

Universidade Federal de Lavras

Inferência Bayesiana



Democratizando
Conhecimento

Ben Dêivide de O. Batista

1	Fundamentos	1
1.1	Paradigma da Estatística Clássica	1
1.2	Paradigma Bayesiano	1
1.3	Principal diferença entre a inferência clássica e Bayesiana	2
1.4	Teorema de Bayes para o caso contínuo	3
2	Escolha de <i>priori</i>	7
2.1	Informação a <i>priori</i>	7
2.2	Conceitos de probabilidade	7
2.3	Distribuição a <i>priori</i> subjetiva	7
2.4	Método para obtenção de <i>priori</i>	8
2.4.1	Método estrutural	8
2.5	Tipos de distribuição a <i>priori</i>	9
2.5.1	Distribuição a <i>priori</i> não-informativa	9
2.6	Distribuições impróprias e suas implicações	13
3	Inferência	15
3.1	Estimação pontual	15
3.2	Estimativa por região	16
3.3	Teste de hipótese	17
3.4	Comparação de modelos	17
3.5	Inferência Bayesiana para um modelo linear (PAULINO; TURKMAN; MURTEIRA, 2003, p. 216)	18
3.5.1	Teorema de Bayes	18
3.5.2	Informação a <i>priori</i>	18
3.5.3	Análise a <i>posteriori</i>	19
3.5.4	Algumas considerações	22
3.6	Análise preditiva	22
4	Aproximações Analíticas e Numéricas	24
4.1	Métodos analíticos	25
4.1.1	Aproximação da distribuição a <i>posteriori</i> por uma distribuição multivariada	25
4.1.2	Método de Laplace	26
4.2	Quadratura Gaussiana ou integração numérica (Método numérico)	27
4.3	Métodos de Simulação	27
4.3.1	Métodos baseados em simulação estocástica	27
4.3.2	Métodos de Monte Carlo Ordinário	28
4.3.3	Monte Carlo com amostragem por importância	29
4.3.4	Método de Monte Carlo via Cadeia de Markov (MCMC)	30
4.4	Crítérios de convergência	33

4.4.1	Critério de Gelman e Rubin (1992)	33
4.4.2	Critério de Raftery e Lewis (1992)	34
4.4.3	Critério de Geweke	34
4.4.4	Critério de Heidelberger e Welch	34
5	Validação e seleção de modelos	35
5.1	Medidas de Diagnóstico - Adequabilidade de um modelo	35
5.2	Comparação de Modelos	36
5.2.1	Alternativas	37
6	Listas de exercício	38
6.1	Lista 1	38
6.2	Lista 2	46
6.3	Lista 3	54
6.4	Lista 4	59
6.5	Lista 5	69
6.6	Lista 6	78
6.7	Lista 7	89
	Referências Bibliográficas	99

1.1 Paradigma da Estatística Clássica

O principal objetivo da estatística continua a se reformular do seguinte modo: determinar que generalização sobre a população pode ser feita a partir da amostra que foi recolhida.

Um aspecto importante da influência clássica consiste em observar a variabilidade que se verifica de amostra para amostra. Assim, os procedimentos clássicos são julgados à luz de **repetições de amostragem**.

Uma das faces do princípio reside principalmente na **interpretação frequentista de probabilidade**, isto é, na utilização de frequências como medida de incerteza. A outra face reside na avaliação dos procedimentos estatísticos em torno da frequência com que produzem respostas corretas ou bons resultados.

1.2 Paradigma Bayesiano

No que toca propriamente as ideias bayesianas e a sua aplicação à estatística, tem de citar Harold Jeffreys que conseguiu resurgir o Bayesianismo em 1939, dando-lhes status lógico e avançado com solução de problemas estatísticos que naquele tempo persistiam sem solução.

O teorema de Bayes é uma proposição sobre probabilidades condicionadas indiscutível desde que se aceite as leis de probabilidade ou o axioma de Kolmogorov. O que tem dado lugar a grande controvérsia é a sua interpretação e a sua aplicação a problemas de inferência estatística.

“Toda a análise Bayesiana está centrada no teorema de Bayes”.

Considere uma partição finita de um espaço de resultados de Ω (Espaço amostral).

$$\begin{aligned}
 &A_1, A_2, \dots, A_n \\
 &P(A_i) > 0, \quad i = 1, 2, \dots, n \\
 &A \cap A_j \neq 0, \forall \quad i \neq j \\
 &\bigcup_i A_i = \Omega
 \end{aligned}$$

Dado um certo acontecimento B qualquer, com $P(B) > 0$, é fácil verificar que

$$B = \bigcup_i (A_i \cap B)$$

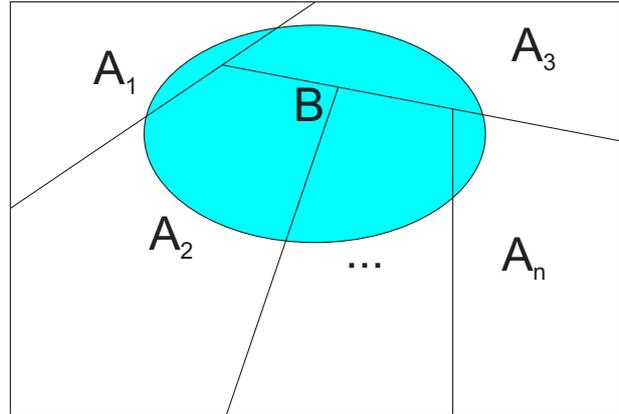


Figura 1.1: Partição do espaço amostral ($k = n$).

Portanto, da definição de probabilidade condicionada e a aditividade da função de probabilidade, temos

$$\begin{aligned}
 P(B) &= \sum_i P(A_i \cap B) = \sum_i P(B|A_i)P(A_i), \\
 \text{ou ainda} &= \sum_i P(A_i|B)P(B).
 \end{aligned}$$

Então,

$$P(A_i \cap B) = P(B|A_i)P(A_i) = P(A_i|B)P(B), \tag{1.1}$$

$$\begin{aligned}
 \Rightarrow P(A_i|B) &= \frac{P(B|A_i)P(A_i)}{P(B)}, \\
 &= \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)} \quad (\text{Fórmula de Bayes})
 \end{aligned}$$

Interpretação:

1. Considere os A_i como antecedentes, causas ou hipóteses a que o investigador atribua graus de credibilidade (caso discreto) ou função densidade de probabilidade (caso contínuo) *a priori* $P(A_i)$ de natureza **SUBJETIVA**. Nota-se que a determinação e a interpretação da distribuição *a priori* são entre os pontos mais melindrosos e controversos da teoria bayesiana e constituem um dos principais obstáculos à respectiva implementação.
2. Depois da informação adicional que consiste em saber que o acontecimento B se realizou, o investigador revê suas probabilidades *a priori* através da fórmula de Bayes e passa a atribuir aos A_i as probabilidades *a posteriori* $P(A_i|B)$, $i = 1, 2, \dots, n$.

O teorema de Bayes é um dos poucos resultados da matemática que se propõe caracterizar a aprendizagem com a experiência, isto é, modificação de atitude inicial em relação aos “antecedentes” depois de ter a informação adicional de que certo acontecimento ou acontecimentos se realizaram.

“O teorema de Bayes pode ser visto como uma atualização da opinião”.

1.3 Principal diferença entre a inferência clássica e Bayesiana

No modelo clássico o parâmetro θ a ser estimado é fixo, isto é, igual ao valor do parâmetro particular. No modelo Bayesiano, o parâmetro θ é aleatório, ou seja, θ é desconhecido e incerto. Toda incerteza deve ser quantificada em termos de probabilidade.

Se θ é um parâmetro discreto, então $P(\theta)$ é a função de probabilidade *a priori*. Assim, $P(\theta)$ exprime o grau de credibilidade que o indivíduo que procede a análise atribui ao θ considerado. Se

θ é um parâmetro contínuo, $P(\theta)$ é a função densidade de probabilidade *a priori*. Assim, $P(\theta)d\theta$ exprime o grau de credibilidade no intervalo $\theta, \theta + d\theta$.

A família \mathfrak{F} de distribuição também faz parte do modelo bayesiano, ou seja, a componente amostral ou experimental é comum aos modelos clássicos e bayesianos, embora para este os elementos de $f(X|\theta)$ de \mathfrak{F} em geral tem uma interpretação subjetiva.

1.4 Teorema de Bayes para o caso contínuo

Suponha que se observa $X = x$. Considerando um elemento qualquer de \mathfrak{F} , ou seja, $f(x|\theta)$, e a distribuição $P(\theta)$, o teorema de Bayes para densidades, conduz à relação

$$P(\theta|x) = \frac{f(x|\theta)P(\theta)}{\int_{\Theta} f(x|\theta)P(\theta)d\theta},$$

onde $\theta \in \Theta$, e $P(\theta|x)$ é a distribuição de θ depois de saber que se obteve $X = x$.

Observando-se uma amostra casual, $(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$, tem-se

$$P(\theta|x_1, x_2, \dots, x_n) = \frac{\prod_i f(x_i|\theta)P(\theta)}{\int_{\Theta} \prod_i f(x_i|\theta)P(\theta)d\theta}, \quad \theta \in \Theta,$$

em que $P(\theta|x_1, x_2, \dots, x_n)$ é a distribuição *a posteriori* de θ depois de conhecida a particular amostra (x_1, x_2, \dots, x_n) .

Assim, tendo em conta a informação contida nos dados x , a atitude inicial do investigador, $P(\theta)$, é modificada passando a nova atitude para $P(\theta|x)$. Observando a expressão da *posteriori*, nota-se que o denominador não depende de θ , funciona como uma constante normalizadora de $P(\theta|x)$. Assim reescrevendo

$$P(\theta|x) \propto f(x|\theta)P(\theta), \quad \theta \in \Theta.$$

Sendo a análise condicionada pelo valor x observado com θ variando sobre Θ , o fator $f(x|\theta)$ identifica-se como a função de verossimilhança *a priori* enquanto $P(\theta)$ é chamada distribuição *a priori* de θ , ou seja,

$$P(\theta|x) \propto L(\theta|x)P(\theta).$$

Em outras palavras tem-se que

$$\boxed{\text{distribuição } a \text{ posteriori} \propto \text{verossimilhança} \times \text{distribuição } a \text{ priori}}$$

Note que ao omitir o termo $\int_{\Theta} f(x|\theta)P(\theta)d\theta$, a igualdade foi substituída por uma proporcionalidade.

A função de verossimilhança, tem importante papel na fórmula de Bayes, pois representa o meio através do qual os dados x , transformam o conhecimento *a priori* sobre θ , ou seja, a verossimilhança pode ser interpretada como expressão da informação sobre θ fornecida pelos dados x .

As Figuras 1.2 e 1.3 esboçam o esquema clássico e bayesiano, respectivamente, para a inferência na estatística.

Exercício resolvido 1.1: Paulino, Turkman e Murteira (2003), Ex.1.4 Pág. 16

Considere-se a observação de uma variável aleatória com distribuição normal $X \sim N(\theta, \sigma^2)$, com σ^2 conhecido e admita-se *a priori* $\theta \sim N(a, b^2)$ com a e b também conhecidos. Encontre a distribuição *a posteriori* para θ .

Solução:

Seja $X \sim N(\theta, \sigma^2)$, então

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}, \\ &\propto \frac{1}{(\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}, \end{aligned}$$

$$\begin{aligned} P(\theta) &= \frac{1}{\sqrt{2\pi b^2}} \exp\left\{-\frac{1}{2b^2}(\theta-a)^2\right\}, \\ &\propto \frac{1}{(b^2)^{1/2}} \exp\left\{-\frac{1}{2b^2}(\theta-a)^2\right\}, \end{aligned}$$

$$\begin{aligned} f(x|\theta)P(\theta) &\propto \frac{1}{(\sigma^2)^{1/2}} \frac{1}{(b^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2 - \frac{1}{2b^2}(\theta-a)^2\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-a)^2}{b^2}\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{1}{2}\left[\frac{b^2(x^2 - 2x\theta + \theta^2) + \sigma^2(\theta^2 - 2a\theta + a^2)}{\sigma^2 b^2}\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{1}{2}\left[\frac{b^2 x^2 - 2b^2 x\theta + b^2 \theta^2 + \sigma^2 \theta^2 - 2\sigma^2 a\theta + \sigma^2 a^2}{\sigma^2 b^2}\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2 b^2}\left[\theta^2(b^2 + \sigma^2) - \theta(2b^2 x + 2\sigma^2 a) + b^2 x^2 + \sigma^2 a^2\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2 b^2}\left[\theta^2(b^2 + \sigma^2) - \theta(2b^2 x + 2\sigma^2 a)\right]\right\} \times \\ &\quad \times \underbrace{\exp\left\{-\frac{1}{2\sigma^2 b^2}\left[b^2 x^2 + \sigma^2 a^2\right]\right\}}_{\text{Constante, retira-se da proporcionalidade}}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2 b^2}\left[\theta^2(b^2 + \sigma^2) - \theta(2b^2 x + 2\sigma^2 a)\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{(b^2 + \sigma^2)}{2\sigma^2 b^2}\left[\theta^2 - 2\theta\frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)} + \right.\right. \\ &\quad \left.\left. + \underbrace{\left(\frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)}\right)^2 - \left(\frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)}\right)^2}_{=0}\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{(b^2 + \sigma^2)}{2\sigma^2 b^2}\left[\left(\theta - \frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)}\right)^2 - \left(\frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)}\right)^2\right]\right\}, \\ &\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp\left\{-\frac{(b^2 + \sigma^2)}{2\sigma^2 b^2}\left[\left(\theta - \frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)}\right)^2\right]\right\} \times \\ &\quad \times \underbrace{\exp\left\{\frac{(b^2 + \sigma^2)}{2\sigma^2 b^2}\left(\frac{(b^2 x + \sigma^2 a)}{(b^2 + \sigma^2)}\right)^2\right\}}_{\text{Constante, retira-se da proporcionalidade}}, \end{aligned}$$

$$\propto \frac{1}{(\sigma^2 b^2)^{1/2}} \exp \left\{ -\frac{(b^2 + \sigma^2)}{2\sigma^2 b^2} \left[\left(\theta - \frac{b^2 x + \sigma^2 a}{b^2 + \sigma^2} \right)^2 \right] \right\}.$$

Portanto,

$$\theta|x \sim N \left(\frac{b^2 x + \sigma^2 a}{b^2 + \sigma^2}, \frac{b^2 \sigma^2}{b^2 + \sigma^2} \right).$$

Exercício resolvido 1.2

Suponha que as v.a. $X_i, i = 1, 2, \dots, n$ iid $N(0, \sigma^2)$ com precisão $\lambda = 1/\sigma^2$ e que a *priori* para λ é dada por $\lambda \sim \text{Gama}(\alpha, \alpha\lambda_0^{-1})$, isto é,

$$p(\lambda) = \frac{(\alpha\lambda_0)^\alpha}{\Gamma(\alpha)} \exp \left\{ -\alpha \frac{\lambda}{\lambda_0} \right\} \lambda^{\alpha-1}.$$

Determine $\lambda|x_1, x_2, \dots, x_n$.

Solução:

Seja $X_i \sim N(0, \sigma^2)$, considerando $\lambda = 1/\sigma^2$, então

$$\begin{aligned} f(x_1, x_2, \dots, x_n|\lambda) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} x_i^2 \right\}, \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma^2} \right)^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} x_i^2 \right\}, \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \lambda^{1/2} \exp \left\{ -\frac{\lambda}{2} x_i^2 \right\}, \\ &\propto \prod_{i=1}^n \lambda^{1/2} \exp \left\{ -\frac{\lambda}{2} x_i^2 \right\}, \\ &\propto \lambda^{n/2} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^n x_i^2 \right\}, \end{aligned}$$

$$\begin{aligned} p(\lambda) &= \frac{(\alpha\lambda_0)^\alpha}{\Gamma(\alpha)} \exp \left\{ -\alpha \frac{\lambda}{\lambda_0} \right\} \lambda^{\alpha-1}, \\ &\propto \exp \left\{ -\alpha \frac{\lambda}{\lambda_0} \right\} \lambda^{\alpha-1}, \end{aligned}$$

$$\begin{aligned} p(\lambda|x_1, \dots, x_n) &\propto f(x_1, x_2, \dots, x_n|\lambda)p(\lambda) \\ &\propto \lambda^{n/2} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^n x_i^2 \right\} \exp \left\{ -\alpha \frac{\lambda}{\lambda_0} \right\} \lambda^{\alpha-1}, \\ &\propto \lambda^{\alpha+n/2} \exp \left\{ -\lambda \left(\frac{\alpha}{\lambda_0} + \frac{\sum_{i=1}^n x_i^2}{2} \right) \right\} \end{aligned}$$

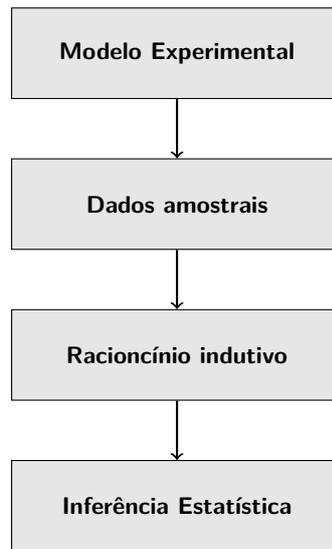


Figura 1.2: Esquema clássico.

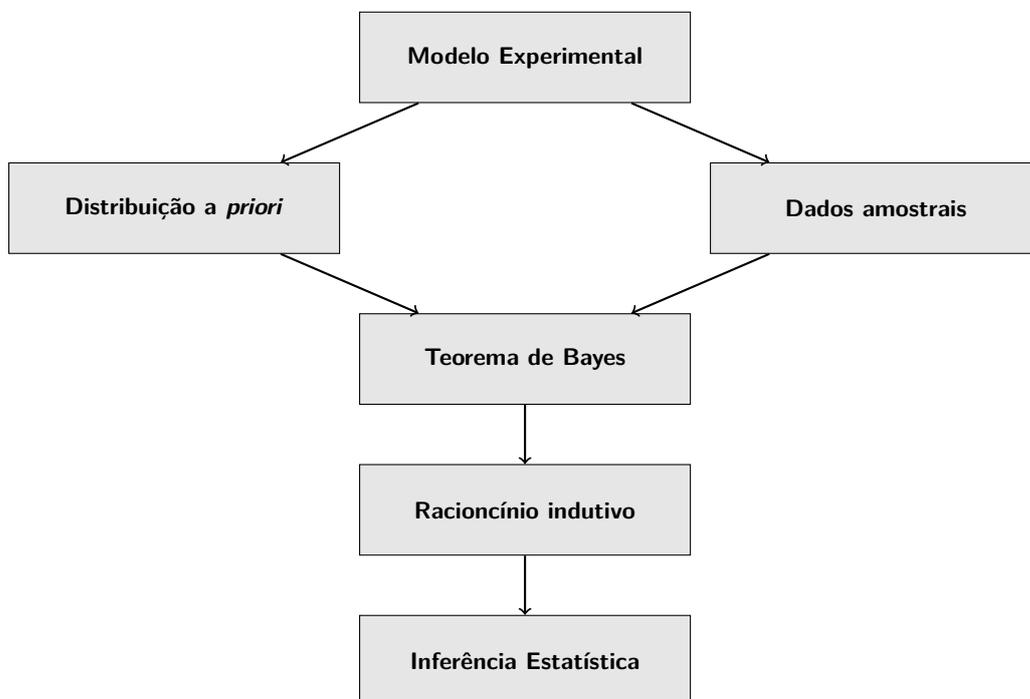


Figura 1.3: Esquema Bayesiano.



2.1 Informação a *priori*

A informação a *priori* que se pretende incorporar à análise é a “opinião”, “crença” possuída por alguém que pode ser identificada como o especialista do problema concreto e contém elementos subjetivos que muitas vezes são obtidos de fontes objetivas como dados históricos ou de problemas análogos.

2.2 Conceitos de probabilidade

Para a inferência clássica, a probabilidade de um evento é a frequência relativa de ocorrência do evento numa sequência infinita de repetição do experimento em questão.

Os conceitos subjetivos e lógicos de probabilidade tem como ponto comum a característica de representarem **graus de crença** condicionadas à informação disponível seja num acontecimento ou em um conjunto de afirmação sobre a situação em análise.

A **probabilidade subjetiva** é uma medida do **grau de crença específico de um indivíduo**. Pode variar de indivíduo para indivíduo, até porque a informação que cada um possui é geral e rigorosamente diferenciada.

A **probabilidade lógica** é uma medida do **grau de crença objetivo** (não pessoal), que todo indivíduo racional deve ter e está ligado à informação disponível.

2.3 Distribuição a *priori* subjetiva

Existem situações práticas em que há informação *a priori* mais ou menos substancial sobre os parâmetros do modelo, seja por parte do pesquisador ou seja por parte de outros indivíduos a quem o

pesquisador pode recorrer. A questão de como obter e quantificar esta informação de natureza subjetiva, de modo a transformá-la em uma distribuição *a priori* que possa ser utilizada para a metodologia Bayesiana é também tratada na literatura.

Exemplo 2.1: Paulino, Turkman e Murteira (2003), Pág. 76

Uma fábrica produz certo tipo de peças eletrônicas de grande precisão, estas peças para serem vendidas são agrupadas em lotes de tamanho N (N grande). Suponha que sempre existe uma certa porcentagem θ dessas peças que podem ser consideradas defeituosas por não obedecerem a certos critérios de precisão. O preço do lote depende desta porcentagem e conseqüentemente há necessidade de inferir sobre o seu valor através, por exemplo, da especificação de um limite superior C para esta porcentagem com uma probabilidade elevada de não ser ultrapassado. **Há bastante informação acumulada** sobre o valor provável dessa porcentagem, quer de lotes anteriormente vendidos, quer pelas razões pelas quais são produzidas peças defeituosas.

Para estabelecer este limite o estatístico vai recolher de especialistas na área, informação *a priori* que existem sobre θ .

Pontos a serem levantados:

1. Que perguntas poderiam ser feitas ao especialista?
2. Como ajudar os especialistas, que não necessitam ter conhecimento de probabilidade e estatística, a responderem coerente e consistentemente a essas perguntas?
3. Se recolher informação a mais de um especialista, como combinar a informação recebida?
4. Como usar a informação recebida de modo a construir uma distribuição de probabilidade para o parâmetro ou acontecimento de interesse?

O estatístico pode optar por não realizar a experiência, fixar um ou mais valores de C e pedir aos especialistas para atribuírem uma probabilidade *a priori* para o acontecimento $A = \{\theta > C\}$.

- ✦ Como ajudá-los a estabelecer esta probabilidade?
- ✦ Como o estatístico pode fazer uso da informação recolhida de um especialista para formular uma probabilidade *a posteriori* para um acontecimento de interesse?

O estatístico pode admitir um modelo amostral para o número X de peças defeituosas numa amostra x e, portanto para obter C baseando-se na distribuição *a posteriori* de θ dado o vetor observado de X . Para tanto necessita novamente de uma *priori* para θ .

2.4 Método para obtenção de *priori*

2.4.1 Método estrutural

Qualquer método onde a escolha da *priori* para θ é baseada em questão relacionadas diretamente com o parâmetro.

A natureza das questões depende do tipo do parâmetro que se quer inferir, do conhecimento do especialista, etc.

O'Hagan (1998) cita dois princípios que acredita serem importantes:

1. Fazer perguntas que sejam entendidas pelo especialista numa linguagem simples e que seja familiar;
2. Identificar e pesquisar separadamente as principais fontes de incerteza.

2.4.1.1 Método do Histograma - Paulino, Turkman e Murteira (2003) pág. 83

Se θ é contínuo, considere uma partição de Θ em R intervalos, $\Theta = \bigcup_{i=1}^R \theta_i$, e questione o especialista sobre qual probabilidade ele atribuiria a θ pertencer a cada um dos intervalos.

Com base nesta informação construa um histograma e a partir daí estime uma densidade de probabilidade adequada. Após a obtenção desta *priori*, consultar o especialista para ver se está de acordo com a “crença” ou não.

Um processo iterativo permitiria obter uma distribuição final de acordo com o conhecimento do especialista.

Se θ é discreto, e $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ ou seja, há apenas um número finito de valores possível para θ , o problema fica reduzido a determinar a probabilidade para cada elemento de Θ .

2.4.1.2 Método verossimilhança relativa - Paulino, Turkman e Murteira (2003) pág. 85

Este método é particularmente útil quando θ é um intervalo de \mathbb{R} . Consiste em comparar as verossimilhanças de vários pontos do espaço de parâmetro e a partir daí construir função $g(\theta)$ que passe por estes pontos. Esta função, não é necessariamente uma densidade, mas pode ser multiplicada por uma constante de modo que integre 1.

2.4.1.3 Escolha de hiperparâmetros

Em muitas situações, há consenso sobre a forma que a distribuição *a priori* deve ter e o problema para sua completa especialização consiste em determinar seus parâmetros, denominados hiperparâmetros.

Exemplo 2.2

Se a priori é uma $Gama(\alpha, \beta)$, o problema seria dar valores para α e β .

2.5 Tipos de distribuição *a priori*

A simplicidade da operação bayesiana fica garantida se se impuser que a família de distribuição *a priori* H seja fechada sob amostragem de $\mathfrak{F} = f(x|\theta)$, $\theta \in \Theta$, isto é, que se $P(\theta) \in H$ então $P(\theta|X) \propto f(x|\theta)P(\theta) \in H$.

Exemplo 2.3

Para $f(x|\mu, \sigma^2) \sim Normal$:

✘ se $\mu \sim Normal \Rightarrow (\mu|x) \sim Normal$;

✘ se $\sigma^2 \sim GI \Rightarrow (\sigma^2|x) \sim GI$.

2.5.1 Distribuição *a priori* não-informativa

Quando não existe informação *a priori* palpável, de natureza quer objetiva, quer subjetiva (“ignorância *a priori*”) ou quando se tem pouca informação relativamente à informação amostral (“conhecimento vago”) é comum o uso de *priors* não-informativa.

Este tipo de distribuição pode desempenhar um papel de referência, mesmo que se disponha de fortes crenças *a priori*, como forma de:

- ✘ deduzir as crenças *a posteriori* para quem parte de um conhecimento escasso (i.e., quando a amostra fornece o grosso da informação sobre o parâmetro) e, nessa medida, se acha incapaz de determinar subjetivamente uma distribuição razoável. Assim reconhecendo a própria ignorância;

✦ permite a comparação com os resultados da inferência clássica que só usa a informação amostral.

2.5.1.1 Métodos para encontrar *prioris* não-informativas

✦ **Bayes-Laplace** - Princípio da razão insuficiente. Na ausência da razão insuficiente para privilegiar uma probabilidade em detrimento de outras, decorrente da escassez informativa *a priori*, deve-se adotar que todas têm a mesma de ocorrer.

Exemplo 2.4

Utilizar como *priori* a distribuição uniforme.

✦ **Método de Jeffreys** - A *priori* de Jeffreys é obtida a partir da matriz de informação de Fisher $I(\theta)$. Para o caso **uniparamétrico**, a priori de Jeffreys é dada por

$$P(\theta) = \sqrt{I(\theta)},$$

sendo

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \ln f(x|\theta) \right)^2 \middle| \theta \right].$$

No caso **multiparamétrico**, a matriz de informação de Fisher é dada por $I(\theta)$, na qual o ij -ésimo elemento da matriz é

$$I_{ij}(\theta) = E \left(\frac{\partial \ln f(x|\theta)}{\partial \theta_i} \frac{\partial \ln f(x|\theta)}{\partial \theta_j} \middle| \theta \right)$$

Um resultado interessante sobre a informação de Fisher:

$$E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(x_i, \theta) \right)^2 \right] = -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \log \prod_{i=1}^n f(x_i, \theta) \right].$$

Seja,

$$\begin{aligned} \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta) &\stackrel{\text{iid}}{=} \frac{\partial}{\partial \theta} \sum_{i=1}^n \log f(x_i, \theta), \\ &= \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(x_i, \theta), \\ &= \sum_{i=1}^n \frac{f'(x_i, \theta)}{f(x_i, \theta)}, \end{aligned}$$

então

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \prod_{i=1}^n f(x_i, \theta) &= \frac{\partial}{\partial \theta} \left(\sum_{i=1}^n \frac{f'(x_i, \theta)}{f(x_i, \theta)} \right), \\ &= \sum_{i=1}^n \frac{\partial}{\partial \theta} \frac{f'(x_i, \theta)}{f(x_i, \theta)}, \\ &= \sum_{i=1}^n \frac{f''(x_i, \theta)f(x_i, \theta) - f'(x_i, \theta)f'(x_i, \theta)}{(f(x_i, \theta))^2}, \\ &= \sum_{i=1}^n \left[\frac{f''(x_i, \theta)}{f(x_i, \theta)} - \left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right]. \end{aligned}$$

Logo,

$$\begin{aligned}
 -E \left[\frac{\partial^2}{\partial \theta^2} \log \prod_{i=1}^n f(x_i, \theta) \right] &= -E \left\{ \sum_{i=1}^n \left[\frac{f''(x_i, \theta)}{f(x_i, \theta)} - \left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right] \right\}, \\
 &= -\sum_{i=1}^n E \left[\frac{f''(x_i, \theta)}{f(x_i, \theta)} - \left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right], \\
 &= -\sum_{i=1}^n \left[\int \left(\frac{f''(x_i, \theta)}{f(x_i, \theta)} - \left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right) f(x_i, \theta) dx \right], \\
 &= -\sum_{i=1}^n \left[\int f''(x_i, \theta) dx - \int \left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 f(x_i, \theta) dx \right], \\
 &= -\sum_{i=1}^n \left\{ \int \frac{\partial^2}{\partial \theta^2} f(x_i, \theta) dx - E \left[\left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right] \right\}, \\
 &= -\sum_{i=1}^n \left\{ \underbrace{\frac{\partial^2}{\partial \theta^2} \int f(x_i, \theta) dx}_{=1} - E \left[\left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right] \right\}, \\
 &= -\sum_{i=1}^n \left\{ 0 - E \left[\left(\frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right] \right\}, \\
 &= E \left[\left(\sum_{i=1}^n \frac{f'(x_i, \theta)}{f(x_i, \theta)} \right)^2 \right], \\
 &= E \left[\left(\frac{\partial^2}{\partial \theta^2} \log \prod_{i=1}^n f(x_i, \theta) \right)^2 \right].
 \end{aligned}$$

Vale notar entretanto que a *priori* não informativa de Jeffreys viola o princípio da verossimilhança, já que a informação de Fisher depende da distribuição amostral.

