

# Estimação de parâmetros de populações de plantas daninhas usando inferência Bayesiana

MARCEL RODRIGUES LOPES

*Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica*

**Orientadora: Profa. Dra. Vilma Alves de Oliveira**

# **Livros Grátis**

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

Milhares de livros grátis para download.

**À meus pais Maria Dulce e Diógenes, por toda força e carinho que sempre me deram...**

**A minha avó Dulce Dalmiglio, pelo exemplo de vida e dedicação...**

**Ao mestre Takamori que iluminou meu caminho esclarecendo o verdadeiro objetivo da vida....**

## Agradecimentos

---

**Todo trabalho científico é resultante de um trabalho em equipe. Assim, algum mérito que este trabalho possa apresentar é dedicado àqueles que direta e indiretamente colaboraram em sua realização, em especial...**

**À minha família, pelo amor, carinho, dedicação e por terem me ajudado nos bons e maus momentos da minha vida.**

**À minha orientadora Professora Vilma Alves de Oliveira por toda paciência e dedicação.**

**A todos amigos do Laboratório de Contole (LAC), em especial a Glaucia, o Fabio, a Lilian, o Wilson, o Alexandre, o Maurílio e o Rafael.**

**Aos meus amigos da minha cidade, no qual eu tenho o orgulho de ter a amizade, Marinho, Fábio, Luciano, Fernando Frare, Fernando Fonseca, Daniel, Nelson, Nélio e**

# Conteúdo

<b>Resumo</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>1 Introdução</b>	<b>1</b>
. Dinâmica de comunidades de plantas daninhas . . . . .	
. Representação matemática da dinâmica populacional . . . . .	
. Objetivos e organização do trabalho . . . . .	
<b>2 Inferência Estatística</b>	<b>7</b>
. O pensamento clássico . . . . .	
. O pensamento Bayesiano . . . . .	
. Distribuições a priori informativas . . . . .	6
. . . Informação a priori com base em um especialista . . . . .	
. Distribuições a priori não-informativas . . . . .	
. . . Método de Bayes-Laplace . . . . .	6
. . . Método de Jeffreys . . . . .	
. Métodos de Monte Carlo e Simpson . . . . .	
. . . Monte Carlo simples . . . . .	
. . . Monte Carlo via função de importância . . . . .	
. . . Método numérico de Simpson . . . . .	6
. . . Regra de Simpson . . . . .	
.6 Métodos de Monte Carlo via cadeias de Markov . . . . .	
.6. Cadeias de Markov . . . . .	6
.6. Algoritmo de Metropolis-Hastings . . . . .	
.6. Amostrador de Gibbs . . . . .	
. Estimador de máxima verossimilhança via MCMC . . . . .	
. . . Famílias de densidades normalizadas . . . . .	

. Validação e seleção de modelos . . . . .	
. . . Medidas de diagnóstico - medida de adequabilidade de um modelo ‘‘	
. . . Comparação de modelos . . . . .	
. . . Seleção de modelos via métodos MCMC . . . . .	
<b>3 Modelos para populações de plantas daninhas</b>	<b>43</b>
. Modelo de regressão não-linear . . . . .	
. Densidade a priori para os parâmetros . . . . .	6
. Densidades a posteriori para os parâmetros . . . . .	
. Distribuição a posteriori do modelo de parâmetros . . . . .	
. . . Uma densidade a priori não-informativa para $\theta = (s, a)$ e $\sigma$ . . . . .	
. . . Densidade a posteriori conjunta para $\theta$ e $\sigma$ . . . . .	‘‘
. . . Densidades a posteriori marginais para $s$ e $a$ . . . . .	‘‘
. Distribuição a posteriori do modelo de parâmetros . . . . .	
. . . Uma densidade a priori não-informativa para $\theta = (s, a, b)$ e $\sigma$ . . . . .	
. . . Densidade a posteriori conjunta para $\theta(s, a, b)$ e $\sigma$ . . . . .	
. . . Densidades a posteriori marginais para $s, a$ e $b$ . . . . .	
.6 MCMC para os parâmetros e seleção dos modelos . . . . .	
.6. Algoritmo MCMC . . . . .	
.6. Densidades preditivas ordenadas (DPO) . . . . .	
.6. Métodos de seleção AIC e BIC . . . . .	
.6. MCMC em estimador de máxima verossimilhança . . . . .	6
<b>4 Resultados</b>	<b>57</b>
. Inferência Bayesiana via método de Simpson . . . . .	
. Inferência Bayesiana via método MCMC . . . . .	
. Inferência clássica via método de máxima verossimilhança usando MCMC	
. Discussão dos resultados . . . . .	
. . . Método numérico de Simpson . . . . .	
. . . Taxas de rejeição e convergência das cadeias . . . . .	‘‘
. . . Seleção de modelos . . . . .	‘‘
. . . Estimador de máxima verossimilhança usando o MCMC . . . . .	6
<b>Conclusão</b>	<b>73</b>
<b>A Noções de probabilidade</b>	<b>74</b>
A. Eventos . . . . .	

---

A.	Frequência relativa . . . . .	
A.	Probabilidade condicionada e eventos independentes . . . . .	
A. .	Eventos independentes . . . . .	
<b>B</b>	<b>Algoritmo para o cálculo das posterioris pelo método de Simpson</b>	<b>79</b>
B.	Modelo de parâmetros . . . . .	
B. .	Parâmetro $a$ . . . . .	
<b>Bibliografia</b>		<b>88</b>

# Resumo

O banco de sementes de plantas daninhas pode vir a ser um sério problema para a atividade agrícola por manter infestações por longos períodos. A dinâmica da população de plantas daninhas pode ser representada por modelos matemáticos que relaciona as densidades de sementes produzidas e de plântulas em áreas de cultivo. Os valores dos parâmetros dos modelos podem ser inferidos diretamente de experimentação e análise estatística. O presente trabalho tem por objetivo estimar parâmetros de populações das plantas daninhas anuais denominadas *Digitaria ciliares*, *Panicum maximum* e *Euphorbia heterophylla* e avaliar um modelo espacial com dois parâmetros, a partir de um experimento conduzido em cultura de *Zea mays* (milho) usando inferência Bayesiana.

**Palavras-chave:** Dinâmica populacional, Inferência Bayesiana, Priori não-informativa, Densidade preditiva, Monte Carlo em cadeia de Markov.

# Abstract

*The seeds bank of weed could be a serious problem to the agricultural activity because it maintains infestations for long periods. The dynamics of weed populations can be described by mathematical models that relates the produced seeds and seedlings densities in areas of crop. The values of the parameters of the models can be inferred from experimentation and statistics analysis. The objective of this work is to estimate parameters of anual weed populations denoted *Digitaria ciliares*, *Panicum maximum* e *Euphorbia heterophylla* and evaluate a spatial model with 2 and 3 parameters from experimental data of *Zea mays* (corn) fields using Bayesian inference.*

# Capítulo 1

## Introdução

### 1.1 Dinâmica de comunidades de plantas daninhas

Tem havido um grande interesse na simulação do impacto do uso de herbicidas para controle de populações de plantas daninhas visando a identificação do nível de aplicação adequado. E, os modelos matemáticos têm se tornado ferramentas valiosas para a simulação de diferentes situações das variáveis do problema (Doyle, 1982). Existem diversos exemplos de modelagem de população ecológica em termos de dinâmica não-linear. A população de plantas daninhas apresenta um ciclo de vida que começa com a germinação no solo, seguida por crescimento vegetativo, floração, produção de semente, difusão de semente, sobrevivência ou mortalidade de sementes no solo, e, germinação no ano seguinte (Sakai, 1996). Devido à existência de sementes dormentes no solo, a população enterrada possui sementes de diferentes gerações. A densidade das plântulas (plantas que germinaram) daninhas afeta fatores ecológicos como taxa de mortalidade e nascimento, fecundidade e produção. Em plantas, um crescimento na densidade inicial resulta em alta taxa de mortalidade e baixa taxa de crescimento, baixa fecundidade e decréscimo de produção de sementes. Isto acontece porque aumenta a competição por recursos como água, luz e nutrientes.

