -1994. Um deles sugere que se despreze o termo que envolve o determinante, $|\mathbf{M}_p^{(p)}|$ no logaritmo da função de verossimilhança. Os outros dois propõe aproximação para a derivada desse termo. Um método proposto recentemente em [Muller - 1995] sugere uma aproximação computacionalmente mais eficiente para a derivada desse termo.

No entanto, a experiência mostra que quando o número de observações " N " é grande em relação a ordem do modelo " p " o uso da verossimilhança aproximada (ver Cap.4) mostra-se bastante adequado apresentando valores muito próximos dos estimadores obtidos através da verossimilhança exata.

Neste trabalho, ilustraremos essa propriedade com dois exemplos apresentados no Capítulo 7 de resultados.

6.2 Abordagem Bayesiana e uso do MCMC

Cosiderando agora os parâmetros $\phi = (\phi_1, \dots, \phi_p)'$ e τ como *v.a.* independentes, a densidade a priori não informativa de Jeffreys [Box & Tiao - 1973] para esse parâmetros é:

$$\Pi_0(\phi, \tau) \propto 1/\tau \quad , \tau > 0 \quad (6.19)$$

Combinando a verossimilhança dada em (6.11) com (6.19) temos a densidade conjunta a posteriori dada por:

$$\begin{aligned} \Pi(\phi, \tau) \propto \tau^{\frac{N}{2}-1} |\mathbf{M}_p^{(p)}|^{\frac{1}{2}} \exp\left\{ -\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p] \right\} \\ \exp\left\{ -\frac{\tau}{2} [\mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})] \right\} \end{aligned} \quad (6.20)$$

As densidades condicionais para ϕ e τ sendo que $\Pi(\phi|\tau)$ é uma densidade p -variada. São dadas por:

$$\begin{aligned} \Pi(\phi|\tau, \mathbf{Z}) &\propto \tau^{\frac{p}{2}} |\mathbf{M}_p^{(p)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p]\right\} \\ &\exp\left\{-\frac{\tau}{2} [(\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\} \end{aligned} \quad (6.21)$$

ou seja $\Pi(\phi|\tau)$ é proporcional a um núcleo normal p -dimensional de forma que podemos escrever:

$$\Pi(\phi|\tau, \mathbf{Z}) \propto \Psi(\phi, \tau) N[\hat{\phi}, (\tau \mathbf{X}' \mathbf{X})^{-1}]$$

onde:

$$\Psi(\phi, \tau) = |\mathbf{M}_p^{(p)}|^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p]\right\}$$

A densidade condicional de τ é dada por:

$$\Pi(\tau|\phi, \mathbf{Z}) \propto \tau^{\frac{N}{2}-1} \exp\left\{-\frac{\tau}{2} [\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p + \mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\} \quad (6.22)$$

Portanto temos que:

$$\Pi(\tau|\phi, \mathbf{Z}) \propto \text{Gama}\left(\frac{N}{2}, \frac{\mathcal{B}}{2}\right)$$

$$\text{onde: } \mathcal{B} = \mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p + \mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})$$

Aplicando essa abordagem para os modelos AR(1) e AR(2), temos:

Para os modelos AR(1):

$$\begin{aligned} \Pi(\phi, \tau|\mathbf{Z}) &\propto \tau^{\frac{N}{2}-1} (1 - \phi^2)^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} [Z_1^2 (1 - \phi^2)]\right\} \\ &\exp\left\{-\frac{\tau}{2} [\mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\} \end{aligned} \quad (6.23)$$

As condicionais são:

$$\Pi(\phi|\tau, \mathbf{Z}) \propto \Psi(\phi, \tau) N[\hat{\phi}, (\tau \mathbf{X}' \mathbf{X})^{-1}]$$

$$\Psi(\phi, \tau) = \sqrt{(1 - \phi^2)} \exp\left\{-\frac{\tau}{2} [Z_1^2 (1 - \phi^2)]\right\}$$

$$\Pi(\tau|\phi, \mathbf{Z}) \propto \text{Gama}\left(\frac{N}{2}, \frac{\mathcal{B}}{2}\right)$$

$$\mathcal{B} = Z_1^2 (1 - \phi^2) + \mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})$$

Temos ainda que:

$$\mathcal{R} = (\mathbf{Z} - \hat{\mathbf{Z}})' (\mathbf{Z} - \hat{\mathbf{Z}});$$

$$\hat{\phi} = (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Z}) \text{ e}$$

$$\hat{\mathbf{Z}} = \mathbf{X} \hat{\phi}$$

sendo:

$$\mathbf{X} = (Z_1, Z_2, \dots, Z_{N-1})' \text{ e } \mathbf{Z} = (Z_2, \dots, Z_N)'$$

Para os modelos AR(2):

$$\begin{aligned} \Pi(\phi, \tau|\mathbf{Z}) &\propto \tau^{\frac{N}{2}-1} (1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2] \\ &\exp\left\{-\frac{\tau}{2} [Z_1 (1 - \phi_2^2) - 2Z_1 Z_2 \phi_1 (1 + \phi_2) + Z_2^2 (1 - \phi_2^2)]\right\} \\ &\exp\left\{-\frac{\tau}{2} [\mathcal{R} + (\phi - \hat{\phi})' (\mathbf{X}' \mathbf{X}) (\phi - \hat{\phi})]\right\} \end{aligned} \quad (6.24)$$

As densidades condicionais são:

$$\Pi(\phi|\tau, \mathbf{Z}) \propto \Psi(\phi, \tau) N[\hat{\phi}, (\tau \mathbf{X}' \mathbf{X})^{-1}]$$

$$\Psi(\phi, \tau) = (1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2] \\ \exp\left\{-\frac{\tau}{2}[Z_1(1 - \phi_2^2) - 2Z_1Z_2\phi_1(1 + \phi_2) + Z_2^2(1 - \phi_2^2)]\right\}$$

$$\Pi(\tau|\phi, \mathbf{Z}) \propto \text{Gama}\left(\frac{N}{2}, \frac{\mathcal{B}}{2}\right)$$

$$\mathcal{B} = Z_1(1 - \phi_2^2) - 2Z_1Z_2\phi_1(1 + \phi_2) + Z_2^2(1 - \phi_2^2) \\ + \mathcal{R} + (\phi - \hat{\phi})'(X'X)(\phi - \hat{\phi})$$

No caso do AR(2) temos:

$$\mathcal{R} = (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}});$$

$$\hat{\phi} = (X'X)^{-1}(X'Z) \quad \text{e} \quad \hat{\mathbf{Z}} = X\hat{\phi}$$

com: $X = (Z_2, \dots, Z_{N-1})'$ e $Z = (Z_3, \dots, Z_N)'$.

6.3 Uso do Algoritmo MCMC

Com as condicionais:

$$\Pi(\phi|\tau, \mathbf{Z}) \propto \Psi(\phi, \tau) N[\hat{\phi}, (\tau X'X)^{-1}]$$

onde: $\Psi(\phi, \tau) = |\mathbf{M}_p^{(p)}|^{-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}[\mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p]\right\}$

$$\Pi(\tau|\phi, \mathbf{Z}) \propto \text{Gama}\left(\frac{N}{2}, \frac{\mathcal{B}}{2}\right)$$

$$\mathcal{B} = \mathbf{Z}'_p \mathbf{M}_p^{(p)} \mathbf{Z}_p + \mathcal{R} + (\phi - \hat{\phi})'(X'X)(\phi - \hat{\phi})$$

$$\mathcal{R} = (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})$$

Usamos para gerar os valores dos parâmetros como segue:

passo 1: Condições iniciais

$$\boldsymbol{\phi}^{(0)} = (\phi_1^{(0)}, \dots, \phi_p^{(0)}) \text{ e } \tau^{(0)}$$

passo 2: Gerar: $\tau^{(j+1)} \sim \text{Gama}(\frac{N}{2}, \frac{B}{2})$,

$$\boldsymbol{y}^{(j+1)} \sim N[\hat{\boldsymbol{\phi}}, (\tau \boldsymbol{X}' \boldsymbol{X})^{-1}]$$

passo 3: Gerar u de Uniforme[0, 1] e calcular:

$$\alpha(\boldsymbol{y}^{(j+1)}, \boldsymbol{\phi}^{(j)}) = \min\left\{1, \frac{\psi(\boldsymbol{y}^{(j+1)}, \tau^{(j+1)})}{\psi(\boldsymbol{\phi}^{(j)}, \tau^{(j+1)})}\right\}$$

$$\text{se } u \leq \alpha(\boldsymbol{y}^{(j+1)}, \boldsymbol{\phi}^{(j)})$$

$$\boldsymbol{\phi}^{(j+1)} = \boldsymbol{y}^{(j+1)}$$

caso contrário:

$$\boldsymbol{\phi}^{(j+1)} = \boldsymbol{\phi}^{(j)}$$

passo 4: Repetir os passos 2 e 3 até a convergência.