Exemplo 2.5

Seja $X = \{x_1, \dots, x_n\}$ uma amostra aleatória, $X_i \sim N(\mu, \sigma^2)$, μ e σ^2 desconhecidos.

$$\begin{aligned}
 f(x|\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right\}, \\
 &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}.
 \end{aligned}$$

$$\ln f(x|\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

$$\begin{aligned}
 \frac{\partial}{\partial \mu} \ln f(x|\mu, \sigma^2) &= -\frac{1}{2\sigma^2} \left[2 \sum_{i=1}^n (x_i - \mu)(-1) \right], \\
 &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu), \\
 &= \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - n\mu \right).
 \end{aligned}$$

$$\frac{\partial^2}{\partial^2 \mu} \ln f(x|\mu, \sigma^2) = -\frac{n}{\sigma^2}.$$

$$E \left[\left(\frac{\partial}{\partial \mu} \ln f(x|\mu, \sigma^2) \right)^2 \right] = -E \left[\left(\frac{\partial^2}{\partial^2 \mu} \ln f(x|\mu, \sigma^2) \right) \right] = \frac{n}{\sigma^2}$$

$$\frac{\partial}{\partial \mu \partial \sigma^2} \ln f(x|\mu, \sigma^2) = -\frac{(\sum_{i=1}^n x_i - n\mu)}{(\sigma^2)^2}$$

$$E \left[\left(\frac{\partial}{\partial \mu \partial \sigma^2} \ln f(x|\mu, \sigma^2) \right)^2 \right] = -E \left[\left(\frac{\partial^2}{\partial^2 \mu \partial^2 \sigma^2} \ln f(x|\mu, \sigma^2) \right) \right] = 0$$

$$\begin{aligned} \frac{\partial^2}{\partial \sigma^2} \ln f(x|\mu, \sigma^2) &= -\frac{n2\pi}{4\pi\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2}, \\ &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2}. \end{aligned}$$

$$\frac{\partial^2}{\partial \sigma^2} \ln f(x|\mu, \sigma^2) = \frac{n}{2(\sigma^2)^2} - \frac{2\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^3}$$

$$\begin{aligned} E \left[\left(\frac{\partial}{\partial \sigma^2} \ln f(x|\mu, \sigma^2) \right)^2 \right] &= -E \left[\left(\frac{\partial^2}{\partial^2 \sigma^2} \ln f(x|\mu, \sigma^2) \right) \right] \\ &= -\frac{n}{2(\sigma^2)^2} + \frac{2E \left[\sum_{i=1}^n (x_i - \mu)^2 \right]}{2(\sigma^2)^3} \\ &= -\frac{n}{2(\sigma^2)^2} + \frac{2nE \left[(x - \mu)^2 \right]}{2(\sigma^2)^3} \\ &= -\frac{n}{2(\sigma^2)^2} + \frac{2n\sigma^2}{2(\sigma^2)^3} \\ &= -\frac{n}{2(\sigma^2)^2} + \frac{2n}{2(\sigma^2)^2} \\ &= \frac{n}{2(\sigma^2)^2} \end{aligned}$$

$$I(\mu, \sigma^2) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/[2(\sigma^2)^2] \end{pmatrix}$$

A priori de Jeffreys para o caso hiperparamétrico é

$$\begin{aligned} P(\mu, \sigma^2) &= \left(\begin{matrix} n/\sigma^2 & 0 \\ 0 & n/[2(\sigma^2)^2] \end{matrix} \right)^{1/2}, \\ &= \left(\frac{n^2}{2(\sigma^2)^3} \right)^{1/2}, \\ &\propto \sigma^{-3}, \quad \mu \in \mathbb{R}, \quad \sigma^2 > 0. \end{aligned}$$

Essa distribuição conjugada, com a verossimilhança, origina uma distribuição a *posteriori* conjunta caracterizada por

$$\begin{aligned}\mu|\sigma^2, x &\sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) \\ \sigma^2|x &\sim GI\left(\frac{n}{2}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2}\right)\end{aligned}$$

Quando μ é conhecido, se usa a distribuição uniforme paramétrica de Jeffreys, isto é,

$$P(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \sigma^2 > 0.$$

Exemplo 2.6

Verifique a afirmação:

“Quando μ é conhecido, se usa a distribuição uniforme paramétrica de Jeffreys, isto é,

$$P(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \sigma^2 > 0.”$$

Exemplo 2.7: Paulino, Turkman e Murteira (2003), Exerc. 2.12 - pág. 102

Num esquema de observação de variáveis aleatórias *i.i.d.* $Ber(\theta)$, considere-se, ao invés de uma amostragem binomial, uma amostragem binomial inversa em que o número de sucessos, t , é fixo e o número de observações, N , é aleatório (com distribuição Binomial Negativa de valor médio t/θ).

Como a medida de informação de Fisher é $I(\theta) = t\theta^{-2}(1 - \theta)^{-1}$, conclui-se que a distribuição a *priori* de Jeffreys é a distribuição imprópria “ $Be(0, 1/2)$ ”,

$$h(\theta) \propto \theta^{-1}(1 - \theta)^{1/2}, \theta \in (0, 1),$$

que é consistente com uma distribuição “Uniforme” para a função $\psi = \ln \frac{1 - \sqrt{1 - \theta}}{1 + \sqrt{1 - \theta}}$. A diferença entre as distribuições a *priori* deve-se às distribuições a *posteriori*, que são $Be(t, n - t + 1/2)$ e $Be(t + 1/2, n - 1 + 1/2)$ para os esquemas de amostragem binomial inversa e binomial, respectivamente, conduzindo a que as inferências em geral não sejam idênticas.

Imaginando agora que os dados podem ser obtidos por aplicação sequencial destes dois delineamentos, percebe-se que a distribuição a *posteriori* varia consoante a ordem em que eles são aplicados. Isto ilustra outra consequência da dependência das distribuições de Jeffreys para com o espaço amostral, a de que a distribuição a *posteriori* depende da ordem em que os dados vão sendo obtidos.

2.6 Distribuições impróprias e suas implicações

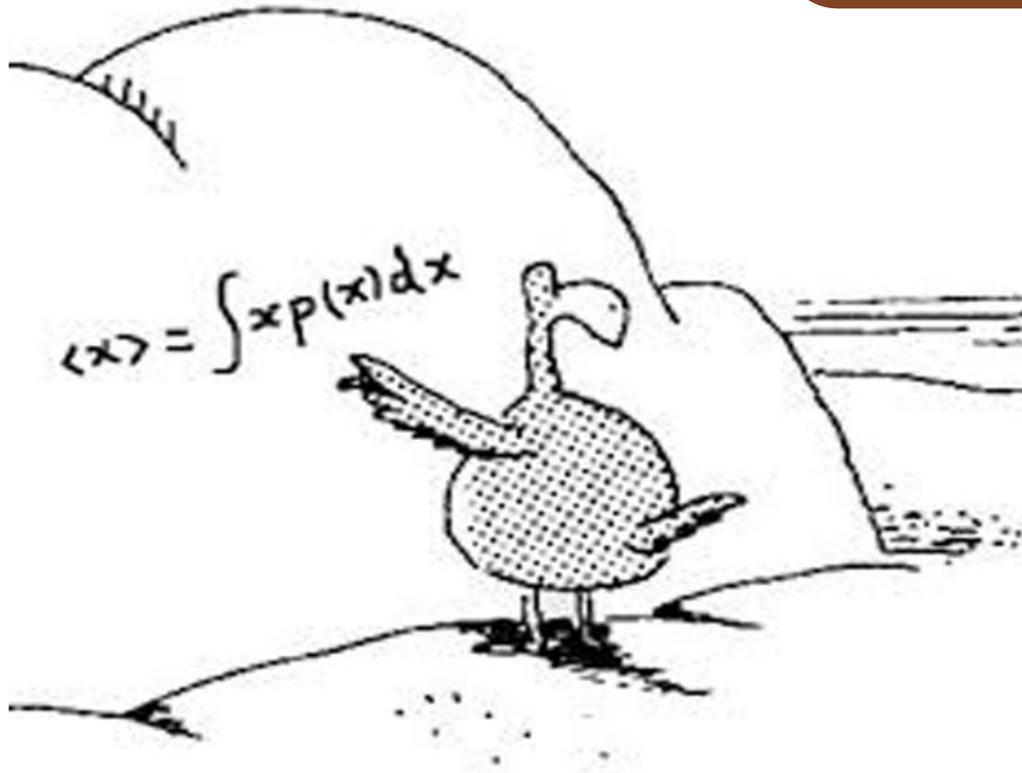
“Distribuições impróprias” são distribuições que satisfaça, o axioma de probabilidade unitária.

Quando a informação a *priori* é fraca à informação amostral frequentemente as *posteriores* são robustas perante especificações da distribuição a *priori*. Neste caso, o uso de *prioris* impróprias¹ podem ser aceitáveis para o subjetivista no sentido de construírem aproximações razoáveis a distribuições próprias² que representam informação a *priori*. O problema mais grave do uso de *prioris* impróprias

¹Não é uma fdp.

²É uma fdp.

é obter *a posteriori* imprópria inviabilizando a realização da inferência. Uma solução utilizada, será no lugar de usar uma *priori* imprópria, utilizar uma distribuição própria difusa que as aproximem, por exemplo usar distribuição normal com variância grande.



A distribuição *a posteriori* $P(\theta|Y)$ é a descrição completa do conhecimento atualizado de θ obtido da quantificação da informação *priori* e da informação amostral em $L(\theta|Y) \propto f(Y|\theta)$ dada pela expressão

$$P(\theta|Y) = \frac{L(\theta|Y)P(\theta)}{\int_{\Theta} L(\theta|Y)P(\theta)d\theta}, \quad \theta \in \Theta.$$

O denominador $P(Y)$ representa a distribuição marginal de Y . Dentro da abordagem Bayesiana existem dois caminhos para se estudar os parâmetros: inferência e teoria da decisão.

A **inferência** consiste em plotar a distribuição *a posteriori* em determinar características como média, moda e mediana da distribuição.

Pode-se desejar estimar os parâmetros por estimadores pontuais ou por intervalos ou ainda testar hipóteses sobre os parâmetros.

A **teoria da decisão** utiliza a função perda $L(d, \phi)$, a qual mede a perda quando d é a decisão e ϕ é o valor do parâmetro. A perda d^* que minimiza a perda média $E[L(d, \phi)]$ com respeito *a posteriori* de ϕ é chamada decisão de Bayes.

3.1 Estimação pontual

A escolha da estimativas pontuais de θ depende naturalmente da forma de $P(\theta|Y)$, assim como dos objetivos do seu uso.

As estimativas mais usadas são moda *a posteriori*, média *a posteriori* e a mediana *a posteriori* (para parâmetro escalar).

Seja $\theta = (\theta_1, \dots, \theta_R)$, ENTÃO

- a) Vetor das medianas *a posteriori* $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_R)$ tal que $P(\theta_i \geq \hat{\theta}_i|Y) \geq 1/2$ e $P(\theta_i \leq \hat{\theta}_i|Y) \leq 1/2$, $i = 1, \dots, R$.
- b) Média *a posteriori* $\hat{\theta} = E[\theta|Y]$ com $E[\theta_i|Y] = \int_{\Theta} \theta_i P(\theta|Y) d\theta$, $i = 1, \dots, R$.
- c) Moda *a posteriori* $\tilde{\theta}$ tal que $P(\tilde{\theta}|Y) = \max_{\theta \in \Theta} P(\theta|Y) = \max_{\theta \in \Theta} \{P(\theta)L(\theta|Y)\}$.

A definição da moda *a posteriori* mostra que:

- 1) basta conhecer o núcleo da distribuição *a posteriori*;
- 2) se $P(\theta)$ é constante, coincide (aproximadamente) com a estimativa de máxima verossimilhança. Ou seja, permite interpretar o valor de θ que tem mais **credibilidade *a posteriori*** e não o valor de θ que torna a amostra mais plausível, que é a interpretação clássica.
- 3) O ponto máximo da distribuição conjunta dos θ_i não é necessariamente o vetor das modas *a posteriori* marginais dos θ_i , determinadas pela maximização separada de $P(\theta_i|Y) = \int P(\theta|Y) \prod_{j \neq i} d\theta_j$, $i = 1, \dots, R$, onde a integral é estendida no conjunto dos valores possíveis para o subvetor de θ obtido por remoção de θ_i .

A escolha de uma estimativa poderá basear-se na relevância de cada uma dessas quantidades para o problema e/ou na facilidade de seu cálculo. A escolha pode ser baseada na função perda e assim ter o risco mínimo *posteriori*.

De acordo com o critério de Bayes, θ^B é a estimativa de Bayes de θ se o seu risco *a posteriori* é o menor possível, isto é, $r(\theta^B|Y) = E[L(\theta, \theta^B)|Y] = \min_{\theta \in \Theta} E[L(\theta, \hat{\theta})|Y]$, sendo $L()$ a função perda. Ver os tipos de função perda em Paulino, Turkman e Murteira (2003, p. 140).

3.2 Estimativa por região

Um resumo de $P(\theta|Y)$ mais informativo que qualquer estimativa pontual é obtido de uma região de Θ que contenha uma parte substancial da massa probabilística *a posteriori*. O paradigma bayesiano de região de confiança.

Definição 3.1

$R(Y)$ é uma região de credibilidade γ para θ se

$$P(\theta \in R(Y)|Y) \equiv \int_{R(Y)} P(\theta|Y) d\theta \geq \gamma.$$

Dada a infinidade de região de credibilidade com o mesmo grau de γ , interessa selecionar aquele que contenha todos os valores de θ mais creíveis *a posteriori*, ou seja, aquele que não exclua nenhum valor mais plausível do que aqueles nela incluídos, satisfazendo assim a condição:

$$P(\theta_1|Y) \geq P(\theta_2|Y) \quad \forall \quad \theta_1 \in R(Y), \quad \theta_2 \notin R(Y).$$

Definição 3.2

$R(Y)$ é a região de credibilidade γ com densidade *a posteriori* máxima HPD (highest posterior density) se $\sup_{\theta \notin R(Y)} P(\theta|Y) \leq c_j \leq \inf_{\theta \in R(Y)} P(\theta|Y)$ para algum c_j tal que $P(\theta \in R(Y)|Y) \geq \gamma$ ou equivalentemente, se $R(Y) = \{\theta; P(\theta|Y) \geq c_j\}$ com $c_j > 0$ a maior constante tal que $\int_{R(Y)} P(\theta|Y) \lambda d\theta \geq \gamma$.

Exemplo 3.1: Pacote R para HPD - library(TeachingDemo)

Se a *a posteriori* do parâmetro é uma $Beta(4, 10)$ o HPD com $\alpha = 0,95$ é

```
hpd(qbeta, conf=0.95, shape1=4, shape2=10)
```

Se os valores são gerados de uma $Beta(4, 10)$,

```
theta=rbeta(10^5, 4, 10)
emp.hpd(theta, conf=0.95)
```

(HPD empírico)

3.3 Teste de hipótese

Problemas:

- Testar $H_0 : \theta \in \Theta_0$
- Contra $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$

O que se necessita é calcular as respectivas probabilidades *a posteriori* e caso se pretenda optar por uma delas em função de algum critério que indique sua grandeza relativa.

Esta é traduzida pela razão de vantagens *a posteriori* a favor de H_0 contra H_1 dada pela quantidade

$$O(H_0, H_1|Y) = \frac{P(H_0|Y)}{P(H_1|Y)}.$$

Com o objetivo de **medir a influência dos dados Y na alteração da credibilidade relativa** de H_0 e H_1 , o que se faz frequentemente é verificar a razão entre a razão das vantagens *a posteriori* e a razão das probabilidade *a priori*

$$O(H_0, H_1) = \frac{P(H_0|Y)}{P(H_1)}.$$

A razão dessas razões é denominada **Fator de Bayes** a favor de H_0 (contra H_1) e é denotada por

$$B(Y) = \frac{O(H_0, H_1|Y)}{O(H_0, H_1)}.$$

Um fator de Bayes muito grande ou muito pequeno relativamente a 1 (esse valor seria como um fator de comparação) representa uma tendência bastante forte nos dados a favor de uma hipótese contra a outra, no sentido que uma hipótese é muito mais ou muito menos favorável do que era *a priori*.

Pode-se adotar a regra de rejeição de H_0 se $P(H_0|Y) < P(H_1|Y)$ ou seja, se $O(H_0, H_1|Y) < 1$.

3.4 Comparação de modelos

Representa-se por $B = \{f_i(Y|\theta_i), P_i(\theta_i), i \in I\}$ um conjunto de modelos para dados e parâmetros.

Um caso particular, $B = \{f_i(Y|\theta), P_i(\theta), i \in I\}$, situação em que há um ou mais modelos para os dados dotados do mesmo parâmetro, variando apenas na forma distribucional, e uma ou mais distribuições *a priori*. Outro caso para $B = \{f_i(Y|\theta), P(\theta), i \in I\}$ é utilizado na análise de **sensibilidade ou de robustez** de inferência na variação da distribuição *a priori*.

A classe B mais geral de modelos dá origem a uma classe correspondente de modelos preditivos *a priori*, $\mu = \{M_i, i \in I\}$ cujas distribuições são definidas por

$$P(Y|M_i) = \int f_i(Y|\theta_i)P_i(\theta_i)d\theta_i, i \in I.$$

Sendo I um conjunto, descrito que contém o verdadeiro modelo, a distribuição preditiva global corresponde a

$$P(Y) = \sum_i P(M_i)P(Y|M_i),$$

onde $P(M_i)$ é a probabilidade *a priori* de M_i ser o modelo verdadeiro que é, pois atualizada em face aos dados Y por

$$P(M_i|Y) = \frac{P(M_i)P(Y|M_i)}{P(Y)}, i \in I.$$

O fator de Bayes a favor de M_i contra M_j é a razão de chances

$$\begin{aligned} B_{ij} &= \frac{\left(\frac{P(M_i|Y)}{P(M_i)}\right)}{\left(\frac{P(M_j|Y)}{P(M_j)}\right)}, \\ &= \frac{\left(\frac{P(M_i|Y)}{P(M_j|Y)}\right)}{\left(\frac{P(M_i)}{P(M_j)}\right)}, \\ &= \frac{P(Y|M_i)}{P(Y|M_j)}. \end{aligned}$$

3.5 Inferência Bayesiana para um modelo linear (PAULINO; TURKMAN; MURTEIRA, 2003, p. 216)

Seja $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, onde $\boldsymbol{\theta}_{(p \times 1)}$ é o vetor de parâmetros $\mathbf{Y} = (y_1, \dots, y_n)'_{(n \times 1)}$ vetor de observações, $\mathbf{X}_{(n \times p)}$ matriz conhecida e $\boldsymbol{\epsilon} \sim N(0, \tau^{-1}\mathbf{I}_n)$ sendo $\mathbf{P} = \tau^{-1}\mathbf{I}_n$ matriz de precisão de $\boldsymbol{\epsilon}_{(n \times 1)} = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]'$, a qual tem covariância $\sigma^2\mathbf{I}_n$ e $\sigma^2 = \tau^{-1} > 0$ desconhecida. O objetivo é fazer inferência sobre $\boldsymbol{\theta}$ e τ depois de observar a amostra $\mathbf{Y} = \{y_1, \dots, y_n\}$.

A palavra inferência é usualmente empregada como um procedimento que extrai informação sobre $\boldsymbol{\theta}$ e τ de uma amostra.

3.5.1 Teorema de Bayes

A informação *a priori* sobre $\boldsymbol{\theta}$ e τ é dada pela distribuição conjunta *a priori* $P(\boldsymbol{\theta}, \tau)$, $\boldsymbol{\theta} \in \mathbb{R}^p$ e $\tau > 0$.

A função de verossimilhança para $\boldsymbol{\theta}$ e τ é

$$L(\boldsymbol{\theta}, \tau|\mathbf{Y}) \propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \right\}.$$

A função de verossimilhança é uma informação da amostra sobre os parâmetros e é a função densidade condicional de uma amostra de variáveis aleatórias dado $\boldsymbol{\theta}$ e τ .

O teorema de Bayes nos dá a distribuição de $\boldsymbol{\theta}$ e τ dado \mathbf{Y} ,

$$P(\boldsymbol{\theta}, \tau|\mathbf{Y}) \propto L(\boldsymbol{\theta}, \tau|\mathbf{Y})P(\boldsymbol{\theta}, \tau),$$

que é a *posteriori* de $\boldsymbol{\theta}$ e τ , representando nosso conhecimento de $\boldsymbol{\theta}$ e τ depois de observar \mathbf{Y} .

3.5.2 Informação *a priori*

A informação *a priori* sobre os parâmetros $\boldsymbol{\theta}$ e τ é dada de duas maneiras: *priori* hierárquica e *prioris* independentes.

3.5.2.1 *Priori* hierárquica

$$P(\boldsymbol{\theta}, \tau) = P(\boldsymbol{\theta}|\tau)P(\tau),$$

é uma normal-gama, isto é,

$$P(\boldsymbol{\theta}|\tau) \propto \tau^{p/2} \exp \left\{ -\frac{\tau}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{P} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\},$$

ou seja,

$$\boldsymbol{\theta}|\tau \sim N_p(\boldsymbol{\mu}, (\tau \mathbf{P})^{-1}),$$

com $\boldsymbol{\mu}_{(p \times 1)}$ - vetor de médias e $\mathbf{P}_{(p \times p)}$ - matriz positiva definida, e

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

isto é,

$$\tau \sim \text{Gama}(\alpha, \beta).$$

3.5.2.2 *Prioris* independentes

$$P(\boldsymbol{\theta}, \tau) = P(\boldsymbol{\theta})P(\tau)$$

Neste caso, vamos considerar

$$P(\tau) \propto \frac{1}{\tau} (\text{priori de Jeffreys})$$

e

$$P(\boldsymbol{\theta}) \propto \text{constante}.$$

A *priori* de Jeffreys satisfaz certas regras de invariância e leva “pouca” informação sobre os parâmetros. Essa “densidade” embora imprópria, produz uma *posteriori* para $\boldsymbol{\theta}$ e τ que é própria (normal-gama). Assim, uma *priori* imprópria pode resultar numa densidade *a posteriori* própria, mas não é o que ocorre em geral.

3.5.3 Análise *a posteriori*

Usando o teorema de Bayes para a *priori* hierárquica, temos

$$\begin{aligned} P(\boldsymbol{\theta}, \tau|\mathbf{Y}) &= L(\boldsymbol{\theta}, \tau|\mathbf{Y})P(\boldsymbol{\theta}|\tau)P(\tau) \\ &\propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \right\} \times \\ &\quad \times \tau^{p/2} \exp \left\{ -\frac{\tau}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{P} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\} \tau^{\alpha-1} \exp\{-\beta\tau\} \\ &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [2\beta + (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) + (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{P} (\boldsymbol{\theta} - \boldsymbol{\mu})] \right\}. \end{aligned}$$

Vamos encontrar $P(\boldsymbol{\theta}|\tau, \mathbf{Y})$, $P(\tau|\mathbf{Y})$ e $P(\boldsymbol{\theta}|\mathbf{Y})$. Completando o quadrado para $\boldsymbol{\theta}$, obtemos:

$$\begin{aligned}
 P(\boldsymbol{\theta}, \tau | \mathbf{Y}) &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [2\beta + \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\mu} - \right. \\
 &\quad \left. - \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\theta} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu}] \right\} \\
 &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} (2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} + [\boldsymbol{\theta}'(\mathbf{X}'\mathbf{X} + \mathbf{P})\boldsymbol{\theta} - \boldsymbol{\theta}'(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}) - \right. \\
 &\quad \left. - (\mathbf{Y}'\mathbf{X} + \boldsymbol{\mu}'\mathbf{P})\boldsymbol{\theta}]) \right\} \\
 &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} + \right. \\
 &\quad \left. + (\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}))'(\mathbf{X}'\mathbf{X} + \mathbf{P})(\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})) - \right. \\
 &\quad \left. - \underbrace{(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}_{\text{Essa foi a parte que sobrou quando completamos o quadrado}} \right\},
 \end{aligned}$$

que é uma normal-gama. A condicional dado que se conhece τ é

$$\begin{aligned}
 P(\boldsymbol{\theta} | \tau, \mathbf{Y}) &\propto \tau^{p/2} \exp \left\{ -\frac{\tau}{2} [(\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}))'(\mathbf{X}'\mathbf{X} + \mathbf{P}) \times \right. \\
 &\quad \left. \times (\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})) \right\},
 \end{aligned}$$

sendo $\boldsymbol{\theta} | \tau, \mathbf{Y} \sim N [(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}), (\tau(\mathbf{X}'\mathbf{X} + \mathbf{P}))^{-1}]$, e a distribuição condicional $P(\tau | \boldsymbol{\theta}, \mathbf{Y})$ é

$$\begin{aligned}
 P(\tau | \boldsymbol{\theta}, \mathbf{Y}) &\propto \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} + \right. \\
 &\quad \left. + (\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}))'(\mathbf{X}'\mathbf{X} + \mathbf{P})(\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})) - \right. \\
 &\quad \left. - (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}) \right\}
 \end{aligned}$$

sendo

$$\begin{aligned}
 \tau | \boldsymbol{\theta}, \mathbf{Y} &\sim \text{Gama} \left(\frac{n+2\alpha}{2}, \frac{2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} + (\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}))' \times \right. \\
 &\quad \left. \frac{\times (\mathbf{X}'\mathbf{X} + \mathbf{P})(\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})) -}{2} \right. \\
 &\quad \left. \frac{- (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}{2} \right).
 \end{aligned}$$

$$\begin{aligned}
 P(\tau | \mathbf{Y}) &\propto \int_{-\infty}^{\infty} P(\boldsymbol{\theta}, \tau | \mathbf{Y}) d\boldsymbol{\theta} \\
 &\propto \int_{-\infty}^{\infty} \tau^{p/2} \exp \left\{ -\frac{\tau}{2} [(\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}))'(\mathbf{X}'\mathbf{X} + \mathbf{P})(\boldsymbol{\theta} - (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1} \times \right. \\
 &\quad \left. \times (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})) \right\} d\boldsymbol{\theta} \times \\
 &\quad \times \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} - (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})] \right\} \\
 &\propto \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} - (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})] \right\}
 \end{aligned}$$

sendo $\tau | \mathbf{Y} \sim \text{Gama} \left(\frac{n+2\alpha}{2}, \frac{2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} - (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}{2} \right)$.

A marginal *a posteriori* de $\boldsymbol{\theta}$, $\boldsymbol{\theta}|\mathbf{Y}$ é encontrada por integração de $P(\boldsymbol{\theta}, \tau|\mathbf{Y})$ em relação a τ , isto é

$$\begin{aligned} P(\boldsymbol{\theta}, \tau|\mathbf{Y}) &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} \left[\underbrace{2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} - (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}_{F_{\boldsymbol{\theta}}} + \right. \right. \\ &\quad \left. \left. + (\boldsymbol{\theta} - \underbrace{(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}_{\boldsymbol{\mu}_{\boldsymbol{\theta}}})' \underbrace{(\mathbf{X}'\mathbf{X} + \mathbf{P})}_{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}} \boldsymbol{\theta} - \underbrace{(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}_{\boldsymbol{\mu}_{\boldsymbol{\theta}}} \right] \right\} \\ &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [F_{\boldsymbol{\theta}} + (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})] \right\} \end{aligned}$$

Completando a expressão acima para obtermos uma distribuição Gama(λ, ω), vamos considerar $\lambda = \frac{n+p+2\alpha}{2}$ e $\omega = F_{\boldsymbol{\theta}} + (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})$. Assim,

$$P(\boldsymbol{\theta}, \tau|\mathbf{Y}) = \frac{\Gamma(\lambda) \omega^{\lambda}}{\Gamma(\lambda) \omega^{\lambda}} \tau^{\lambda-1} \exp \left\{ -\frac{\tau}{2} \omega \right\}.$$

Integrando em relação a τ , temos

$$\begin{aligned} P(\boldsymbol{\theta}|\mathbf{Y}) &= \int_0^{\infty} P(\boldsymbol{\theta}, \tau|\mathbf{Y}) d\tau \\ &= \int_0^{\infty} \frac{\Gamma(\lambda) \omega^{\lambda}}{\Gamma(\lambda) \omega^{\lambda}} \tau^{\lambda-1} \exp \left\{ -\frac{\tau}{2} \omega \right\} d\tau \\ &= \frac{\Gamma(\lambda)}{\omega^{\lambda}} \underbrace{\int_0^{\infty} \frac{\omega^{\lambda}}{\Gamma(\lambda)} \tau^{\lambda-1} \exp \left\{ -\frac{\tau}{2} \omega \right\} d\tau}_{=1} \\ &= \frac{\Gamma(\lambda)}{\omega^{\lambda}} \\ &\propto \frac{1}{\omega^{\lambda}}. \end{aligned}$$

Sabendo que $\lambda = \frac{n+p+2\alpha}{2}$ e $\omega = F_{\boldsymbol{\theta}} + (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})$, então

$$\begin{aligned} P(\boldsymbol{\theta}|\mathbf{Y}) &\propto [F_{\boldsymbol{\theta}} + (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})]^{-\left(\frac{n+p+2\alpha}{2}\right)}, \quad \nu = n + 2\alpha, \\ &\propto \left[\frac{F_{\boldsymbol{\theta}}}{F_{\boldsymbol{\theta}}} \right]^{-\left(\frac{\nu+p}{2}\right)} [F_{\boldsymbol{\theta}} + (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})]^{-\left(\frac{\nu+p}{2}\right)} \\ &\propto [F_{\boldsymbol{\theta}}]^{-\left(\frac{\nu+p}{2}\right)} \left[\frac{F_{\boldsymbol{\theta}}}{F_{\boldsymbol{\theta}}} + \frac{\nu}{\nu} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}^{-1})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) \right]^{-\left(\frac{\nu+p}{2}\right)} \\ &\propto \left[1 + \frac{1}{\nu} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})'(\boldsymbol{\Sigma}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}^{-1} \nu)(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) \right]^{-\left(\frac{\nu+p}{2}\right)}, \quad \mathbf{T} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}^{-1} \nu \\ &\propto \left[1 + \frac{1}{\nu} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})' \mathbf{T} (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) \right]^{-\left(\frac{\nu+p}{2}\right)} \end{aligned}$$

Assim, $\boldsymbol{\theta}|\mathbf{Y} \sim t$ de Student de dimensão p , com $\nu = n + 2\alpha$ graus de liberdade, média $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ e matriz de precisão \mathbf{T} .

Teorema 3.1

Se θ e τ são parâmetros do modelo linear $\mathbf{Y} = \mathbf{X}\theta + \epsilon$, em que $\mathbf{X}'\mathbf{X}$ tem posto completo e se a priori de θ e τ é a priori de Jeffreys $P(\theta, \tau) \propto \frac{1}{\tau}$, então a posteriori de θ e τ é uma normal-gama, em que a marginal a posteriori de τ é uma gama com parâmetros $\frac{n}{2}$ e $\frac{(\mathbf{Y}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})\mathbf{X}'\mathbf{Y})}{2}$ e a posteriori condicional de θ dado τ é uma normal multivariada com média $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ e matriz de precisão $[\tau(\mathbf{X}'\mathbf{X})^{-1}]$. Além disso, a marginal a posteriori de θ tem distribuição t - p dimensional com $n - p$ graus de liberdade, média $E[\theta|\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ e matriz de precisão $D(\theta) = (n - p)^{-1}(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}]$.