### 1.2 Representação matemática da dinâmica populacional

Sakai (1996) descreveu a dinâmica populacional de plantas daninhas através de fatores dependentes e independentes da densidade de plantas daninhas e obteve o número de

sementes por área (densidade de sementes) em sucessivos anos a partir do número de sementes do ano inicial observado. A densidade de sementes existentes no ciclo  $t + 1$  é determinada pela densidade de sementes do ciclo anterior  $t$ , sendo descrita por

$$x_{t+1} = gosx_t + (1 - g)vx_t \quad ( . )$$

onde  $t$  é o ciclo de vida das plantas,  $x$  é o número de sementes por área,  $g$ ,  $o$ ,  $s$  e  $v$  são as taxas de germinação, mortalidade, produtividade (número de sementes produzidas por planta) e de sementes vivas no solo no ano seguinte, respectivamente, com  $g$ ,  $o$  e  $s$  constantes. Com exceção da taxa de produtividade  $s$ , as taxas  $o$  e probabilidades  $e$  encontram-se entre zero e um. O segundo termo de ( . ) representa as gerações sobrepostas e quando as sementes no solo não persistirem durante um ano tem-se  $v = 0$ . O modelo ( . ) considera somente a densidade da população em intervalos de uma geração individual a cada ciclo e é por isto denominado modelo de um ciclo estacionário. A densidade de plantas de daninhas denotada  $y$  pode ser descrita por

$$y_t = gy_t \quad ( . )$$

onde  $y$  é o número de plantas por área e  $g$  parâmetro que representa a taxa de germinação de sementes introduzidas. A densidade de plantas influencia parâmetros ecológicos devido ao efeito densidade, isto é, a taxa de crescimento da população, a população em pleno vigor vegetativo, a mortalidade e o número de sementes produzidas são funções da densidade de plantas de daninhas, já a taxa de germinação é independente da densidade. Assim,  $o$  e  $s$  em ( . ) não são constantes, mas, funções da densidade de plantas de daninhas. Desta forma, um modelo para a densidade de sementes produzidas pode ser expresso, por exemplo, como função da densidade de plantas de daninhas

$$x_{t+1} = F(y_t), t = 1, 2, \dots \quad ( . )$$

onde  $F: \mathfrak{R} \rightarrow \mathfrak{R}$ . Logo, a dinâmica populacional de plantas daninhas pode ser descrita por ( . ) e ( . ) que relacionam as densidades de sementes produzidas e de plantas. Note que os sistemas ( . ) e ( . ) representam a independência e dependência da taxa de crescimento populacional da densidade de plantas, respectivamente. Baseado nestas equações, a relação entre a densidade de sementes  $x_t$  e  $x_{t+1}$  pode ser expresso por um

sistema dinâmico discreto com o número de sementes no solo sendo a variável de estado  $x$

$$x_{t+1} = F(gx_t), t = 1, 2, \dots \quad ( . )$$

Wu (1966) apresenta a dinâmica populacional de plantas daninhas em função da densidade de plantas controladas com aplicação de herbicida, cuja forma funcional utilizada para a função resposta de taxa de herbicida é a exponencial. Neste caso, o modelo de densidade de plantas ( . ) é modificado para

$$y_t = gx_t e^{-hH_t} \quad ( . )$$

onde agora  $y$  é a densidade de plantas de daninhas sobreviventes à aplicação de herbicida,  $h$  a taxa de eficiência do herbicida e  $H$  a dosagem de herbicida que pode ser interpretada como a toxicidade do herbicida usado. Das daninhas sobreviventes ao herbicida algumas produzem sementes e um novo ciclo é iniciado. Para simplificar, Wu (1966) considerou que todas as sementes ou germinaram ( $g = 1$ ) ou morreram ( $v = 0$ ), isto é, se  $g = 1$  ou  $v = 0$ , então,  $(1 - g)v x_t = 0$  e que pode haver produção de sementes a partir das daninhas sobreviventes à aplicação de herbicida ( $sy_t$ ). Como Wu (1966) não se refere à taxa de mortalidade, considera-se  $o = 1$  em ( . ). Sob estas condições, a densidade de sementes no banco é dada por

$$x_{t+1} = sy_t = sgx_t e^{-hH_t}. \quad (.6)$$

Note que Wu (1966) incluiu uma variável e dois parâmetros ao modelo ( . ):  $H$ ,  $s$  e  $h$ , respectivamente, enquanto que Sakai (1966) considerou  $s$  como uma função da densidade de plantas. Em geral,  $g$ ,  $h$  e  $s$  são variáveis aleatórias porque são influenciadas por tempestades, temperatura e outros fatores incertos. Cousens e Mortimer ( ) apresentam um modelo da dinâmica populacional de plantas daninhas baseado na trajetória da densidade da população (plantas ou sementes). Neste contexto, um modelo para a produção de sementes sob baixas densidades de plantas daninhas é da forma

$$x_{t+1} = Rx_t, R > 1 \quad ( . )$$

onde  $R$  é a taxa de crescimento populacional em baixa densidade. Com a competição intra-específica, a taxa de crescimento populacional proposta por Hassell ( ) e

Edelstein-Keshet ( ) é da forma

$$\phi(x_t) = \frac{R}{(1 + dx_t)^b} \quad ( . )$$

em que

$$x_{t+1} = \phi(x_t) \quad ( . )$$

com  $R$  a taxa de crescimento em baixa densidade,  $d$  a taxa de declínio de  $\phi$  com o aumento da densidade e  $b$  a forma da trajetória populacional (taxa que reflete a intensidade do efeito da densidade de plantas daninhas na produção de sementes). Observe que o modelo ( . ) com a taxa de crescimento dada por ( . ) possui a forma geral de ( . ) com  $R$ ,  $d$  e  $b$  constantes ou dependentes da densidade. Considerando-se  $R = sg$  e  $d = ag$  e combinando-se ( . ), ( . ) e ( . ) obtém-se finalmente

$$x_{t+1} = \frac{sy_t}{(1 + ay_t)^b}. \quad ( . )$$

A solução de equilíbrio de ( . ) com a taxa de crescimento como em ( . ) é

$$x_e = \frac{R^{\frac{1}{b}} - 1}{d}. \quad ( . )$$

Assim, a trajetória da densidade da população pode ser caracterizada pela atuação de  $\phi$  com o tempo sobre as gerações da população. O tipo de trajetória depende dos valores de  $R$  e  $b$ . Se  $b = 1$  ou  $R \leq e$  ( $e$  é o número de Euler ou neperiano) a densidade da população converge monotonicamente e/ou assintoticamente para o equilíbrio. Se  $1 < b < 2$  e  $e < R < e^2$  a densidade converge para o equilíbrio com oscilações amortecidas. Se  $b > 2$  e  $R > e^2$  podem ocorrer comportamentos complexos, tais como o caos. Quando  $b = 1$  e  $R = 1$  a densidade da população se extingue. Lembrando que  $e$  é o número neperiano e seu valor aproximado é de 2,718, tem-se que a trajetória populacional de  $x$  em quatro zonas no gráfico  $R$  versus  $b$  (Figura . ).

A dinâmica populacional de uma espécie pode ser afetada por  $n$  vezes populacionais de outras espécies, assim, se uma das espécies é uma daninha e a outra é a cultura semeadas em uma mesma densidade a cada ciclo, a densidade de sementes de daninhas pode ser dada por

$$x_{t+1} = \frac{Rx_t}{[1 + d(x_t + \alpha x_c)]^b} \quad ( . )$$

ε

e

Bayesiana.

O trabalho encontra-se dividido em capítulos. No Capítulo 1, seguindo esta introdução, são apresentados alguns fundamentos do problema de inferência estatística. Aborda o problema fundamental da estatística, considerando a visão clássica em contrapartida à visão Bayesiana. Neste contexto, é explicado um pouco mais detalhadamente a teoria Bayesiana, centrada no teorema de Bayes e por fim é apresentada uma breve discussão sobre alguns métodos de análise de adequabilidade de modelos. Neste capítulo existe um particular interesse em mostrar aspectos teóricos que auxiliem na estimação de parâmetros dos modelos adotados neste trabalho, levando em consideração a abordagem Bayesiana, métodos numéricos e os métodos MCMC (Monte Carlo Markov Chains - Monte Carlo em cadeias de Markov). No Capítulo 2 são apresentadas algumas aplicações da priori Jeffreys ao modelo de plantas daninhas, considerado. No Capítulo 3 são apresentados os resultados em forma de tabelas, gráficos e figuras e são feitas algumas considerações e discussões acerca do que foi obtido. No Capítulo 4 são apresentadas as considerações finais e também são levantadas algumas propostas relevantes para a sequência do trabalho.