Capítulo 7

Aplicação

7.1 Estudo de Casos

A comparação entre as abordagens clássica e Bayesiana é feita neste trabalho considerando-se duas séries de dados gerados e duas séries reais correspondente a componente estocástica dos dados correspondente as vazões médias mensais que chegam ao reservatório de Furnas no sudeste e Itumbiara no centro oeste do Brasil. Uma descrição dos quatro componentes de dados é apresentado na Tabela 1.

Tabela 1- Dados usados para ajuste dos modelos

	Descrição	ϕ_1	ϕ_2	τ	N
série 1	AR(2)	0.8	- 0.5	0.5	200
série 2	AR(2)	- 0.5	0.3	1.0	200
série 3 (Furnas)	dados reais	-	-	-	552
série 4 (Itumbiara)	dados reais	-	-	-	672

No caso das séries reais vamos assumir que a série de vazões é descrita pelo modelo $X_{t(r,m)} = \mu_m + \sigma_m Z_{t(r,m)}$ onde m representa o número de meses $m = 1, 2, \dots, 12$, r o número de anos $r = 1, 2, \dots$ e $t(r, m) = (r - 1)12 + m$, μ_m e σ_m são as médias e desvio-padrão estimado dos dados para cada mês do ano, a componente estocástica $Z_{t(r,m)}$ é calculada como:

$$Z_{t(r,m)} = \frac{X_{t(r,m)} - \mu_m}{\sigma_m}$$

Para as componentes $Z_{t(r,m)}$ ajustamos um modelo $AR(p)$:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$$

As séries de vazões tem 672 e 552 observações correspondendo a vazões médias mensais medidas de Janeiro de 1935 a Dezembro de 1990 e de Janeiro de 1945 a Dezembro de 1990 respectivamente. Um gráfico com as séries 1, 2 e com os últimos 200 pontos das séries 3(Furnas) e 4(Itumbiara) é mostrado nas figuras 1a, 1b, 1c e 1d.

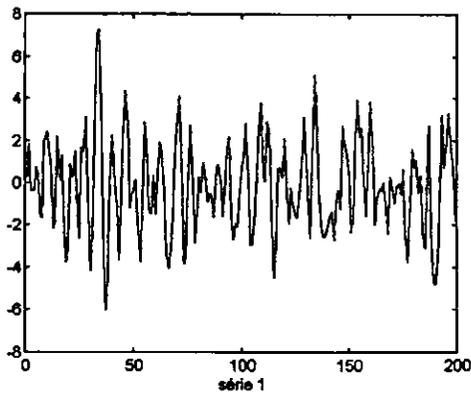


Figura 1a: série 1

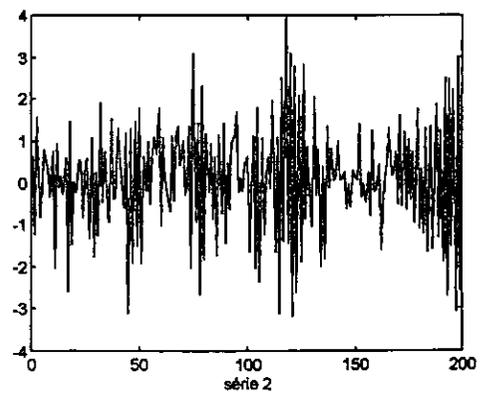


Figura 1b: série 2

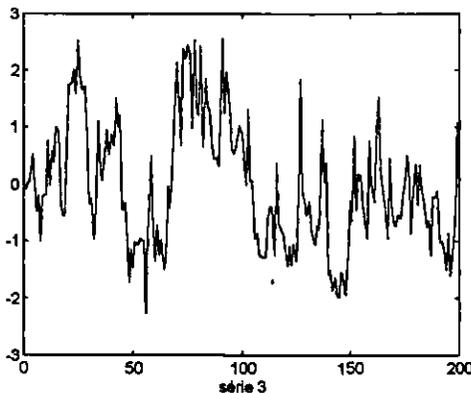


Figura 1c: série 3(Furnas)

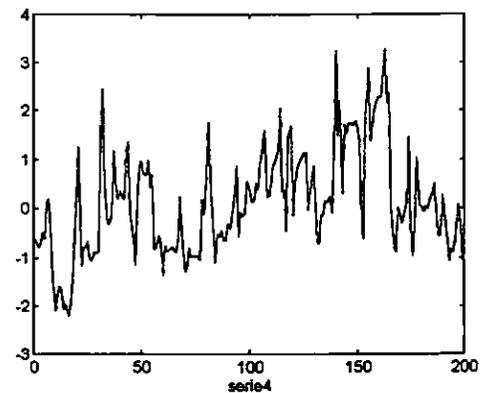


Figura 1d: série 4(Itumbiara).

A Figura 2 a seguir mostra a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP), para os dados da série 3 (Furnas).

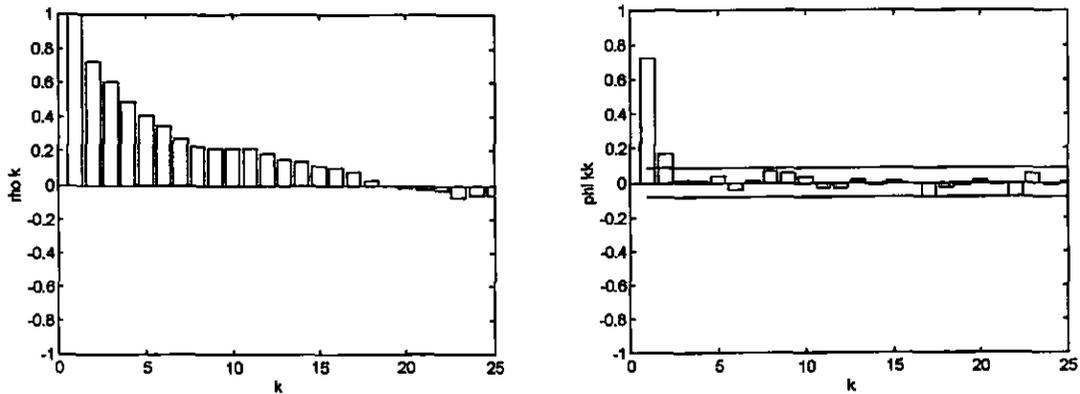


Figura 2: FAC e FACP, para os dados da série 3.

A Figura 3 a seguir mostra a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP), para os dados da série 4 (Itumbiara).

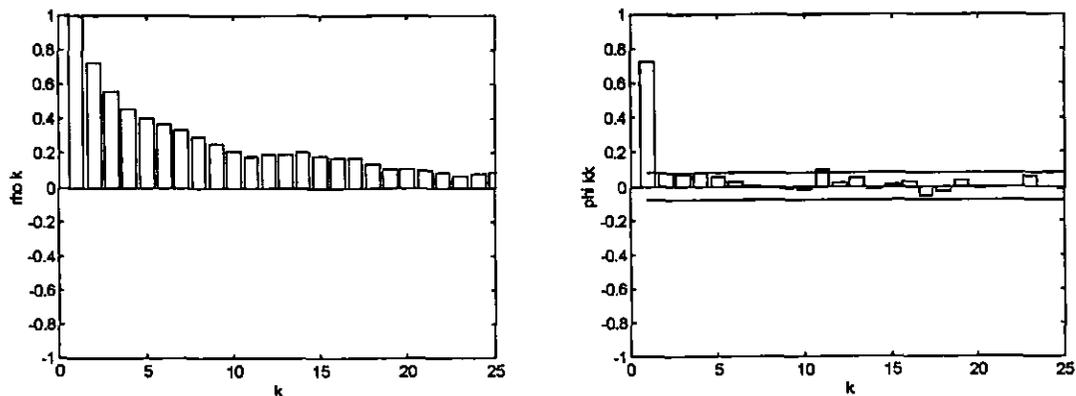


Figura 3: FAC e FACP, para os dados da série 4 (Itumbiara)

Através da análise da FAC e FACP podemos escolher um modelo AR(2) para a série 3 (Furnas) e AR(1) para a série 4 (Itumbiara) e usando-se o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC), (ver apêndice 1), esses modelos mostram-se como os mais parcimoniosos (menor número de parâmetros possíveis com menor variância dos resíduos). Neste capítulo vamos comparar esses critérios e os critérios de seleção de modelos que utilizam a densidade preditiva ordenada, para os modelos ajustados, os resultados são apresentados nas Tabelas 2 e 3.