3.5.4 Algumas considerações

A estimação pontual de θ e τ será baseada na distribuição a posteriori conjunta de θ e τ .

Se θ é o parâmetro de interesse e τ é o parâmetro “sem importância” a estimativa de θ será dada pela marginal a posteriori de θ . Da mesma maneira, se τ é o parâmetro de interesse, θ é um parâmetro “sem importância”, a estimativa de τ será feita a partir da marginal a posteriori de τ .

A conjunta a posteriori de τ e θ é uma normal-gama com parâmetros

$$\alpha^* = \frac{n + 2\alpha}{2}, \quad \beta^* = \frac{2\beta + \mathbf{Y}'\mathbf{Y} + \boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu} - (\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})'(\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu})}{2},$$

$$\boldsymbol{\mu}^* = (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{P}\boldsymbol{\mu}), \quad \text{e} \quad \mathbf{P}^* = (\tau(\mathbf{X}'\mathbf{X} + \mathbf{P}))^{-1}.$$

A marginal a posteriori de τ é uma Gama(α^* , β^*). Como a distribuição gama é assimétrica, se vai ser utilizado a média ou a moda é uma escolha pessoal.

No caso da marginal a posteriori de θ com distribuição simétrica com média $\boldsymbol{\mu}_\theta$ e a matriz de precisão \mathbf{T} . Neste caso, a média, moda e mediana coincidem. Dessa forma, $\boldsymbol{\mu}_\theta$ é uma escolha natural para estimar θ .

Observação:

- (1) Não interessa as probabilidades amostrais do estimador, já que a distribuição a posteriori dos parâmetros é condicionado a valores da amostra \mathbf{Y} .
- (2) Propriedades amostrais dos estimadores como variância, erro quadrático médio, desvio, correlação, etc., não serão buscados.

Existem conexões interessantes entre os estimadores utilizados pelos não Bayesianos em certos momentos da posteriori de θ e τ quando é utilizada a priori de Jeffreys.

Por exemplo, um estimador de mínimos quadrados para θ é $\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ o qual tem distribuição $N(\mathbf{0}, [\tau(\mathbf{X}'\mathbf{X})]^{-1})$. Entretanto, a marginal a posteriori de θ tem distribuição t de Student com média $\hat{\theta}$ e matriz de precisão $(n - p)^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{Y} - \mathbf{X}\hat{\theta})'(\mathbf{Y} - \mathbf{X}\hat{\theta})$.

Em outras palavras, o estimador de mínimos quadrados é a média da distribuição marginal a posteriori de θ e a matriz de precisão da distribuição amostral de θ .

A inferência Bayesiana utilizando a priori de Jeffreys é mais ou menos tolerada pelos não-Bayesianos dada a proximidade entre os estimadores. Entretanto, a média que se utiliza *prioris* próprias, os estimadores são vistos com mais suspeita desde que eles diferem mais e mais dos estimadores tradicionais.

3.6 Análise preditiva

A análise preditiva é desenvolvida para fazer previsão de observações futuras. Naturalmente é de grande importância para economia e finanças onde são usados sofisticadas técnicas de séries temporais.

Os Bayesianos usam a densidade preditiva Bayesiana para obterem valores (observações) futuras. A abordagem Bayesiana para fazer previsão é baseada na distribuição condicional de valores futuros

dado o passado. Por exemplo: se $\omega = (\omega_1, \omega_2, \dots, \omega_k)'$ são observações futuras e $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ é uma amostra. Então a densidade preditiva é $P(\omega|\mathbf{Y})$.

Considerando o modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \epsilon$, a densidade de $\omega|\boldsymbol{\theta}, \tau, \mathbf{Y}$ é a função de verossimilhança para ω . A *posteriori* conjunta de $\omega, \boldsymbol{\theta}$ e τ dado \mathbf{Y} é

$$\underbrace{P(\omega, \boldsymbol{\theta}, \tau|\mathbf{Y})}_{\text{Post. conj. p/ } \omega, \boldsymbol{\theta} \text{ e } \tau} \propto \underbrace{f(\omega|\boldsymbol{\theta}, \tau, \mathbf{Y})}_{\text{verossim. p/ } \omega} \times \underbrace{P(\boldsymbol{\theta}, \tau|\mathbf{Y})}_{\text{Post. p/ } \boldsymbol{\theta} \text{ e } \tau}.$$

Observe que ω é considerado como um parâmetro a ser estimado. Integrando $P(\omega, \boldsymbol{\theta}, \tau|\mathbf{Y})$ em relação a $\boldsymbol{\theta}$ e τ , obtém-se a densidade preditiva para $\omega|\mathbf{Y}$. Isto é,

$$P(\omega|\mathbf{Y}) = \int_{\boldsymbol{\theta}} \int_{\tau} P(\omega, \boldsymbol{\theta}, \tau|\mathbf{Y}) d\boldsymbol{\theta} d\tau.$$

Inserir a dedução para o caso de uma *posteriori* normal-gama. Ver Paulino, Turkman e Murteira (2003) pág.177-178.

Aproximações Analíticas e Numéricas



“Is this needed for a Bayesian analysis?”

Esse material é referente ao capítulo 5 de (PAULINO; TURKMAN; MURTEIRA, 2003, p. 223). Como já foi dito, toda a inferência sobre um parâmetro θ é feita na análise Bayesiana através do Teorema de Bayes,

$$P(\theta|\mathbf{Y}) = \frac{L(\theta|\mathbf{Y})P(\theta)}{P(\mathbf{Y})}.$$

Inferências podem ser obtidas na forma de valores esperados *a posteriori*

$$E[g(\theta)|\mathbf{Y}] = \int_{\Theta} g(\theta)P(g(\theta)|\mathbf{Y})d\theta,$$

para escolhas apropriadas de $g(\cdot)$.

Uma alternativa pontual para θ , por exemplo, é obtida fazendo $g(\theta) = \theta$. Da mesma forma, se w for uma observação futura de um modelo indicado pelo mesmo parâmetro θ , inferências preditivas sobre w são baseados na distribuição preditiva

$$P(W|\mathbf{Y}) = \int L(W|\theta, \mathbf{Y})P(\theta|\mathbf{Y})d\theta,$$

onde $L(W|\theta, \mathbf{Y})$ é a distribuição de W sob o modelo paramétrico considerado, o qual pode ser ou não estocasticamente dependente de \mathbf{Y} .

Predições sumárias podem ser obtidas na forma de esperança preditiva

$$E[g(W)|\mathbf{Y}] = \int g(W)P(W|\mathbf{Y})d\mathbf{Y},$$

para escolhas adequadas de $g(\cdot)$.

Assim, no caso contínuo, a geração de integração desempenha um papel fundamental na estatística Bayesiana. Contudo, raramente é possível obter expressão explícita para as integrais.

Várias estratégias foram seguidas para solucionar esses problemas técnicos, entre elas:

- (1) Aproximação da distribuição *a posteriori* por uma distribuição multivariada;
- (2) Abordagem de Laplace;
- (3) Métodos de quadratura numérica (integração numérica);
- (4) Método Monte Carlo;
- (5) Método Monte Carlo via na Cadeia de Markov (MCMC).

(1) e (2) são métodos analíticos e (3), método numérico.

4.1 Métodos analíticos

4.1.1 Aproximação da distribuição *a posteriori* por uma distribuição multivariada

Para valores grandes de n e sob certas condições de regularidade, a distribuição *a posteriori* de um vetor $\boldsymbol{\theta}$ k -dimensional aproximadamente normal multivariada é dada por:

$$\begin{aligned} P(\boldsymbol{\theta}|\mathbf{Y}) &\propto \exp\{\ln(P(\mathbf{Y}|\boldsymbol{\theta})P(\boldsymbol{\theta}))\} \\ &\propto \exp\{\ln(P(\mathbf{Y}|\boldsymbol{\theta})) + \ln(P(\boldsymbol{\theta}))\}. \end{aligned}$$

Desenvolvendo $P(\mathbf{Y}|\boldsymbol{\theta})$ e $P(\boldsymbol{\theta})$ em série de Taylor de 2^a ordem em torno dos respectivos máximos (\mathbf{m}_0), temos:

$$\ln(P(\boldsymbol{\theta})) = \ln(P(\mathbf{m}_0)) - \frac{1}{2}(\boldsymbol{\theta} - \mathbf{m}_0)'\mathbf{H}_0(\boldsymbol{\theta} - \mathbf{m}_0) + R_0,$$

e

$$\ln(P(\mathbf{Y}|\boldsymbol{\theta})) = \ln(P(\mathbf{Y}|\hat{\boldsymbol{\theta}}_n)) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)'\mathbf{H}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) + R_\theta,$$

em que \mathbf{m}_0 é a moda da distribuição *a priori*, $\hat{\boldsymbol{\theta}}_n$ é a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ baseada nos dados \mathbf{Y} , $\mathbf{H}_0 = [-\partial^2 \ln(P(\boldsymbol{\theta})) / (\partial\theta_i \partial\theta_j)] \Big|_{\boldsymbol{\theta}=\mathbf{m}_0}$, $\mathbf{H}(\hat{\boldsymbol{\theta}}_n) = [-\partial^2 \ln(P(\mathbf{Y}|\boldsymbol{\theta})) / (\partial\theta_i \partial\theta_j)] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}$, R_0 e R_n , os restos respectivos desenvolvimentos em série.

Assumindo n grande, temos

$$P(\boldsymbol{\theta}|\mathbf{Y}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \mathbf{m}_n)'\mathbf{H}_n(\boldsymbol{\theta} - \mathbf{m}_n)\right\},$$

para $\mathbf{H}_n = \mathbf{H}_0 + \mathbf{H}(\hat{\boldsymbol{\theta}}_n)$ e $\mathbf{m}_n = \mathbf{H}_n^{-1}(\mathbf{H}_0\mathbf{m}_0 + \mathbf{H}(\hat{\boldsymbol{\theta}}_n)\hat{\boldsymbol{\theta}}_n)$.

Este desenvolvimento sugere que a distribuição *a posteriori* possa ser aproximada, para amostras de dimensão elevada e em condições gerais de regularidade, por uma distribuição Normal multivariada com valor médio \mathbf{m}_n , e matriz de covariâncias \mathbf{H}_n^{-1} , que pode ser aproximada pelo inverso da matriz de informação observada ($N(\mathbf{m}_n, \mathbf{H}_n^{-1})$).

À medida que a dimensão da amostra aumenta é de se esperar que a precisão da distribuição *a priori*, representada por \mathbf{H}_0 , seja totalmente dominada pela precisão $\mathbf{H}(\hat{\boldsymbol{\theta}}_n)$ formada pelos dados, ou seja, $\mathbf{H}_n \approx \mathbf{H}(\hat{\boldsymbol{\theta}}_n)$. Assim, também $\mathbf{m}_0 \approx \hat{\boldsymbol{\theta}}_n$.

Uma outra alternativa é considerar o desenvolvimento de $P(\boldsymbol{\theta}|\mathbf{Y})$ em série de Taylor em torno de seu máximo, isto é, em torno da moda *a posteriori*.

Exemplo 4.1: Paulino, Turkman e Murteira (2003, p. 226)

Suponha que X tem uma distribuição Binomial de parâmetros (n, θ) e que se adotou para θ uma distribuição *a priori* conjugada Beta de parâmetros (a_0, b_0) . Sabe-se que a distribuição *a posteriori* para θ é ainda uma Beta de parâmetros (a_n, b_n) , em que $a_n = a_0 + x$ e $b_n = b_0 + n - x$, ou seja,

$$L(Y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y},$$

$$P(\theta) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \theta^{a_0-1} (1 - \theta)^{b_0-1}, \quad 0 \leq \theta \leq 1.$$

Assim,

$$P(\theta|Y) \propto \theta^{y+a_0-1} (1 - \theta)^{n+b_0-y-1}$$

$$\propto \theta^{a_n-1} (1 - \theta)^{b_n-1},$$

em que $a_n = y + a_0$ e $b_n = n + b_0 - y$. Portanto, $\theta|y \sim \text{Beta}(a_n, b_n)$.

$$L(\theta) = \ln(P(\theta|Y))$$

$$\propto (a_n - 1) \ln(\theta) + (b_n - 1) \ln(1 - \theta),$$

$$L'(\theta) \propto \frac{(a_n - 1)}{\theta} + \frac{b_n - 1}{1 - \theta} (-1)$$

$$\propto \frac{a_n - 1}{\theta} - \frac{b_n - 1}{1 - \theta},$$

$$L''(\theta) \propto -\frac{(a_n - 1)}{\theta^2} - \frac{(b_n - 1)}{(1 - \theta)^2},$$

e portanto

$$m_n = \frac{a_n - 1}{a_n + b_n - 1} \quad \text{e} \quad \{L''(m_n)\}^{-1} = \frac{(a_n - 1)(b_n - 1)}{(a_n + b_n - 2)^2},$$

que é obtido substituindo m_n em θ para $L''(\theta)$.

Assim, para n grande a distribuição *a posteriori* pode ser bem aproximada por uma normal com valor médio $\mu = m_n$ e $\sigma_n^2 = \{L''(m_n)\}^{-1}$.

4.1.2 Método de Laplace

Este método consiste em utilizar decomposição em séries para aproximar integrais da forma

$$E[g(\theta|\mathbf{Y})] = \int g(\theta) P(\theta|\mathbf{Y}) d\theta.$$

Suponha que Ψ é uma função regular de um parâmetro θ k -dimensional e que $-\Psi$ tem máximo em $\hat{\theta}$. O método de Laplace aproxima uma integral da forma

$$I = \int f(\theta) \exp \{-n\Psi(\theta)\} d\theta,$$

através do desenvolvimento em séries de Ψ e f em torno de $\hat{\theta}$. Em geral, é suficiente a decomposição até termos de 2ª ordem.

No caso de se considerar o valor esperado da *posteriori* de uma função $g(\boldsymbol{\theta})$ do parâmetro, tendo-se que

$$E[g(\boldsymbol{\theta}|\mathbf{Y})] = \frac{\int g(\boldsymbol{\theta})L(\mathbf{Y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{\int L(\mathbf{Y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}d\boldsymbol{\theta}.$$

O que se faz é aplicar separadamente o método de Laplace à integrais do numerador e do denominador e considerar a razão das aproximações obtidas.

Tierney e Kadane (1986) sugerem considerar $\exp\{-n\Psi(\boldsymbol{\theta})\} = L(\boldsymbol{\theta}|\mathbf{Y})P(\boldsymbol{\theta})$ e $f(\boldsymbol{\theta}) = g(\boldsymbol{\theta})$ no numerador e $f(\boldsymbol{\theta}) = 1$ no denominador.

4.2 Quadratura Gaussiana ou integração numérica (Método numérico)

A integração numérica é uma solução para resolver integrais para fazer inferência Bayesiana. Considere a integral

$$I = \int_a^b g(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Assim, a regra de quadratura aproxima I por $\hat{I} = \sum_{i=1}^n w_i g(\theta_i)$ para alguns pesos w_i e gride de pontos θ_i , $i = 1, 2, \dots, n$. A regra mais simples é tomar n pontos igualmente espaçados e $w_i = c = (b - a)/n$.

Para a maioria das integrais de dimensão 1, tomar n de ordem 10^2 é uma boa aproximação. Outras regras de quadratura são a regra do trapézio e a regra de Simpson.

A regra gaussiana foi desenvolvida e tabelado quando o integrando é bem aproximado pela forma $L(\boldsymbol{\theta})P(\boldsymbol{\theta})$, em que $L(\boldsymbol{\theta})$ é uma função polinomial de $\boldsymbol{\theta}$ e P é uma função densidade.

4.3 Métodos de Simulação

O termo simulação refere-se a tratamentos de problemas reais através de reprodução em ambientes controlados pelo pesquisador.

Em algumas situações, alguns ou todos os comportamentos estão sujeitos a comportamentos aleatórios, ou seja, não podem ser descritos de forma exata por uma regra matemática, mas apenas através do uso de informação probabilística.

Este processo de simulação é estocástico, isto é, baseado em distribuição de probabilidade.

Do ponto de vista da estatística, esses sistemas podem ser vistos como uma coleção de variáveis aleatórias. O ponto de partida para simulação estocástica é a construção de uma geração de números aleatórios.

4.3.1 Métodos baseados em simulação estocástica

A inferência baseada em técnicas de simulação utiliza amostra da *posteriori* P para extrair informação a ser repetida.

Obviamente, como uma amostra é sempre um substituto parcial da informação contida em uma densidade, métodos baseados em simulação são aproximados e devem ser usados apenas quando for constatada a impossibilidade de extração analítica de informação da *posteriori*.

Os métodos baseados em simulação não dependem de quantas observações foram tomadas e fornecem aproximação que serão tão melhores quanto maior for o número de valores gerados.

A única restrição para que seja atingido qualquer nível de precisão desejada é o custo computacional. Para a maioria dos problemas de relevância prática é complicado fazer uma geração da *posteriori* P .

Antes de descrever os métodos de obtenção de uma amostra da *posteriori*, vale a pena descrever como se procede a inferência Bayesiana na presença de uma amostra $\theta_1, \theta_2, \dots, \theta_n$ de uma densidade $P(\boldsymbol{\theta}|\mathbf{Y})$ de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$.

Para se obter uma amostra marginal, a i -ésima componente de $\boldsymbol{\theta}$, basta tomar as i -ésimas componentes dos valores gerados.

$\theta_{i1}, \theta_{i2}, \dots, \theta_{in}$ constituindo a amostra marginal de $P(\theta_i|Y)$. Essa eliminação de θ_{-i} é trivial e extremamente mais simples que a operação de integração necessária para a obtenção analítica da marginal.

De posse de uma amostra é possível obter estimativas pontuais, intervalos de credibilidade é até mesmo reconstruir densidades marginais univariadas e bivariadas usando técnicas gráficas ou de suavização.

Exemplo 4.2

Um exemplo é $E[\theta_i|Y] = \frac{1}{n} \sum_{j=1}^n \theta_{ij}$, em que é a média da i -ésima componente dos valores amostrados.

Outro exemplo é a mediana amostral $(\theta_i|Y) = \theta_{[\frac{n+1}{2}],i}$, da i -ésima componente dos valores amostrados, onde $[m]$ é o menor inteiro maior ou igual a m .

Intervalo de confiança $100(1 - \alpha)\%$ para θ_i $[\theta_{[n\alpha_1],i}, \theta_{[n(1-\alpha_2)],i}]$ com $\alpha_1 + \alpha_2 = \alpha$. Em particular, regiões equiprováveis são deixados de fora tomando $\alpha_1 = \alpha_2 = \alpha/2$.

4.3.2 Métodos de Monte Carlo Ordinário

Considere o problema de aproximar uma integral da forma

$$\int g(\theta)P(\theta|Y)d\theta = E[g(\theta|Y)],$$

em que θ e Y podem ser vetores, admitindo sua existência.

Muitas quantidades *a posteriori* de interesse são dadas por esta função $g(\theta)$ integrável.

É o caso das médias e covariáveis *a posteriori* de componentes de θ , onde $g(\theta)$ é representada por θ_i e $(\theta_i - E[\theta_i|Y])(\theta_j - E[\theta_j|Y])$ para cada i e j fixados.

Se puder simular uma amostra aleatória $\theta_1, \dots, \theta_n$ da densidade *a posteriori* $P(\theta|Y)$, o método Monte Carlo ordinário aproxima a integral

$$\int g(\theta)P(\theta|Y)d\theta \approx \frac{1}{n} \sum_{i=1}^n g(\theta_i)$$

a qual pela Lei forte dos Grandes Números, converge quase certamente para $E[g(\theta|Y)]$.

A precisão desta aproximação pode ser medida pelo erro padrão (estimado) de Monte Carlo dado por

$$\frac{1}{\sqrt{n(n-1)}} \left\{ \sum_{i=1}^n \left(g(\theta_i) - \frac{1}{n} \sum_{i=1}^n g(\theta_i) \right)^2 \right\}^{1/2}.$$

O nível de precisão destas aproximações está sob o controle do analista que, teoricamente, pode aumentar tanto quanto queira da amostra simulada.

Resumindo, se conseguir amostrar da distribuição *a posteriori* $P(\theta|Y)$ a aplicação do método de Monte Carlo é trivial para resolver integrais do tipo $\int g(\theta)P(\theta|Y)d\theta$.

4.3.2.1 Preditiva

Sabendo que a densidade preditiva *a posteriori* é o valor esperado

$$P(W|Y) = E_{\theta|Y}(L(W|\theta, Y)),$$

facilmente se obtém a aproximação de Monte Carlo

$$\hat{P}(W|Y) = \frac{1}{n} \sum_{i=1}^n L(W|\theta_i, Y)$$

baseado nos valores *iid* simulados de $P(\theta|Y)$

$$E_{\theta|Y}(L(W|\theta_i, Y)) = \int L(W|\theta_i, Y)P(\theta|Y)d\theta.$$

Para a estimação Monte Carlo de quantidade associadas com a distribuição preditiva $P(W|Y)$, é necessário obter-se uma amostra aleatória desta distribuição. Isto é possível através do método de composição, caso saiba amostrar da distribuição amostral de W , obtendo-se a amostra (w_1, w_2, \dots, w_n) de $P(W|Y)$ do seguinte modo:

1. Retire uma amostra $\theta_1, \theta_2, \dots, \theta_n$ de $P(\theta|Y)$;
2. Para cada i , retire w_i de $L(W|\theta_i, Y)$, $i = 1, 2, \dots, n$.

A partir desta amostra, tem-se as informações necessárias.

$P(\theta Y)$	$L(W \theta, Y)$
θ_1	W_1 de $W \theta_1, Y$
θ_2	W_2 de $W \theta_2, Y$
\vdots	\vdots
θ_n	W_n de $W \theta_n, Y$

4.3.3 Monte Carlo com amostragem por importância

Geralmente não é possível obter uma amostra *iid* diretamente da distribuição *a posteriori* $P(\theta|Y)$ e assim, deve-se encontrar alternativas.

Uma das estratégias é simular de uma distribuição “semelhante” à distribuição *a posteriori*. Vamos definir a **amostragem na função de importância**.

Seja $h(\theta)$ uma densidade da qual seja fácil de simular valores e que aproximem

$$P(\theta|Y) = \frac{L(Y|\theta)P(\theta)}{\int L(Y|\theta)P(\theta)d\theta},$$

então

$$\begin{aligned} \int g(\theta)P(\theta|Y)d\theta &= \frac{\int g(\theta)L(Y|\theta)P(\theta)d\theta}{\int L(Y|\theta)P(\theta)d\theta} \\ &= \frac{\int g(\theta)\frac{L(Y|\theta)P(\theta)}{h(\theta)}d\theta \times h(\theta)}{\int \frac{L(Y|\theta)P(\theta)}{h(\theta)}d\theta \times h(\theta)} \\ &= \frac{\int g(\theta)w(\theta)h(\theta)d\theta}{\int w(\theta)h(\theta)d\theta}. \end{aligned}$$

Se obtiver uma amostra $\theta_1, \theta_2, \dots, \theta_n$ de $h(\theta)$, pode-se aplicar o método de Monte Carlo, obtendo-se como aproximação de $E[g(\theta)|Y] = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i g(\theta_i)$, onde $w_i = \underbrace{L(Y|\theta_i)P(\theta_i)}_{P(\theta|Y)} / h(\theta_i)$.

O método de amostragem via função de importância atribui mais peso a região onde $h(\theta) < P(\theta|Y)$ e menos peso a região onde $h(\theta) > P(\theta|Y)$.

O erro padrão de Monte Carlo para esta aproximação é dado por

$$\frac{1}{\sum_{j=1}^n w_j} \left[\sum_{i=1}^n \left\{ g(\theta_i) - \frac{1}{\sum_{j=1}^n w_j} \sum_{i=1}^n w_i g(\theta_i) \right\}^2 \right]^{1/2}.$$

A razão de convergência depende de quão bem $h(\theta)$ imita $P(\theta|Y)$. Boas propriedades de importância são:

- i) Simplicidade na geração de números aleatórios;
- ii) Ter caudas mais pesadas $P(\cdot|Y)$;
- iii) Ser uma boa aproximação para $P(\cdot|Y)$.

Note que para aplicar esta metodologia há a necessidade que $P(\theta|Y)$ seja conhecida a menos da constante de proporcionalidade, isto é basta conhecer $L(Y|\theta)P(\theta)$.

4.3.4 Método de Monte Carlo via Cadeia de Markov (MCMC)

O método de Monte Carlo baseado na simulação de cadeias de Markov baseia-se na distribuição estacionária que é a distribuição *a posteriori* de interesse, diretamente da qual não é possível gerar uma amostra para a inferência via Monte Carlo.

Assim, a ideia básica é transformar o problema estático em um problema de natureza dinâmica, construindo um processo temporal estocástico, artificial, que seja fácil de simular, e que convirja, fracamente para a distribuição original. Este processo temporal é, em geral, uma cadeia de Markov homogênea cuja distribuição de equilíbrio é a distribuição que se pretende simular.

O processo estocástico $\{X_t\}$ é uma cadeia de Markov se a distribuição de X_t for independente de todos os estados prévios em que se encontra a cadeia, com exceção do estado imediatamente anterior, isto é, $P(X_t = x_j | X_{t-1} = x_{i_{t-1}}, X_{t-2} = x_{i_{t-2}}, \dots, X_2 = x_{i_2}, X_1 = x_{i_1}) = P(X_t = x_j | X_{t-1} = x_{i_{t-1}})$. Quando esta probabilidade for independente de t , a cadeia diz-se homogênea.

Serão abordados dois métodos de Monte Carlo via cadeias de Markov: o amostrador de Gibbs e o algoritmo de Metropolis-Hastings.

4.3.4.1 Amostrador de Gibbs (AG)

O amostrador de Gibbs é uma técnica para gerar amostras aleatórias de uma distribuição (marginal) sem que se conheça a sua densidade. Embora o AG tenha sido aplicado em metodologias bayesianas, ele também é extremamente útil para se amostrar a função de verossimilhança nos métodos frequentistas.

Em modelos complicados, raramente se consegue obter amostras diretamente das distribuições *a posteriori*. A ideia do método de Monte Carlo via Cadeia de Markov (MCMC) é simular um processo aleatório no espaço do parâmetro θ , o qual converge para uma distribuição estacionária, que é a distribuição *a posteriori* $P(\theta|Y)$.

Pode-se pensar no AG como uma implementação prática de fato que o conhecimento das distribuições condicionais é suficiente para determinar a distribuição conjunta (se ela existir, que é o caso de condicionais impróprias que não garantem que a conjunta é uma fdp).

O algoritmo é baseado no fato que se a distribuição conjunta $P(\theta|Y)$ for positiva em $\theta_1 \times \theta_2 \times \dots \times \theta_k$, com θ_{-i} suporte da distribuição de θ_i para $i = 1, \dots, k$, então ela é unicamente determinada pelas distribuições condicionais completas $P(\theta_i | \theta_{-i}, Y)$ $i = 1, \dots, k$. O algoritmo é um esquema markoviano dinâmico que requer a amostragem dessas distribuições condicionais.

Descrição do Algoritmo Gibbs:

Suponhamos que o vetor de parâmetros θ seja dividido em k subvetores $(\theta_1, \theta_2, \dots, \theta_k)'$ e que as distribuições condicionais de cada parâmetro θ_i dado todos os outros, seja conhecida. Essas distribuições são chamadas distribuições condicionais completas e denotadas por

$$\begin{aligned}
 &P(\theta_1 | \theta_2, \dots, \theta_R, Y) \\
 &P(\theta_2 | \theta_1, \theta_3, \dots, \theta_R, Y) \\
 &\quad \vdots \\
 &P(\theta_R | \theta_1, \theta_2, \dots, \theta_{R-1}, Y).
 \end{aligned}$$

Seja $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_R^{(0)})$ um valor arbitrário para o vetor $\boldsymbol{\theta}$. Proceda-se da seguinte forma:

1. Obtém-se

$$\begin{aligned} \theta_1^{(1)} &\text{ de } P(\theta_1^{(1)} | \theta_2^{(0)}, \dots, \theta_R^{(0)}, Y) \\ \theta_2^{(1)} &\text{ de } P(\theta_2^{(1)} | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_R^{(0)}, Y) \\ &\vdots \\ \theta_R^{(1)} &\text{ de } P(\theta_R^{(1)} | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{R-1}^{(1)}, Y). \end{aligned}$$

2. Completa-se assim uma iteração do esquema e uma transição de $\boldsymbol{\theta}^{(0)}$ para $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_R^{(1)})$.

3. O esquema é repetido M vezes produzindo $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(M)}$.

4. A sucessão $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(M)}, \dots$ é uma realização de uma cadeia de Markov.

Pode-se provar que quando $M \rightarrow \infty$, $(\theta_1^{(M)}, \theta_2^{(M)}, \dots, \theta_R^{(M)})$ tende em distribuição para um vetor aleatório cuja função densidade de probabilidade conjunta é $P(\boldsymbol{\theta}|Y)$.

Em particular, $\theta_i^{(M)}$ tende em distribuição para uma quantidade aleatória cuja densidade é $P(\theta_i|Y)$ (densidades marginal *a posteriori* de θ_i) e $\frac{1}{M} \sum_{i=1}^M g(\theta_i) \rightarrow E(g(\theta)|Y)$ quase certamente para qualquer função $g(\cdot)$, em que $E(g(\theta)|Y)$ representa o valor esperado de $g(\theta)$ em relação à distribuição *a posteriori* $P(\boldsymbol{\theta}|Y)$.

Uma questão muito natural é como se pode usar a amostra resultante do procedimento de Gibbs para fazer inferência sobre $\boldsymbol{\theta}$? Questões encontradas:

1. Obviamente os vetores $\boldsymbol{\theta}^{(M)}$ que vão sendo gerados são correlacionados. Como é que então se deve proceder para obter uma amostra *iid* da distribuição *a posteriori*? Como utilizar a amostra obtida para fazer inferência?
2. Como é que se pode controlar a convergência da cadeia de Markov para o estado de equilíbrio?
3. O que é que se pode dizer a velocidade de convergência para a distribuição de interesse e sobre a precisão das estimativas obtidas?
4. Finalmente, que fazer se uma (ou mais) das distribuições condicionais não forem de fácil simulação, ou se forem apenas conhecidas a menos de uma constante normalizadora?