## Capítulo 2

# Inferência Estatística

### 2.1 O pensamento clássico

Segundo Paulino et al. (2006) a inferência clássica surgiu praticamente na metade deste século impulsionada pelos seus fundadores Karl Pearson, Ronald A. Fisher e Jerzy Neyman. E dentro da inferência estatística, a clássica é explicada da seguinte forma: Fazer com que generalizações (se possível) sobre a população possam surgir a partir de amostras retiradas dessa mesma. As amostras são observações ou dados estatísticos resultantes de experiências, em um processo de repetição e condições constantes (ou amostragem aleatória simples). (Paulino et al., 2006)

toria  $X$  ou de  $n$  variáveis aleatórias  $X = (X_1, X_2, X_3, \dots, X_n)$  com função de distribuição  $\mathcal{F}_0$  que representa a variabilidade ou incerteza na observação de  $X$ .

A função de distribuição  $\mathcal{F}_0$ , em geral, não é perfeitamente conhecida. No entanto, tem-se algum tipo de informação inicial sobre a natureza do fenômeno aleatório em estudo ou existe algum conhecimento sobre a coleta de dados que leva a propor ou conjecturar uma família de distribuições  $\mathcal{F}$  a que pertence  $\mathcal{F}_0$  e que representa o modelo estatístico para  $X$ . Desse modo, através de especificação surge uma proposta para o modelo que é uma fase essencial no processo de inferências.

Se, como é usual, as distribuições de  $\mathcal{F}$  são representadas pelas respectivas densidades (função densidade de probabilidade ou função de probabilidade) e a estas estiverem associado um parâmetro  $\theta$  com domínio num conjunto  $\Theta$  (espaço-parâmetro), o modelo estatístico é descrito por

$$\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}, x \in X.$$

onde  $f(x|\theta)$  é a distribuição amostral.

Em muitos casos, quando se trabalha com  $X = (X_1, X_2, X_3, \dots, X_n)$  as  $n$  variáveis aleatórias  $(X_1, X_2, X_3, \dots, X_n)$  são supostas independentes condicionalmente em  $\theta$  e o modelo estatístico pode ser representado em termos das densidades marginais das variáveis  $X_i$ ,  $f_i(x_i|\theta)$ ,  $i = 1, 2, \dots, n$

$$\mathcal{F} = \left\{ f(x|\theta) = \prod_{i=1}^n f_i(x_i|\theta) : \theta \in \Theta \right\}, x \in X.$$

Caso as variáveis  $X_i$ ,  $i = 1, 2, \dots, n$ , sejam independentes e identicamente distribuídas (i.i.d.), fala-se em amostragem casual ou aleatória e o modelo acaba por simplificar-se com  $f_i(\cdot|\theta) = f(\cdot|\theta)$  para  $i = 1, 2, \dots, n$

$$\mathcal{F} = \left\{ f(x|\theta) = \prod_{i=1}^n f(x_i|\theta) : \theta \in \Theta \right\}, x \in X.$$

A escolha da família  $\mathcal{F}$  pode-se dizer que resulta de uma complexa síntese entre vários fatores dos quais se destacam:

- (a) a evidência experimental obtida anteriormente no tratamento de fenômenos semelhantes;
- (b) as considerações teóricas sobre os objetivos do estudo e sobre a natureza dos fenômenos envolvidos;
- (c) as considerações teóricas a respeito das técnicas experimentais aplicadas;

(d) a preocupação de recorrer a modelos de forma tão simples quanto o possível, com parâmetros suscetíveis de clara interpretação.

Ultrapassada a importante e nem sempre fácil tarefa de modelagem e parametrização, a inferência clássica contém uma vasta gama de procedimentos que se destinam a explorar os tipos de conclusões que podem ser extraídas da amostra sobre as características do modelo representativo da população. Mais concretamente, a inferência clássica procura responder a questões como estas:

(a) Os dados  $x$  suportam ou são compatíveis com a família  $\mathcal{F}$ ?

(b) Supondo que a especificação esteja correta e que os dados surgem de uma das distribuições da família  $\mathcal{F}$ , quais as conclusões que se pode tirar a respeito do parâmetro particular  $\theta_0$  (valor verdadeiro do parâmetro) referente à função de distribuição  $\mathcal{F}_0$  que descreve 'aproximadamente' as condições em que se observam as variáveis aleatórias?

Os procedimentos clássicos são julgados à luz da amostragem repetida, princípio que os clássicos adotam no processo de interpretação das conclusões do trabalho inferencial. Com base nesse princípio de amostragem repetida os métodos estatísticos se aperfeiçoam considerando um número elevado de repetições de ensaios, todos efetuados nas mesmas condições. Uma das faces do princípio reside precisamente na interpretação frequentista de probabilidade, isto é, as medidas de incertezas são frequências geradas; a outra face reside na ideia de avaliação, em termos frequentistas, que possam vir a produzir respostas corretas ou bons resultados.

## 2.2 O pensamento Bayesiano

A origem da abordagem Bayesiana pode ser creditada a Richard Price quando em 1763 publicou a obra postuma do Rev. Thomas Bayes intitulada 'An Essay Towards Solving a Problem in the Doctrine of Chances'. A ideia da probabilidade como grau de credibilidade, tão importante para entender a filosofia Bayesiana, tem uma longa história. Parece ter sido Bernoulli, na sua obra *Ars Conjectandi* (1713), um dos primeiros autores a definir probabilidade como o grau de confiança numa dada proposição que não se sabe se é verdadeira ou falsa. De Morgan, na *Formal Logic* (1837), afirma: (1) a probabilidade identifica-se com um grau de credibilidade; (2) os graus de credibilidade podem ser medidos; (3) os graus de credibilidade podem identificar-se com um certo

complexo de sentimentos.

No que toca propriamente às idéias Bayesianas e à sua aplicação estatística, tem de citar-se Harold Jeffreys que, reagindo violentamente contra a predominante posição clássica em meados do século, embora solitário e sem apoio, conseguiu ressuscitar o Bayesianismo, dar-lhe status lógico e avançar com soluções de problemas estatísticos que naquele tempo persistiam sem uma solução. A partir daí a lista de Bayesianos foi engrossando sucessivamente e, na impossibilidade de citar todos, merecem realce os nomes de Good (1952), Savage (1967) e Lindley (1967).

O teorema de Bayes é uma proposição sobre probabilidades condicionais indiscutível desde que se aceitem as leis tradicionais do século de probabilidades ou a axiomática de Kolmogorov. O que tem dado lugar a grande controvérsia é a sua interpretação e a sua aplicação a problemas de inferência estatística. Ocupa, como é óbvio, lugar crucial na inferência Bayesiana.

Nos problemas de probabilidade, nomeadamente quando expostos por matemáticos, costuma partir-se de um espaço de probabilidade  $(\Omega, A, P)$ , terno onde:

$\Omega$  é um espaço fundamental não vazio (que nas aplicações coincide, em regra, com o espaço de resultados ou espaço-amostra) com elementos  $\omega$ ,  $\omega \in \Omega$ , chamados acontecimentos elementares e subconjuntos  $A$ ,  $A \subseteq \Omega$ , designados acontecimentos;

$A$  é a família (álgebra ou  $\sigma$ -álgebra) dos acontecimentos dotados de probabilidade;

$P$  — medida de probabilidade definida para os acontecimentos  $A \subseteq \Omega$ ,  $A \in A$ , em que  $P(A)$  é a probabilidade do acontecimento  $A$ .

Considere-se uma partição finita (ou infinita) de  $\Omega$

$$A_1, A_2, \dots, A_m, P(A_i) > 0, A_i \cap A_j = \emptyset, i \neq j, \cup_i A_i = \Omega.$$

Dado um outro acontecimento  $B$  qualquer, com  $P(B) \geq 0$ , é fácil verificar a decomposição de  $B$  na união de conjuntos disjuntos

$$B = \cup_i (A_i \cap B).$$

Consequentemente, atendendo no caso presente à aditividade da função  $P$  e à definição de probabilidade condicionada, tem-se,

$$P(B) = \sum_i P(A_i \cap B) = \sum_i P(B|A_i)P(A_i).$$

Finalmente, notando,

$$P(A_i \cap B) = P(B|A_i)P(A_i) = P(A_i|B)P(B),$$

e resolvendo em ordem a  $P(A_i|B)$ , chega-se ao teorema de Bayes

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}. \quad ( . )$$

Trata-se de uma proposiç o extremamente simples, mas suscet vel, a grandes contro-vérsias.