Tabela 2 identificação do modelo

série 1		AR(1)	AR(2)	AR(3)
	AIC		+ 1.77800	+ 0.89652
BIC		+ 1.30967	+ 0.94402	+ 0.96051
série 2		AR(1)	AR(2)	AR(3)
	AIC		- 0.10979	- 0.17145
BIC		- 0.07812	- 0.12395	- 0.10440
série 3 (Furnas)		AR(1)	AR(2)	AR(3)
	AIC		- 0.76617	- 0.81162
BIC		- 0.73862	- 0.78777	- 0.77656
série 4 (Itumbiara)		AR(1)	AR(2)	AR(3)
	AIC		- 0.65157	- 0.64859
BIC		- 0.63814	- 0.62846	- 0.61877

Tabela 3 seleção de modelos

série 1		AR(1)	AR(2)	AR(3)
	\bar{c}_k		0.3726	0.4379
$c(l)$		1.8114e-024	1.7825e-022	6.8961e-023
série 2		AR(1)	AR(2)	AR(3)
	\bar{c}_k		0.7450	0.7824
$c(l)$		2.0413e-008	1.1434e-007	1.0365e-007
série 3 (Furnas)		AR(1)	AR(2)	AR(3)
	\bar{c}_k		0.9887	1.0107
$c(l)$		4.7841e-005	5.6579e-005	5.4942e-005
série 4 (Itumbiara)		AR(1)	AR(2)	AR(3)
	\bar{c}_k		0.9777	0.9812
$c(l)$		4.4694e-006	2.1485e-006	3.8376e-006

Usando o procedimento Bayesiano para verificação da adequabilidade do modelo através da densidade preditiva $\Pi(Z_{N+k}|Z_{N+k-1})$, os valores médios de \bar{c}_k e os produtos $c(l)$ para as séries estudadas neste trabalho, indicam a escolha de modelos AR(2) para as séries 1, 2 e 3 e para a série 4 o modelo recomendado é o AR(1). Portanto os resultados obtidos pelo critério de adequabilidade do modelo baseado na preditiva ordenada coincidem com o AIC e BIC.

As figuras 3a, 3b, 3c e 3d mostra o gráfico de c_k versus k ($k = 1, 2, \dots, n$) para diferentes modelos, das séries Furnas, Itumbiara, série 1 e série 2 respectivamente.

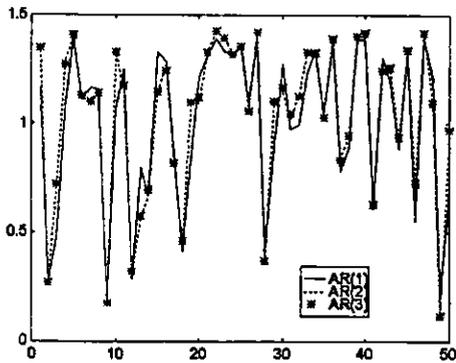


Figura 3a: $c_k \times k$ para dif. modelos, Furnas

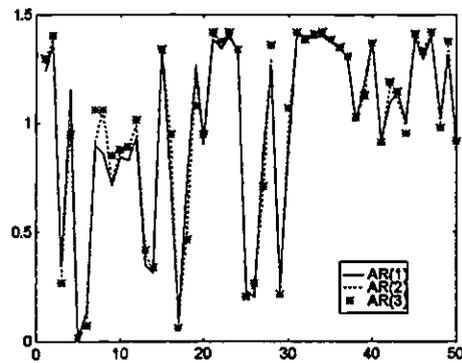


Figura 3b: $c_k \times k$ para dif. modelos, Itumbiara

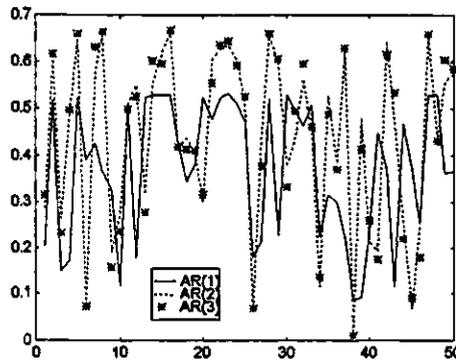


Figura 3c: $c_k \times k$ para dif. modelos, série 1

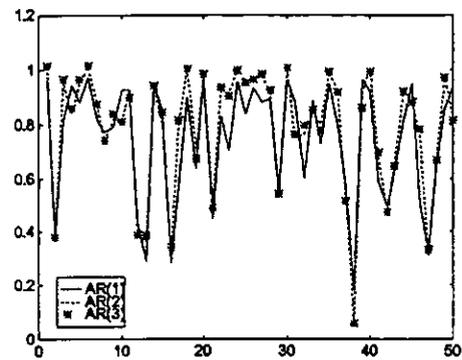


Figura 3d: $c_k \times k$ para dif. modelos, série 2

Nas tabelas 2 e 3 observamos que o modelo AR(2) é o mais adequado, para as séries 1, 2 e 3, e para a série 4 o modelo mais adequado é um AR(1).

7.2 Inferência Bayesiana dos parâmetros dos Modelos

Apresentamos nesta seção os histogramas e distribuição acumuladas para os parâmetros dos modelos ajustados às séries 3(Furnas) e 4(Itumbiara). Os resultados obtidos para as séries geradas (série 1 e 2) são semelhantes, por isso são omitidos aqui.

Na Figura 4 apresentamos os histogramas estimado por simulação para os parâmetros ϕ_1 , ϕ_2 e τ , quando consideramos priori Normal-Gama, para a série 3(Furnas) estudada neste trabalho.