Uma solução possível para a questão 1, é a construção de uma longa cadeia utilizando vetores espaçados, de modo a anular o efeito da correlação. Os problemas mais complicados dizem respeito ao estudo da convergência da cadeia de Markov para o estado de equilíbrio.

Entre várias propostas de análise de convergência temos:

- i) Gelfand et al (1990) propõe métodos adhoc baseados em representação gráfica das estimativas das densidades e na monitoração da convergência das médias ergóticas para o parâmetro de interesse.
- ii) Gilman e Rubin (1991) sugerem métodos baseados na análise de variância (cadeias múltiplas).
- iii) Raftery e Lewis (1992) apresentam fórmulas que relacionam a dimensão da cadeia de Markov a ser construído o espaçamento entre os vetores e a dimensão da amostra a ser utilizada.

Não há no entanto, um método que se possa dizer ser o melhor ou o mais eficiente e a utilização de métodos diferentes para o mesmo problema pode conduzir a respostas diferentes.

Respondendo a questão 2, após um número suficientemente grande M de iteração, forma-se uma amostra de $\boldsymbol{\theta}$ e pode-se construir um histograma de qualquer um de seus componentes. O mesmo procedimento pode ser repetido após $M + k$ iterações. Não havendo diferenças perceptíveis a olho nu

entre os gráficos obtidos com relação as M e $M + k$ iterações, conclui-se pela convergência da cadeia. Lembrando que o valor k não pode ser muito pequeno, pois a correlação inerente a cadeia de Markov estará exercendo sua influência e não se poderá dizer que a similaridade é devida a convergência ou à correlação do processo.

Valores de k muito grandes são desnecessários, pois se há suspeita de convergência após M iterações, não há necessidade de se ir muito além na cadeia apenas para sua confirmação. Valores para k entre 10 e 50 são apropriados.

Um outro gráfico é o traço da cadeia, que dá uma ideia se a cadeia atingiu o equilíbrio.

4.3.4.2 Algoritmo de Metropolis-Hastings

O amostrador de Gibbs seleciona amostras aleatórias de todas as distribuições condicionais obtidas da distribuição *a posteriori* conjunta $P(\theta|Y)$. Se for considerado as distribuições *a priori* conjugadas, as distribuições são marginal, da forma Gama, Normal, Beta, Poisson, etc. e as amostras geradas dessas distribuições são facilmente obtidas, pois estão disponíveis em muitos softwares.

Quando essas distribuições não são identificáveis ou se a geração não iterativa for complicada ou custosa (computacionalmente), usamos o algoritmo de Metropolis-Hastings.

Seja $P(\theta|Y)$ uma distribuição não-identificada. O algoritmo de Metropolis-Hastings cria sequências de pontos aleatórios $(\theta^1, \theta^2, \dots)$ cuja distribuição converge para $P(\theta|Y)$. Cada sequência pode ser considerada como um passeio aleatório com distribuição estacionária $P(\theta|Y)$. O algoritmo de Metropolis-Hastings é especificado por uma densidade $q(\theta, \beta)$ denominada núcleo de transição. Essa densidade é selecionada de uma família de distribuições que requer a especificação de alguma característica do parâmetro, como escala ou posição.

O núcleo de transição $q(\theta, \beta)$ representa a probabilidade de movimento de θ para β . Geralmente, o processo se move com mais frequência de θ para β do que de β para θ , contradizendo, a condição de reversibilidade $P(\theta|Y)q(\theta, \beta) = P(\beta|Y)q(\beta, \theta)$. Esse fato pode ser corrigido, diminuindo o número de movimento de θ para β , e para isto introduz-se uma probabilidade de movimento p . O algoritmo é como segue:

- i) inicialize com um valor (arbitrário) θ^0 , $j = 0$;
- ii) gere β de $q(\theta^{(j)}, \cdot)$ e u de uma uniforme(0,1);
- iii) seja $p = \min \left\{ 1, \frac{P(\beta|Y)q(\theta^{(j)}, \beta)}{P(\theta^{(j)}|Y)q(\beta, \theta^{(j)})} \right\}$;
- iv) se $u \leq p$, faça $\theta^{(j+1)} = \beta$, senão $\theta^{(j+1)} = \theta^{(j)}$;
- v) Repita (ii) e (iii) até que a distribuição estacionária seja obtida.

O núcleo de transição q apenas define uma proposta de movimento, a qual deve (ou não) ser confirmada por p .

Para que o algoritmo Metropolis-Hastings seja implementado, necessitamos de uma densidade q adequada. Um fator crítico na eficiência do algoritmo é a escolha do fator de escala ou dispersão de q . O fator escala da densidade q afeta o comportamento da cadeia em pelo menos duas dimensões, uma é a taxa de aceitação e a outra é a região do espaço amostral que é coberta pela cadeia.

Para vermos isto, considere uma situação na qual a cadeia convergiu e a densidade esta sendo amostra em torno da moda. Se a dispersão da densidade q é muito grande, alguns pontos de β estarão longe dos valores atuais de θ e por isto terão probabilidade de serem aceitos.

Por outro lado, se o fator escala for muito pequeno a cadeia não cobrirá o suporte da densidade e região de baixa probabilidades serão amostradas.

Escolha de $q(\theta, \beta)$ (Chib and Greenberg,1995)

Várias são as possibilidades para a escolha de $q(\theta, \beta)$, uma que é uma solução eficiente, quando possível, é explorar a forma de $P(\theta|Y)$ para especificar a densidade de q . Por exemplo, se $P(\theta|Y)$ pode ser escrita como $P(\theta|Y) \propto \psi(\theta|Y)h(\theta|Y)$ onde $h(\theta|Y)$ é uma densidade que pode ser amostrada e $\psi(\theta|Y)$ é uniformemente limitada. Então faça $h(\theta|Y) = q(\theta|\beta)$ para amostrar candidatos. Nesse caso, a probabilidade de movimento necessita apenas do valor de $\phi(\theta)$ e é dado por $p = \left\{1, \frac{\psi(\beta)}{\phi(\theta)}\right\}$.

Taxa de aceitação (Roberto, Gilman e Gilks, 1994)

Ex.: Se q é uma densidade normal, a taxa de aceitação é de 45% para problemas unidimensionais e 25% para $dim < 6$.

Amostragem Aceitação-rejeição (Acceptation - Rejection Sampling)

Em contraste aos métodos de MCMC, o método é uma técnica clássica para gera amostras não-Markov (usualmente independentes).

O objetivo é gerar amostras de uma densidade de interesse contínua $\pi = \frac{f(x)}{k}$, onde $x \in \mathbb{R}^n$, $f(x)$ é uma densidade não-normalizada e k é a constante normalizadora (possivelmente desconhecida).

Seja $h(x)$ uma densidade que pode ser simulada por algum método conhecido e suponha que exista uma constante c tal que $f(x) \leq ch(x)$, $\forall x$. Então, para se obter uma amostra aleatória de $\pi(\cdot)$,

- i) gere z de $h(\cdot)$ e um valor u de $U(0, 1)$;
- ii) se $u \leq \frac{f(z)}{ch(z)}$, faça $y = z$. Senão volte (i).

O valor aceito y é uma variável aleatória de $\pi(\cdot)$. Para este método ser eficiente, c deve ser cuidadosamente escolhido. O método é otimizado fazendo $c = \sup_x \frac{f(x)}{h(x)}$. Mesmo para esta escolha de c , pode-se obter um número grande de rejeição.

4.4 Critérios de convergência

O termo convergência está relacionado com a amostra da distribuição *a posteriori* gerada pela cadeia de Markov, e não com um número ou uma distribuição em específico. O processo antes de chegar a convergência, segundo Nogueira (2004) envolve simulações provindas de uma complexa e, geralmente, multivariada distribuição de interesse, $P(\theta)$, indiretamente gerado de uma Cadeia de Markov com a densidade de interesse, uma densidade estacionária. Geralmente são simuladas $m \geq 1$ sequências de simulações, cada uma de tamanho n , $(\theta_{j1}, \theta_{j2}, \dots, \theta_{jn})$, para $j = 1, \dots, m$ (sendo cada θ_{jt} um vetor). Se $m > 1$, as m sequências são usualmente, mas nem sempre, simuladas independentemente. Pode-se verificar que em muitos casos a convergência pode ser facilmente determinada provinda de sequências múltiplas independentes em paralelo, mas não pode ser diagnosticada usando o resultado da simulação provida de qualquer sequência simples. Daí a necessidade de usar critérios para para essa verificação.

4.4.1 Critério de Gelman e Rubin (1992)

O método pressupõe que m sequências tenham sido simuladas em paralelo, cada uma partindo de diferentes pontos iniciais. Tendo obtido os pontos iniciais satisfatórios, ou seja, pertencentes ao espaço paramétrico da *posteriori*, as sequências são geradas para $2n$ iterações, das quais as primeiras n são descartadas para evitar o período de aquecimento (“burn-in”), ou seja, para evitar a influência dos valores iniciais. As m sequências rendem m possíveis inferências. Se estas inferências são bastante similares, é indicativo de que a convergência foi alcançada ou está próxima.

A limitação principal para o método é a suposição de normalidade da distribuição marginal de cada parâmetro. Para mais detalhes, ver Nogueira (2004).

4.4.2 Critério de Raftery e Lewis (1992)

O critério sugerido por Raftery e Lewis (1992) é um método que estima quantas iterações são necessárias para que o Amostrador de Gibbs apresente convergência à distribuição estacionária, propondo especificações de quantidades necessárias para isto. Deve-se ter uma sequência do Amostrador de Gibbs com um N mínimo, que é o número mínimo de iterações requeridas para se obter a precisão necessária de estimação. O método fornece as estimativas do “burn-in” (B), que é o número de iterações que devem ser descartadas, o número de iterações que devem ser computadas N (total) e o k , que é a distância mínima de uma iteração à outra para se obter a subamostra aproximadamente independente (“thin”). Outra saída importante do critério, é o fator de desempenho que é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência, devido a autocorrelação dentro de sequência. Segundo Raftery e Lewis (1992), se esse fator for maior que 5,0, pode-se dizer que a convergência ainda não foi obtida, necessitando reparametrizar o modelo em estudo. Para mais detalhes, ver Nogueira (2004).

4.4.3 Critério de Geweke

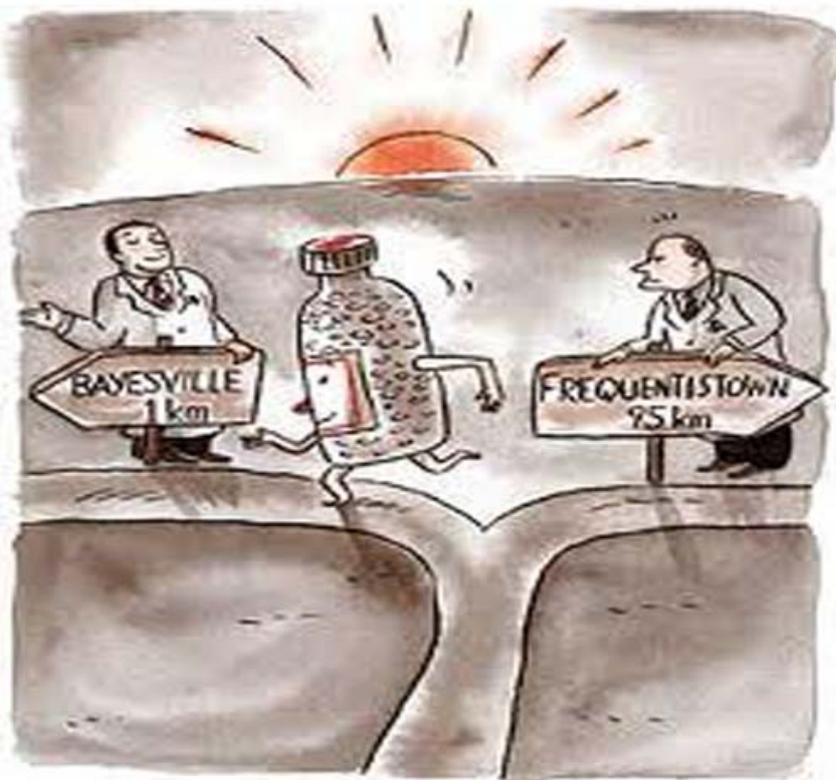
Esse critério é um método de análise espectral usado em série temporal para testar a convergência, mas que também é muito utilizado para os métodos MCMC, já que este é um caso particular de série temporal estacionária.

Na prática, deve-se notar que esta tentativa de diagnóstico para verificar uma necessária, mas não suficiente condição de convergência, informa ao pesquisador somente se a convergência não foi alcançada, e não se realmente convergiu. Para mais detalhes, ver Nogueira (2004).

4.4.4 Critério de Heidelberger e Welch

Esses dois autores apresentaram um método para determinar o valor inicial de uma sequência discreta de eventos, que é também apropriada para ser utilizada para o diagnóstico de convergência do Amostrador de Gibbs. Suas ideias são baseadas na teoria Browniana.

O diagnóstico tem muito pouco poder para detectar a falta de convergência se a amostra é de tamanho muito pequeno, podendo ser problema também em obter estimativas fidedignas da densidade espectral, o que ocorre com o método de Geweke. Para mais detalhes, ver Nogueira (2004).



Uma análise cuidadosa dos dados deve levar em conta a avaliação e escolha do modelo que melhor represente os dados.

A escolha do melhor modelo deve levar em conta o objetivo do estudo, ou seja, que uso será dado ao modelo e quais modelos estão em competição.

Deseja-se responder perguntas tais como:

1. O modelo em consideração é adequado?
2. Dada uma coleção de modelos, qual é o melhor?

5.1 Medidas de Diagnóstico - Adequabilidade de um modelo

Seja $x = (x_1, x_2, \dots, x_n)$ a amostra observada, isto é, a amostra usada para construir, a partir do modelo em consideração, a distribuição *a posteriori* $P(\theta|x)$.

Suponha-se que se tem $y = (y_1, y_2, \dots, y_n)$ uma amostra de validação, isto é, uma amostra independente de x usada para validar o modelo em estudo.

A distribuição preditiva de y é

$$P(y|x) = \int \underbrace{L(y|\theta)P(\theta|x)}_{P(y,\theta|x)} d\theta.$$

Esta distribuição preditiva permite uma avaliação do modelo no sentido em que caso os dados y não estejam de acordo com $P(y|x)$, então não é de se esperar que o modelo seja adequado (PAULINO; TURKMAN; MURTEIRA, 2003, p. 349).

O valor médio e a variância preditiva de cada componente de y é

$$E[y_i|x] = \int y_i P(y|x) dy,$$

e

$$Var[y_i|x] = \int (y_i - E[y_i|x])^2 P(y|x) dy,$$

são úteis para definir os resíduos bayesianos padronizados,

$$d_i = \frac{y_i - E[y_i|x]}{\sqrt{Var[y_i|x]}}, \quad i = 1, 2, \dots, n.$$

Estes resíduos podem ser usados, à semelhança do que é feito na análise frequentista, para averiguar informalmente a validade do modelo. Por exemplo um gráfico destes resíduos contra os valores preditos (valores médios preditos) pode revelar inadequabilidade do modelo, assim como um gráfico contra a ordem das observações para revelar falta de independência.

Esta discussão assume a existência de uma amostra de observação independente, e que não acontece com frequência na prática.

Se uma amostra inicial for de dimensão elevada, há sempre uma possibilidade de particionar em duas de modo que uma parte sirva de amostra observada para a construção da distribuição *a posteriori* e a outra de amostra de validação para se obter a distribuição preditiva.

Não sendo possível particionar a amostra para fazer a validade cruzada, pode-se optar por um tipo JackKnife (leave one out). Assim, se se designar por $x_{(-i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, o vetor constituído de todas as observações a exceção de x_i se pode obter a distribuição preditivas condicionais

$$P(x_i|x_{(-i)}) = \frac{p(x)}{p(x_{(-i)})} = \int L(x_i|\theta, x_{(-i)})P(\theta|x_{(-i)})d\theta,$$

e conseqüentemente os resíduos bayesianos de eliminação padronizados

$$d'_i = \frac{x_i - E[X_i|x_{(-i)}]}{\sqrt{Var(X_i|x_{(-i)})}}, \quad i = 1, 2, \dots, n,$$

onde os valores médios e variâncias são calculados, obviamente a custa das correspondentes distribuições condicionais.

Gelman e Meng (1996) sugerem o seguinte procedimento para averiguar a adequabilidade de um modelo. Considere-se as distribuições preditivas condicionais $p(x_i|x_{(-i)})$, simule-se, para cada $i = 1, 2, \dots, n$, uma amostra $x_{i1}^*, \dots, x_{im}^*$, de dimensão m , da distribuição preditiva correspondente e construa-se à custa da amostra, um intervalo preditivo de probabilidade $(1 - \alpha)$, por exemplo, com caudas equiprováveis. Então se o modelo for adequado é de se esperar que $(1 - \alpha) \times 100\%$ das $x_{i.obs}$ caíam dentro dos respectivos intervalos.

Uma vantagem em usar as distribuições preditivas $P(x_i|x_{(-i)})$, está no fato de estas existirem mesmo que $p(x)$ não seja finita (no caso de *priori's* impróprias).

5.2 Comparação de Modelos

Várias metodologias têm sido seguidas para a relação de modelos. A abordagem inicial sugere a utilização do fator de Bayes para a comparação de modelos.

Assim, se tiver, por exemplo, dois modelos paramétricos em competição M_1 e M_2 para os dados X , com parâmetros θ_1 e θ_2 e distribuição *a priori* $P_1(\theta_1)$ e $P_2(\theta_2)$, respectivamente, as distribuições marginais de X são dadas por

$$P(X|M_i) = \int L(X|\theta_i, M_i)P(\theta_i)d\theta_i, \quad i = 1, 2,$$

sendo o fator de Bayes dado por

$$B = \frac{P(X|M_1)}{P(X|M_2)}.$$

OBS.: O uso de *priori*'s impróprias pode levar a *posteriori*'s impróprias e atrapalhar o fator de Bayes.

Um problema sério com o fator de Bayes é que para muitos modelos de interesse, pelo menos parte da especificação da *priori* é vaga de modo que a distribuição *a priori* $P(\theta)$ é imprópria.

Assim, mesmo que a distribuição *a posteriori* seja própria, a distribuição preditiva *a priori* (distribuição marginal dos dados) usada no cálculo do fator de Bayes é imprópria.

Tabela 5.1: Interpretação do fator de Bayes

Valor de B	Conclusão
$B < 1$	Evidência a favor de M_2
$1 < B < 3,2$	Evidência muito fraca a favor de M_1
$3,2 < B < 10$	Evidência fraca a favor de M_1
$10 < B < 100$	Evidência forte a favor de M_1
$B \geq 100$	Evidência muito forte a favor de M_1

Para ultrapassar estas dificuldades várias modificações ao fator de Bayes foram propostas.

5.2.1 Alternativas

1. O produto $\prod_{i=1}^n P(x_i|x_{(-i)})$ tem sido proposto como substituto de $P(x)$ no cálculo do fator de Bayes para comparar modelos, dando origem ao pseudo fator de Bayes

$$\prod_{i=1}^n \frac{p(x_{i.obs}|x_{(-i)}, M_1)}{p(x_{i.obs}|x_{(-i)}, M_2)}$$

Se este fator for maior que 1, prefere-se M_1 e M_2 , caso contrário, prefere-se M_2 .

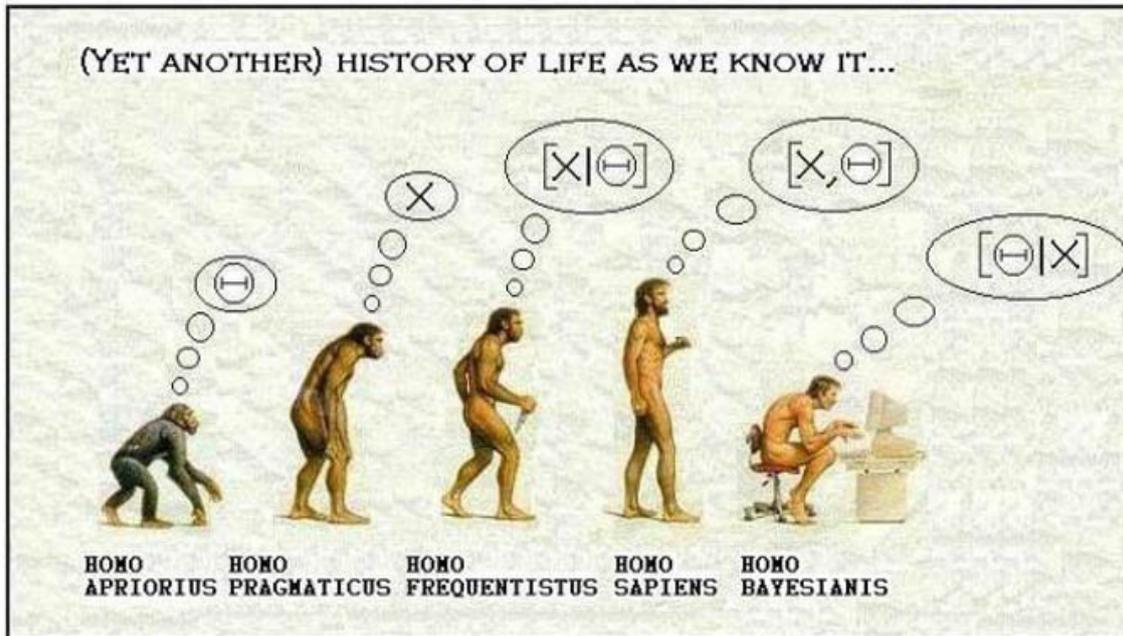
No caso de haver vários modelos em competição, pode-se calcular para cada modelo,

$$\prod_{i=1}^n P(x_{i.obs}|x_{(-i)}),$$

ou, equivalentemente a soma de algoritmos das ordenadas preditivas condicionais, e escolher o modelo que apresente o maior valor.

2. A soma dos quadrados (ou dos valores absolutos) dos resíduos padronizados (quer obtidos por validação cruzada ou por JackKnife) pode também ser usada na relação de modelos. Sugere-se o modelo que apresente o menor valor.

Listas de exercício



6.1 Lista 1

Exercício proposto 6.1: Paulino, Turkman e Murteira (2003), Exerc. 2.12 - pág. 132

Seja X o número de peixes num viveiro em condição de serem enviados para o mercado entre N peixes capturados supostamente distribuídos segundo uma lei $Bin(N, \theta)$, em que não só θ mas também N são desconhecidos. Usando as distribuições *a priori* uniformes independentes para (N, θ) , mostre que as distribuições *a posteriori* condicionais são próprias (e identifique-as), contrariamente ao que ocorre com as distribuições *a posteriori* marginais.

Solução:

$$\begin{aligned} P(N, \theta | x) &\propto k \binom{N}{x} \theta^x (1 - \theta)^{N-x} I_{\{1, \dots, N\}}(x) \\ &\propto \binom{N}{x} \theta^x (1 - \theta)^{N-x}. \end{aligned}$$

Assim, $N, \theta | x \sim Bin(N, x)$.

Antes de passarmos para os próximos passos, MAGALHÃES (2006) define que alguns autores definem a Binomial Negativa como número de ensaios anteriores ao r -ésimo sucesso. Isto cor-

responde a um troca de variável na definição da distribuição binomial negativa $X \sim BN(r, p)$, ou seja,

$$\binom{x+r-1}{x} p^r (1-p)^x.$$

Tomando $y = x + r - 1$ temos a quantidade desejada e seus valores variam de $(r - 1)$ em diante. Assim, sendo Y o número de ensaios anteriores ao r -ésimo sucesso, temos

$$P(Y = y) = \binom{y}{r-1} p^r (1-p)^{y-r+1}, \quad y = r-1, r, \dots \quad (6.1)$$

Assim,

Posteriores condicionais para $N|\theta, x$:

$$\begin{aligned} P(N|\theta, x) &\propto \binom{N}{x} \theta^x (1-\theta)^{N-x} \\ &\propto \binom{N}{x+1-1} \theta^x (1-\theta)^{N-x} \\ &\propto \binom{N}{(x+1)-1} \theta^x (1-\theta)^{N-(x+1)+1}, \end{aligned}$$

percebe-se que esse resultado comparado com a expressão (6.1), $y = N$, $r = x + 1$, $p = \theta$ e $q = 1 - p = 1 - \theta$. Então, $N|\theta, x \stackrel{d}{=} M - 1|\theta, x$ com $M|\theta, x \sim \text{Binomial Negativa}(x + 1, \theta)$, que é uma distribuição própria.

Posteriores condicionais para $\theta|N, x$:

$$\begin{aligned} P(\theta|N, x) &\propto \binom{N}{x} \theta^x (1-\theta)^{N-x} \\ &\propto \theta^{x+1-1} (1-\theta)^{N-x+1-1}. \end{aligned}$$

Observamos que $\theta^{x+1-1} (1-\theta)^{N-x+1-1}$ é o núcleo da distribuição beta, isto é, $\theta|N, x \sim \text{Beta}(x + 1, N - x + 1)$, que também é uma distribuição própria.

Marginais *a posteriori*:

- Para θ , desenvolveremos por meio de $P(N, \theta|x)$ com relação a N . Como N é um parâmetro discreto em $\{0, 1, 2, \dots, a\}$, sendo a um número positivo finito, então

$$\begin{aligned} P(\theta|x) &= \sum_{N=0}^a P(\theta, N|x) \\ &\propto \sum_{N=0}^a \binom{N}{x} \theta^x (1-\theta)^{N-x} \\ &\propto \sum_{N=0}^a \binom{N}{x} \theta^{x+1-1} (1-\theta)^{N-x} \\ &\propto \frac{1}{\theta} \sum_{N=0}^a \binom{N}{x} \theta^{x+1} (1-\theta)^{N-x}, \quad y = x + 1 \end{aligned}$$

$$\begin{aligned} &\propto \frac{1}{\theta} \underbrace{\sum_{N=0}^a \binom{N}{y-1} \theta^y (1-\theta)^{N-y+1}}_{=1, \text{ soma de binomial negativa}} \\ &\propto \frac{1}{\theta}. \end{aligned}$$

Tem-se que $P(\theta) \propto \frac{1}{\theta} I_{[0,1]}(\theta)$, que é uma distribuição imprópria, visto que $\int_0^1 \frac{1}{\theta} d\theta = -\infty$.

- Para N , desenvolveremos por meio de $P(N|\theta, x)$ integrando com relação a θ . Como θ é um parâmetro contínuo em $[0, 1]$, então

$$\begin{aligned} P(N|x) &= \int_0^1 P(\theta, N|x) d\theta \\ &\propto \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} d\theta \\ &\propto \binom{N}{x} \int_0^1 \theta^x (1-\theta)^{N-x} d\theta \\ &\propto \frac{N!}{x!(N-x)!} \frac{\text{Beta}(x+1, N-x+1)}{\text{Beta}(x+1, N-x+1)} \int_0^1 \theta^{(x+1)-1} (1-\theta)^{(N-x+1)-1} d\theta, \end{aligned}$$

Considerando $a = x + 1$, $b = N - x + 1$, $\text{Beta}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ e $\Gamma(n+1) = n!$, então

$$\begin{aligned} P(N|x) &\propto \frac{N!}{x!(N-x)!} \frac{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}}{\text{Beta}(a, b)} \int_0^1 \theta^{a-1} (1-\theta)^{b-1} d\theta \\ &\propto \frac{N!}{x!(N-x)!} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \underbrace{\int_0^1 \frac{1}{\text{Beta}(a, b)} \theta^{(a)-1} (1-\theta)^{(b)-1} d\theta}_{=1} \\ &\propto \frac{N!}{x!(N-x)!} \frac{\Gamma(x+1)\Gamma(N-x+1)}{\Gamma(N+2)} \\ &\propto \frac{N!}{x!(N-x)!} \frac{x!(N-x)!}{(N+1)!} \\ &\propto \frac{N!}{(N+1)N!} \\ &\propto \frac{1}{(N+1)}. \end{aligned}$$

Logo, para $N \geq x$, $P(N|x)$ é uma distribuição imprópria.

Exercício proposto 6.2: Paulino, Turkman e Murteira (2003), Exerc. 2.11 - pág. 132

Considere que Y_i , $1 \leq i \leq n$ são variáveis aleatórias independentes com distribuição $N(x'_i \beta, \sigma^2)$, onde $x'_i = (x'_{i1}, \dots, x'_{ik})$ é um vetor de k variáveis explicativas e $\beta \in \mathbb{R}^k$.

- a) Verifique se a distribuição conjunta para $\theta = (\beta, \sigma^2)$ definindo $\sigma^2 \sim GI(c, d)$ e $\beta|\sigma \sim N_k(a, \sigma^2 \mathbf{V})$, com \mathbf{V} uma matriz simétrica definida positiva, defina a distribuição conjugada natural do modelo amostral indicado.

- b) Mostre que a distribuição *a priori* para θ obtida pela regra de Jeffreys multiparamétrico é

$$P(\beta, \sigma^2) \propto (\sigma^2)^{-\frac{k+2}{2}}.$$

c) Que distribuição obtém pelo método de Jeffreys se admitir independência *a priori* entre β e σ^2 ?