Uma interpretaç o pouco ‘ortodoxa’, mas importante, consiste em considerar, relativamente a dada situaç o ou conjuntura, os  $A_i = 1, 2, \dots, m$ , com ‘antecedentes’, ou probabilidades a priori  $P(A_i), i = 1, \dots, m$ , de natureza subjetiva. Depois da informaç o adicional que consiste em saber que o acontecimento  $B$  se realizou (o acontecimento  $B$  pode ser a observaç o de um conjunto de dados), o investigador revê as suas probabilidades a priori através da formula de Bayes e passa a atribuir aos  $A_i, i = 1, 2, \dots, m$  as probabilidades a posteriori  $P(A_i|B), i = 1, 2, \dots, m$ .

O teorema de Bayes é, para muitos, um dos poucos resultados da matem tica que se propõe caracterizar a aprendizagem com a experiênc ia, isto é, a modi caç o da atitude inicial em relaç o aos ‘antecedentes’, ‘causas’, ‘hipoteses’ ou ‘estados’ depois de ter a informaç o adicional de que certo acontecimento ou acontecimentos se realizaram (depois de conhecer os dados da experiênc ia ou da observaç o). Quando o investigador est na completa ignorância em relaç o aos  $A_i, i = 1, 2, \dots, m, P(A_i) = 1/m$ , obtendo a express o abaixo, aos invés de ( . )

$$P(A_i|B) = \frac{P(B|A_i)}{\sum_i P(B|A_i)}. \quad ( . )$$

Os métodos Bayesianos passam, em certo sentido, por uma extens o do modelo cl ssico, extens o que tem raiz na seguinte divergência fundamental.

No modelo cl ssico o parâmetro  $\theta, \theta \in \Theta$ , é um escalar ou vetor desconhecido, mas x o, isto é, igual ao valor particular que indexa a distribuiç o da fam lia  $\mathcal{F}$  que descreve ‘apropriadamente’ o processo ou sistema f sico que gera as observaç ões. No modelo Bayesiano o parâmetro  $\theta, \theta \in \Theta$ , é tomado como um escalar ou vetor aleatorio (n o observ vel). A loso a Bayesiana é, nesse ponto, a seguinte: o que é desconhecido (no caso em quest o o parâmetro  $\theta$ ) é incerto e toda a incerteza deve ser quanti cada em

termos de probabilidade.

Correlativamente, os Bayesianos defendem que a informação inicial ou a priori (anterior ou externa em relação à experiência mas demasiadamente importante para ser ignorada ou tratada ad hoc) pode traduzir-se formalmente em uma distribuição de probabilidade, geralmente subjetiva, para  $\theta$ , seja  $p(\theta)$ , designada distribuição a priori muitos autores, escrevem  $p(\theta|I)$  para ter presente que tal distribuição depende de  $I$  (informação inicial do indivíduo), o que tem vantagem sobretudo quando há alteração na informação inicial (ex:  $I_1$  em vez de  $I_0$ ) ou quando se comparam soluções propostas por diferentes indivíduos (ex:  $I_A$  versus  $I_B$ ).

Assim, se  $\theta$  é um parâmetro discreto, designando  $p(\theta)$  a função de probabilidade a priori, tem-se que  $p(\theta)$  exprime o grau de credibilidade que o indivíduo que procede à análise atribui ao particular  $\theta$  considerado; se  $\theta$  é um parâmetro contínuo, caso mais corrente, designando  $p(\theta)$  a função densidade de probabilidade a priori, tem-se que  $p(\theta)d\theta$  exprime o grau de credibilidade que o mesmo indivíduo atribui ao intervalo  $\theta, \theta + d\theta$ . Note-se que a determinação e a interpretação da distribuição a priori estão entre os pontos mais delicados e controversos da teoria Bayesiana e constituem um dos principais obstáculos à respectiva implementação.

A família  $\mathcal{F}$  também faz parte do modelo Bayesiano; quer dizer, a componente amostral ou experimental é comum aos modelos clássicos e Bayesiano, embora para estes os elementos  $f(x|\theta)$  de  $\mathcal{F}$  em geral são supostos ter, tal como  $p(\theta)$ , uma interpretação subjetiva.

A discussão das distribuições a priori e da sua natureza subjetiva ilustra muito claramente alguns aspectos do confronto entre Bayesianos e clássicos. Para os primeiros a escolha subjetiva do modelo (família  $\mathcal{F}$ ) traduz muitas vezes um uso mais drástico da informação a priori do que o emprego de uma distribuição a priori para o parâmetro  $\theta$  do modelo. Para os segundos há uma importante diferença entre a modelagem de  $\mathcal{F}$  e a modelagem de  $p(\theta)$ , pois, enquanto se dispõe de um conjunto de observações,  $(x_1, x_2, x_3, \dots, x_n)$ , geradas por um membro de  $\mathcal{F}$  que se pode empregar para testar a forma da distribuição, o valor de  $\theta$  é apenas uma observação da distribuição  $p(\theta)$ .

Os Bayesianos afirmam que na sua modelagem os clássicos atendem à informação a priori quando muito informalmente, atitude que consideram algo limitada porquanto, para eles, a informação inicial ou a priori detida por um dado investigador deve traduzir-

se formalmente em uma distribuição de probabilidade para a variável aleatória  $\theta$ .

Para entender o ponto de vista Bayesiano repare-se que um clássico em todos os problemas que envolvem uma variável  $X$  com distribuição Binomial recorre sempre ao mesmo modelo, nomeadamente, a função de probabilidade

$$f(x|\theta) = \theta^x(1 - \theta)^{n-x}$$

em que o parâmetro  $\theta$  representa a probabilidade de um 'sucesso'.

Para os Bayesianos cada problema é único e tem um contexto real próprio onde  $\theta$  é uma quantidade significativa acerca da qual existem, em geral, graus de conhecimento que variam de problema para problema e de investigador para investigador. Assim, para os Bayesianos, a distribuição de probabilidade que capta essa variabilidade é baseada na informação a priori (distribuição a priori) e é de natureza subjetiva, quer dizer, específica de um dado problema e de um dado investigador.

Suponha-se que se observa  $X = x$ . Considerando um qualquer elemento de  $\mathcal{F}$ , seja  $f(x|\theta)$ , e a distribuição a priori do investigador  $p(\theta)$ , o teorema de Bayes para densidades - compara-se com ( . )- conduz à relação,

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

onde  $p(\theta|x)$  é a distribuição a posteriori de  $\theta$  depois de saber que saiu  $X = x$ . Assim, tendo em conta a informação contida nos dados  $x$  a atitude inicial do investigador, caracterizada por  $p(\theta)$ , é modificada passando a nova atitude a traduzir-se por  $p(\theta|x)$ .

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

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

onde  $p(\theta|x_1, x_2, \dots, x_n)$  é a distribuição a posteriori de  $\theta$  depois de conhecida a amostra particular  $(x_1, x_2, \dots, x_n)$ .

É importante notar que a distribuição a posteriori é o elemento fundamental que serve de base ao desenvolvimento de toda inferência Bayesiana. Os denominadores de ( . ) e ( . ),

$$\int_{\Theta} f(x|\theta)p(\theta)d\theta \quad ( . )$$

$$\int_{\Theta} \prod_i f(x_i|\theta)p(\theta)d\theta \quad (.6)$$

s o as distribuições marginais ou a priori respectivamente de  $X$  e de  $(X_1, X_2, \dots, X_n)$ . Dizem respeito a observação de  $X$  ou de  $(X_1, X_2, \dots, X_n)$  qualquer que seja  $\theta$ .

O cálculo das integrais ( . ) ou ( .6) nem sempre é fácil havendo muitas vezes necessidade de recorrer ao cálculo numérico. Quando  $\theta$  é um vetor o trabalho pode naturalmente tornar-se mais pesado. Os aspectos computacionais associados com os métodos Bayesianos são estudados mais adiante.