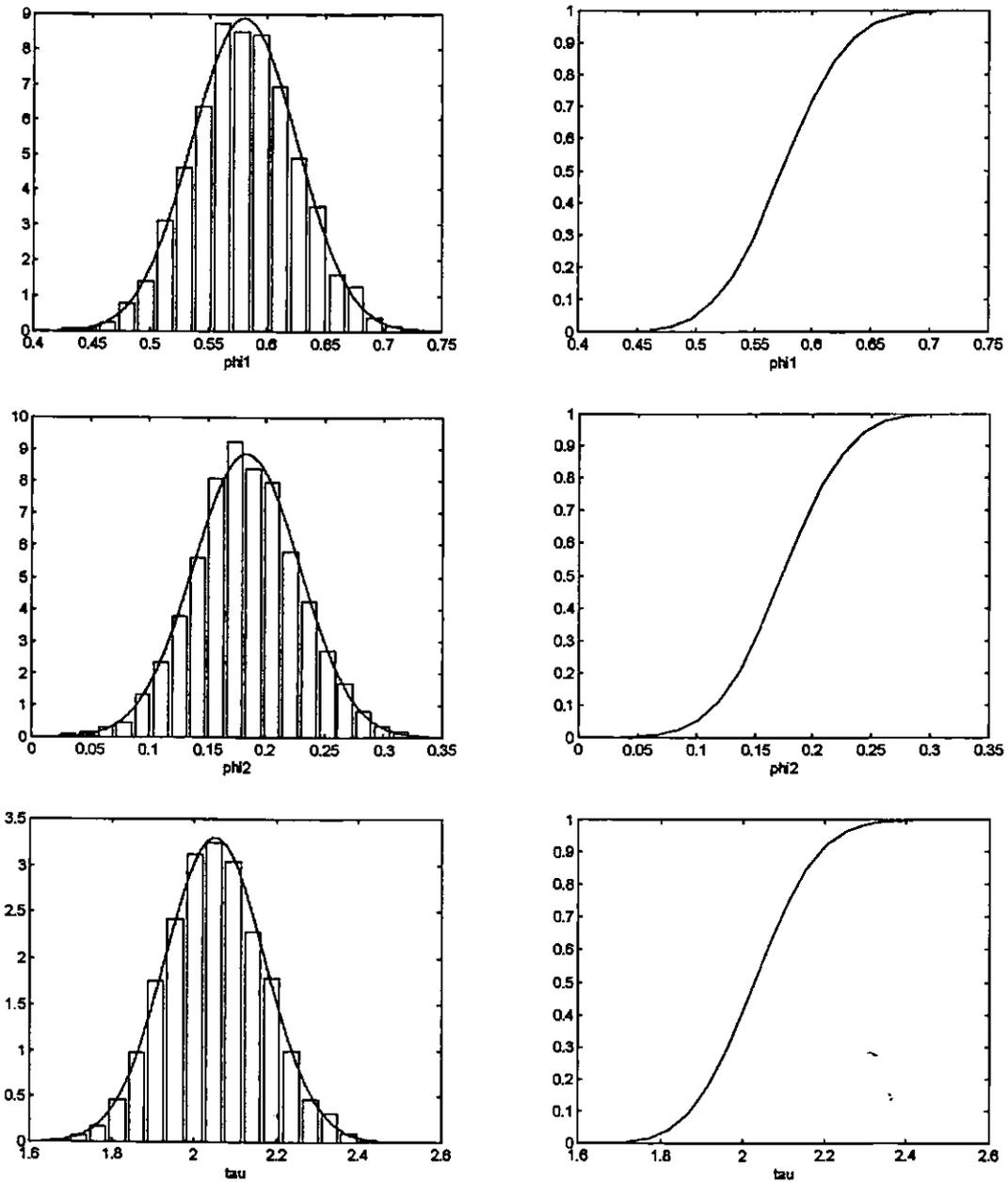


Figura 4: histograma estimado a posteriori para a série 3 (Furnas) considerando priori Normal-Gama para os parâmetros ϕ_1 , ϕ_2 e τ respectivamente.

Observamos na Figura 4 que a forma simétrica da posteriori é a forma esperada teoricamente como foi deduzido na seção 4.2.2 onde vimos que a distribuição conjunta para (ϕ_1, ϕ_2, τ) a posteriori é uma normal-gama.

Na Figura 5 apresentamos os histogramas estimado por simulação para os parâmetros ϕ_1 , ϕ_2 e τ , quando consideramos priori t -Student, para a série 3(Furnas).

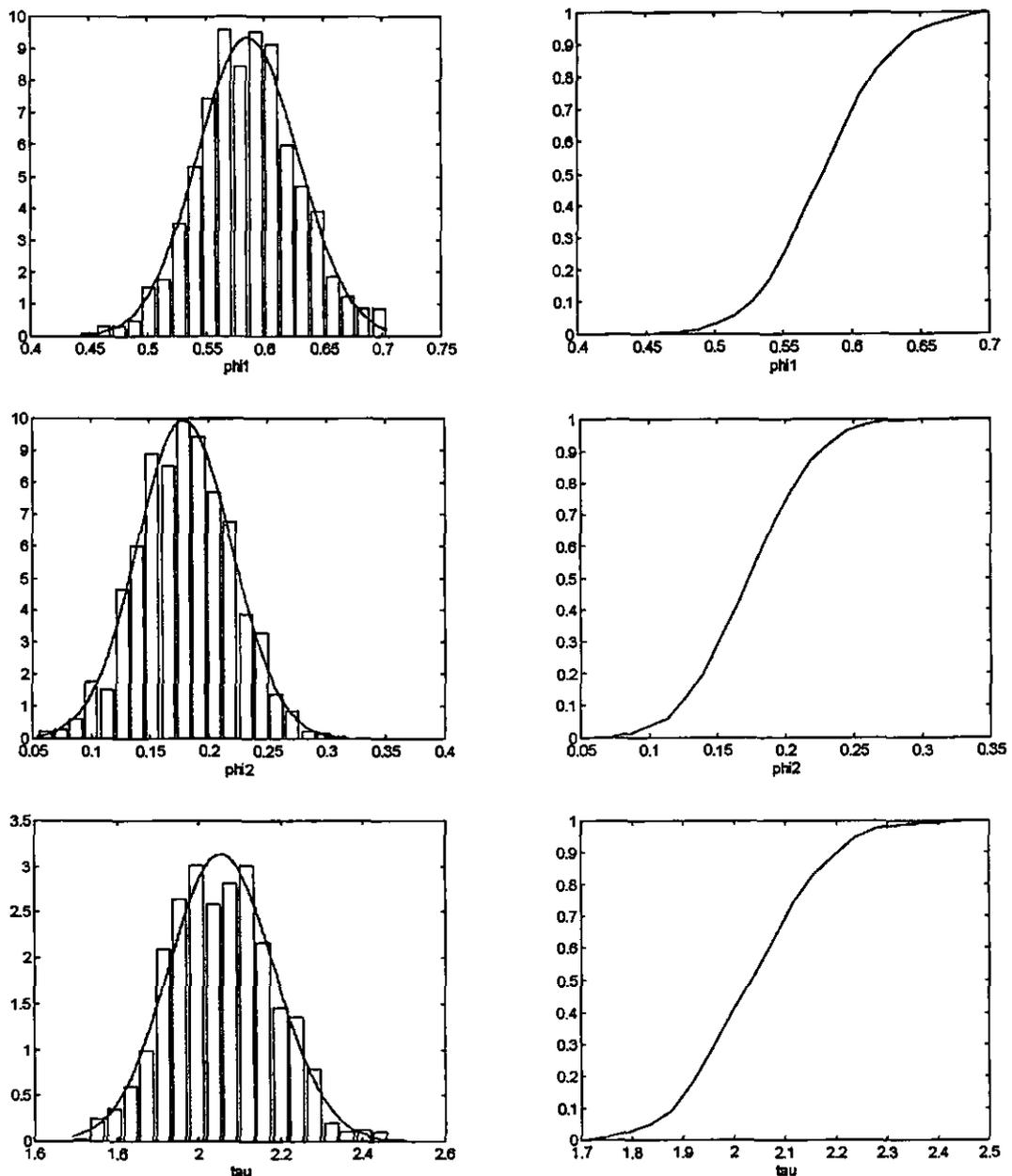


Figura 5: histograma estimado a posteriori para a série 3 (Furnas) considerando priori t -Student para os parâmetros ϕ_1 , ϕ_2 e τ respectivamente.

Notamos na Figura 5 que as distribuições a posteriori são simétricas, esses resultados só podem ser obtidos por simulação pois não é possível deduzir explicitamente a expressão dessas distribuições.

Na Figura 6 apresentamos os histogramas estimado por simulação para os parâmetros ϕ_1 e τ , quando consideramos priori Normal-Gama, para a série 4(Itumbiara).

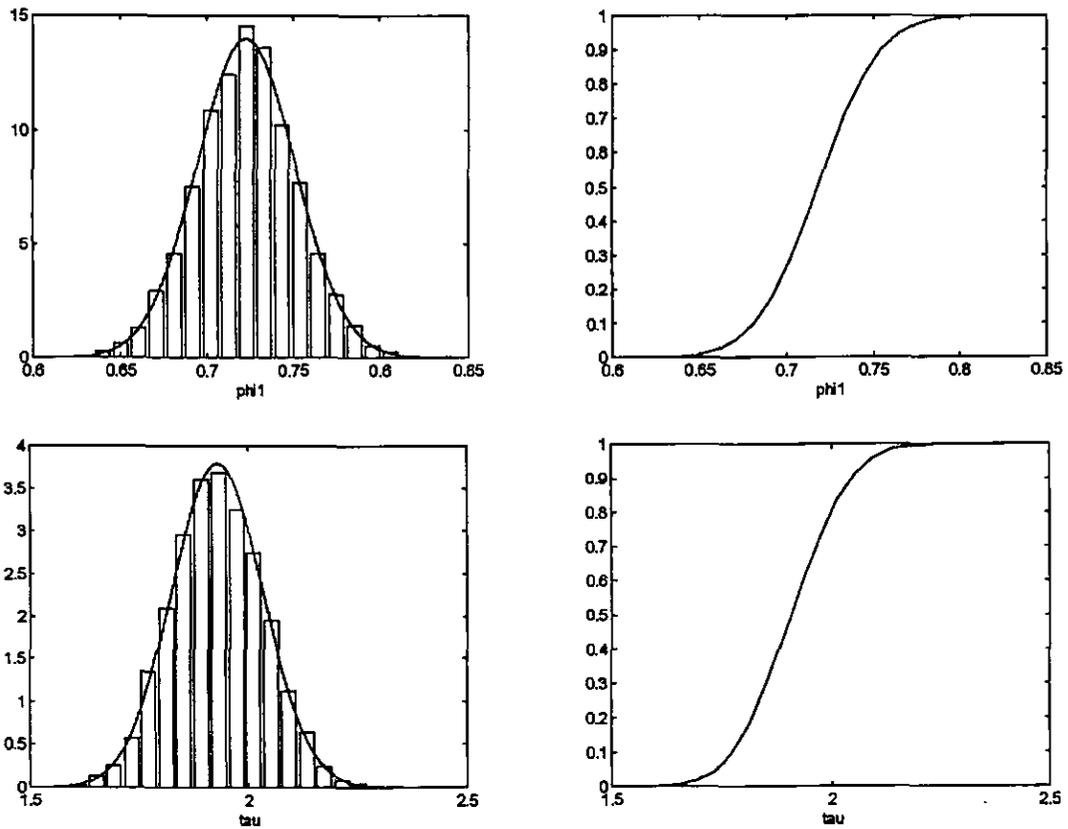


Figura 6: histograma estimado a posteriori para a série 4 (Itumbiara) considerando priori Normal-Gama para os parâmetros ϕ_1 e τ respectivamente.