Solução: (a)

Vamos definir as variáveis e os parâmetros:

- $\mathbf{Y}_{n \times 1} = [y_1, y_2, \dots, y_n]'$ - vetor de observações;
- $\mathbf{X}_{n \times k} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$ - matriz conhecida (variáveis explicativas);
- $\beta_{k \times 1} = [\beta_1, \beta_2, \dots, \beta_k]'$ - parâmetro;
- $\mathbf{a}_{k \times 1} = [a_1, a_2, \dots, a_k]'$ - parâmetro de médias;
- σ^2, c e d - parâmetros;
- $\mathbf{V}_{k \times k} = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1k} \\ v_{21} & v_{22} & \dots & v_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \dots & v_{kk} \end{pmatrix}$ - matriz positiva definida;

Tendo *a priori* hierárquica:

$$P(\beta, \sigma^2) = P(\beta|\sigma^2)P(\sigma^2),$$

com

$$\begin{aligned} P(\beta|\sigma^2) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{k/2} (\sigma^2 \mathbf{V})^{-k/2} \exp\left\{-\frac{1}{2}(\beta - \mathbf{a})'(\sigma^2 \mathbf{V})^{-1}(\beta - \mathbf{a})\right\} \\ &\propto (\sigma^2 \mathbf{V})^{-k/2} \exp\left\{-\frac{1}{2}(\beta - \mathbf{a})'(\sigma^2 \mathbf{V})^{-1}(\beta - \mathbf{a})\right\} \end{aligned}$$

e

$$\begin{aligned} P(\sigma^2) &= \frac{d^c}{\Gamma(c)} (\sigma^2)^{-(c+1)} \exp\{-d/\sigma^2\} \\ &\propto (\sigma^2)^{-(c+1)} \exp\{-d/\sigma^2\}. \end{aligned}$$

A distribuição conjunta entre β e σ^2

$$\begin{aligned} P(\beta, \sigma^2) &\propto \left[(\sigma^2 \mathbf{V})^{-k/2} \exp\left\{-\frac{1}{2}(\beta - \mathbf{a})'(\sigma^2 \mathbf{V})^{-1}(\beta - \mathbf{a})\right\} \right] \times \\ &\quad \times \left[(\sigma^2)^{-(c+1)} \exp\{-d/\sigma^2\} \right] \\ &\propto (\sigma^2)^{-k/2} (\sigma^2)^{-(c+1)} \exp\left\{-\frac{1}{2}(\beta - \mathbf{a})'(\sigma^2 \mathbf{V})^{-1}(\beta - \mathbf{a}) - d/\sigma^2\right\} \\ &\propto (\sigma^2)^{-k/2} (\sigma^2)^{-(c+1)} \exp\left\{-\frac{1}{2\sigma^2} [(\beta - \mathbf{a})' \mathbf{V}^{-1}(\beta - \mathbf{a})] - d/\sigma^2\right\}. \end{aligned}$$

A função de verossimilhança será

$$\begin{aligned}
 P(\mathbf{Y}|\beta, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n/2} (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)'(\sigma^2)^{-1}(\mathbf{Y} - \mathbf{X}\beta)\right\} \\
 &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\right\}.
 \end{aligned}$$

Calculando a distribuição *a posteriori*:

$$\begin{aligned}
 P(\beta, \sigma^2|\mathbf{Y}) &= P(\mathbf{Y}|\beta, \sigma^2)P(\beta, \sigma^2) \\
 &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\right\} \times \\
 &\quad \times (\sigma^2)^{-k/2} (\sigma^2)^{-(c+1)} \exp\left\{-\frac{1}{2\sigma^2}[(\beta - \mathbf{a})'\mathbf{V}^{-1}(\beta - \mathbf{a})] - d/\sigma^2\right\} \\
 &\propto (\sigma^2)^{-\left(\frac{n+k}{2}\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\underbrace{(\beta - \mathbf{a})'\mathbf{V}^{-1}(\beta - \mathbf{a}) + (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)}_{(**)}\right]\right\} \times \\
 &\quad (\sigma^2)^{-(c+1)} \exp\{-d/\sigma^2\}.
 \end{aligned}$$

Desenvolvendo o termo no expoente, tem-se:

$$\begin{aligned}
 (**) &= \beta'\mathbf{V}^{-1}\beta - \beta'\mathbf{V}^{-1}\mathbf{a} - \mathbf{a}'\mathbf{V}^{-1}\beta + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{Y} + \beta'\mathbf{X}'\mathbf{X}\beta \\
 &= \beta'(\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})\beta - \beta'(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}) - (\mathbf{a}\mathbf{V}^{-1} + \mathbf{Y}'\mathbf{X})\beta + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y}.
 \end{aligned} \tag{6.2}$$

Para uma normal multivariada^a, temos o seguinte expoente:

$$(\beta - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\beta - \boldsymbol{\mu}) = \beta'\boldsymbol{\Sigma}^{-1}\beta - \beta'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\beta + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{6.3}$$

Comparando as expressões (6.2) e (6.3), tem-se:

$$\beta'\boldsymbol{\Sigma}^{-1}\beta = \beta'(\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})\beta \tag{6.4}$$

$$\beta'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \beta'(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}). \tag{6.5}$$

Por meio da equação (6.4), sabe-se que $\boldsymbol{\Sigma}^{-1} = (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})$. Sabemos pela expressão (6.5) que

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = (\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}). \tag{6.6}$$

Para encontrar $\boldsymbol{\mu}$ na equação (6.6), vamos pré-multiplicar os dois membros da expressão por $\boldsymbol{\Sigma} = (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$ temos-se:

$$\begin{aligned}
 \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} &= (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}) \\
 \boldsymbol{\mu} &= (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}).
 \end{aligned}$$

Com os resultados de $\boldsymbol{\Sigma}$ e $\boldsymbol{\mu}$ obtidos, e comparando com as expressões (6.2) e (6.3), temos-se:

$$(**) = \underbrace{\beta'(\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})\beta}_{\beta'\boldsymbol{\Sigma}^{-1}\beta} - \underbrace{\beta'(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y})}_{\beta'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} - \underbrace{(\mathbf{a}\mathbf{V}^{-1} + \mathbf{Y}'\mathbf{X})\beta}_{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\beta} + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y}.$$

Assim, podemos dizer que

$$\begin{aligned}
 (**) &= \left[\boldsymbol{\beta} - (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}) \right]' (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X}) \times \\
 &\quad \times \left[\boldsymbol{\beta} - (\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{V}^{-1}\mathbf{a} + \mathbf{X}'\mathbf{Y}) \right] - \underbrace{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y}}_{(\star)} \\
 &= (\boldsymbol{\beta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) - \underbrace{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y}}_{(\star)}.
 \end{aligned}$$

(\star) Essa parte foi a que sobrou para completar o quadrado entre (6.2) e (6.3). Assim, precisamos retirar.

Retornando a distribuição *a posteriori*, temos

$$\begin{aligned}
 P(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y}) &\propto (\sigma^2)^{-\left(\frac{n+k}{2}\right)} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \mathbf{a})'\mathbf{V}^{-1}(\boldsymbol{\beta} - \mathbf{a}) + (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})] \right\} \times \\
 &\quad (\sigma^2)^{-(c+1)} \exp \{ -d/\sigma^2 \} \\
 &\propto (\sigma^2)^{-\left(\frac{2c+n+k}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \right. \\
 &\quad \left. + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y} + 2d] \right\}
 \end{aligned}$$

Observamos que $P(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y})$ é uma Normal-Gama Inversa. Assim essa distribuição define a distribuição conjugada natural do modelo amostral indicado.

Se considerarmos σ^2 conhecido, concluímos que

$$P(\boldsymbol{\beta} | \sigma^2, \mathbf{Y}) \propto (\sigma^2)^{-\left(\frac{k}{2}\right)} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})] \right\},$$

ou seja, $\boldsymbol{\beta} | \sigma^2, \mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Se considerarmos $\boldsymbol{\beta}$ conhecido, concluímos que

$$P(\sigma^2 | \boldsymbol{\beta}, \mathbf{Y}) \propto (\sigma^2)^{-\left(\frac{2c+n+k}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\beta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \right. \\
 \left. + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y} + 2d] \right\},$$

isto é, $\sigma^2 | \boldsymbol{\beta}, \mathbf{Y} \sim GI \left(\frac{2c+n+k}{2}, \frac{(\boldsymbol{\beta} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{a}'\mathbf{V}^{-1}\mathbf{a} + \mathbf{Y}'\mathbf{Y} + 2d}{2} \right)$

^aPara nosso caso, estamos considerando $\boldsymbol{\Sigma}$ uma matriz ($k \times k$) e $\boldsymbol{\mu}$ uma vetor ($k \times 1$).

Solução: (b)

Sendo a função de verossimilhança

$$\begin{aligned}
 P(\mathbf{Y} | \boldsymbol{\beta}, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi}} \right)^{n/2} (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\sigma^2)^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\
 &\propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\},
 \end{aligned}$$

e a função suporte

$$\begin{aligned}\ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2) &\propto -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &\propto -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}),\end{aligned}$$

então calculando as derivadas parciais em relação aos parâmetros σ^2 e $\boldsymbol{\beta}$ e as respectivas esperanças, temos

$$\begin{aligned}\frac{\partial \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial^2 (\sigma^2)} &= -\frac{n}{2(\sigma^2)^2} + \frac{1}{2(\sigma^2)^3} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma^2} (-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial^2 \boldsymbol{\beta}} &= -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \\ \frac{\partial \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \partial \boldsymbol{\beta}} &= \frac{1}{2(\sigma^2)^2} (-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta} \partial \sigma^2} &= \frac{1}{2(\sigma^2)^2} (-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\ -E\left(\frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial^2 (\sigma^2)}\right) &= \frac{n}{2(\sigma^2)^2} \\ -E\left(\frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}^2}\right) &= \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \\ -E\left(\frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \partial \boldsymbol{\beta}}\right) &= 0.\end{aligned}$$

A *priori* de Jeffreys é dada por $P(\sigma^2, \boldsymbol{\beta}) \propto |I(\sigma^2, \boldsymbol{\beta})|^{1/2}$, em que $I(\sigma^2, \boldsymbol{\beta})$ representa a matriz de informação de Fisher. Assim, tem-se

$$I(\sigma^2, \boldsymbol{\beta}) = \begin{pmatrix} \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{pmatrix}$$

$$\begin{aligned}I(\sigma^2, \boldsymbol{\beta}) &= \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \times \frac{n}{2(\sigma^2)^2} \\ &= \left(\frac{1}{\sigma^2}\right)^k (\mathbf{X}'\mathbf{X}) \frac{n}{2(\sigma^2)^2}.\end{aligned}$$

Portanto,

$$\begin{aligned}P(\sigma^2, \boldsymbol{\beta}) &\propto \left[\left(\frac{1}{\sigma^2}\right)^k (\mathbf{X}'\mathbf{X}) \frac{n}{2(\sigma^2)^2} \right]^{1/2} \\ &\propto (\sigma^2)^{-\frac{k+2}{2}}.\end{aligned}$$

Solução: (c)

Sendo a função de verossimilhança

$$\begin{aligned} P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n/2} (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\sigma^2)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\} \\ &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\}, \end{aligned}$$

e a função suporte

$$\begin{aligned} \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2) &\propto -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &\propto -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}), \end{aligned}$$

então *a priori* de Jeffreys para σ^2 :

$$\begin{aligned} \frac{\partial \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial^2 (\sigma^2)} &= -\frac{n}{2(\sigma^2)^2} + \frac{1}{2(\sigma^2)^3}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

Informação de Fisher: $I(\sigma^2) = \left(\frac{n}{2(\sigma^2)^2}\right)$,

$$\begin{aligned} h_1(\sigma^2) &\propto [I(\sigma^2)]^{1/2} \\ &\propto \left(\frac{n}{2(\sigma^2)^2}\right)^{1/2} \\ &\propto \frac{1}{\sigma^2}. \end{aligned}$$

Priori de Jeffreys para $\boldsymbol{\beta}$:

$$\begin{aligned} \frac{\partial \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma^2}(-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial^2 \ln P(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2)}{\partial^2 \boldsymbol{\beta}} &= -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2}. \end{aligned}$$

Informação de Fisher: $I(\boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\mathbf{X}}{\sigma^2}\right)$,

$$\begin{aligned} h_2(\boldsymbol{\beta}) &\propto [I(\boldsymbol{\beta})]^{1/2} \\ &\propto \left(\frac{\mathbf{X}'\mathbf{X}}{\sigma^2}\right)^{1/2} \\ &\propto \phi \quad (\text{uma constante}). \end{aligned}$$

Logo *a priori* de Jeffreys conjunta σ^2 e $\boldsymbol{\beta}$ é dado por

$$\begin{aligned} P(\sigma^2, \boldsymbol{\beta}) &\propto \frac{1}{\sigma^2} \times \phi \\ &\propto \frac{1}{\sigma^2}. \end{aligned}$$

6.2 Lista 2

Exercício proposto 6.3: Paulino, Turkman e Murteira (2003, p. 216)

Aplicação à análise de regressão linear simples

Medição da radiação de fundo, através de contagens Y num detector de Geiger por minutos, em 39 locais com altitudes x (em metros) diferenciadas.

Y	11.0	12.7	11.6	12.8	11.8	11.5	11.1	13.6
x	213.4	426.8	487.8	518.3	579.3	640.2	701.2	731.7
\bar{Y}	12.5	13.9	15.7	17.8	17.2	14.4	15.0	13.8
x	884.1	884.1	1097.6	1097.6	1128.0	1158.5	1158.5	1219.5
Y	15.8	12.4	13.9	18.0	14.9	16.8	17.3	17.9
x	1250.0	1311.0	1463.4	1524.4	1554.9	1554.9	1554.9	1554.9
\bar{Y}	18.4	19.6	20.1	16.6	19.3	22.0	19.8	27.8
x	1554.9	1554.9	1554.9	1646.3	1646.3	1890.2	1951.2	2103.7
Y	29.0	20.4	25.0	26.8	22.7	24.9	26.0	
x	2103.7	2164.6	2256.1	2256.1	2530.5	3231.7	3231.7	

- Admitindo-se o modelo de regressão linear simples sumarie-se as inferências sobre os parâmetros do modelo relativas ao uso da distribuição *a priori* não informativa usual.
- Compare-se sob o modelo anterior inferências sobre a radiação média num local a 200m com as predições sobre a média das radiações a obter em m locais com a mesma altitude nos casos $m = 1$ e $m = 5$.

Solução: (a)

Denote-se o modelo por

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \stackrel{\text{ind.}}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$$

com $n = 39$ e $P(\beta_0, \beta_1, \sigma^2) \propto \frac{1}{\sigma^2}$ no conjunto $\mathbb{R}^2 \times \mathbb{R}_+$.

A seguir indicaremos os principais resultados relativos às distribuições *a posteriori* para os parâmetros do modelo:

$$\text{Declive da reta de regressão: } \beta_1 | y \sim t_1(37; 6, 2 \times 10^{-3}; 3, 2 \times 10^{-7}).$$

A média e o desvio padrão *a posteriori* de β_1 são respectivamente dados por $E[\beta_1 | y] = 6, 2 \times 10^{-3}$ e $\sqrt{\text{var}(\beta_1 | y)} = 5, 8 \times 10^{-4}$. Destaca-se a grande concentração da densidade *a posteriori* retratada no pequeníssimo valor do desvio padrão *a posteriori* derivado da elevada dispersão das altitudes. O correspondente intervalo de credibilidade HPD exato a 95% é $(5, 1 \times 10^{-3}, 7, 4 \times 10^{-3})$ indicando, assim, fortes evidências contra a ausência do efeito de altitude na radiação ($\beta_1 = 0$).

$$\text{Ordenada na origem da reta: } \beta_0 | y \sim t_1(37; 8, 55; 0, 82).$$

Este parâmetro representa *a posteriori* uma média e um desvio padrão dados por $E[\beta_0 | y] = 8, 55$ e $\sqrt{\text{var}(\beta_0 | y)} = 0, 93$. O seu intervalo de credibilidade HPD exato a 95% dado por $(6, 67; 10, 44)$, contém fortes evidências contra a ausência da radiação média ao nível do mar ($\beta_0 = 0$) se se admitir também nesta gama de altitudes a validade do modelo. A reta ajustada fica definida por $\hat{Y} = 8, 55 + 0, 0062x$.

$$\text{Variância do modelo amostral: } \sigma^2 | y \sim GI(37/2; 224/2).$$

A esperança é dada por $E[\sigma^2|y] = 6,40$ e o desvio padrão $\sqrt{\text{var}(\sigma^2|y)} = 1,58$. O intervalo de credibilidade HPD a 95% é dado por (3,74;9,56).

As medidas de localização comuns para esta distribuição seguem a relação

$$\text{Moda} = 5,74 < \text{Mediana} = 6,16 < E[\sigma^2|Y] = 6,40.$$

Para encontrar esses resultados, foi utilizado o software R (R DEVELOPMENT CORE TEAM, 2011), usando as seguintes expressões:

✘ Parâmetro $\boldsymbol{\theta} = [\beta_0, \beta_1]'$: Sabemos que $\boldsymbol{\theta} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, sendo $\nu = n - p$, $\boldsymbol{\mu} = \hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ e $\boldsymbol{\Sigma} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, sendo um estimador para σ^2 , $s^2 = (n - p)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})$. Assim,

$$E[\boldsymbol{\theta}|y] = \boldsymbol{\mu} \text{ e } \text{Var}[\boldsymbol{\theta}|y] = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma}.$$

✘ Para σ^2 : Sabemos que $\sigma^2|y \sim GI^a(\alpha, \beta)$, sendo $\alpha = \nu/2$ e $\beta = \nu s^2/2$. Assim,

$$E[\sigma^2|y] = \frac{\beta}{\alpha - 1} \text{ para } \alpha > 1, \beta > 0 \text{ e } \text{Var}[\sigma^2|y] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \text{ para } \alpha > 2, \beta > 0.$$

✘ Para encontrar o HPD para os estimadores dos parâmetros $\boldsymbol{\theta}$ e σ^2 , geramos amostras a partir da distribuição desses parâmetros. Após isso, usamos funções específicas do software R, como pode ser detalhado no código abaixo.

^aGama Invertida

Código R:

```
> #####
> #Aplicacao a analise de regresao linear simples
> #Paulino et.al. (2003) - pag. 216
> #####
>
> # ALTERNATIVA (A)
>
> y=c(11.0,12.7,11.6,12.8,11.8,11.5,11.1,13.6,
+ 12.5,13.9,15.7,17.8,17.2,14.4,15,13.8,
+ 15.8,12.4,13.9,18,14.9,16.8,17.3,17.9,
+ 18.4,19.6,20.1,16.6,19.3,22,19.8,27.8,
+ 29,20.4,25,26.8,22.7,24.9,26)
>
>
> x=c(213.4,426.8,487.8,518.3,579.3,640.2,701.2,731.7,
+ 884.1,884.1,1097.6,1097.6,1128,1158.5,1158.5,1219.5,
+ 1250,1311,1311,1463.4,1524.4,1554.9,1554.9,1554.9,
+ 1554.9,1554.9,1554.9,1646.3,1646.3,1890.2,1951.2,2103.7,
+ 2103.7,2164.6,2256.1,2256.1,2530.5,3231.7,3231.7)
>
> #####
> reg <- lm(y~x);reg #regressao do modelo
```

Call:

```

lm(formula = y ~ x)

Coefficients:
(Intercept)          x
  8.553502         0.006204

> #####
> #matrizes do modelo:
> xmat <- model.matrix(reg)
> xlx1 <- solve(t(xmat)%*%xmat)
> xlx <- t(xmat)%*%xmat
> xly <- t(xmat)%*%y
> yly <- t(y)%*%y
> #####
> #INFERENCIAS PARA THETA
> #####
> #vetor theta:
> theta <- xlx1%*%xly #vetor theta
> beta0 <- theta[1];beta0 #parametro beta 0
[1] 8.553502
> beta1 <- theta[2];beta1 #parametro beta 1
[1] 0.006203927
> #####
> #matriz de precisao:
> nu <- length(y)-length(theta) # graus de liberdade
> s2 <- as.vector(t(y-xmat)%*%theta)%*(y-xmat)%*%theta/nu) #var amostral
> matprec <- diag(s2*solve(xlx));matprec #matriz de precisao
  (Intercept)          x
8.178972e-01 3.199291e-07
> #####
> #matriz de variancia e covariancia:
> vari <- (nu*s2)/(nu-2)*solve(xlx);vari #variancia
  (Intercept)          x
(Intercept)  0.8646341867 -4.867452e-04
x            -0.0004867452  3.382108e-07
> desv <- diag(((nu*s2)/(nu-2)*solve(xlx))^{1/2});desv #desvio padrao
  (Intercept)          x
0.9298570786 0.0005815589
> #####
> #Intervalos de credibilidade para theta (EXATO):
>
> #beta0:
> t <- qt(0.975,df=nu)
> ICbeta0 <- c(beta0-desv[1]*t,beta0+desv[1]*t);ICbeta0
(Intercept) (Intercept)
  6.669433   10.437572
>
> #beta1:
> t <- qt(0.975,df=nu)
> ICbeta1 <- c(beta1-desv[2]*t,beta1+desv[2]*t);ICbeta1
  x          x

```

```

0.005025577 0.007382278
>
> #####
> #Intervalos de credibilidade para theta (APROXIMADO):
>
> #OBS.: Nesse caso, foi gerado uma amostra t_nu(0,1) univariada
> #      para cada parametro do vetor theta, e posteriormente,
> #      feito a transformacao t-tranf=media + desv x t_nu(0,1).
> #      Essa mudanca torna a amostra gerada para mesma media e
> #      variancia da distribuicao t do parametro em estudo.
>
> library(TeachingDemos)#pacote para carregar o comando ''emp.hpd()''
> #beta0:
> set.seed(10) #semente
> ran.beta0=beta0+desv[1]*rt(10^5,nu) #amostra da t_nu(E[beta0],desv[]^2)
> mean(ran.beta0) # esperanca para beta0
[1] 8.552841
> emp.hpd(ran.beta0,conf=0.95) #intervalo de credibilidade
[1] 6.653687 10.422423
>
> #beta1:
> set.seed(10) #semente
> ran.beta1=beta1+desv[2]*rt(10^5,nu) #amostra da t_nu(E[beta1],desv[]^2)
> mean(ran.beta1) # esperanca para beta1
[1] 0.006203514
> emp.hpd(ran.beta1,conf=0.95) #intervalo de credibilidade
[1] 0.005015729 0.007372803
>
> #####
> #Outra forma de Intervalo de credibilidade para theta (APROXIMADO):
>
> #OBS.: Nesse caso, geramos a amostra da propria distribuicao
> #      t multivariada, gerando assim, uma matriz com duas
> #      colunas. A primeira, vem de uma amostra com dist
> #      t_nu(beta_0,var0). A segunda vem de uma amostra com dist
> #      t_nu(beta_1,var1).
>
> library(mvtnorm) # pacote para carregar o comando rmvt()
>
> #matriz diagonal de variancia:
> sig=matrix(c(vari[1,1],0,0,vari[2,2]),nrow = 2, ncol = 2)
>
> #amostra da t multivariada:
> set.seed(10) #semente
> ran.tmult=rmvt(10^5,sigma=sig,delta=theta,df=nu) #amostra t-multivariada
>
> #beta0:
> mean(ran.tmult[,1]) #esperanca para beta0
[1] 8.550616
> emp.hpd(ran.tmult[,1],conf=0.95) #intervalo de credibilidade
[1] 6.653576 10.438435

```

```

>
> #beta1:
> mean(ran.tmult[,2]) #esperanca para beta1
[1] 0.006201223
> emp.hpd(ran.tmult[,2],conf=0.95) #intervalo de credibilidade
[1] 0.005031976 0.007383050
>
>
> #####
> #INFERENCIAS PARA SIGMA
> #####
> #Parametros da gama-invertida:
> a=nu/2
> b=nu*s2/2
> #####
> mean.sigma <- (nu*s2)/(nu-2)
> mean.sigma <- b/(a-1);mean.sigma #esperanca de sigma^2
[1] 6.400748
> var.sigma <- b^2/(((a-1)^2)*(a-2));var.sigma #variancia de sigma^2
[1] 2.483005
> desv.sigma <- sqrt(b^2/(((a-1)^2)*(a-2)));desv.sigma #desv de sigma^2
[1] 1.575755
> #####
>
> #Intervalos de credibilidade para sigma:
> set.seed(10) #semente
> gamainv=1/rgamma(10^5,a,b) #amostra da gama invertida
> mean(gamainv) #esperanca de sigma^2
[1] 6.413991
> emp.hpd(gamainv,conf=0.95) #intervalo de credibilidade
[1] 3.738792 9.559910
>
> #####
>
> #Estatisticas descritivas:
>
> media <- b/(a-1);media
[1] 6.400748
> mediana <- 1/qgamma(0.5,a,b);mediana
[1] 6.165489
> moda <- b/(a+1);moda
[1] 5.744261

```

Solução: (b)

Como queremos prever a média, \bar{Y} para altitudes de m locais, a correspondente densidade preditiva *a posteriori* é a mistura das densidades amostrais $\bar{Y}|\mu, \sigma^2, x \sim N(\mu, \sigma^2/m)$ pela densidade *a posteriori* Normal-Gama Invertida de (μ, σ^2) .

De acordo com Charnet et al. (1999), o valor esperado para a predição $2000m$ é $\phi = \beta_0 + \beta_1 \times 2000 = E[\bar{Y}_0|\beta_0, \beta_1, \sigma^2] = 20,96136$. Para o caso do desvio padrão, vamos considerar um vetor $\mathbf{x}_0 = [1, 2000]'$ e $s^2 = (n - p)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})$. Assim, $\sqrt{\text{var}(\phi|y)} = s^2 \times$

$[\mathbf{x}_0(\mathbf{X}'\mathbf{X})(\mathbf{x}_0)'] = 0,5058411$. O respectivo intervalo de credibilidade HPD a 95% é dado por $IC = (\phi - s \times q_t; \phi + s \times q_t) = (19,93643; 21,98629)$.

Para a predição para $m = k$ elemento fora da amostra, temos que $var_k = s^2 \times [1/k + \mathbf{x}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0']$, e $desv_k = \sqrt{var_k}$. Assim, o intervalo de credibilidade é dado por $IC_k = (\phi - desv_k \times q_t; \phi + desv_k \times q_t)$. Dessa forma, temos as médias das radiações em m locais a 2000m:

$$\begin{aligned}
 m = 1 & \quad E[\bar{Y}_0|y] = 20,96; \quad \sqrt{var(\bar{Y}_0|y)} = 2,512098 \\
 & \quad IC \text{ HPD } 95\% : (15,87136; 26,05135) \\
 m = 5 & \quad E[\bar{Y}_0|y] = 20,96; \quad \sqrt{var(\bar{Y}_0|y)} = 1,211127 \\
 & \quad IC \text{ HPD } 95\% : (18,50738; 23,41533).
 \end{aligned}$$

De acordo com Paulino, Turkman e Murteira (2003, p. 178), temos que

$$\bar{Y}|x \sim t_k \left(\bar{x}, \sqrt{\frac{m+n}{mn}} s^2 \right),$$

evidenciando que a predição média de \bar{Y} é a média *a posteriori*, e seu desvio padrão preditivo

$$\sigma(\bar{Y}|x) = \sqrt{\frac{n-1}{n-3}} \frac{s}{\sqrt{\frac{mn}{m+n}}}.$$

Seja $\phi = \beta_0 + \beta_1 \times 2000 = E[\bar{Y}_0|\beta_0, \beta_1, \sigma^2] = 20,96136$.

Radiação esperada a 2000m: $\phi|y \sim t_1(37; 20,96; 0,5524951)$.

Médias das radiações em m locais a 2000m:

$$\begin{aligned}
 m = 1 & \quad \bar{Y}_0|y \sim t_1(37; 20,96; 2,49199) \\
 & \quad E[\bar{Y}_0|y] = 20,96; \quad \sqrt{var(\bar{Y}_0|y)} = 2,560276 \\
 & \quad IC \text{ HPD } 95\% : (15,73039; 26,10726) \\
 m = 5 & \quad \bar{Y}_0|y \sim t_1(37; 20,96; 1,168847) \\
 & \quad E[\bar{Y}_0|y] = 20,96; \quad \sqrt{var(\bar{Y}_0|y)} = 1,200876 \\
 & \quad IC \text{ HPD } 95\% : (18,50782; 23,37500).
 \end{aligned}$$

Observe-se a coincidência da média *a posteriori* de ϕ e da média preditiva (*a posteriori*) de \bar{Y}_0 e o aumento das correspondentes variâncias distribucionais quando se passa de ϕ para \bar{Y}_0 quando $m = 5$ e $m = 1$, por esta ordem, com seus reflexos nos intervalos HPD.

Código R:

```

> #####
>
> #Análise preditiva:
>
>
> #Valor esperado da radiacao para a altitude 2000m:

```

```

> phi=beta0+beta1*2000;phi
[1] 20.96136
>
> #####
> # Parametros da distribuicao t-univariada:
> # (Charnet et. al, 1999, pág. 104)
> #####
>
> #INTERVALOS DE CREDIBILIDADE EXATOS:
>
> #####
> #Altitude 2000m:
> mean.pred <- phi;mean.pred
[1] 20.96136
>
> s2 #variância amostral
[1] 6.054762
> x0 <- matrix(c(1,2000),ncol=2)
> var <- s2*(x0%*%solve(t(xmat)%*%xmat)%*%t(x0));var
      [,1]
[1,] 0.2558752
> desvpad.y <- sqrt(var);desvpad.y
      [,1]
[1,] 0.5058411
>
> #IC exato:
> t<-qt(0.975,df=nu)
> IC <- c(mean.pred-desvpad.y*t,mean.pred+desvpad.y*t);IC
[1] 19.93643 21.98629
>
> #####
> # m=1:
> mean.pred <- phi;mean.pred
[1] 20.96136
>
> s2 #variância amostral
[1] 6.054762
> x0 <- matrix(c(1,2000),ncol=2)
> var <- s2*(1+x0%*%solve(t(xmat)%*%xmat)%*%t(x0));var
      [,1]
[1,] 6.310637
> desvpad.y <- sqrt(var);desvpad.y
      [,1]
[1,] 2.512098
>
> #IC exato:
> t<-qt(0.975,df=nu)
> IC <- c(mean.pred-desvpad.y*t,mean.pred+desvpad.y*t);IC
[1] 15.87136 26.05135
>
> #####

```

```

> # m=5:
> mean.pred <- phi;mean.pred
[1] 20.96136
>
> s2 #variancia amostral
[1] 6.054762
> x0 <- matrix(c(1,2000),ncol=2)
> var <- s2*(1/5+x0%*%solve(t(xmat)%*%xmat)%*%t(x0));var
      [,1]
[1,] 1.466828
> desvpad.y <- sqrt(var);desvpad.y
      [,1]
[1,] 1.211127
>
> #IC exato:
> t<-qt(0.975,df=nu)
> IC <- c(mean.pred-desvpad.y*t,mean.pred+desvpad.y*t);IC
[1] 18.50738 23.41533
>
> #####
> #Parametros da distribuicao t-multivariada (APROXIMADOS):
> #Paulino et.al (2003)
> #####
>
> #Para m=1:
> m=1 #numero de m locais a prever o valor medio da radiacao
> n=length(y);n # tamanho da amostra original
[1] 39
>
>
> #Parametro de localizacao:
> mean.pred <- phi;mean.pred
[1] 20.96136
>
> #Parametro de escala:
> s2 #variancia amostral
[1] 6.054762
> esc.pred <-sqrt((m+n)/(m*n)*s2);esc.pred
[1] 2.49199
>
> #desvio padrao:
> div1 <- sqrt((n-1)/(n-3)) # func aux1
> div2 <- sqrt((m*n)/(m+n)) # func aux2
>
> desvpad.y <- div1*sqrt(s2)/div2;desvpad.y #desvio padrao
[1] 2.560276
>
> set.seed(10) #semente
> ran.y1=phi+desvpad.y*rt(10^5,nu) #amostra da t_nu(E[],desv[]^2)
> mean(ran.y1) # esperanca a 2000m
[1] 20.95954

```

```

> library(TeachingDemos)#pacote para carregar o comando ''emp.hpd()''
> emp.hpd(ran.y1,conf=0.95) #intervalo de credibilidade
[1] 15.73039 26.10726
>
> #####
> #Para m=5:
> m=5 #numero de m locais a prever o valor medio da radiacao
> n=length(y);n # tamanho da amostra original
[1] 39
>
>
> #Parametro de localizacao:
> mean.pred <- phi;mean.pred
[1] 20.96136
>
> #Parametro de escala:
> s2 #variancia amostral
[1] 6.054762
> esc.pred <-sqrt((m+n)/(m*n)*s2);esc.pred
[1] 1.168847
>
> #desvio padrao:
> div1 <- sqrt((n-1)/(n-3)) # func aux1
> div2 <- sqrt((m*n)/(m+n)) # func aux2
>
> desvpad.y <- div1*sqrt(s2)/div2;desvpad.y #desvio padrao
[1] 1.200876
>
> set.seed(10) #semente
> ran.y5=phi+desvpad.y*rt(10^5,nu) #amostra da t_nu(E[],desv[]^2)
> mean(ran.y5) # esperanca a 2000m
[1] 20.9605
> library(TeachingDemos)#pacote para carregar o comando ''emp.hpd()''
> emp.hpd(ran.y5,conf=0.95) #intervalo de credibilidade
[1] 18.50782 23.37500

```

6.3 Lista 3

Exercício proposto 6.4: Séries temporais - Inferência Bayesiana

Seja,

$$\begin{aligned}
 Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t, & \epsilon_t &\sim N(0, 1), \\
 \underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} &= \underbrace{\begin{pmatrix} Y_0 & Y_{-1} \\ Y_1 & Y_0 \\ \vdots & \vdots \\ Y_{n-1} & Y_{n-2} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}}_{\boldsymbol{\phi}} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{\boldsymbol{\epsilon}} \\
 \mathbf{Y} &= \mathbf{X}\boldsymbol{\phi} + \boldsymbol{\epsilon},
 \end{aligned}$$

um modelo autorregressivo $AR(2)$ de uma série temporal.