**Exemplo 2.1** (Paulino et al., 2003) Um investigador Bayesiano estuda determinada população e está interessado na verdadeira proporção de fumadores denotada  $\theta$ ,  $0 \leq \theta \leq 1$ . Como não tem qualquer informação inicial que lhe permita distinguir entre os diferentes valores do intervalo  $[0, 1]$ , parece natural (mas não é incontroverso) tomar uma distribuição a priori uniforme,

$$p(\theta) = \begin{cases} 1 & \text{se } \theta \in [0, 1] \\ 0 & \text{se } \theta \notin [0, 1]. \end{cases} \quad (. )$$

Colhida uma amostra casual de dimensão  $n$  e considerada a variável aleatória  $X$  que designa o numero de fumadores  $X \in \{0, 1, 2, \dots, n\}$ , o modelo experimental é, correntemente, a distribuição binomial,

$$P(X = x|\theta) = f(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad 0 \leq \theta \leq 1, \quad x \in \{0, 1, 2, \dots, n\}. \quad (. )$$

Observando  $X = x$ ,  $x \in \{0, 1, 2, \dots, n\}$ , a expressão (2.3) fornece a distribuição a posteriori de  $\theta$ ,

$$p(\theta|x) = \frac{\binom{n}{x} \theta^x (1 - \theta)^{n-x}}{\int_0^1 \binom{n}{x} \theta^x (1 - \theta)^{n-x} d\theta} = \frac{\theta^x (1 - \theta)^{n-x}}{B(x + 1, n - x + 1)}, \quad 0 \leq \theta \leq 1 \quad (. )$$

Para deduzir a expressão acima recorreu-se à conhecida função Beta que, com  $\alpha, \beta$

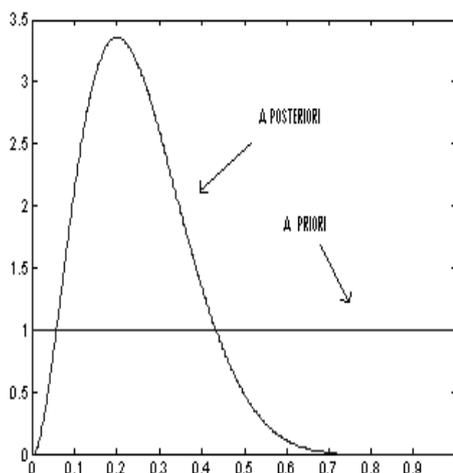


Figura 2.1: Comparação da distribuição a priori com a distribuição a posteriori.

números reais positivos, é definida por,

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx. \quad (2.6)$$

A distribuição a posteriori traduz a nova atitude do investigador depois de observar  $X = x$ . Suponha-se, para concretização, que foi observada uma amostra de 10 pessoas e que se registrou a presença de 2 fumadores. Tem-se,

$$p(\theta|x = 2) = \frac{1}{B(3,9)} \theta^2(1-\theta)^8, 0 \leq \theta \leq 1. \quad (2.7)$$

Na Figura (2.1) comparam-se as distribuições (2.7) e (2.11). A observação de 0,2 fumadores leva o investigador a passar da distribuição uniforme para uma distribuição que atribui credibilidade mais elevada aos valores não muito afastados de 0.2.

O conceito de função de verossimilhança é estudada no contexto da inferência clássica mas não é menos importante no contexto Bayesiano. Na respectiva definição convém manter a distinção entre os casos discretos e contínuos. Considere inicialmente o caso discreto. Se as variáveis aleatórias  $X_i, i = 1, \dots, n$ , são i.i.d. com função de probabilidade  $f(x|\theta)$ ,  $A := (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$  é um acontecimento com probabilidade  $\prod_i f(x_i|\theta)$ . Fixando  $A$  e fazendo variar  $\theta$  em  $\Theta$ , obtém-se a função de verossimilhança,

$$L(\theta) = L(\theta|A) = \prod_i f(x_i|\theta), \quad (2.8)$$

com o domínio em  $\Theta$  e que para cada  $\theta \in \Theta$  exprime a verossimilhança ou plausibilidade que lhe é atribuída quando se sabe que o acontecimento  $A$  se realizou, isto é, quando se observa  $(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ .

A verossimilhança não é uma probabilidade; por exemplo, não faz sentido adicionar verossimilhanças. Somente a razão de verossimilhanças tem significado: o quociente,

$$L(\theta)/L(\theta^*) = \prod_i f(x_i|\theta)/\prod_i f(x_i|\theta^*)$$

mede o peso da evidência ou plausibilidade de  $\theta$  contra  $\theta^*$  decorrente da observação, de  $(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ . Por este fato, a função de verossimilhança é definida a menos de um fator constante (i.e., independente de  $\theta$ ) positivo da seguinte forma:

$$L(\theta) = K \prod_i f(x_i|\theta), \theta \in \Theta. \tag{2.1}$$

Retomando-se a expressão (2.1), observa-se que o denominador (a distribuição preditiva de  $f(x)$ ) não depende de  $\theta$  e a expressão da posteriori, se torna

$$p(\theta|x) \propto f(x|\theta)p(\theta), \theta \in \Theta. \tag{2.2}$$

Sendo a análise condicionada pelo valor  $x$  observado (que, portanto, permanece fixo) com  $\theta$  variando sobre  $\Theta$ , o fator  $f(x|\theta)$  identifica-se com a função de verossimilhança, isto é,

$$\text{Distribuição a posteriori} \propto \text{Verossimilhança} \times \text{Distribuição a Priori}$$

### 2.3 Distribuições a priori informativas

Existem situações práticas em que há informação a priori mais ou menos substancial sobre os parâmetros do modelo, quer por parte do investigador, quer por parte de outras pessoas que o ele pode vir a recorrer. A informação a priori é de natureza subjetiva, dependendo das pessoas que a interpretam segundo o problema em questão. Como quantificar e reproduzir a informação a priori é o alvo da literatura científica no que diz respeito a metodologia Bayesiana. Para entender isso melhor recorre-se ao seguinte problema (Paulino et al., 2006).

Uma determinada fábrica produz um certo tipo de peças eletrônicas de grande preci-

s o, as quais s o agrupadas em lotes de  $N$  unidades (com  $N$  elevado) para venda futura. Admite-se que existe sempre uma porcentagem  $\theta$  dessas peças que n o obedece a um critério espec co de precis o e como tal s o consideradas defeituosas. O preço do lote depende dessa porcentagem e conseqüentemente h necessidade de inferir sobre seu valor através, por exemplo, da especi caç o de um limite superior para essa porcentagem com uma probabilidade elevada de n o ser ultrapassado. H bastante informaç o acumulada sobre o valor prov vel dessa porcentagem, quer de lotes anteriormente vendidos, quer pelas razões pelas quais s o produzidas peças menos boas.

Para poder estabelecer esse limite o estat stico vai recolher, de especialistas na matéria, informaç o a priori que existe sobre  $\theta$ . Pode us -la para tomar uma decis o sem recorrer a uma experiênci a, ou pode optar por conciliar essa informaç o com informaç o obtida testando, por exemplo, um pequeno n mero  $n$  dessas peças retiradas ao acaso do lote. Qual o procedimento que o estat stico deve adotar para uma abordagem mais correta do problema? Surgem, ent o, perguntas a serem colocadas, tais como:

- (a) Que perguntas deve-se fazer aos(s) especialista(s)?
- (b) Como ajudar o(s) especialista(s), que n o necessita(m) de ter conhecimentos de probabilidade e estat stica, a responder coerente e consistentemente a essas perguntas?
- (c) Se recorrer a mais de um especialista, como combinar a informaç o recebida?
- (d) Como usar a informaç o recebida de forma a construir uma distribuç o de probabilidade para o parâmetro ou acontecimento de interesse?

O estat stico pode optar por n o realizar a experiênci a, xar um ou mais valores de  $c$ , e pedir ao(s) especialista(s) para atribuir(em) uma probabilidade a priori para o acontecimento  $A = \{\theta > c\}$ . Como ajud -lo(s) a estabelecer(em) essa probabilidade? Este é um dos problemas bastante tratado na literatura, inclusive com a ajuda de exemplos espec cos. É importante entender como é que o estat stico pode fazer uso da informaç o recolhida de um especialista para formular uma probabilidade a posteriori para um acontecimento de interesse. A generalizaç o da mesma teoria para quando h mais do que um acontecimento de interesse n o ser tratada aqui.

Por outro lado, o estat stico pode admitir um modelo amostral para o n mero  $X$  de peças defeituosas numa amostra de  $n$  e portanto, para obter  $c$ , toma-se como base a distribuç o a posteriori de  $\theta$  dado o valor observado de  $X$ . Desse modo necessita-se,

primeiramente, formular uma distribuição a priori para  $\theta$ . Surge então, o problema da questão (a). Pode-se optar por responder questões diretamente relacionadas com o parâmetro de interesse (método estrutural) ou responder questões relacionadas com estatísticas baseadas em amostras do processo experimental (método preditivo), sendo possível uma combinação desses dois casos.

Nesse contexto será dada uma exposição (através de alguns métodos de elicitação), de generalidade variável, de forma não exaustiva. Isso de fato nem poderia ser uma vez que a tarefa de elicitação está muito dependente da aplicação concreta e do indivíduo a inquirir (Paulino et al., 2006).