Na Figura 7 apresentamos os histogramas estimado por simulação para os parâmetros ϕ_1 e τ , quando consideramos priori t -Student, para a série 4(Itumbiara).

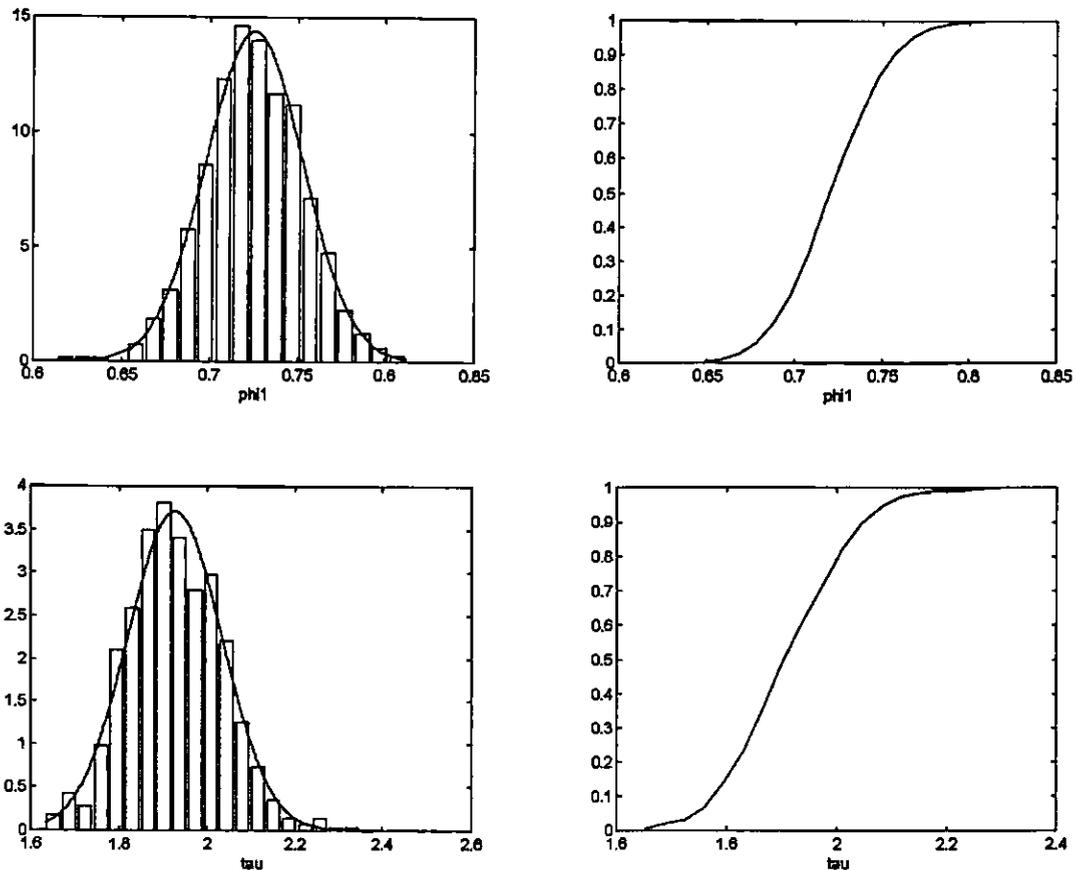


Figura 7: histograma estimado a posteriori para a série 4(Itumbiara) considerando priori t -Student para os parâmetros ϕ_1 e τ respectivamente.

Observamos nas Figuras 6 e 7, que o comportamento das distribuições a posteriori é aquele esperado na seção 4.2.2, e que com o uso da priori t -Student, apesar de não conhecermos a distribuição a posteriori, os histogramas também se apresentam de uma forma simétrica.

No caso particular de priori t -Student, onde utilizamos o algoritmo metropolis-hastings foram geradas duas amostras de tamanho $N = 6.000$, desprezamos 3.000 e selecionamos para estimação uma amostra de tamanho 200 tomando-se um valor a cada 15 gerado na última sequência de 3.000 valores. Esse procedimento mostrou-se necessário para evitar a correlação entre os valores gerados devido a rejeição de novos valores.

A Figura 8 a seguir mostra a correlação entre os valores gerados quando não se adota a seleção espaçada, para a série Furnas.

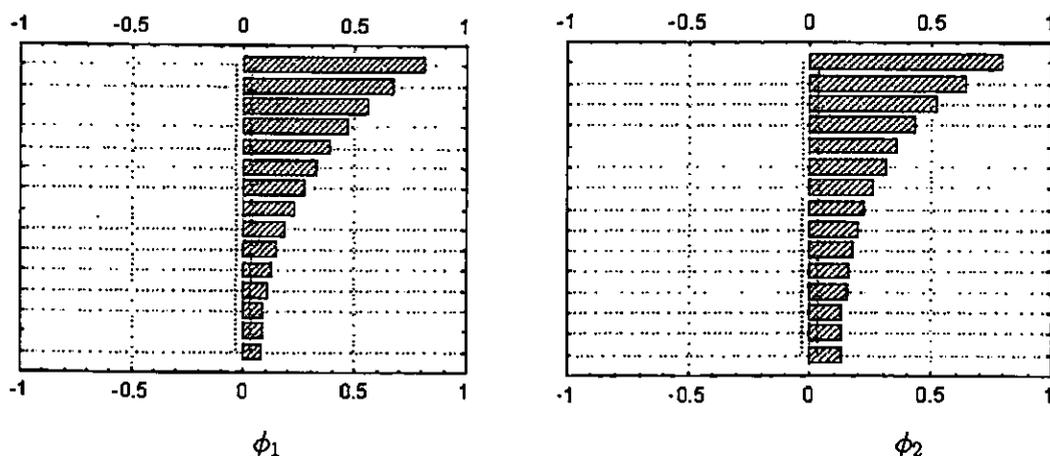


Figura 8 função de autocorrelação entre os valores gerados para ϕ_1 e ϕ_2 ; $N=3.000$.

A Figura 9 mostra a correlação entre os valores gerados quando se adota a seleção espaçada, procedendo assim diminui-se a correlação entre os dados gerados.

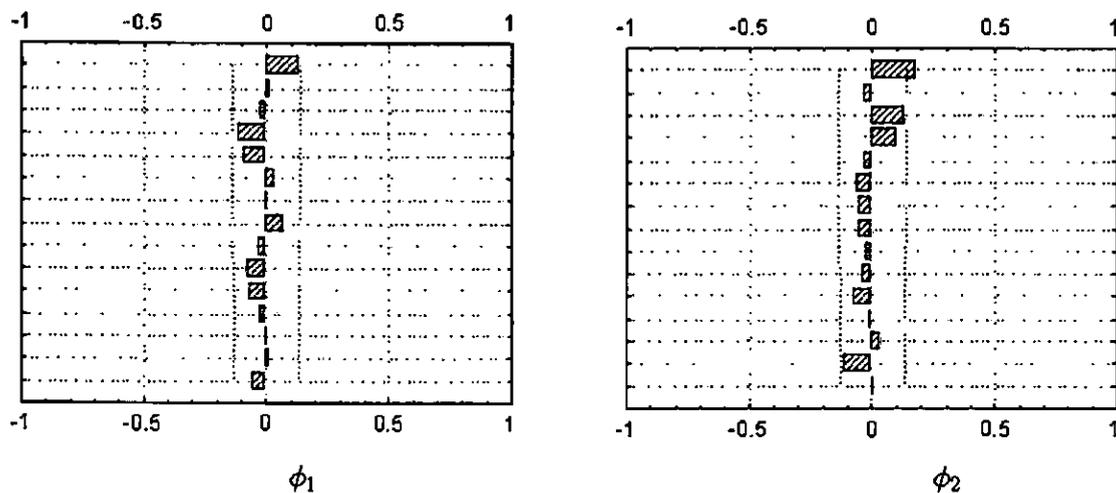


Figura 9- função de autocorrelação entre os valores gerados para ϕ_1 e ϕ_2 ; $N=200$.