A condição para estacionariedade para a série é

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

Encontre a distribuição *a posteriori* para $\phi = (\phi_1, \phi_2)$ e σ^2 e o HPD para esses parâmetros. Considere que $n = 1000$.

Solução:

Considere $\phi_2 = -0,6$. Então,

$$\boxtimes \phi_1 > \phi_2 - 1 \Rightarrow \phi_1 > -0,6 - 1 \Rightarrow \phi > -1,6;$$

$$\boxtimes \phi_1 < 1 - \phi_2 \Rightarrow \phi < 1 + 0,6 \Rightarrow \phi < 1,6.$$

Assim, os valores de ϕ_1 para que o modelo seja estacionário é $-1,6 < \phi_1 < 1,6$. Portanto, vamos considerar que $\phi_1 = 0,8$.

A seguir indicaremos os principais resultados relativos às distribuições *a posteriori* para os parâmetros do modelo:

$$\text{Modelo para } \phi_1: \phi_1|y \sim t_1(998; 0,8301865; 0,0006109596).$$

A média e o desvio padrão *a posteriori* de ϕ_1 são respectivamente dados por $E[\phi_1|y] = 0,8301865$ e $\sqrt{\text{var}(\phi_1|y)} = 0,0247424$. O correspondente intervalo de credibilidade HPD a 95% é $(0,7815556; 0,8788982)$.

$$\text{Modelo para } \phi_2: \phi_2|y \sim t_1(998; -0,6255618; 0,0006109746).$$

Este parâmetro representa *a posteriori* uma média e um desvio padrão dados por $E[\phi_2|y] = -0,6255618$ e $\sqrt{\text{var}(\phi_2|y)} = 0,0247427$. O seu intervalo de credibilidade HPD a 95% dado por $(-0,6741933; -0,5768495)$.

$$\text{Variância do modelo amostral: } \sigma^2|y \sim GI(449; 490,5859).$$

A esperança é dada por $E[\sigma^2|y] = 0,985122$ e o desvio padrão $\sqrt{\text{var}(\sigma^2|y)} = 0,001952608$. O intervalo de credibilidade HPD a 95% é dado por $(0,8998899; 1,0726488)$.

Para encontrar esses resultados, foi utilizado o software R (R DEVELOPMENT CORE TEAM, 2011), usando as seguintes expressões:

\boxtimes Parâmetro $\phi = [\phi_1, \phi_2]'$: Sabemos que $\phi \sim t_\nu(\mu, \Sigma)$, sendo $\nu = n - p$, $\mu = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ e $\Sigma = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, sendo um estimador para σ^2 , $s^2 = (\mathbf{X}'\mathbf{X})^{-1}(n - p)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\phi})'(\mathbf{Y} - \mathbf{X}\hat{\phi})$. Assim,

$$E[\phi|y] = \mu \text{ e } \text{Var}[\phi|y] = \frac{\nu}{\nu - 2}\Sigma.$$

\boxtimes Para σ^2 : Sabemos que $\sigma^2|y \sim GI^a(\alpha, \beta)$, sendo $\alpha = \nu/2$ e $\beta = \nu s^2/2$. Assim,

$$E[\sigma^2|y] = \frac{\beta}{\alpha - 1} \text{ para } \alpha > 1, \beta > 0 \text{ e } \text{Var}[\sigma^2|y] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \text{ para } \alpha > 2, \beta > 0.$$

\boxtimes Para encontrar o HPD para os estimadores dos parâmetros ϕ e σ^2 , geramos amostras a partir da distribuição desses parâmetros. Após isso, usamos funções específicas do software R, como pode ser detalhado no código abaixo.

^aGama Invertida

Código R:

```

> #gerar valores de um AR(2):  $Y(t) = \phi_1 \times Y(t-1) + \phi_2 \times Y(t-2) + e(t)$ 
>
> phi1=0.8 #parametro do modelo AR(2)
> phi2=-0.6 #parametro do modelo AR(2)
> p=2 # num de parametros do modelo AR(2)
> n=1000 # tamanho da amostra
>
>
>
> #erro:
> set.seed(10) #semente
> erro=rnorm(n) #normal padrao
>
> #gerando a serie
> y=numeric(n)
>
> for(t in (p+1):n){
+   y[t]=phi1*y[t-1]+phi2*y[t-2]+erro[t]
+ }
>
> #matriz X:
> X=matrix(0,(n-p),p)
> for(j in 1:p) {
+ for(i in (p+1):n){
+   X[(i-p),j] <- y[(i-j)]
+ }
+ }
>
> Y=y[(p+1):n] #vetor Y
>
> #####
> #matrizes do modelo:
> xlx1 <- solve(t(X)%*%X)
> xlx <- t(X)%*%X
> xly <- t(X)%*%Y
> yly <- t(Y)%*%Y
>
> #####
> #INFERENCIAS PARA PHI:
> #####
> #vetor phi:
> phi <- xlx1%*%xly #vetor phi
> phi1 <- phi[1];phi1 #parametro phi 1
[1] 0.8301865
> phi2 <- phi[2];phi2 #parametro phi 2
[1] -0.6255618
> #####
> #matriz de precisao:
> nu <- length(y)-length(phi);nu # graus de liberdade
[1] 998

```

```

> s2 <- as.vector(t(Y-X%*%phi)%*(Y-X%*%phi)/nu);s2 #var amostral
[1] 0.9831381
> matprec <- diag(s2*solve(xlx));matprec #matriz de precisao
[1] 0.0006109596 0.0006109746
> #####
> #matriz de variancia e covariancia:
> vari <- (nu*s2)/(nu-2)*solve(xlx);vari #variancia
      [,1]      [,2]
[1,] 0.0006121865 -0.0003127031
[2,] -0.0003127031 0.0006122014
> desv <- diag(((nu*s2)/(nu-2)*solve(xlx))^{1/2});desv #desvio padrao
[1] 0.0247424 0.0247427
> #####
> #Intervalos de credibilidade para phi:
>
> #OBS.: Nesse caso, foi gerado uma amostra t_nu(0,1) univariada
> #      para cada parametro do vetor phi, e posteriormente,
> #      feito a transformacao t-tranf=media + desv x t_nu(0,1).
> #      Essa mudanca torna a amostra gerada para mesma media e
> #      variancia da distribuicao t do parametro em estudo.
>
> library(TeachingDemos)#pacote para carregar o comando ''emp.hpd()''
> #phi1:
> set.seed(10) #semente
> ran.phi1=phi1+desv[1]*rt(10^5,nu) #amostra da t_nu(E[beta0],desv[]^2)
> mean(ran.phi1) # esperanca para phi1
[1] 0.8301312
> emp.hpd(ran.phi1,conf=0.95) #intervalo de credibilidade
[1] 0.7815556 0.8788982
>
> #phi2:
> set.seed(10) #semente
> ran.phi2=phi2+desv[2]*rt(10^5,nu) #amostra da t_nu(E[beta1],desv[]^2)
> mean(ran.phi2) # esperanca para phi2
[1] -0.6256171
> emp.hpd(ran.phi2,conf=0.95) #intervalo de credibilidade
[1] -0.6741933 -0.5768495
>
> #####
> #Outra forma de Intervalo de credibilidade para phi:
>
> #OBS.: Nesse caso, geramos a amostra da propria distribuicao
> #      t multivariada, gerando assim, uma matriz com duas
> #      colunas. A primeira, vem de uma amostra com dist
> #      t_nu(phi_1,var1). A segunda vem de uma amostra com dist
> #      t_nu(phi_1,var2).
>
> library(mvtnorm) # pacote para carregar o comando rmvt()
>
> #matriz diagonal de variancia:
> sig=matrix(c(vari[1,1],0,0,vari[2,2]),nrow = 2, ncol = 2)

```

```

>
> #amostra da t multivariada:
> set.seed(10) #semente
> ran.tmult=rmvt(10^5,sigma=sig,delta=phi,df=nu) #amostra t-multivariada
>
> #phi1:
> mean(ran.tmult[,1]) #esperanca para phi1
[1] 0.8301055
> emp.hpd(ran.tmult[,1],conf=0.95) #intervalo de credibilidade
[1] 0.7814045 0.8785823
>
> #phi2:
> mean(ran.tmult[,2]) #esperanca para phi2
[1] -0.6256773
> emp.hpd(ran.tmult[,2],conf=0.95) #intervalo de credibilidade
[1] -0.6744183 -0.5776214
>
> #####
> #INFERENCIAS PARA SIGMA
> #####
> #Parametros da gama-invertida:
> a=nu/2;a
[1] 499
> b=nu*s2/2;b
[1] 490.5859
> #####
> mean.sigma <- (nu*s2)/(nu-2)
> mean.sigma <- b/(a-1);mean.sigma #esperanca de sigma^2
[1] 0.9851122
> var.sigma <- b^2/(((a-1)^2)*(a-2));var.sigma #variancia de sigma^2
[1] 0.001952608
> desv.sigma <- sqrt(b^2/(((a-1)^2)*(a-2)));desv.sigma #desv de sigma^2
[1] 0.04418832
> #####
>
> #Intervalos de credibilidade para sigma:
>
> set.seed(10) #semente
> gamainv=1/rgamma(10^5,a,b) #amostra da gama invertida
> mean(gamainv) #esperanca de sigma^2
[1] 0.9855017
> emp.hpd(gamainv,conf=0.95) #intervalo de credibilidade
[1] 0.8998899 1.0726488

```

6.4 Lista 4

Exercício proposto 6.5: Séries temporais - Inferência Bayesiana

Seja,

$$\begin{aligned}
 Y_t &= \theta_1 Y_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} N(0, \tau^{-1}), \\
 \tau &= 2 = \sigma^{-2} > 0 \stackrel{\text{iid}}{\sim} Ga(\alpha = 1, \beta = 2), \\
 \theta_1 &\stackrel{\text{iid}}{\sim} N(a = 1, b = 2)
 \end{aligned}$$

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_{n-1} \end{pmatrix}}_{\mathbf{X}} \theta_1 + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{\boldsymbol{\epsilon}}$$

$$\mathbf{Y} = \mathbf{X}\theta_1 + \boldsymbol{\epsilon},$$

um modelo autorregressivo $AR(1)$ de uma série temporal.

Seja,

$$\begin{aligned}
 Y_t &= \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} N(0, \tau^{-1}), \\
 \tau &= 2 = \sigma^{-2} > 0 \stackrel{\text{iid}}{\sim} Ga(\alpha = 1, \beta = 2), \\
 \boldsymbol{\theta} &\stackrel{\text{iid}}{\sim} N_2\left(\mathbf{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)
 \end{aligned}$$

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} Y_0 & Y_{-1} \\ Y_1 & Y_0 \\ \vdots & \vdots \\ Y_{n-1} & Y_{n-2} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{\boldsymbol{\epsilon}}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

um modelo autorregressivo $AR(2)$ de uma série temporal.

Use o Amostrador de Gibbs, e estime os parâmetros dos modelos e obtenha representações gráficas das distribuições *a posteriori*.

Solução:

O modelo $AR(1)$ é autorregressivo de primeira ordem e é estacionário se $|\theta_1| < 1$. O modelo $AR(2)$ é autorregressivo de segunda ordem, é estacionário se

$$\begin{cases} \theta_1 + \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1. \end{cases}$$

Para a solução analítica desse problema, vamos usar o modelo autorregressivo geral $AR(p)$, dado por

$$\theta(B)Y_t = \epsilon_t, \quad t = 1, 2, \dots$$

em que ϵ_t é o ruído branco com precisão τ ,

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p,$$

e $B^i Y_t = Y_{t-i}$, $i = 1, 2, \dots, n$. O processo é estacionário se as raízes de $\theta(B) = 0$ estão fora do círculo unitário.

Suponha que há n observações Y_1, Y_2, \dots, Y_n de um processo $AR(p)$, então

$$\theta(B)Y_t = \epsilon_t, \quad t = 1, 2, \dots, n$$

em que y_0, y_{-1}, y_{1-p} são observações iniciais assumidas constantes conhecidas. Desde que $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ sejam *i.i.d.* $N(0, \tau^{-1})$, a função de verossimilhança das n observações é

$$\begin{aligned} f(y_1, y_2, \dots, y_n | \theta, \tau) &\propto \prod_{i=1}^n \tau^{1/2} \exp \left\{ -\frac{\tau}{2} (y_t - \theta_1 y_{t-1} - \dots - \theta_p y_{t-p})^2 \right\} \\ &\propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^n (y_t - \theta_1 y_{t-1} - \dots - \theta_p y_{t-p})^2 \right\}, \end{aligned}$$

em que $\theta_i \in \mathbb{R}$, $i = 1, 2, \dots, p$, $\tau > 0$, e $y_t \in \mathbb{R}$, $t = 1, 2, \dots, n$.

Usando a forma matricial, temos:

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} Y_0 & Y_{-1} & Y_{-2} & \dots & Y_{-p} \\ Y_1 & Y_0 & Y_{-3} & \dots & Y_{1-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{n-1} & Y_{n-2} & Y_{n-3} & \dots & Y_{n-p} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{\boldsymbol{\epsilon}}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}.$$

Assim, a função de verossimilhança de forma matricial é

$$f(\mathbf{Y} | \boldsymbol{\theta}, \tau) \propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \right\}.$$

Considerando as *priori*'s:

$$p(\boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{P}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\}, \quad \boldsymbol{\theta} \in \mathbb{R}^p$$

densidade marginal *a priori* de $\boldsymbol{\theta}$, em que \mathbf{P} a matriz de covariância, e

$$p(\tau) \propto \tau^{\alpha-1} e^{-\tau\beta}, \quad \tau > 0$$

é a densidade marginal *a priori* de τ . Os hiperparâmetros conhecidos são $\boldsymbol{\mu}$, \mathbf{P} , α e β , em que $\boldsymbol{\mu} \in \mathbb{R}^p$, $\alpha > 0$, $\beta > 0$, e \mathbf{P} é a matriz de covariância positiva definida de ordem p .

Considerando ainda que $\boldsymbol{\theta}$ e τ são independentes a distribuição conjunta $P(\boldsymbol{\theta}, \tau) = P(\boldsymbol{\theta}) \times P(\tau)$.

Usando o Teorema de Bayes, a densidade *posteriori* dos parâmetros $\boldsymbol{\theta}$ e τ é

$$\begin{aligned} P(\boldsymbol{\theta}, \tau | \mathbf{Y}) &= f(\mathbf{Y} | \boldsymbol{\theta}, \tau) P(\boldsymbol{\theta}, \tau) \\ &\propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \right\} \times \\ &\quad \times \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{P}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\} \tau^{\alpha-1} e^{-\tau\beta} \\ &\propto \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) + 2\beta] - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \mathbf{P}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right\} \end{aligned}$$

A distribuição condicional $P(\tau | \boldsymbol{\theta}, \mathbf{Y})$ é

$$P(\tau | \boldsymbol{\theta}, \mathbf{Y}) \propto \tau^{\frac{n+2\alpha}{2}-1} \exp \left\{ -\frac{\tau}{2} [(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) + 2\beta] \right\}$$

sendo

$$\tau|\boldsymbol{\theta}, \mathbf{Y} \sim \text{Gama}\left(\frac{n+2\alpha}{2}, \frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})+2\beta}{2}\right).$$

A distribuição condicional $P(\boldsymbol{\theta}|\tau, \mathbf{Y})$ é

$$\begin{aligned} P(\boldsymbol{\theta}|\tau, \mathbf{Y}) &\propto \tau^{\frac{n+p+2\alpha}{2}-1} \exp\left\{-\frac{\tau}{2}[2\beta + \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}] - \frac{1}{2}[\boldsymbol{\theta}'\mathbf{P}^{-1}\boldsymbol{\theta} - \right. \\ &\quad \left. - \boldsymbol{\theta}'\mathbf{P}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{P}^{-1}\boldsymbol{\theta} + \boldsymbol{\mu}'\mathbf{P}^{-1}\boldsymbol{\mu}]\right\} \\ &\propto \exp\left\{-\frac{1}{2}[2\tau\beta + \tau\mathbf{Y}'\mathbf{Y} - (\tau\mathbf{Y}'\mathbf{X})\boldsymbol{\theta} - \boldsymbol{\theta}'(\tau\mathbf{X}'\mathbf{Y}) + \boldsymbol{\theta}'(\tau\mathbf{X}'\mathbf{X})\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{P}^{-1}\boldsymbol{\theta} - \right. \\ &\quad \left. - \boldsymbol{\theta}'\mathbf{P}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{P}^{-1}\boldsymbol{\theta} + \boldsymbol{\mu}'\mathbf{P}^{-1}\boldsymbol{\mu}]\right\} \\ &\propto \exp\left\{-\frac{1}{2}[-(\tau\mathbf{Y}'\mathbf{X})\boldsymbol{\theta} - \boldsymbol{\theta}'(\tau\mathbf{X}'\mathbf{Y}) + \boldsymbol{\theta}'(\tau\mathbf{X}'\mathbf{X})\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{P}^{-1}\boldsymbol{\theta} - \right. \\ &\quad \left. - \boldsymbol{\theta}'\mathbf{P}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{P}^{-1}\boldsymbol{\theta}]\right\} \\ &\propto \exp\left\{-\frac{1}{2}[-(\tau\mathbf{Y}'\mathbf{X} + \boldsymbol{\mu}'\mathbf{P}^{-1})\boldsymbol{\theta} - \boldsymbol{\theta}'(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu}) + \boldsymbol{\theta}'(\tau\mathbf{X}'\mathbf{X} + \mathbf{P}^{-1})\boldsymbol{\theta}]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[[\boldsymbol{\theta} - (\tau(\mathbf{X}'\mathbf{X}) + \mathbf{P}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu})]'[\tau(\mathbf{X}'\mathbf{X} + \mathbf{P}^{-1})] \times \right. \right. \\ &\quad \times [\boldsymbol{\theta} - (\tau(\mathbf{X}'\mathbf{X}) + \mathbf{P}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu})] - \\ &\quad \left. \left. - \underbrace{(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu})(\tau(\mathbf{X}'\mathbf{X}) + \mathbf{P}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu})}_{\text{Como não depende de } \boldsymbol{\theta}, \text{ será eliminado!}}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[[\boldsymbol{\theta} - (\tau(\mathbf{X}'\mathbf{X}) + \mathbf{P}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu})]'[\tau(\mathbf{X}'\mathbf{X} + \mathbf{P}^{-1})] \times \right. \right. \\ &\quad \left. \left. \times [\boldsymbol{\theta} - (\tau(\mathbf{X}'\mathbf{X}) + \mathbf{P}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu})]\right]\right\}, \end{aligned}$$

sendo

$$\boldsymbol{\theta}|\tau, \mathbf{Y} \sim N_p\left((\tau(\mathbf{X}'\mathbf{X}) + \mathbf{P}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{Y} + \mathbf{P}^{-1}\boldsymbol{\mu}), [\tau(\mathbf{X}'\mathbf{X} + \mathbf{P}^{-1})]^{-1}\right)$$

Desses resultados, poderemos usá-los para desenvolver as rotinas do algoritmo amostrador de Gibbs para qualquer p do processo $AR(p)$. A seguir, faremos as rotinas em R para os casos $AR(1)$ e $AR(2)$.

Código R: AR(1)

```
> #####
> #Amostrador de Gibbs para o modelo AR(1)
> #####
> #
> #
> #####
> #pacotes necessarios
> #####
> library(MASS) #carregando mvnrm()
> library(TeachingDemos)#pacote para carregar o comando ''emp.hpd()''
> #####
> #
```

```

> #
> n=1000 # tamanho da série
> theta=0.8 #parametro da serie temporal
> tau=2 #precisão
> p=1 # um parametro para o modelo AR(1)
>
> #gerando a serie do processo AR(1)
> y=numeric(n)
> #
> #####
> #erro da serie estimada
> #####
> #
> erro=rnorm(n,0,sqrt(1/tau))
> for(t in (p+1):n){
+   y[t]=theta*y[t-1]+erro[t]
+ }
> #
> #####
> #matriz X:
> X=matrix(0,(n-p),p)
> for(j in 1:p) {
+   for(i in (p+1):n){
+     X[(i-p),j] <- y[(i-j)]
+   }
+ }
>
> Y=y[(p+1):n] #vetor Y
> #####
> #
> #####
> #parametros da distribuicao normal p/ theta
> #####
> mu=1 #media
> P=2*diag(p);P #matriz covariancia - positiva definida
      [,1]
[1,]    2
> #
> #####
> #parametros da distribuicao gama p/ tau
> #####
> alpha=1
> beta=2
> #
> #
> #####
> #ALGORITMO DE GIBBS
> #####
> #
> #M: num de iteracoes
> M=10000

```

```

> #
> #taunew: valor a ser gerado
> taunew <- 1.5 #valor inicial
> #
> #thetaneew: valor a ser gerado
> thetaneew <- 0.3 #valor inicial
> #
> #
> for(i in 2:M){
+
+   #parametros da condicional de tau (pág. 57):
+   alphanew <- (n+2*alpha)/2
+   betaneew <- (2*beta+t(Y-X**%thetaneew[i-1])**%(Y-X**%thetaneew[i-1]))/2
+
+   #Gerando iterativamente:
+   taunew <- rbind(taunew,rgamma(1,alphanew,betaneew))
+
+   #parametros da Condicional de theta:
+   munew <- solve(taunew[i]*(t(X)**%X)+solve(P)**%
+                 (taunew[i]*(t(X)**%Y)+solve(P)**%mu)
+   sigma <- solve(taunew[i]*(t(X)**%X)+solve(P))
+
+   #Gerando iterativamente:
+   thetaneew <- rbind(thetaneew,mvrnorm(1,munew,sigma))
+
+ }
> #
> #encontrando o valor esperado para tau
> mean(taunew)
[1] 2.047159
> #
> #intervalo de credibilidade para tau
> emp.hpd(taunew,conf=0.95)
[1] 1.869342 2.229360
> #
> #
> #encontrando o valor esperado para theta
> mean(thetaneew)
[1] 0.7895878
> #
> #intervalo de credibilidade para theta
> emp.hpd(thetaneew,conf=0.95)
[1] 0.7496022 0.8266618

```

Código R: AR(1)

```

> #####
> #grafico dos valores estimados dos parametros
> #####
> plot.ts(taunew)

```

Figura 6.1

```
> plot.ts(thetanew)
```

Figura 6.2

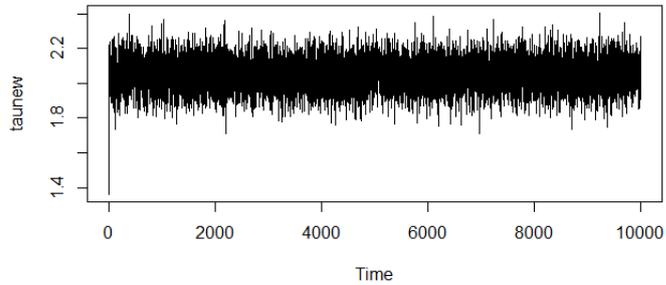


Figura 6.1: Gráfico da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro τ .

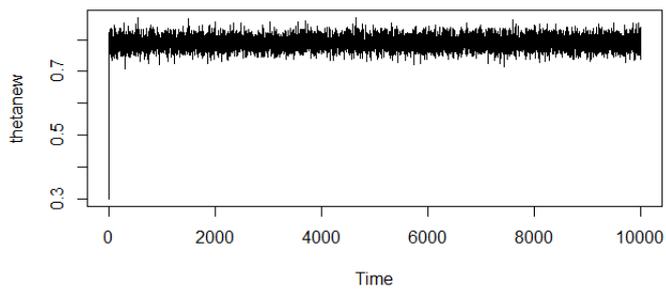


Figura 6.2: Gráfico da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro θ .

Código R: AR(1)

```
#####
#grafico indicando os lags significativos (autocorrelacao)
#####
acf(taunew)
```

Figura 6.3

```
acf(thetanew)
```

Figura 6.4

Código R: AR(2)

```
> #####
> #Amostrador de Gibbs para o modelo AR(2)
> #####
> #
```

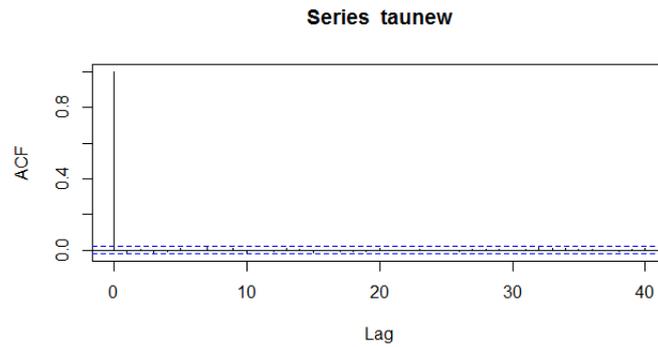


Figura 6.3: Gráfico de autocorrelação da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro τ .

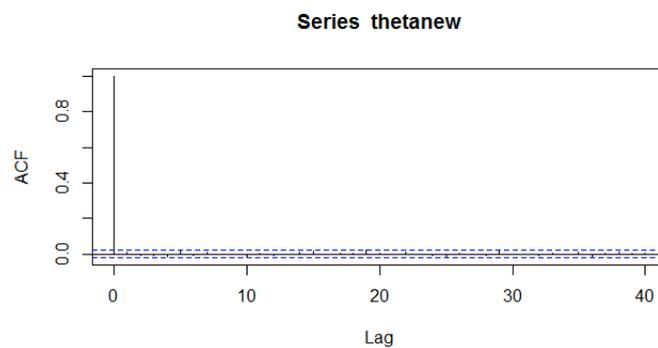


Figura 6.4: Gráfico de autocorrelação da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro θ .

```
> #
> #####
> #pacotes necessarios
> #####
> library(MASS) #carregando mvrnorm()
> library(TeachingDemos)#pacote para carregar o comando ''emp.hpdc()''
> #####
> #
> #
> n=1000 # tamanho da série
> theta1=0.8 #parametro do modelo AR(2)
> theta2=-0.6 #parametro do modelo AR(2)
> tau=3 #precisão
> p=2 # um parametro para o modelo AR(2)
```

Código R: $AR(2)$

```
> #####
> #erro da serie estimada
> #####
> erro=rnorm(n,0,sqrt(1/tau))
> #
```

```

> #####
> #gerando a serie do processo AR(2)
> #####
> y=numeric(n)
> for(t in (p+1):n){
+   y[t]=theta1*y[t-1]+theta2*y[t-2]+erro[t]
+ }
> #
> #####
> #matriz X:
> X=matrix(0,(n-p),p)
> for(j in 1:p) {
+   for(i in (p+1):n){
+     X[(i-p),j] <- y[(i-j)]
+   }
+ }
>
> Y=y[(p+1):n] #vetor Y
> #####
> #
> #####
> #parametros da distribuicao normal p/ theta
> #####
> mu=c(1,1) #media
> P=2*diag(p);P #matriz covariancia - positiva definida
      [,1] [,2]
[1,]    2    0
[2,]    0    2
> #
> #####
> #parametros da distribuicao gama p/ tau
> #####
> alpha=1
> beta=2
> #
> #
> #####
> #ALGORITMO DE GIBBS
> #####
> #
> #M: num de iteracoes
> M=10000
> #
> #taunew: valor a ser gerado
> taunew <- 4 #valor inicial
> #
> #thetaneu: valor a ser gerado
> thetaneu <- rbind(c(2,-1)) #valor inicial
> #
> #
> for(i in 2:M){

```

```

+
+ #parametros da condicional de tau (pág. 57):
+ alphanew <- (n+2*alpha)/2
+ betanew <- (2*beta+t(Y-X%*%thetane[i-1,])%*%
+           (Y-X%*%thetane[i-1,]))*(1/2)
+
+ #Gerando iterativame
+ taunew <- rbind(taunew,rgamma(1,alphanew,betanew))
+
+
+ #parametros da Condicional de theta:
+ munew <- solve(taunew[i]*(t(X)%*%X)+solve(P))%*%
+           (taunew[i]*(t(X)%*%Y)+solve(P))%*%mu
+ sigma <- solve(taunew[i]*(t(X)%*%X)+solve(P))
+
+ #Gerando iterativamente:
+ thetanew <- rbind(thetane, mvrnorm(1, munew, sigma))
+
+ }
> #
> #encontrando o valor esperado para tau
> mean(taunew)
[1] 3.033506
> #
> #intervalo de credibilidade para tau
> emp.hpd(taunew,conf=0.95)
[1] 2.761919 3.289148
> #
> #
> #encontrando o valor esperado para phi
> apply(thetane,2,mean)
[1] 0.8085643 -0.5885005
> apply(thetane,2,emp.hpd,conf=0.95)
      [theta1]      [theta 2]
[1,] 0.7575882   -0.6409099 #Limite inferior
[2,] 0.8582705   -0.5390941 #Limite superior

```

Código R: AR(2)

```

> #####
> #grafico dos valores estimados dos parametros
> #####
> plot.ts(taunew)

Figura 6.5

> plot.ts(thetane)

Figura 6.6

```

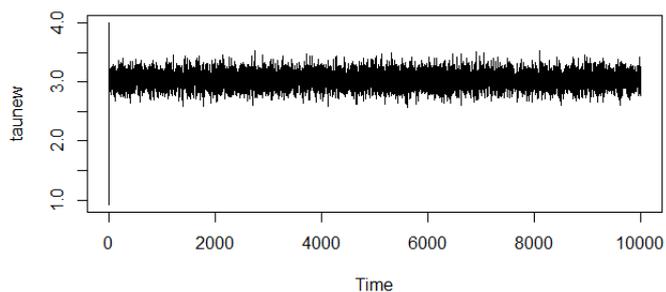


Figura 6.5: Gráfico da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro τ do modelo AR(2).