### 2.3.1 Informação a priori com base em um especialista

Seja  $A$  um certo acontecimento de interesse e seja  $p$  o valor que o estatístico atribui à probabilidade da realização desse acontecimento, ou seja  $P(A) = p$ . O estatístico deseja, no entanto, usar a opinião de um especialista - o qual especifique como sendo  $q$  a credibilidade atribuída à realização de  $A$  - para rever a sua probabilidade a priori de ocorrência de  $A$ . Designe-se essa probabilidade, após revisão, por  $P(A|q)$ .

Uma abordagem, intuitiva e simples, consiste numa ponderação linear das duas conjecturas relativas à probabilidade de realização de  $A$ , ou seja,

$$P(A|q) = \omega q + (1 - \omega)p, \quad (2.5)$$

sendo  $\omega \in [0, 1]$  um coeficiente de ponderação, o qual deve refletir a importância que o estatístico atribui à previsão sugerida pelo especialista. Alguns autores designam essa abordagem por 'linear opinion pool'. Outros sugerem a chamada 'logarithmic opinion pool' em que

$$P(A|q) = q^\omega + p^{(1-\omega)}. \quad (2.6)$$

Qualquer destas duas abordagens baseia-se na premissa de que a conjectura do especialista sobre a realização de  $A$  pode ser encarada como uma probabilidade de ocorrência de  $A$ . Esta ideia é criticada por vários autores que sugerem, alternativamente, usar a informação do especialista como sendo um dado, permitindo obter, via teorema de Bayes, uma revisão da probabilidade a priori de  $A$  em conformidade com o pensamento Bayesiano. Assim, um conceito fundamental na formulação que se segue é que qualquer

forma de informação inferencial fornecida pelo especialista deve ser encarada como uma fonte de ‘dados’.

De acordo com o conceito fundamental estabelecido, o estatístico pode encarar a quantidade  $q$  fornecida pelo especialista como uma observação de uma variável aleatória  $Q$  com suporte em  $[0, 1]$ . Usando o teorema de Bayes o estatístico atualiza a sua probabilidade de ocorrência do acontecimento obtendo

$$p^* = P(A|q) = \frac{pf(q|A)}{pf(q|A) + (1-p)f(q|A^c)} \quad ( . . )$$

onde  $f(q|A)$  e  $f(q|A^c)$  são, respectivamente, as distribuições de probabilidade de  $Q$  atribuídas pelo estatístico dada a ocorrência de  $A$  e do seu complementar  $A^c$ . Estas distribuições dependem, obviamente, da credibilidade do especialista aos olhos do estatístico. Como

$$\frac{P(A|q)}{P(A^c|q)} = \frac{P(A)f(q|A)}{P(A^c)f(q|A^c)},$$

o fator de Bayes  $B = \frac{f(q|A)}{f(q|A^c)}$  determina o modo como as credibilidades do estatístico acerca da realização de  $A$  são influenciadas pela opinião. (o)- . . a (m) . . (e)

*auomof d edppoilidiaia ermpdam erees.aptraosaapf(q|A) f(q|A) f(q|A^c) f(q|A^c) . . od.j. Rψψf ψψd Antootenψf o.Rψψf ψψ*

chamado estado de ‘ignorância a priori’, t o caro aos adeptos da corrente objetivista das probabilidades lógicas) ou em que o conhecimento a priori é pouco significativo relativamente à informação amostral (o estado de conhecimento ‘vago’ ou ‘difuso’). Focar-se-ão os principais métodos advogados para o efeito conduzindo a distribuições a priori minimamente informativas em algum sentido, em regra, são denominadas de **distribuições não-informativas**.

Estas distribuições começaram por ser predominantemente interpretadas como representações formais de ignorância, mas há hoje uma tendência (motivada pela não aceitação de representações objetivas próprias da ignorância) para encarar estas representações como opções alternativas a que se recorre em caso de informação a priori insuficiente que torne difícil elicitar uma distribuição subjetiva considerada adequada. Independentemente da interpretação, este tipo de distribuição pode desempenhar ainda um papel de referência, mesmo que se disponha de fortes crenças a priori, como forma de:

(a) 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, determinar subjetivamente uma distribuição razoável - assim se reconhecendo a sua própria ignorância;

(b) permitir a comparação com os resultados da inferência clássica que só usa a informação amostral (no todo ou na parte);

(c) averiguar a influência nas inferências da distribuição a priori subjetiva que descreve a informação realmente existente, quando confrontada com as que resultam do uso da distribuição a priori de referência.

### 2.4.1 Método de Bayes-Laplace

O argumento primeiramente invocado para gerar distribuições não informativas foi o Princípio da Razão Insuficiente devido a Bayes e Laplace. De acordo com este princípio, na ausência de razão suficiente para privilegiar umas possibilidades em detrimento de outras, decorrente da escassez informativa a priori, deve-se adotar a equiprobabilidade. A consideração deste argumento como objetivista não é totalmente pacífica já que é questionável falar-se de ‘ausência de razão suficiente’ sem envolvimento de juízos subjetivos.

No caso em que  $\Theta$  é finito, diga-se  $\Theta = \{\theta_1, \dots, \theta_h\}$ , a distribuição não-informativa gerada por este argumento é a distribuição Uniforme Discreta

$$p(\theta) = \frac{1}{h}, \theta \in \Theta. \quad (2.4.1)$$

Quando  $\Theta$  é infinito numerável, é conhecido que não existe nenhuma genuína distribuição de probabilidades compatível com a equiprobabilidade de todos os valores possíveis de  $\theta$ . Por outras palavras, a distribuição Uniforme Discreta neste caso não satisfaz o axioma de probabilidade total unitária, sendo por isso tipicamente denominada **distribuição imprópria**. O uso destas distribuições no mecanismo Bayesiano é, por isso, suscetível de várias críticas.

No caso em que  $\Theta$  é infinito não numerável, o Princípio da Razão Insuficiente é interpretado de modo a conduzir a distribuição Uniforme Contínua, que é, de novo, imprópria se  $\Theta$  não for limitado.

A principal objecção que se veio a levantar a ideia de representação da ignorância por distribuições uniformes é a sua inconsistência no sentido em que se  $\psi = \psi(\theta)$  é uma transformação injetiva de um parâmetro  $\theta$ , que assume uma gama contínua de valores possíveis, as distribuições uniformes para  $\theta$  e  $\psi$  não são, em regra, probabilisticamente compatíveis. De fato, sendo  $p(\theta)$  uma distribuição a priori para  $\theta$ ,

$$p(\psi) = p[\theta(\psi)] \left| \frac{d\theta}{d\psi} \right| \quad (2.4.2)$$

deve ser a correspondente distribuição para a reparametrização injetiva  $\psi$ , que não é necessariamente uniforme quando  $p(\theta)$  o é.

### 2.4.2 Método de Jeffreys

A crítica da inconsistência da distribuição uniforme na representação formal de ignorância suscita que esta deva ser invariante sob transformações injetivas. Entre os procedimentos que asseguram essa invariância está aquele advogado por Jeffreys e que se baseia no uso da medida de informação de Fisher sobre  $\theta \in \mathfrak{R}$ ,

$$I(\theta) = E \left[ \left( \frac{\partial \ln f(X|\theta)}{\partial \theta} \right)^2 \middle| \theta \right]. \quad (2.4.3)$$

Com efeito, o fato de para qualquer transformaç o real injetiva de  $\theta \in \mathfrak{R}$  se ter

$$I(\psi) = I(\theta(\psi)) \left( \frac{d\theta}{d\psi} \right)^2 \quad ( . . )$$

mostra que a distribuiç o proposta por Jeffreys para o caso uniparamétrico,

$$I = \int_a^b (b-a)g(x) \frac{1}{(b-a)} dx = (b-a)E[g(x)]$$

identificando  $X$  como uma variável aleatória com distribuição  $U(a, b)$ . Assim, transformase o problema de avaliar a integral no problema estatístico de estimar uma média,  $E[g(X)]$ . Tendo a disposição uma amostra aleatória de tamanho  $n$ ,  $x_1, x_2, x_3, \dots, x_n$  da distribuição uniforme no intervalo  $(a, b)$  será possível conseguir também uma amostra de valores  $g(x_1), \dots, g(x_n)$  da função  $g(x)$  e a integral acima pode ser estimada pela média amostral, isto é

$$\hat{I} = (b-a) \frac{1}{n} \sum_i g(x_i).$$

Não é difícil verificar que esta estimativa é não viesada já que  $E[g(x_i)] = E[g(x)]$  para todo  $i$  e portanto

$$E(\hat{I}) = \frac{(b-a)}{n} \sum_1^n E[g(x_i)] = (b-a)E[g(x)] = \int_a^b g(x)dx.$$

Pode-se então usar um algoritmo com os seguintes passos.