As Figuras 10a e 10b a seguir mostram a convergência, para os dados da série 3(Furnas) e série 4(Itumbiara) respectivamente, partindo-se de uma condição inicial arbitrária, mostrando de forma gráfica a convergência do algoritmo MCMC.

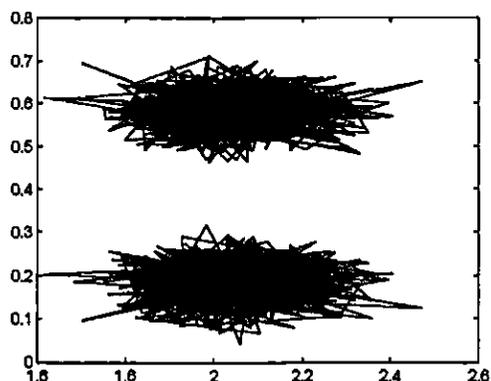


Figura 10a: gráfico da convergência para ϕ_1 , ϕ_2 e τ com os dados da série 3(Furnas).

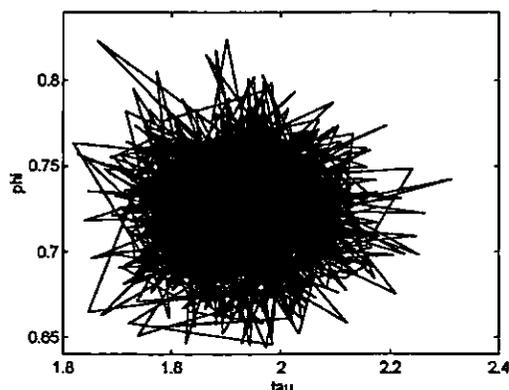


Figura 10b: gráfico da convergência para ϕ_1 e τ com os dados da série 4(Itumbiara).

As Tabelas 4, 5, 6 e 7 apresentam os estimadores dos parâmetros, para cada uma das quatro séries analisadas neste trabalho. EMV corresponde ao estimador de máxima verossimilhança; PNI(·) corresponde ao estimador quando adotamos priori não-informativa, EX significa que esse é o estimador exato obtido pelas expressões do valor esperado da densidade a posteriori e GS significa o estimador obtido por simulação usando-se o algoritmo Gibbs Sampling; PNG(·) são os estimadores obtidos usando priori conjugada Normal-Gama; PTS são os estimadores obtidos usando-se priori t -Student com o algoritmo Metropolis-Hastings. Nestas tabelas apresentamos também a variância do estimador e o critério de convergência proposto por Gelman & Rubin (1992) ver apêndice 2, quando os estimadores são obtidos por simulação. Este critério assegura a convergência quando $R < 1.1$.

Tabela 4 Valores estimados dos parâmetros ϕ_1, ϕ_2 e τ para a série 1

	ϕ_1	SD	R	ϕ_2	SD	R	τ	SD	R
EMV	0.8156	0.0580	—	-0.5773	0.0581	—	0.4119	—	—
PNI(EX)	0.8156	0.0586	—	-0.5773	0.0587	—	0.4077	0.0412	—
PNI(GS)	0.8158	0.0581	1.0006	-0.5773	0.0578	1.0001	0.4085	0.0411	1.0001
PNG(EX)	0.8141	0.0578	—	-0.5759	0.0579	—	0.4178	0.0400	—
PNG(GS)	0.8130	0.0588	1.0002	-0.5756	0.0580	1.0003	0.4186	0.0398	1.0001
PTS(MH)	0.8181	0.0550	1.0003	-0.5806	0.0557	1.0078	0.4154	0.0385	1.0035

Tabela 5 Valores estimados dos parâmetros ϕ_1, ϕ_2 e τ para a série 2

	ϕ_1	SD	R	ϕ_2	SD	R	τ	SD	R
EMV	-0.5158	0.0679	—	0.2946	0.0684	—	1.2550	—	—
PNI(EX)	-0.5158	0.0686	—	0.2946	0.0691	—	1.2423	0.1255	—
PNI(GS)	-0.5163	0.0690	1.0000	0.2937	0.0685	1.0001	1.2443	0.1257	1.0001
PNG(EX)	-0.5141	0.0725	—	0.2951	0.0731	—	1.1003	0.1054	—
PNG(GS)	-0.5149	0.0737	1.0005	0.2947	0.0741	1.0001	1.1025	0.1058	1.0001
PTS(MH)	-0.5127	0.0620	1.0010	0.2965	0.0633	1.0053	1.0961	0.0987	1.0024

Tabela 6 Valores estimados dos parâmetros ϕ_1, ϕ_2 e τ para a série 3(Furnas)

	ϕ_1	SD	R	ϕ_2	SD	R	τ	SD	R
EMV	0.5834	0.0420	—	0.1815	0.0421	—	2.3244	—	—
PNI(EX)	0.5834	0.0421	—	0.1815	0.0421	—	2.3244	0.1417	—
PNI(GS)	0.5835	0.0421	1.0002	0.1809	0.0419	1.0000	2.3161	0.1409	1.0004
PNG(EX)	0.5815	0.0447	—	0.1825	0.0447	—	2.0528	0.1229	—
PNG(GS)	0.5821	0.0447	1.0000	0.1820	0.0446	1.0000	2.0531	0.1229	1.0002
PTS(MH)	0.5811	0.0406	1.0032	0.1824	0.0413	1.0003	2.0505	0.1223	1.0051

Tabela 7 Valores estimados dos parâmetros ϕ_1 e τ para a série 4(Itumbiara)

	ϕ_1	SD	R	τ	SD	R
EMV	0.7237	0.0269	—	2.1146	—	—
PNI(EX)	0.7238	0.0270	—	2.1114	0.1165	—
PNI(GS)	0.7238	0.0270	1.0002	2.1114	0.1152	1.0000
PNG(EX)	0.7237	0.0269	—	2.1114	0.1165	—
PNG(GS)	0.7225	0.0283	1.0008	1.9291	0.1049	1.0002
PTS(MH)	0.7253	0.0278	1.0004	1.9267	0.1075	1.0070

Observando os resultados obtidos concluímos que o uso dos algoritmos Gibbs Sampling e Metropolis-Hastings tem um bom desempenho na estimativa dos parâmetros.

No caso particular de priori não informativa ou priori conjugada Normal-Gama as posteriori dos parâmetros $\phi_i, i = 1, 2, \dots, p$ é uma t -Student então podemos usar a variância estimada $Var(\hat{\phi}_i)$ para construir intervalos de confiança $\phi_i = \hat{\phi}_i \pm t_{\alpha/2} \sqrt{Var(\hat{\phi}_i)}$, ou testar hipóteses $H_0 : \phi_i = 0, i = 1, 2$. Porém no caso de priori t -Student a posteriori dos parâmetros

não tem uma distribuição conhecida, portanto o intervalo de credibilidade é calculado empiricamente dos dados simulados, e os resultados obtidos são apresentados na Tabela 7.

Tabela 7 Intervalo de Credibilidade com os percentis 2,5% e 97,5%, para ϕ e τ

	ϕ_1	IC		ϕ_2	IC		τ	IC	
série 1	0.8181	0.6906;	0.9026	-0.5783	-0.6834;	-0.4617	0.4147	0.3583;	0.4967
série 2	-0.5127	-0.6390;	-0.3959	0.2965	0.1681;	0.4226	1.0961	0.9017;	1.2823
série 3	0.5811	0.4966;	0.6672	0.1824	0.0922;	0.2629	2.0505	1.8187;	2.2747
série 4	0.7253	0.6703;	0.7780	—	—		1.9267	1.7368;	2.1300

7.3 Previsões

As Tabelas 8a, 8b, 8c e 8d apresentam as previsões 5 passos a frente, através das abordagens clássica e Bayesiana com o uso da priori Normal-Gama, para cada uma das quatro séries analisadas neste trabalho.