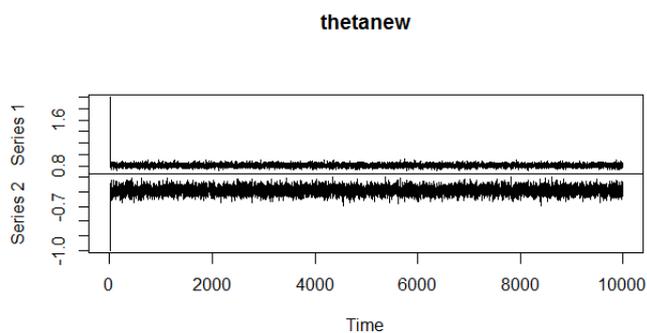


Figura 6.6: Gráfico da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro θ do modelo AR(2).

Código R: AR(2)

```
> #####
> #grafico indicando os lags significativos (autocorrelacao)
> #####
> acf(taunew)
```

Figura 6.7

```
> acf(thetaneu[,1])
```

Figura 6.8

Código R: AR(2)

```
> acf(thetaneu[,2])
```

Figura 6.9

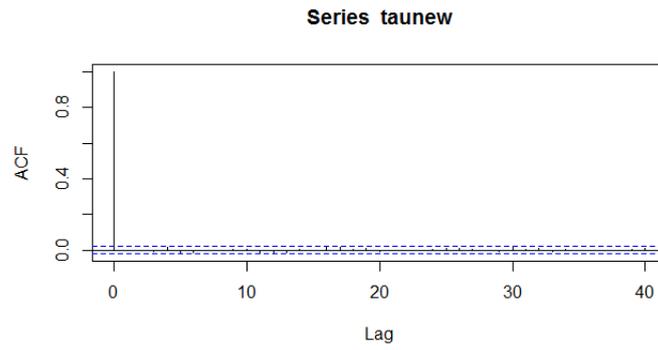


Figura 6.7: Gráfico de autocorrelação da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro τ do modelo AR(2).

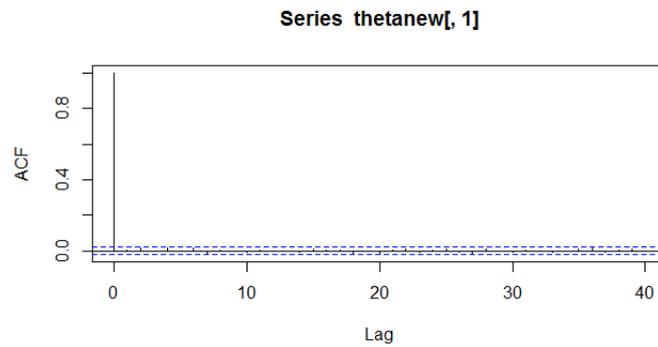


Figura 6.8: Gráfico de autocorrelação da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro θ_1 do modelo AR(2).

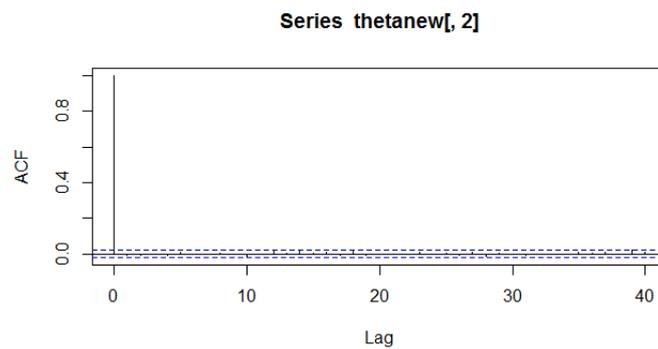


Figura 6.9: Gráfico de autocorrelação da cadeia gerada pelo algoritmo Amostrador de Gibbs para o parâmetro θ_2 do modelo AR(2).

6.5 Lista 5

Exercício proposto 6.6

Encontre as marginais:

$$P(\boldsymbol{\theta}|\mathbf{Y}) = \int_D P(\boldsymbol{\theta}, \tau|\mathbf{Y})d\tau$$

$$P(\tau|\mathbf{Y}) = \int_D P(\boldsymbol{\theta}, \tau|\mathbf{Y})d\boldsymbol{\theta}$$

distribuições dos parâmetros do modelo $AR(p)$, utilizando os algoritmos MCMC para gerar amostras dessas distribuições por meio das condicionais completas $P(\boldsymbol{\theta}|\tau, \mathbf{Y})$ e $P(\tau|\boldsymbol{\theta}, \mathbf{Y})$ encontradas no **Lista 4**. Assim, use o algoritmo de Gibbs para gerar a cadeia de Markov da qual a distribuição de equilíbrio é $P(\boldsymbol{\theta}|\mathbf{Y})$ e o algoritmo de Metropolis Hastings para gerar a cadeia de Markov da qual a distribuição de equilíbrio é $P(\tau|\mathbf{Y})$ (Estamos considerando a título de ilustração que $P(\tau|\boldsymbol{\theta}, \mathbf{Y})$ não tem distribuição conhecida e utilizaremos a distribuição $N(0, 1)$ como distribuição auxiliar).

Código R: $AR(2)$

```
> #####
> #Amostrador de Gibbs e Hastings Metropolis para o modelo AR(2)
> #####
> #OBS.: No algoritmo quando n é muito grande, exige que o num
> # de iterações (M) seja também. A escolha da função auxiliar
> # é extremamente importante. Observe que meu tau=1. Assim
> # estou usando a func aux |X|~N(0,1). Se esse tau fosse
> # 3, a N(0,1) não seria a mais adequada, teríamos que
> # variar os parâmetros para varrer o dominio de forma
> # que ficasse o mais prox possivel do valor real para tau
> #
> #
> #####
> # limpando a memória
> rm(list=ls())
> #####
> #
> #####
> #pacotes necessarios
> #####
> library(MASS) #carregando mvrnorm()
> library(TeachingDemos)#pacote para carregar o comando ''emp.hpd()''
> #####
> #
> #
> n=100 # tamanho da série
> theta1=0.8 #parametro do modelo AR(2)
> theta2=-0.6 #parametro do modelo AR(2)
> tau=1 #precisão
> p=2 # um parametro para o modelo AR(2)
> #
> #####
> #erro da serie estimada
> #####
```

```

> #
> set.seed(1) #semente
> erro=rnorm(n,0,sqrt(1/tau))
> #
> #####
> #gerando a serie do processo AR(2)
> #####
> y=numeric(n)
> for(t in (p+1):n){
+   y[t]=theta1*y[t-1]+theta2*y[t-2]+erro[t]
+ }
> #
> #####
> #matriz X:
> X=matrix(0,(n-p),p)
> for(j in 1:p) {
+   for(i in (p+1):n){
+     X[(i-p),j] <- y[(i-j)]
+   }
+ }
>
> Y=y[(p+1):n] #vetor Y
> #####
> #
> #####
> #parametros da distribuicao normal p/ theta
> #####
> mu <- c(1,1) #media
> P <- 2*diag(p);P #matriz covariancia - positiva definida
      [,1] [,2]
[1,]    2    0
[2,]    0    2
> #
> #####
> #parametros da distribuicao gama p/ tau
> #####
> alpha <- 1
> beta <- 2
> #
> #####
> #INICIO DO ALGORITMO
> #####
> #M: num de iteracoes
> M=10000
> #
> #taunew: valor a ser gerado
> taunew <- 2 #valor inicial
> #
> #thetaneu: valor a ser gerado
> thetanew <- rbind(c(2,-1)) #valor inicial (não precisa ser gerado)
> #

```

```

> #
> #Taxa:
> taxa <- 0
> #
> for (i in 2:M) {
+ #####
+ #ALGORITMO DE GIBBS
+ #####
+ #
+ #parametros da Condicional de theta (pág. 58):
+ munew <- solve(taunew[i-1]*(t(X)%*%X)+solve(P))%*%
+   (taunew[i-1]*(t(X)%*%Y)+solve(P))%*%mu
+ sigma <- solve(taunew[i-1]*(t(X)%*%X)+solve(P))
+ #Gerando iterativamente:
+ thetanew <- rbind(thetanew,mvrnorm(1,munew,sigma))
+ #
+ #####
+ #ALGORITMO DE HASTINGS E METOPOLIS
+ #####
+ #
+ #soma de quadrado
+ sqd <- function(theta){
+   sq = t(Y-X)%*%theta)%*%(Y-X)%*%theta)
+   return(sq)
+ }
+ #
+ #gerando tau de uma distribuição auxiliar N(0,1)
+ taunewc=abs(rnorm(1)) #tau candidato com protecao |tau|~N(0,1) > 0
+ #
+ #gerando uma uniforme
+ u=runif(1)
+ #
+ #Gerando valores da condicional de tau
+ pscale <- (n+2*alpha)/2
+ pshape <- sqd(thetanew[i,])+2*beta
+ PN= taunewc^(pscale-1)*exp(-taunewc/2*(pshape))
+ PA= taunew[i-1]^(pscale-1)*exp(-taunew[i-1]/2*(pshape))
+ #
+ #Gerando valores da distribuição auxiliar N(0,1)
+ qN = exp((-taunewc^2)/(2))
+ qA = exp((-taunew[i-1]^2)/(2))
+ #
+ a=min(1,(PN*qA)/(PA*qN))
+ #
+ if(u<=a){
+   taunew <- rbind(taunew,taunewc)
+   taxa=taxa+1 # conta quantas vezes a cadeia recebeu um novo valor
+ }
+ else taunew <- rbind(taunew,taunew[i-1])
+ }
> #

```

```

> #####
> #Verificando a taxa de aceites da func aux
> #####
> taxa
[1] 2032
> #
> #
> #####
> #####Média e intervalo de Credibilidade#####
> #####
> #
> #Tau (Média):
> mean(taunew)
[1] 1.187192
> #Tau (IC):
> emp.hpd(taunew,conf=0.95)
[1] 0.8664238 1.5104527
> #
> #Theta (Média):
> apply(thetanew,2,mean)
[1] 0.8563314 -0.6208276
> #Theta (IC:
> apply(thetanew,2,emp.hpd,conf=0.95)
      [,1]      [,2]
[1,] 0.695588 -0.7823651
[2,] 1.017716 -0.4579176
> #

```

Código R: AR(2) - Gráficos para o parâmetro τ

```

> #####
> #####Gráficos#####
> #####
> #
> #Tau (Série):
> plot.ts(taunew)
> #Tau (Densidade):
> plot(density(taunew))

```

Figura 6.10

```

> #Tau (Densidade):
> plot(density(taunew))

```

Figura 6.11

Código R: AR(2) - Gráficos de autocorrelação para τ

```

> #
> #####
> #grafico indicando os lags significativos (autocorrelacao)
> #####
> #Tau:

```

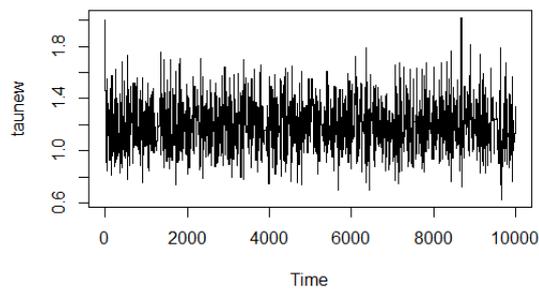


Figura 6.10: Gráfico da cadeia gerada pelo algoritmo para o parâmetro τ do modelo AR(2).

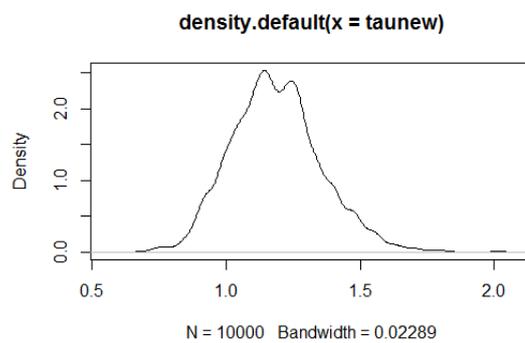


Figura 6.11: Gráfico da densidade da cadeia gerada pelo algoritmo para o parâmetro τ do modelo AR(2).

```
> acf(taunew)
```

Figura 6.12

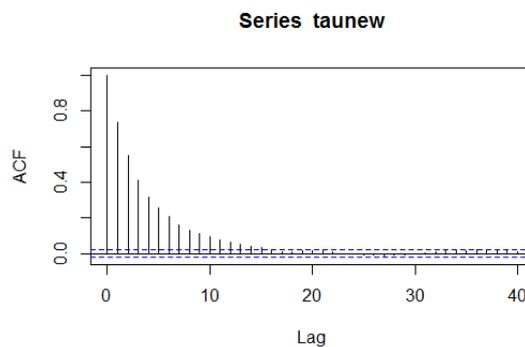


Figura 6.12: Gráfico de autocorrelação da cadeia gerada pelo algoritmo para o parâmetro τ do modelo AR(2).

Código R: AR(2) - Gráficos de autocorrelação para θ

```
> #####
> #Theta (Série):
> plot.ts(thetanew)
```

Figura 6.13

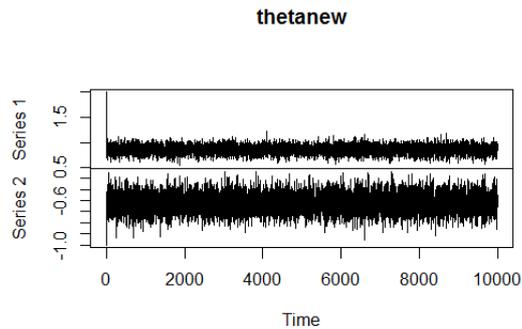


Figura 6.13: Gráfico de autocorrelação da cadeia gerada pelo algoritmo para o parâmetro τ do modelo AR(2).

Código R: AR(2) - Gráfico da densidade de θ

```
> #Theta (Densidade):
> plot(density(thetanew))
```

Figura 6.14

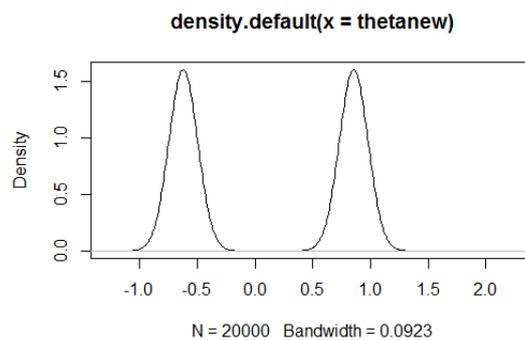


Figura 6.14: Gráfico da densidade da cadeia gerada pelo algoritmo para o parâmetro τ do modelo AR(2).

Código R: AR(2) - Gráficos de autocorrelação de θ_1

```
> #Theta 1:
> acf(thetanew[,1])
```

Figura 6.15

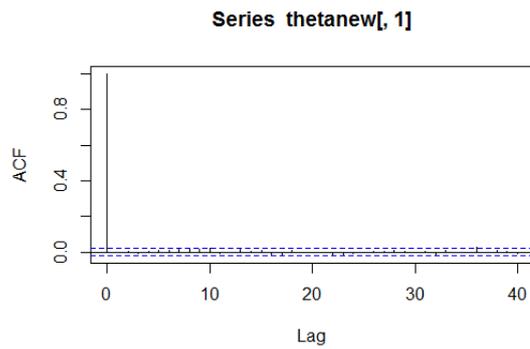


Figura 6.15: Gráfico de autocorrelação da cadeia gerada pelo algoritmo para o parâmetro θ_1 do modelo AR(2).

Código R: AR(2) - Gráfico de autocorrelação para θ_2

```
> #Theta 2:
> acf(thetaneu[,2])
```

Figura 6.16

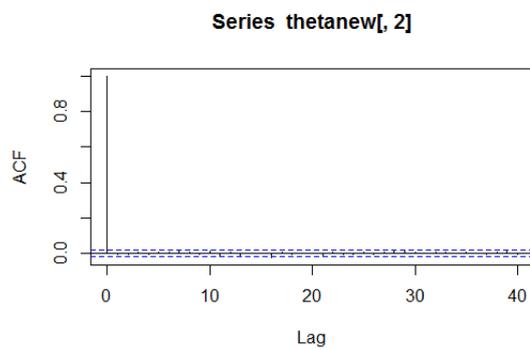


Figura 6.16: Gráfico de autocorrelação da cadeia gerada pelo algoritmo para o parâmetro θ_2 do modelo AR(2).

Código R: AR(2) - Critério de convergência

```
#####
#####Verificando a Convergência#####
#####
parametros <- cbind(thetanew,taunew)
#
#
#Pacote para verificacao da convergencia da cadeia
library(boa)
#Entrando no menu do pacote
boa.menu()

#Passos para inserir os dados
1 #- File
3 #- Import Data >>
5 #- Data Matrix Objects
parametros #Enter object name [none]
#Passos para a convergencia:
1 #back
1 #back
3 #- Analysis
4 #- Convergence Diagnostics >>
6 #Raftery & Lewis
Press <ENTER> to continue
Enter
1 #back
1 #back
#esc
```

Código R: AR(2) - Critério Raftery & Lewis

```
RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:
=====
Quantile = 0.025
Accuracy = +/- 0.005
Probability = 0.95

Chain: parametros
-----

      Thin Burn-in Total Lower Bound Dependence Factor
par1   1       2  3802      3746          1.014949
par2   1       2  3757      3746          1.002936
par3   6      24 25278      3746          6.747998
```

Código R: AR(2) - Critério de convergência

```
> #####
> # 0 maior burn-in=24 foi devido a anlise de convergencia
> #feita por Raftery & Lewis. Retirando os burn-in's
> parburn <- parametros[25:M,]
```

```

> dados <-parburn
> #####
> #Função para a nova cadeia
> sample.thin <- function(dados,thin){
+ samplethin <-matrix(0,1,dim(dados)[2])
+ #amostra em intervalos thin
+ for(i in 1:dim(dados)[1]){
+ if(i*thin > dim(dados)[1]) stop
+ else samplethin <- rbind(samplethin,dados[(i*thin),])
+ }
+ return(samplethin)
+ }
> #####
> # 0 thin=6 foi devido a anlise de convergencia feita
> # por Raftery & Lewis.
> cadeia <- sample.thin(dados,thin=6)

```

Código R: AR(2) - Valor esperado e HPD

```

> #####
> #Valor esperado e HPD:
> #####
> #Tau (Média):
> mean(cadeia[,3])
[1] 1.18546
> #Tau (IC):
> emp.hpd(cadeia[,3],conf=0.95)
[1] 0.867649 1.518575
> #Theta (Média):
> apply(cadeia[,1:2],2,mean)
[1] 0.8571488 -0.6230248
> #Theta (IC):
> apply(cadeia[,1:2],2,emp.hpd,conf=0.95)
      [theta-1]  [theta-2]
[1,] 0.7046454 -0.7766698 #Limite Inferior
[2,] 1.0188207 -0.4558816 #Limite Superior

```

6.6 Lista 6

Exercício proposto 6.7: Exemplo 8.2: (PAULINO; TURKMAN; MURTEIRA, 2003, p. 334)

Seja (Y, X) um par aleatório em que Y condicional a $X = x$ segue uma distribuição Poisson com valor médio $\lambda(x) = \eta\delta^x$ ($\eta, \delta > 0$) e X segue uma distribuição Normal com valor médio μ e precisão $\tau = 1/\sigma^2$ (representa-se neste caso $N_0(\mu, \tau)$).

Solução:

Suponha-se ainda que se têm dados, do modelo considerando, da forma $\mathcal{D} =$

$\{(y_1, x_1), \dots, (y_n, x_n)\}$. A verossimilhança é

$$f(x, y|\boldsymbol{\theta}) = \prod_{i=1}^n \frac{[\eta\delta^{x_i}]^{y_i}}{y_i!} e^{-\eta\delta^{x_i}} \left[\frac{\tau}{2\pi}\right]^{1/2} \exp\left\{-\frac{\tau}{2}(x_i - \mu)^2\right\}$$

onde $\boldsymbol{\theta} = (\eta, \delta, \mu, \tau)'$ para $\eta, \sigma, \tau > 0$ e $-\infty < \mu < +\infty$.

Se se considerar *a priori* não informativa

$$h(\eta, \delta, \mu, \tau) \propto \frac{1}{\eta\delta\tau}$$

então a função densidade de probabilidade *a posteriori* para $\boldsymbol{\theta}$ é:

$$\begin{aligned} h(\boldsymbol{\theta}|\mathcal{D}) &\propto f(x, y|\boldsymbol{\theta}) \times h(\eta, \delta, \mu, \tau) \\ &\propto \prod_{i=1}^n \frac{[\eta\delta^{x_i}]^{y_i}}{y_i!} e^{-\eta\delta^{x_i}} \left[\frac{\tau}{2\pi}\right]^{1/2} \exp\left\{-\frac{\tau}{2}(x_i - \mu)^2\right\} \times \frac{1}{\eta\delta\tau} \\ &\propto \eta^{\sum_{i=1}^n y_i} \delta^{\sum_{i=1}^n x_i y_i} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \tau^{n/2} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} \times \frac{1}{\eta\delta\tau} \\ &\propto \eta^{\sum_{i=1}^n y_i - 1} \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \tau^{n/2 - 1} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} \\ &\propto \tau^{n/2 - 1} \eta^{\sum_{i=1}^n y_i - 1} \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (x_i^2 - 2x_i\mu + \mu^2)\right\} \\ &\propto \tau^{n/2 - 1} \eta^{\sum_{i=1}^n y_i - 1} \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \exp\left\{-\frac{\tau}{2} \left(\sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i\mu + n\mu^2\right)\right\} \\ &\propto \tau^{n/2 - 1} \eta^{\sum_{i=1}^n y_i - 1} \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \exp\left\{-\frac{\tau}{2} \left(\sum_{i=1}^n x_i^2 - 2\frac{n}{n} \sum_{i=1}^n x_i\mu + \right.\right. \\ &\quad \left.\left.+ n\mu^2\right)\right\} \\ &\propto \tau^{n/2 - 1} \eta^{\sum_{i=1}^n y_i - 1} \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \exp\left\{-\frac{\tau}{2} \left(\sum_{i=1}^n x_i^2 - 2n\bar{x}\mu + n\mu^2\right)\right\}. \end{aligned}$$

Considerando, $n(\mu - \bar{x})^2 = n(\mu^2 - 2\mu\bar{x} + \bar{x}^2) = n\mu^2 - 2n\mu\bar{x} + n\bar{x}^2$, então $-2n\bar{x}\mu + n\mu^2 = n(\mu - \bar{x})^2 - n\bar{x}^2$. Assim,

$$h(\boldsymbol{\theta}|\mathcal{D}) \propto \tau^{n/2 - 1} \eta^{\sum_{i=1}^n y_i - 1} \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \exp\left\{-\frac{\tau}{2} \left(\sum_{i=1}^n x_i^2 + n(\mu - \bar{x})^2 - n\bar{x}^2\right)\right\}.$$

As distribuições condicionais completas são:

$$h(\eta|\mathcal{D}, \delta, \mu, \tau) \propto \eta^{\sum_{i=1}^n y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\} \sim Ga\left(\sum_{i=1}^n y_i, \sum_{i=1}^n \delta^{y_i}\right), \quad \eta > 0,$$

$$h(\delta|\mathcal{D}, \eta, \mu, \tau) \propto \delta^{\sum_{i=1}^n x_i y_i - 1} \exp\left\{-\eta \sum_{i=1}^n \delta^{y_i}\right\},$$

$$\begin{aligned}
 h(\mu|\mathcal{D}, \eta, \delta, \tau) &\propto \exp \left\{ -\frac{\tau}{2} \left(\underbrace{\sum_{i=1}^n x_i^2 + n(\mu - \bar{x})^2 - n\bar{x}^2}_{n(\mu - \bar{x})^2} \right) \right\} \\
 &\propto \exp \left\{ -\frac{\tau}{2} (n(\mu - \bar{x})^2) \right\} \\
 &\propto \exp \left\{ -\frac{n\tau}{2} (\mu - \bar{x})^2 \right\} \sim N_o(\bar{x}, n\tau), \quad \infty < \mu < +\infty,
 \end{aligned}$$

$$\begin{aligned}
 h(\tau|\mathcal{D}, \eta, \delta, \mu) &\propto \tau^{n/2-1} \exp \left\{ -\frac{\tau}{2} \left(\underbrace{\sum_{i=1}^n x_i^2 + n(\mu - \bar{x})^2 - n\bar{x}^2}_A \right) \right\} \\
 &\propto \tau^{n/2-1} \exp \left\{ -\tau \left(\frac{A}{2} \right) \right\} \sim Ga \left(\frac{n}{2}, \frac{A}{2} \right), \quad \tau > 0.
 \end{aligned}$$

Como as distribuições de η , μ e τ são conhecidas, será usado o algoritmo de GIBBS. No caso da distribuição do parâmetro σ não ser uma distribuição comum, será usado o algoritmo Metropolis-Hastings.

Código R:

```

> ##### Exercício 2.8 (Paulino et.al. 2009, pág. 316)#####
>
> #Parametros
> eta1=2
> sigma1=3
> mu1=0
> tau1=2
> n=50
> #####
> #Variavel X ~ N(mu,1/tau)
> set.seed(10)
> X <- rnorm(n,mu1,sqrt(1/tau1))
> #
> #Variavel Y Y|X ~ Pois(eta . sigma^x)
> Y <- rpois(n,(eta1*sigma1^X))
>
> #####
> #Rotina:
> #####
>
> # número de iterações da cadeia
> M <- 50000
>
>
> #Valores simulados:
> etas=numeric(M)
> sigmas=numeric(M)

```

```

> mus=numeric(M)
> taus=numeric(M)
>
>
> #calculos necessarios:
> sumy <-sum(Y)
> meanx <- mean(X)
> funceta <- function(etas) sum(etas^X)
> funcmu <- function(mu) (sum(X^2)-n*mean(X)^2+n*(mu-mean(X))^2)/2
>
>
>
> #chute inicial para sigma:
> sigmas[1] <- 5
>
> #chute inicial para mu:
> mus[1] <- 1
>
> #chute inicial para tau:
> taus[1] <- 1
>
>
> #Algoritmo GIBBS e HASTINGS-METROPOLIS:
>
> taxa=0
> for (i in 2:M) {
+   # gerando os valores de eta:
+   etas[i] <- rgamma(1,sumy,sum(sigmas[i-1]^X))
+
+   # gerando os valores de mu:
+   mus[i] <- rnorm(1,mean(X),sqrt(1/(taus[i-1]*n)))
+
+   # gerando os valores de tau:
+   taus[i] <- rgamma(1,n/2,(sum(X^2)-n*mean(X)^2+n*(mus[i]-mean(X))^2)/2)
+
+
+   # gerando os valores de sigma:
+   sigmac=rgamma(1,8,3) #candidato
+   u=runif(1)
+
+   PN=sigmac^(sum(X*Y)-1)*exp(-etas[i]*sum(sigmac^X))
+   qN=dgamma(sigmac,8,3)
+   PA=sigmas[i-1]^(sum(X*Y)-1)*exp(-etas[i]*sum(sigmas[i-1]^X))
+   qA=dgamma(sigmas[i-1],8,3)
+   razao=(PN*qA)/(PA*qN)
+
+   a=min(1,razao)
+
+   if(u<=a)
+     {sigmas[i]<-sigmac

```

```

+   taxa<-taxa + 1 }   # conta quantas vezes a cadeia recebeu um novo valor
+
+   if(u>a)
+     {sigmas[i]<-sigmas[i-1]}
+ }
> #
> #Taxa de aceitação:
> taxa
[1] 25426
> #
> #
> #Verificando a convergência:
> parametros <- cbind(taus,sigmas,mus,etas)
> #
> #
> #Pacote para verificacao da convergencia da cadeia
> library(boa)
> #Entrando no menu do pacote
> boa.menu()
BOA MAIN MENU
*****

1: File           >>
2: Data Management >>
3: Analysis       >>
4: Plot           >>
5: Options        >>
6: Window         >>

Selection: #Passos para inserir os dados
Enter an item from the menu, or 0 to exit
Selection: 1 #- File

FILE MENU
=====

1: Back
2: -----+
3: Import Data       >> |
4: Save Session      |
5: Load Session     |
6: Exit BOA         |
7: -----+

Selection: 3 #- Import Data >>

IMPORT DATA MENU
-----

1: Back
2: -----+

```

```

3: CODA Output Files      |
4: Flat ASCII Files      |
5: Data Matrix Objects   |
6: View Format Specifications |
7: Options...            |
8: -----+

```

Selection: 5 #- Data Matrix Objects

Enter object name [none]

1: parametros #Enter object name [none]

Read 1 item

+++ Object successfully imported +++

IMPORT DATA MENU

```

1: Back
2: -----+
3: CODA Output Files      |
4: Flat ASCII Files      |
5: Data Matrix Objects   |
6: View Format Specifications |
7: Options...            |
8: -----+

```

Selection: 1 #back

FILE MENU

=====

```

1: Back
2: -----+
3: Import Data          >> |
4: Save Session        |
5: Load Session       |
6: Exit BOA           |
7: -----+

```

Selection: 1 #back

BOA MAIN MENU

```

1: File                >>
2: Data Management    >>
3: Analysis           >>
4: Plot               >>
5: Options            >>
6: Window             >>

```

Selection: 3 #- Analysis

ANALYSIS MENU

=====

- 1: Back
- 2: -----+
- 3: Descriptive Statistics >> |
- 4: Convergence Diagnostics >> |
- 5: -----+

Selection: 4 #- Convergence Diagnostics >>

CONVERGENCE DIAGNOSTICS MENU

- 1: Back
- 2: -----+
- 3: Brooks, Gelman, & Rubin |
- 4: Geweke |
- 5: Heidelberger & Welch |
- 6: Raftery & Lewis |
- 7: Options... |
- 8: -----+

Selection: 6 #Raftery & Lewis

RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:

=====

Quantile = 0.025
 Accuracy = +/- 0.005
 Probability = 0.95

Chain: parametros

	Thin	Burn-in	Total	Lower Bound	Dependence Factor
etas	2	4	7960	3746	2.1249333
mus	1	2	3796	3746	1.0133476
sigmas	3	9	11646	3746	3.1089162
taus	1	2	3713	3746	0.9911906

Press <ENTER> to continue

Enter

CONVERGENCE DIAGNOSTICS MENU

- 1: Back
- 2: -----+

```

3: Brooks, Gelman, & Rubin |
4: Geweke                   |
5: Heidelberger & Welch    |
6: Raftery & Lewis         |
7: Options...              |
8: -----+

```

Selection: 1 #back

ANALYSIS MENU

=====

```

1: Back
2: -----+
3: Descriptive Statistics >> |
4: Convergence Diagnostics >> |
5: -----+

```

Selection: 1 #back

BOA MAIN MENU

```

1: File           >>
2: Data Management >>
3: Analysis       >>
4: Plot           >>
5: Options        >>
6: Window         >>

```

Selection:

```

> #Retirando os burn-in's
> # 0 burn-in=9 foi devido a analise de convergencia feita por Raftery & Lewis.
> parburn <- parametros[10:M,]
> dados <- parburn
> #####
> #Função para a nova cadeia
> sample.thin <- function(dados,thin){
+ samplethin <- matrix(0,1,dim(dados)[2])
+ #amostra em intervalos thin
+ for(i in 1:dim(dados)[1]){
+ if(i*thin > dim(dados)[1]) stop
+ else samplethin <- rbind(samplethin,dados[(i*thin),])
+ }
+ return(samplethin)
+ }
> #Outra funcao para a nova cadeia
>
> #####
> # 0 thin=3 foi devido a anlise de convergencia feita por Raftery & Lewis.