- . Gerar  $x_1, x_2, \dots, x_n$  da distribuição  $U(a, b)$ ;
- . Calcular  $g(x_1), g(x_2), \dots, g(x_n)$ ;
- . Calcular a média amostral  $\bar{g} = \frac{\sum_1^n g(x_i)}{n}$ ;
- . Calcular  $\bar{I} = (b-a)\bar{g}$ .

**Exemplo 2.2** (Ehlers, 2003) Suponha que se quer calcular  $\int_1^3 \exp(-x)dx$ . A integral pode ser reescrita como

$$(3-1) \int_1^3 \exp(-x)/(3-1)dx$$

e será aproximada usando 100 valores simulados da distribuição Uniforme no intervalo  $(1, 3)$  e calculando  $y_i = e^{x_i}$ ,  $i = 1, \dots, 100$ . O valor aproximado da integral é  $2 \sum_{i=1}^{100} y_i/100$ . Por outro lado, sabe-se que  $\exp(-x)$  é a função de densidade de uma v.a.  $X \exp(1)$  e portanto a integral pode ser calculada de forma exata,

$$\int_1^3 \exp(-x)dx = Pr(X \leq 3) - Pr(X \leq 1) = 0.3181.$$

A generalização é bem simples para o caso em que a integral é a esperança matemática de uma função  $g(X)$  onde  $X$  tem função de densidade  $p(x)$ , i.e.

$$I = \int_a^b g(x)p(x)dx = E[g(x)]. \quad ( . . )$$

Neste caso, pode-se usar o mesmo algoritmo descrito acima modificando o passo para gerar  $x_1, x_2, \dots, x_n$  da distribuição  $p(x)$  e calculando  $\hat{I} = \bar{g}$ . Ou seja, usa-se uma média amostral para estimar uma média teórica.

### Erro de Monte Carlo

Naturalmente que  $\hat{I}$  é somente uma aproximação para a quantidade que se deseja calcular, portanto precisa-se estudar o erro  $\hat{I} - I$ . Uma vez que as gerações são independentes, então, pela Lei Forte dos Grandes Números segue que  $\hat{I}$  converge quase certamente para  $I$ , i.e.

$$\frac{1}{n} \sum_{i=1}^n g(x_i) \xrightarrow{n \rightarrow \infty} E[g(x)] \text{ quase certamente.}$$

Além disso, definindo  $\sigma^2 = Var[g(x)]$  e assumindo que esta variância existe o Erro Padrão de Monte Carlo é uma estimativa consistente de  $\sigma$  dada por

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (g(x_i) - \hat{g})^2},$$

i.e. a aproximação pode ser tão acurada quanto se deseje bastando aumentar o valor de  $n$ . É importante notar que  $n$  está sob controle aqui, e não se trata do tamanho da amostra de dados. O teorema central do limite também se aplica aqui de modo que para  $n$  grande segue que

$$\frac{\bar{g} - E[g(x)]}{\hat{\sigma}}$$

tem distribuição aproximadamente  $N(0, 1)$ . Pode-se usar este resultado para testar convergência e construir intervalos de confiança do tipo  $\hat{g} \pm z_{\alpha/2} \hat{\sigma}$ . No caso multivariado a extensão também é direta. Seja  $x = (x_1, \dots, x_n)$  um vetor aleatório de dimensão  $k$  com função de densidade  $p(x)$ . Neste caso, os valores gerados serão também vetores  $x_1, \dots, x_n$  e o estimador de Monte Carlo é

$$\hat{I} = \sum_{i=1}^n g(x_i).$$

### 2.5.2 Monte Carlo via função de importância

Em muitas situações pode ser muito custoso ou mesmo impossível simular valores da distribuição de interesse  $p(x)$ . Neste caso, pode-se recorrer a uma função  $q(x)$  que seja de fácil amostragem, usualmente chamada de função de importância. O procedimento é comumente chamado de amostragem por importância. Se  $q(x)$  for uma função de densidade definida no mesmo espaço de variação de  $x$  então a integral (2.1) pode ser

reescrita como

$$I = \int \frac{g(x)p(x)}{q(x)} q(x) dx = E\left[\frac{g(x)p(x)}{q(x)}\right]$$

onde a esperança agora é com respeito à distribuição  $q$ . Assim, dispondo-se de uma amostra aleatória  $x_1, \dots, x_n$  tomada da distribuição  $q$ , o estimador de Monte Carlo da integral acima é

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \frac{g(x_i)p(x_i)}{q(x_i)}$$

o qual tem as mesmas propriedades do estimador de Monte Carlo simples (erro, variância, etc). Em princípio não há restrições quanto na escolha da densidade de importância  $q$ , porém, na prática alguns cuidados devem ser tomados. Pode-se mostrar que a escolha ótima no sentido de minimizar a variância do estimador consiste em tomar  $q(x)/g(x)p(x)$ .

**Exemplo 2.3** (Ehlers, 2003) Para uma única observação  $X$  com distribuição  $N(\theta, 1)$ ,  $\theta$  desconhecido, quer-se fazer inferência sobre  $\theta$  usando uma priori Cauchy  $(0, 1)$ . Neste caso, a função de verossimilhança e densidade a priori são dadas por

$$p(x|\theta) \propto \exp[-(x - \theta)^2/2] \text{ e } p(\theta) = \frac{1}{\pi(1+\theta^2)}.$$

Portanto, a densidade a posteriori de  $\theta$  é dada por

$$p(\theta|x) = \frac{\frac{1}{(1+\theta^2)} \exp[-(x-\theta^2)/2]}{\int \frac{1}{(1+\theta^2)} \exp[-(x-\theta^2)/2] d\theta}.$$

Uma possível estimativa pontual de  $\theta$  consiste em tomar a sua média a posteriori.

Mas, note que

$$E(\theta|x) = \int \theta p(\theta|x) d\theta = \frac{\int \frac{\theta}{(1+\theta^2)} \exp[-(x-\theta^2)/2]}{\int \frac{1}{(1+\theta^2)} \exp[-(x-\theta^2)/2] d\theta}$$

e as integrais no numerador e denominador não têm solução analítica exata. Uma solução aproximada via simulação de Monte Carlo pode ser obtida usando o seguinte algoritmo.

- gerar  $\theta_1, \theta_2, \dots, \theta_n$  independentes da distribuição  $N(x, 1)$ ;
- Calcular  $g_i = \frac{\theta_i}{1+\theta_i^2}$  e  $g_i^* = \frac{1}{1+\theta_i^2}$ .

Este exemplo ilustrou um problema que geralmente ocorre em aplicações Bayesianas. Como a posteriori só é conhecida a menos de uma constante de proporcionalidade as esperanças a posteriori são na verdade uma razão de integrais. Neste caso, a aproximação é baseada na razão dos dois estimadores de Monte Carlo para o numerador e

denominador.

### 2.5.3 Método numérico de Simpson

Segundo Barroso et al. ( ) uma função  $f(x)$  é contínua em um intervalo  $[a, b]$  e sua primitiva  $F(x)$  é conhecida, então, a integral definida desta função neste intervalo é dada por:

$$\int_a^b f(x)dx = F(b) - F(a) \tag{. .}$$

onde  $F'(x) = f(x)$ .

Entretanto, em alguns casos, o valor desta primitiva  $F(x)$  não é conhecido ou difícil obter, o que dificulta ou mesmo impossibilita o cálculo desta integral.

Por outro lado, em situações práticas, nem sempre se tem a função a ser integrada definida por uma fórmula analítica, e sim por meio de tabela de pontos, o que torna inviável a utilização de ( . ).

Para calcular o valor da integral definida de  $f(x)$ , nas duas situações citadas acima ou em qualquer outra, torna-se necessária a utilização de métodos numéricos. A solução numérica de uma integral simples é comumente chamada de quadratura. Os métodos mais utilizados, podem ser classificados em dois grupos:

- (a) As fórmulas de Newton-Cotes que empregam valores de  $f(x)$ , onde os valores de  $x$  são igualmente espaçados.
- (b) A fórmula de quadratura gaussiana que utiliza pontos diferentemente espaçados, onde este espaçamento é determinado por certas propriedades de polinômios ortogonais.