Tabela 8a - previsão 5 passos a frente para série 3 (Furnas)

val. obs.	prev.clássica	prev.bayesiana	int.credibilidade	
-0.9881	-0.4809	-0.4809	-0.5077	-0.4021
-0.7129	-0.3643	-0.3631	-0.3872	-0.2916
-0.3989	-0.2998	-0.3004	-0.3287	-0.2190
-0.1546	-0.2410	-0.2420	-0.2695	-0.1625
0.1625	-0.1950	-0.1970	-0.2236	-0.1209
EMQ	0.1048	0.1052	---	

Tabela 8b - previsão 5 passos a frente para série 4 (Itumbiara)

val. obs.	prev.clássica	prev.bayesiana	int.credibilidade	
-0.5406	-0.1547	-0.1545	-0.1586	-0.1425
0.2719	-0.1119	-0.1118	-0.1177	-0.0950
-0.3115	-0.1929	-0.1929	-0.2050	-0.1584
-0.6772	-0.2206	-0.2211	-0.2395	-0.1689
-1.0663	-0.2993	-0.3006	-0.3299	-0.2182
EMQ	0.2214	0.2210	---	

Tabela 8c - previsão 5 passos a frente para série 1

val. obs.	prev. clássica	prev. bayesiana	int. credibilidade	
-0.9958	-0.3539	-0.3530	-0.4049	-0.1985
0.1286	-0.6941	-0.6917	-0.7641	-0.4844
-1.0206	-0.3618	-0.3579	-0.4001	-0.2382
1.0503	0.1056	0.1098	0.0526	0.2892
0.4561	0.2950	0.2966	0.2276	0.5167
EMQ	0.4883	0.4871	---	

Tabela 8d - previsão 5 passos a frente para série 2

val. obs.	prev. clássica	prev. bayesiana	int. credibilidade	
-2.2663	-2.4515	-2.4487	-2.5581	-2.1280
2.3289	2.1505	2.1509	1.9932	2.6233
-1.5616	-1.8317	-1.8372	-2.0252	-1.2832
0.8875	1.5785	1.5939	1.3621	2.3098
0.4439	-1.3540	-1.3780	-1.5928	-0.7598
EMQ	0.7698	0.7918	---	

Nas tabelas acima, observamos um bom desempenho do método Bayesiano, para previsões k -passos a frente usando-se algoritmos MCMC.

As Figuras 11a, 11b, 11c e 11d a seguir mostra a previsão, para as séries: Furnas, Itumbiara, série1 e série2 respectivamente.

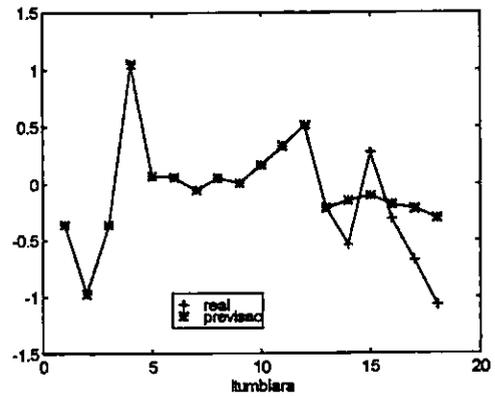
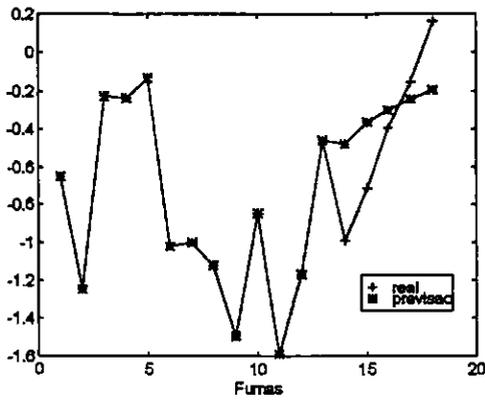


figura 11a: previsão 5 passos a frente(Furnas). figura 11b: previsão 5 passos a frente(Itumbiara).

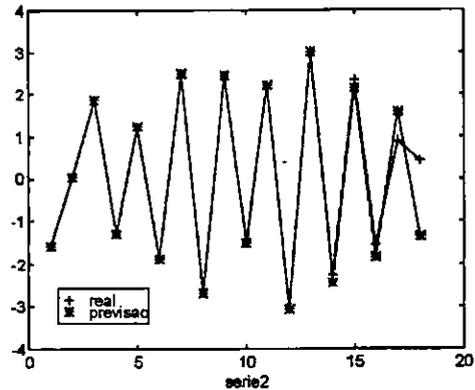
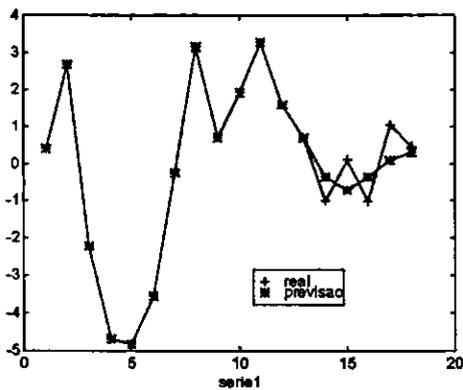


figura 11c: previsão 5 passos a frente(série 1). figura 11d: previsão 5 passos a frente(série 2).

Nas figuras 11a, 11b, 11c e 11d, temos o plot da série e os valores previstos, onde observamos que a previsão acompanha a tendência da série.

7.4 Comparação do uso da Verossimilhança Exata e Aproximada

Um resumo das estimativas dos parâmetros quando usamos Verossimilhança Exata e Verossimilhança Aproximada, é apresentado na Tabela 9.

Tabela 9: estimativas dos parâmetros com verossimilhança exata e aproximada.

série 1		ϕ_1	IC	ϕ_2	IC	τ	IC
	V.Exata	0.8166	0.6888; 0.9309	- 0.5746	- 0.6875; - 0.4494	0.4243	0.3496; 0.5148
V.Aprox.	0.8130	0.7000; 0.9252	- 0.5756	- 0.6893; - 0.4617	0.4186	0.3449; 0.4959	
série 2		ϕ_1	IC	ϕ_2	IC	τ	IC
	V.Exata	- 0.5123	- 0.6512; - 0.3742	0.2910	0.1509; 0.4330	1.1172	0.9124; 1.3457
V.Aprox.	- 0.5149	- 0.6592; - 0.3704	0.2947	0.1510; 0.4434	1.1025	0.9050; 1.3102	
série 3 (Furnas)		ϕ_1	IC	ϕ_2	IC	τ	IC
	V.Exata	0.5848	0.4948; 0.6790	0.1785	0.0858; 0.2618	2.0561	1.8283; 2.2886
V.Aprox.	0.5821	0.4975; 0.6713	0.1820	0.0941; 0.2722	2.0531	1.8226; 2.2763	
série 4 (Itumbiara)		ϕ_1	IC	τ	IC		
	V.Exata	0.7146	0.6583; 0.7745	1.9742	1.7615; 2.1836		
V.Aprox.	0.7225	0.6673; 0.7784	1.9291	1.7321; 2.1327			

Notamos na tabela 9 que os valores das estimativas obtidas usando-se a função de verossimilhança exata e a aproximada são praticamente iguais, portanto isso justifica o uso sa função aproximada, visto que esta facilita muito a implementação computacional dos algoritmos de simulação MCMC.

Conclusão

Mostramos neste trabalho a utilização de Métodos Bayesianos aplicados em séries temporais, em particular para os modelos auto-regressivos $AR(p)$, com o uso de algoritmos de simulação de Monte Carlo em Cadeia de Markov (MCMC).

Nosso interesse foi comparar os métodos de máxima verossimilhança e os métodos Bayesianos quando ambos são utilizados para estimar os parâmetros dos modelos propostos.

A abordagem Bayesiana apresenta resultados semelhantes aos estimadores de máxima verossimilhança. Portanto podem ser usados de forma segura, além de facilitar os cálculos na inferência dos parâmetros.

Na estimação dos parâmetros com o método Bayesiano, os resultados obtidos com o uso da verossimilhança aproximada se mostrou equivalentes aqueles com a verossimilhança exata, aqui ressaltamos a simplicidade de cálculo e programação com a verossimilhança aproximada, são simples e o tempo computacional dispendido é bem menor que aquele com a Verossimilhança Exata.