```

```

> cadeia <- sample.thin(dados,thin=3)
> #
> #####
> #Passos para analises descritivas e HPD:
> #####
> #Pacote para verificacao da convergencia da cadeia
> library(boa)
> #Entrando no menu do pacote
> boa.menu()
BOA MAIN MENU
*****

1: File          >>
2: Data Management >>
3: Analysis     >>
4: Plot         >>
5: Options      >>
6: Window       >>

Selection: 1 #- File

FILE MENU
=====

1: Back
2: -----+
3: Import Data      >> |
4: Save Session    |
5: Load Session    |
6: Exit BOA        |
7: -----+

Selection: 3 #- Import Data >>

IMPORT DATA MENU
-----

1: Back
2: -----+
3: CODA Output Files |
4: Flat ASCII Files  |
5: Data Matrix Objects |
6: View Format Specifications |
7: Options...        |
8: -----+

Selection: 5 #- Data Matrix Objects

Enter object name [none]
1: cadeia #Enter object name [none]
Read 1 item

```

```
+++ Object successfully imported +++
```

```
IMPORT DATA MENU
```

```
-----
```

```
1: Back
2: -----+
3: CODA Output Files      |
4: Flat ASCII Files      |
5: Data Matrix Objects   |
6: View Format Specifications |
7: Options...            |
8: -----+
```

```
Selection: 1 #back
```

```
FILE MENU
```

```
=====
```

```
1: Back
2: -----+
3: Import Data           >> |
4: Save Session         |
5: Load Session        |
6: Exit BOA             |
7: -----+
```

```
Selection: 1 #back
```

```
BOA MAIN MENU
```

```
*****
```

```
1: File                 >>
2: Data Management     >>
3: Analysis             >>
4: Plot                 >>
5: Options              >>
6: Window               >>
```

```
Selection: 3 #- Anlysis
```

```
ANALYSIS MENU
```

```
=====
```

```
1: Back
2: -----+
3: Descriptive Statistics >> |
4: Convergence Diagnostics >> |
5: -----+
```

```
Selection: 3 #- Descriptive Statistics >>
```

DESCRIPTIVE STATISTICS MENU

- ```

-----+
1: Back
2: -----+
3: Autocorrelations |
4: Correlation Matrix |
5: Highest Probability Density Intervals |
6: Summary Statistics |
7: Options... |
8: -----+

```

Selection: 6 #- Summary Statistics

SUMMARY STATISTICS:

=====

Bin size for calculating Batch SE and (Lag 1) ACF = 50

Chain: cadeia

-----

|        | Mean       | SD         | Naive SE     | MC Error     | Batch SE     | Batch ACF    |
|--------|------------|------------|--------------|--------------|--------------|--------------|
| etas   | 1.6351512  | 0.19152416 | 0.0014836585 | 0.0016608097 | 0.0015879089 | 0.0258022770 |
| mus    | -0.1199086 | 0.09141444 | 0.0007081498 | 0.0006313475 | 0.0006665936 | 0.0005568948 |
| sigmas | 2.9078874  | 0.54273469 | 0.0042043412 | 0.0047098527 | 0.0047871262 | 0.0263543192 |
| taus   | 2.4913424  | 0.50348477 | 0.0039002883 | 0.0034665756 | 0.0038314291 | 0.0120354028 |
|        | 0.025      | 0.5        | 0.975        | MinIter      | MaxIter      | Sample       |
| etas   | 1.2837256  | 1.628412   | 2.02721301   | 1            | 16664        | 16664        |
| mus    | -0.3001567 | -0.119012  | 0.05975814   | 1            | 16664        | 16664        |
| sigmas | 1.9964926  | 2.857415   | 4.09950749   | 1            | 16664        | 16664        |
| taus   | 1.6126061  | 2.454618   | 3.58376947   | 1            | 16664        | 16664        |

Press <ENTER> to continue

Enter

DESCRIPTIVE STATISTICS MENU

-----

- ```

-----+
1: Back
2: -----+
3: Autocorrelations |
4: Correlation Matrix |
5: Highest Probability Density Intervals |
6: Summary Statistics |
7: Options... |
8: -----+
    
```

Selection: 5 #- Highest Probability Density Intervals

HIGHEST PROBABILITY DENSITY INTERVALS:

```
=====
```

```
Alpha level = 0.05
```

```
Chain: parametros
```

```
-----
```

```

          Lower Bound Upper Bound
etas      1.277814  2.01589192
mus       -0.297999  0.06049555
sigmas    1.933374  4.01535721
taus      1.571162  3.51358398
```

6.7 Lista 7

Exercício proposto 6.8: Paulino, Turkman e Murteira (2003, p. 216)

Aplicação à análise de regressão linear simples

Medição da radiação de fundo, através de contagens Y num detector de Geiger por minutos, em 39 locais com altitudes x (em metros) diferenciadas.

Y	11.0	12.7	11.6	12.8	11.8	11.5	11.1	13.6
x	213.4	426.8	487.8	518.3	579.3	640.2	701.2	731.7
Y	12.5	13.9	15.7	17.8	17.2	14.4	15.0	13.8
x	884.1	884.1	1097.6	1097.6	1128.0	1158.5	1158.5	1219.5
Y	15.8	12.4	13.9	18.0	14.9	16.8	17.3	17.9
x	1250.0	1311.0	1463.4	1524.4	1554.9	1554.9	1554.9	1554.9
Y	18.4	19.6	20.1	16.6	19.3	22.0	19.8	27.8
x	1554.9	1554.9	1554.9	1646.3	1646.3	1890.2	1951.2	2103.7
Y	29.0	20.4	25.0	26.8	22.7	24.9	26.0	
x	2103.7	2164.6	2256.1	2256.1	2530.5	3231.7	3231.7	

- Admitindo-se o modelo de regressão linear simples sumarie-se as inferências sobre os parâmetros do modelo relativas ao uso da distribuição *a priori* não informativa usual.
- Compare-se sob o modelo anterior inferências sobre a radiação média num local a 200m com as predições sobre a média das radiações a obter em m locais com a mesma altitude nos casos $m = 1$ e $m = 5$.

OBS.: Use o WINBUGS para encontrar os mesmos resultados encontrados na lista 2.

Solução:

No WINBUGS será preciso simplesmente definir as *prioris* dos parâmetros em questão. Paulino, Turkman e Murteira (2003, p. 216) usou uma *priori* não informativa usual

$$h(\beta, \sigma^2) \propto (\sigma^2)^{-1} \quad \beta \in \mathbb{R}^p, \sigma^2 > 0.$$

Só que nesse caso, será usado as seguintes *priori*'s:

$$\begin{aligned}\beta_0 &\sim N(0, \tau), & \tau &= 1/\sigma^2 = 0.000001 \\ \beta_1 &\sim N(0, \tau), & \tau &= 1/\sigma^2 = 0.0001 \\ \tau &\sim \text{Gamma}(0.1, 0.1)\end{aligned}$$

No WINBUGS tem que ser observado que os parâmetros da normal são a média e a precisão (e não o desvio padrão, como de costume no R). Assim, as *priori*'s para os parâmetros β_0 e β_1 , adotou-se uma variância muito alta para que a distribuição não desse muitas informações sobre o valor médio, ou seja, fosse não informativa. Para o parâmetro τ , usou-se a gama.

Código WINBUGS: Modelo

```
model
{
for( i in 1 : N ) {
y[i] ~ dnorm(mu[i],tau)
mu[i] <- beta0 + beta1 * (x[i])
}
beta0 ~ dnorm(0,0.000001)
beta1 ~ dnorm(0,0.0001)
tau ~ dgamma(0.1,0.1)
sigma <- 1 / sqrt(tau)
}
```

Código WINBUGS: Entrando com os dados

```
list(x = c(213.4,426.8,487.8,518.3,579.3,640.2,701.2,731.7,
884.1,884.1,1097.6,1097.6,1128,1158.5,1158.5,1219.5,
1250,1311,1311,1463.4,1524.4,1554.9,1554.9,1554.9,
1554.9,1554.9,1554.9,1646.3,1646.3,1890.2,1951.2,2103.7,
2103.7,2164.6,2256.1,2256.1,2530.5,3231.7,3231.7), N = 39,

y = c(11.0,12.7,11.6,12.8,11.8,11.5,11.1,13.6,
12.5,13.9,15.7,17.8,17.2,14.4,15,13.8,
15.8,12.4,13.9,18,14.9,16.8,17.3,17.9,
18.4,19.6,20.1,16.6,19.3,22,19.8,27.8,
29,20.4,25,26.8,22.7,24.9,26))
```

Para a instalação do WINBUGS, será preciso o arquivo `WinBUGS14_immortality_key.txt`. Após seguir os passos e registrá-lo, o WINBUGS está pronto para o uso.

Como já foi mostrado o modelo e a entrada de dados, já que é um padrão do programa, segue os passos no WINBUGS:

1. Crie um arquivo em **File > New** e insira os códigos WINBUGS: **modelo** e **entrando com os dados**;
2. Salve o arquivo em **File > Save**. Nesse exemplo, o arquivo foi salvo com o nome `Daniel.p.216.odc` no diretório: `D:/DISCIPLINAS - 2013.2/BAYESIAN INFERENCE/WinBUGS`;
3. Após salvo o arquivo, segue (Figura 6.17):
4. Na rotina selecione o nome `model` e clique em `check model` (Figura 6.18):

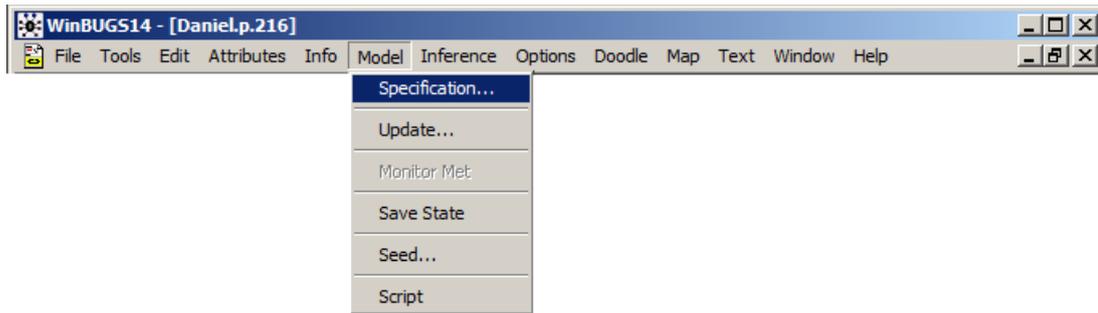


Figura 6.17: Specification

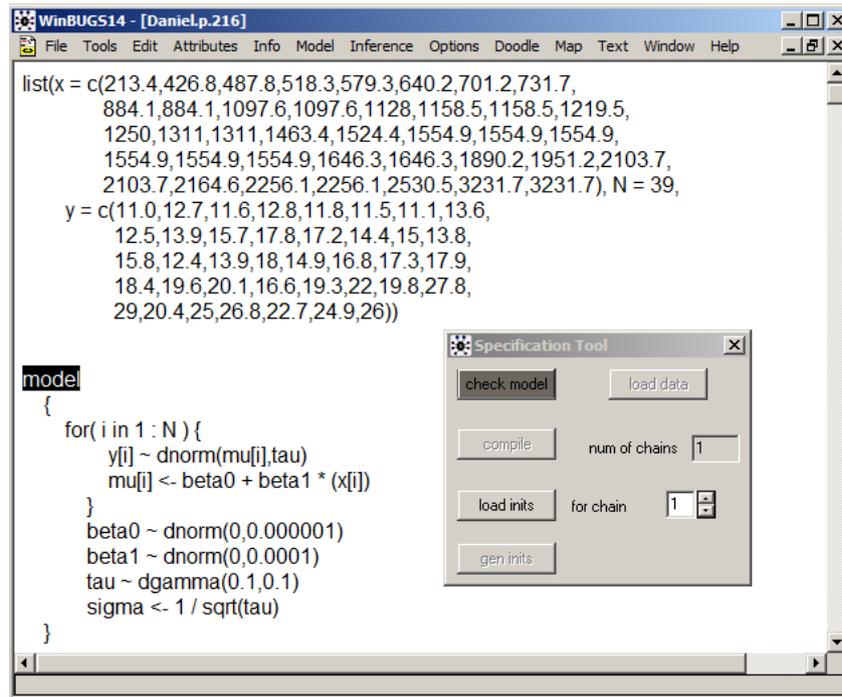


Figura 6.18: Check model

5. Carregue os dados, selecionando `list` na rotina e clicando em `load data` (Figura 6.19):
6. Após isso, compila a rotina e os dados: botão `compile`;
7. Gerando dados iniciais para dar início a cadeia: botão `gen inits`;
8. Sem fechar a janela `specification tool`, clique em `Inference > Samples...` (Figura 6.20);
9. No espaço `node`, acrescente os parâmetros que se deseja gerar a cadeia e após clique em `set`:
 - a) parâmetro β_0 (`beta0`). Exemplo na Figura 6.21: Logo após, clicar em `set`;
 - b) parâmetro β_1 (`beta1`), clica em `set`;
 - c) parâmetro σ^2 (`sigma`), clica em `set`;
10. Sem fechar a janela `Sample Monitor Tool`, no menu em `Model`, clique em `Update` (Figura 6.22);
11. Em `updates` escolhe-se como aquecimento, 1000, e após clique em `update`. Esse valor representa, 1000 elementos em cada cadeia dos três parâmetros em estudo (Figura 6.23);

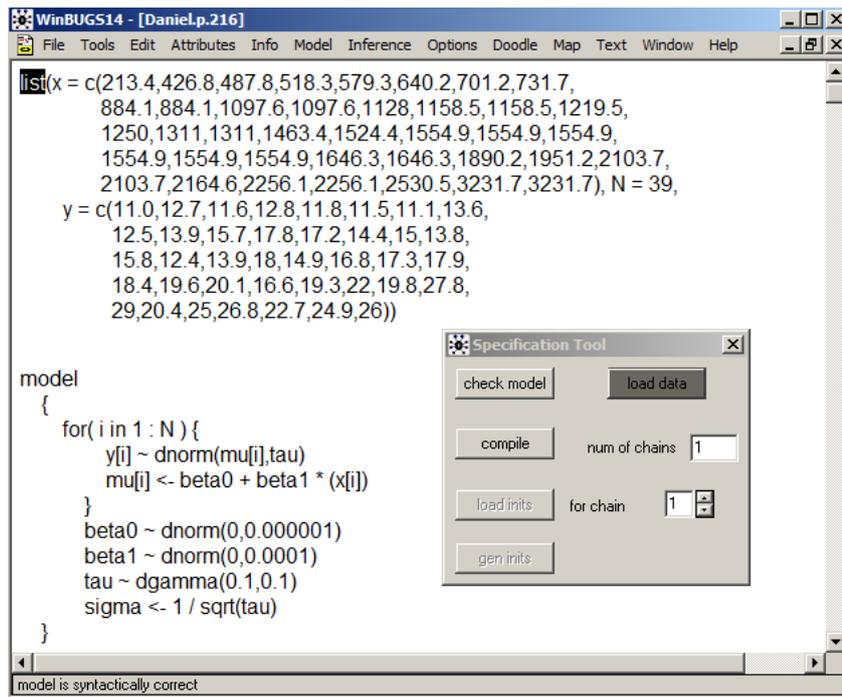


Figura 6.19: Load data

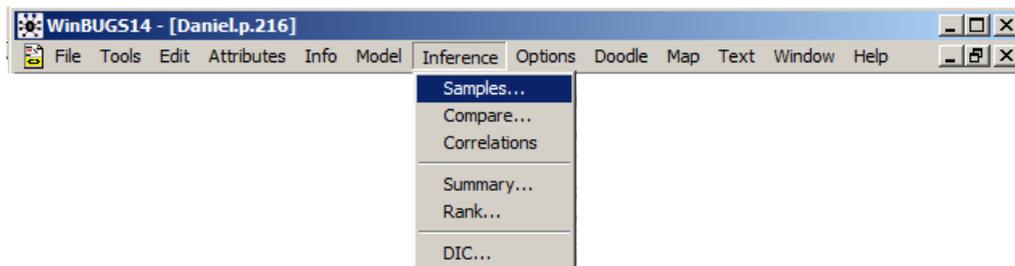


Figura 6.20: Inference

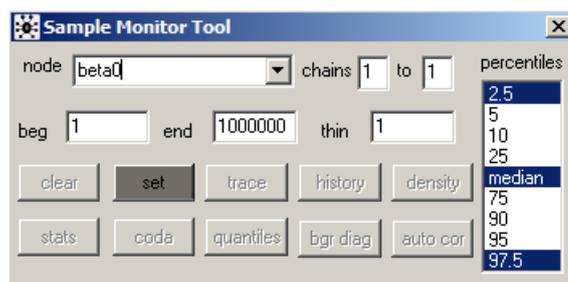


Figura 6.21: Inserindo β_0

12. Posteriormente, acrescenta-se mais 10000 iterações (Figura 6.24):
13. Assim, foi gerado um total de 11000 iterações;
14. Após isso, retornando a janela **Sample Monitor Tool**, na opção **node**, escolha o parâmetro desejado para se obter as estatísticas desejadas (Figura 6.25):
15. Depois de selecionado, clique em **stats**, será fornecido as estatísticas básicas de cada parâmetro (Figura 6.27):

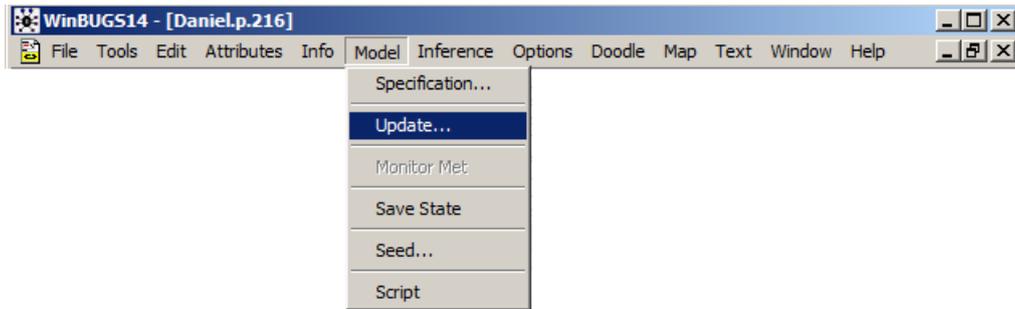


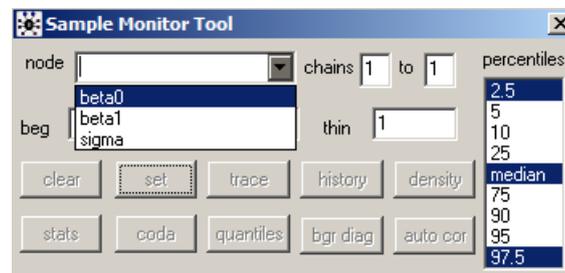
Figura 6.22: Update



Figura 6.23: Aquecimento da cadeia



Figura 6.24: Gerando 10000 iterações

Figura 6.25: Selecionando β_0

node	mean	sd	MC error	95.0%	start	sample
beta0	8.583	2.386	0.0311	10.08	1	11000

node	mean	sd	MC error	95.0%	start	sample
beta1	0.006176	0.002247	2.434E-5	0.00716	1	11000

node	mean	sd	MC error	95.0%	start	sample
sigma	2.539	2.71	0.03508	3.044	1	11000

Figura 6.26: Estatísticas dos parâmetros

16. Esses resultados estão baseados nas 11000 iterações, sendo que não foi verificado a autocorrelação dos elementos, nem a eliminação das iterações iniciais para o aquecimento da cadeia. Assim, será necessário realizar o critério de convergência Raftery e Lewis (1992). Para o seu uso, será utilizado o pacote *boa* do software R. Dessa forma, tem-se que exportar a cadeia do

WINBUGS para o R.

- Para obter a cadeia de cada parâmetro no WINBUGS, na janela **Sample Monitor Tool**, depois de selecionado o parâmetro em **node**, clique em **coda** e será mostrado toda a cadeia (Figura 6.28):

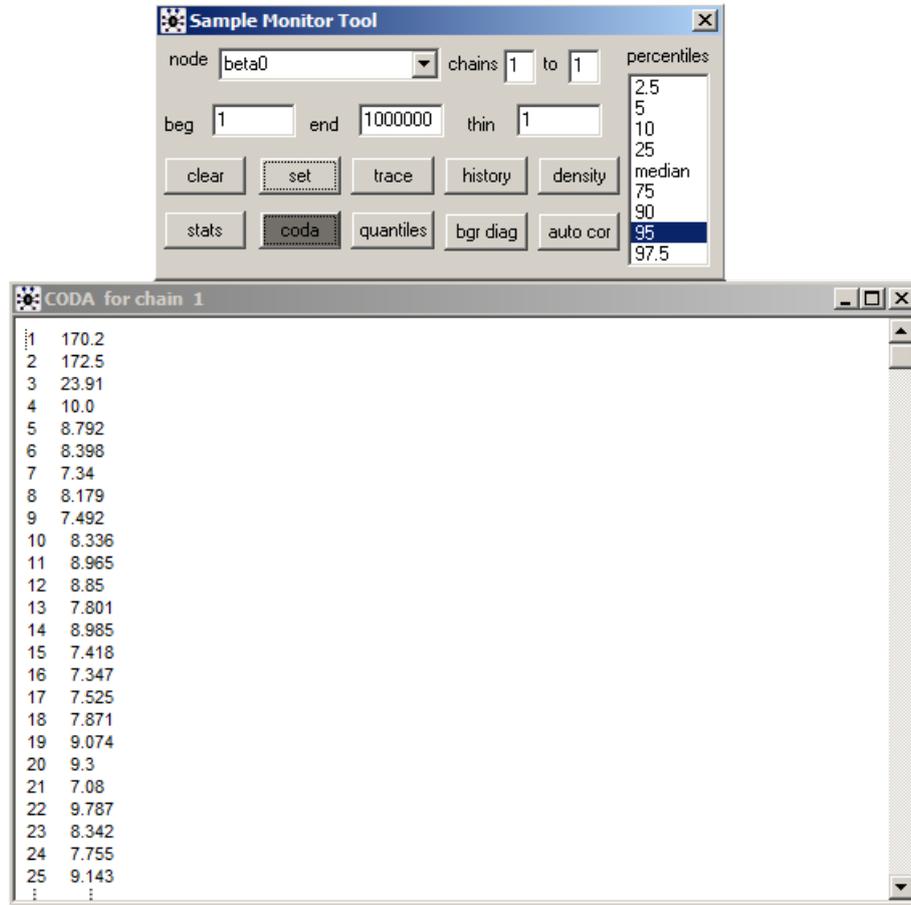


Figura 6.27: Cadeia para o exemplo do parâmetro β_0

- Será utilizado o Excel para armazenar as cadeias num arquivo com a extensão “.txt”, chamado **cadeias.txt**. Nesse arquivo, será armazenado as três cadeias dos parâmetros β_0 , β_1 e σ^2 , como mostrado na Figura 6.28;
- Usando o R...

Código R:

```
> ####Usando R#####
> #Mudando o diretório:
> setwd("D:/DISCIPLINAS - 2013.2/BAYESIAN INFERENCE/WinBUGS")
> #Carregando os dados:
> dados <- read.table("cadeias.txt",h=T)
> #Carregando o pacote ''boa'':
> library(boa)
> #Entrando no menu:
> boa.menu()

BOA MAIN MENU
*****
```

cadeias.txt - Bloco de notas				
Arquivo	Editar	Formatar	Exibir	Ajuda
beta0	beta1		sigma	
7.999	-0.001786		16.06	
8.007	0.003982		5.313	
8.002	0.006275		2.375	
8.001	0.006439		2.482	
8.001	0.006455		2.256	
7.994	0.006332		3.692	
7.994	0.006471		2.467	
8.003	0.006703		2.634	
7.997	0.006673		2.559	
8.002	0.006604		2.171	
8.003	0.006446		2.493	
8.006	0.006607		2.337	
8.003	0.0068		2.337	
7.996	0.006211		2.438	
7.996	0.006673		2.337	
8.002	0.006893		2.396	
7.997	0.006694		2.364	
8.003	0.006786		2.492	
8.003	0.006443		2.754	
8.011	0.006726		2.506	
7.999	0.006796		2.428	
7.998	0.006018		2.791	
7.996	0.006375		3.092	
7.998	0.006591		2.53	
8.006	0.006522		2.72	
8.002	0.006265		2.087	
8.004	0.006294		2.422	
7.999	0.006691		2.457	
⋮	⋮		⋮	

Figura 6.28: Arquivo “cadeias.txt”

```

1: File           >>
2: Data Management >>
3: Analysis       >>
4: Plot           >>
5: Options        >>
6: Window         >>

```

```
Selection: 1
```

```
FILE MENU
```

```
=====
```

```

1: Back
2: -----+
3: Import Data   >> |
4: Save Session |
5: Load Session |
6: Exit BOA     |
7: -----+

```

```
Selection: 3
```

```
IMPORT DATA MENU
```

```
-----
```

```

1: Back
2: -----+

```

```

3: CODA Output Files      |
4: Flat ASCII Files      |
5: Data Matrix Objects   |
6: View Format Specifications |
7: Options...            |
8: -----+

Selection: 5

Enter object name [none]
1: dados
Read 1 item
+++ Object successfully imported +++

IMPORT DATA MENU
-----

1: Back
2: -----+
3: CODA Output Files      |
4: Flat ASCII Files      |
5: Data Matrix Objects   |
6: View Format Specifications |
7: Options...            |
8: -----+

Selection: 1

FILE MENU
=====

1: Back
2: -----+
3: Import Data           >> |
4: Save Session          |
5: Load Session         |
6: Exit BOA              |
7: -----+

Selection: 1

BOA MAIN MENU
*****

1: File                 >>
2: Data Management     >>
3: Analysis             >>
4: Plot                 >>
5: Options              >>
6: Window               >>

```

```

Selection: 3

ANALYSIS MENU
=====

1: Back
2: -----+
3: Descriptive Statistics >> |
4: Convergence Diagnostics >> |
5: -----+

Selection: 4

CONVERGENCE DIAGNOSTICS MENU
-----

1: Back
2: -----+
3: Brooks, Gelman, & Rubin |
4: Geweke |
5: Heidelberger & Welch |
6: Raftery & Lewis |
7: Options... |
8: -----+

Selection: 6

RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:
=====

Quantile = 0.025
Accuracy = +/- 0.005
Probability = 0.95

Chain: dados
-----

      Thin Burn-in Total Lower Bound Dependence Factor
beta0   1       2  6133      3746      1.637213
beta1   1       2  3824      3746      1.020822
sigma   1       2  3804      3746      1.015483

```

20. Observou-se que não houve autocorrelação $\text{Thin}=1$, e deve-se eliminar os 2 primeiros elementos da cadeia $\text{Burn-in}=2$. Depois de verificado essas observações, e retornando ao WINBUGS à janela *Sample Monitor Tool*, inserindo $\text{beg}=3$ (já que se deve eliminar os dois primeiros elementos da cadeia. beg é equivalente ao Burn-in) e $\text{thin}=1$ para cada um dos parâmetros em estudo (Figura 6.29), segue o próximo passo;
21. Clicando em *stats*, pode-se obter as novas estatísticas sem os dados de aquecimento da cadeia e sem autocorrelação (Figura 6.30):



Figura 6.29: Exemplo com o parâmetro β_0

node	mean	sd	MC error	95.0%	start	sample
beta0	8.553	0.9365	0.009267	10.07	3	10998

node	mean	sd	MC error	95.0%	start	sample
beta1	0.0062	5.808E-4	5.836E-6	0.00716	3	10998

node	mean	sd	MC error	95.0%	start	sample
sigma	2.505	0.3128	0.003318	3.043	3	10998

Figura 6.30: Novas estatísticas

Referências Bibliográficas

- CHARNET, R. et al. **Análises de Modelos de Regressão Linear - Com Aplicações**. 2. ed. [S.l.]: Editora UNICAMP, 1999. 356 p.
- GELMAN, A.; MENG, X. L. Model checking and model improvement. In: _____. London: Chapman and Hall, 1996. cap. Markov Chain Monte Carlo in Practice, p. 189–202.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, n. 4, p. 457–511, 1992.
- MAGALHÃES, M. N. **Probabilidade e Variáveis Aleatórias**. 2. ed. [S.l.]: Editora USP, 2006. 411 p.
- NOGUEIRA, D. A. **Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov caso uni e multivariado**. Tese (Mestrado) — Universidade Federal de Lavras, Minas Gerais, 2004.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian/Calouste Gulbenkian, 2003. 442 p.
- R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2011. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.
- RAFTERY, A. E.; LEWIS, S. Comment: One long run with diagnostics - implementation strategies for markov chain monte carlo. **Statistical Science**, v. 7, n. 4, p. 493–497, 1992.
- TIERNEY, L.; KADANE, J. B. Accurate approximations for posterior moments and marginal densities. **Journal of the American Statistical Association**, American Statistical Association, v. 81, n. 393, p. pp. 82–86, 1986. ISSN 01621459. Disponível em: <<http://www.jstor.org/stable/2287970>>.