O uso do método MCMC para previsão k -passos a frente, foi surpreendente pela simplicidade e eficiência do método. Estes resultados indicam uma nova direção a ser tomada na previsão de séries temporais.

O uso do critério Bayesiano de adequabilidade do modelo baseado na densidade preditiva ordenada mostrou-se adequado com resultados coincidentes com os critérios AIC e BIC. Porém devemos recomendar que esses critérios devem ser usados juntamente com os intervalos de credibilidade dos parâmetros para modelos onde o produto das densidades preditivas são muito próximos.

As técnicas utilizadas neste trabalho também poderiam ser estendidas a outros modelos de séries temporais, como médias móveis de ordem q $MA(q)$, e modelos mistos auto-regressivo médias móveis de ordens p e q $ARMA(p, q)$ ajustados pela abordagem clássica e Bayesiana.

Referências

- Akaike, H.**, (1974). A new look at the statistical identification model. *IEEE trans. Auto. Control*, 19, 716-723.
- Box, G.E.P., Jenkins, G.M. & Reinsel, (1994).** Time Séries Analysis: Forecasting and Control, Holden-Day, 3th edition.
- Box, G.E. & Tiao, G.C.**, (1973). Bayesian Inference in Statistical Analysis, New York: Addison Wesley.
- Broemeling, L.D. & Shaarawy, S.**, (1988), Time Series: A Bayesian Analysis in Time Domain, in Bayesian Analysis of TimeSeries an Dynamic Models, *Marcel Dekker*.
- Broemeling, L.D. & Land, M.**, (1984), On Forecasting with univariate autoregressive processes, *Comm. Statist.- Theor. Meth.*, 13(11), 1305-1320.
- Casella, G. & George, E.J.**, (1992), Explaining the Gibbs Sampler, *Amer. Statist. Assoc.*, vol. 46, No.3, 167-174.
- Chang, H.**, (1995) Model Determination Using Predictive Distributions, *PHD These, University of Connecticut*.
- Chib, S. & Greenberg, E.**, (1995), Understanding the Metropolis-Hastings Algorithm, *Amer. Statist. Assoc.*, vol.49, No.4.
- Cowles, M.K. & Carlin, B.P.**, (1996), Markov Chain Monte Carlo Convergence Diagnostics: A comparative Review, *Journal of the American Statistical Association*, 91, 883-904.
- de Alba, Enrique**, (1993), Constrained forecasting in autoregressive time serie models: A Bayesian analysis, *Int. Journal of forecasting* 9, 95-108.
- Gelfand, A.E. & Smith, A.F.M.**, (1990). Sampling - Basead Aproaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, 85, No.410, 398-409.
- Gelfand, A.E., Hills, S.E., Racine-Poon, A. & Smith, A.F.M.**, (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling, *Journal of the American Statistical Association*, 85, No.412, 972-985.
- Gelman, A.E. & Rubin, D.**, (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, 7,457-472.

McCulloch, R.E. & Tsay, R., (1994), Bayesian Analysis of Autoregressive Time Series via The Gibbs Sampler, *Journ. of Time Series Analysis*, vol.15, No.2, 235-250.

Miller, J.W., (1995). Exact Maximum Likelihood Estimation in Autoregressive Process, *Journ. of Time Series Analysis*, vol.16, No.6, 607-615.

Shaarawy, S. & Broemeling, L.D., (1984), Bayesian Inferences and Forecasts with Moving Averages Processes, *Comm. Statist. - Theor. Meth.*, 13(15), 1871-1888.

Shaarawy, S. & Broemeling, L.D., (1985), Inference and Prediction with ARMA Processes, *Comm. Statist.- Theor. Meth.*, 14(10), 2531-2548.

Schwarz, G., (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, vol.6, No.2, 461-464.

Tierney, L., Kass, R.E. & Kadane, J.B., (1986). Fully exponential Laplace approximations for expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84, 710-716.

Apêndice 1

Identificação da Ordem do Modelo

Identificar a ordem de um modelo auto-regressivo consiste em determinar o valor de p , ou seja determinar o número adequado de termos que devemos considerar no modelo auto-regressivo. A escolha de um valor pequeno de p , por exemplo, $p = 1$ pode levar a uma representação inadequada da série, por outro lado a escolha de um valor alto de p , por exemplo, $p = 6$ pode levar a um modelo com um grau de complexidade desnecessário. Na prática a análise da função de autocorrelação (FAC) e autocorrelação parcial (FACP) podem levar a mais de um modelo aparentemente adequados para representar um conjunto de observações de uma série temporal. Nestes casos usa-se os critérios de informação de Akaike (AIC) proposto por [Akaike -1973] ou o critério de informação Bayesiano (BIC), proposto por [Schwarz -1978]. Esses critérios são expressos como:

$$\text{AIC} = \ln(\hat{\sigma}_a^2) + \frac{2r}{N}$$

$$\text{BIC} = \ln(\hat{\sigma}_a^2) + \frac{r \ln(N)}{N}$$

onde $\hat{\sigma}_a^2$ denota o estimador de máxima verossimilhança da variância σ_a^2 e $r = p + 1$ denota o número de parâmetros estimados no modelo incluindo o termo constante. Nos critérios acima o primeiro termo corresponde a menos $2/N$ vezes o \log da máxima verossimilhança enquanto o segundo termo é "fator de penalidade" para a inclusão de parâmetros adicionais no modelo.

A aplicação desses critérios é feita calculando-se o AIC e BIC para vários modelos ajustados e devemos dar preferência ao modelo que apresentar o menor valor desses critérios.

Notando que o BIC impõe uma penalidade maior para o número de parâmetros do modelo que o AIC, o uso do BIC como critério de seleção sempre resultará em uma escolha do modelo cujo número de parâmetros não é maior que o escolhido pelo AIC.

Uma desvantagem imediata do uso desses critérios é que vários modelos devem ser estimados por métodos de máxima verossimilhança o que é computacionalmente custoso.

Apêndice 2

Verificação de Convergência

Uma forma simples de verificar convergência é a utilização de várias cadeias em paralelo começando de diferentes pontos. Com isso, evita-se que as cadeias se concentrem em regiões em torno de uma moda local, no caso de multimodalidade da posteriori. Após convergência, todas as cadeias devem ter o mesmo comportamento qualitativo e quantitativo mas a verificação de convergência pode ser problemática se for analisada apenas uma cadeia.

A verificação visual de similaridade entre as trajetórias das cadeias após algumas iterações certamente é um indício forte de convergência. [Gelman & Rubin -1992] elaboraram sobre essa idéia e propuseram alguns métodos formais para detecção de convergência além dessa detecção visual.

Considerando m cadeias paralelas e uma função real $t(\theta)$, tem-se m trajetórias $\{t_i^{(1)}, t_i^{(2)}, \dots, t_i^{(n)}\}$, $i = 1, \dots, m$ para t . Portanto, podem ser obtidas a variância entre as cadeias E e a variância dentro das cadeias D dadas por,

$$E = \frac{n}{m-1} \sum_{i=1}^m (\bar{t}_i - \bar{t})^2$$

e

$$D = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (t_i^{(j)} - \bar{t}_i)^2$$

onde \bar{t}_i é a média das observações da cadeia i e \bar{t} é a média dessas médias, $i = 1, \dots, m$.

Sob convergência, todos os mn valores serão gerados da posteriori e a variância de t pode ser estimada de forma não-viciada por $\widehat{V}(t(\theta)) = (1 - \frac{1}{n})D + (\frac{1}{n})E$.

Se as cadeias ainda não tiverem convergido então essa estimativa é maior que $V(t(\theta))$ pois os valores iniciais ainda estão sendo influenciados pelos valores dos outros parâmetros da cadeia, de forma que a distribuição de equilíbrio ainda não foi atingida e eles foram escolhidos com dispersão maior que a da distribuição de equilíbrio. Por outro lado, D fornece estimativas menores que $V(t(\theta))$ pois a cadeia não terá coberto toda a variabilidade de $t(\theta)$. Um indicador de convergência é dado pela redução potencial estimada da escala $R = \sqrt{\widehat{V}(t(\theta))/D}$ que é

sempre maior que 1. À medida que n cresce, ambos os estimadores acabarão convergindo para $V(t(\theta))$ e R convergirá para 1. Logo R pode ser usado como indicador de convergência pela avaliação de sua proximidade a 1. [Gelman -1995] sugere aceitar convergência quando o valor de R calculado para todas as funções t de interesse for da ordem de 1.1 à 1.2.