Dentre as fórmulas de Newton-Cotes, será dada ênfase nesse trabalho à regra de Simpson. Para a obtenção das fórmulas de Newton-Cotes, é utilizado o polinômio interpolador de Gregory-Newton:

$$P_n(x) = y_0 + z\Delta y_0 + \frac{z(z-1)}{2!}\Delta^2 y_0 + \frac{z(z-1)(z-2)}{3!}\Delta^3 y_0 + \dots + \frac{z(z-1)(z-2)\dots(z-n+1)}{n!}\Delta^n y_0 + R_n$$

onde  $z = \frac{x-x_0}{h}$  (onde  $h$  será definido mais à frente) e  $R_n$  é o resíduo de interpolação:

$$R_n = \frac{z(z-1)(z-2)\dots(z-n)}{(n+1)!}h^{n+1}f^{(n+1)}(\varepsilon), a \leq \varepsilon \leq b, \tag{. . 6}$$

e  $P_n(x)$  é o polinômio de  $n$ -ésimo grau.

Aproximando a função  $f(x)$  em ( . . . ), pelo polinômio de Gregory-Newton, e integrando-o, obtém-se as formulas de Newton-Côtes. Esta aproximação se justifica, pois este polinômio é de f cil integração.

### 2.5.4 Regra de Simpson

A primeira regra de Simpson é obtida aproximando-se a função  $f(x)$  em ( . . . ) por um polinômio interpolador de segundo grau,  $P_2(x)$

$$f(x) = P_2(x) = y_0 + z\Delta y_0 + \frac{z(z-1)}{2!}\Delta^2 y_0$$

$$I = \int_a^b f(x)dx = \int_a^b P_2(x)dx = \int_a^b \left[ y_0 + z\Delta y_0 + \frac{z(z-1)}{2!}\Delta^2 y_0 \right] dx.$$

Como  $z = \frac{x-x_0}{h}$  tem-se  $dx = h dz$ .

Para aproximar a função  $f(x)$  por um polinômio de segundo grau, são necessários pontos:  $x_0, x_1, x_2$ , que devem estar igualmente espaçados. Sejam  $x_0 = a$  e  $x_2 = b$ . Fazendo uma mudança de variável, tem-se

$$\text{para } x = a \text{ que } z = \frac{a-a}{h} = 0, \text{ e para } x = b \text{ que } z = \frac{b-a}{h} = 2.$$

$$\text{Logo, } I = \int_0^2 \left[ y_0 + z\Delta y_0 + \frac{z(z-1)}{2!}\Delta^2 y_0 \right] h dz.$$

Integrando, obtém-se:

$$I = h \left[ zy_0 + \frac{z^2}{2}\Delta y_0 + \left( \frac{z^3}{6} - \frac{z^2}{4} \right) \Delta^2 y_0 \right]_0^2 = h \left[ 2y_0 + 2\Delta y_0 + 1/3\Delta^2 y_0 \right].$$

Sabe-se que:

$$\Delta y_0 = y_1 - y_0 \text{ e } \Delta^2 y_0 = y_2 - 2y_1 + y_0. \text{ Logo,}$$

$$I = \frac{h}{3} \left[ y_0 + 4y_1 + y_2 \right]. \quad ( . . . )$$

#### Erro de truncamento

Para a determinação do erro cometido na integração, basta integrar o erro de truncamento da aproximação polinomial. Este erro (de truncamento) é cotado pelo resíduo ( . . . 6)

$$E = \int_a^b R_2(x)dx = \int_0^2 \frac{z(z-1)(z-2)}{3!} f^{(3)}(\xi) h^4 dz = 0.$$

Este valor nulo para o erro de integraç o quer dizer que o erro n o depende de  $R_2$  (res duo do segundo grau). Ent o, tem-se que integrar o res duo menor que  $R_2$ , no caso o  $R_3$ . Assim,

$$\begin{aligned}
 E &= \int_0^2 R_3(x)dx = \int_0^2 \frac{z(z-1)(z-2)(z-3)}{4!} f^{(4)}(\varepsilon)h^5 dz \\
 &= \frac{-h^5}{90} f^{(4)}(\varepsilon), a \leq \varepsilon \leq b \quad ( . . )
 \end{aligned}$$

que é a formula de erro da primeira regra de Simpson. Por esta formula pode-se notar que a primeira regra de Simpson fornece valores exatos n o so para a integraç o de polinômios do segundo grau, mas, também para polinômios de terceiro grau (derivada de quarta ordem nula).

**Fórmula composta**

Deve-se, primeiramente, subdividir o intervalo de integraç o  $[a, b]$  em  $n$  subintervalos iguais de amplitude  $h$  e a cada par de subintervalos aplicar a primeira regra de Simpson.

Observaç o importante: Como a regra de Simpson é aplicada em pares de subintervalos, o número  $n$  de subintervalos dever ser sempre par.

Considerando  $n = \frac{b-a}{h}$  e os pontos,  $x_i; i = 0, 1, \dots, n$ . Assim,

$$\begin{aligned}
 I &= \int_a^b f(x)dx = \\
 &\quad \underbrace{\frac{h}{3}[y_0 + 4y_1 + y_2]}_{\text{Aplicado no primeiro par de subintervalos}} + \underbrace{\frac{h}{3}[y_2 + 4y_3 + y_4]}_{\text{Aplicado no segundo par de subintervalos}} + \dots \\
 &\dots + \underbrace{\frac{h}{3}[y_{n-2} + 4y_{n-1} + y_n]}_{\text{Aplicado no ultimo par de subintervalos}}
 \end{aligned}$$

Tem-se nalmente que,

$$I = \frac{h}{3} [y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \dots + 2y_{n-2} + 4y_{n-1} + y_n]. \quad ( . . )$$

**Erro de truncamento**

O erro total cometido ser a soma dos erros cometidos a cada aplicação da primeira regra de Simpson.

$$E = E_1 + E_2 + E_3 + \dots + E_{n/2} = \sum_{i=1}^{n/2} E_i \quad (2.16)$$

onde  $E_i$  é o erro na integração numérica no par de subintervalos cujos extremos são  $[x_{2i-2}, x_{2i-1}]$  e  $[x_{2i-1}, x_{2i}]$ .

Levando (2.15) em (2.16) vem:

$$E = \sum_{i=1}^{n/2} \frac{-h^5}{90} f^{(4)}(\varepsilon_i), x_{2i-2} \leq \varepsilon_i \leq x_{2i}. \quad (2.17)$$

Pela continuidade de  $f^{(4)}(x)$ , existe  $\varepsilon \in [a, b]$ , tal que:

$$\frac{n}{2} f^{(4)}(\varepsilon) = \sum_{i=1}^{n/2} f^{(4)}(\varepsilon_i). \quad (2.18)$$

Levando-se (2.18) em (2.17), tem-se:

$$E = \frac{-h^5}{180} n f^{(4)}(\varepsilon)$$

Como  $h = \frac{b-a}{n}$ , então,

$$E = \frac{-(b-a)^5}{180n^4} f^{(4)}(\varepsilon), a \leq \varepsilon \leq b. \quad (2.19)$$

Pode-se observar que, nesta fórmula, o erro cai com a quarta potência do número de subintervalos.

## 2.6 Métodos de Monte Carlo via cadeias de Markov

Em todos os métodos de simulação apresentados anteriormente obtém-se uma amostra da distribuição de interesse em um único passo. Os valores são gerados de forma independente e não há preocupação com a convergência do algoritmo, bastando que o tamanho da amostra seja suficientemente grande. Por isso estes métodos são chamados não iterativos. No entanto, em muitos problemas pode ser bastante difícil, ou mesmo impossível, encontrar uma densidade de importância que seja simultaneamente uma boa

aproximaç o da densidade de interesse e f cil de ser amostrada. Os métodos de Monte Carlo via cadeias de Markov (MCMC) s o uma alternativa aos métodos n o iterativos em problemas complexos. A idéia ainda é obter uma amostra da distribuiç o de interesse e calcular estimativas amostrais de caracter ticas desta distribuiç o. A diferença é que aqui usa-se técnicas de simulaç o iterativa, baseadas em cadeias de Markov, e assim os valores gerados n o s o mais independentes. Este fato ,porém, n o causa maiores complicaç oes como pode-se ver a seguir. Neste cap tulo s o apresentados os métodos MCMC mais utilizados, o amostrador de Gibbs e o algoritmo de Metropolis-Hastings. A idéia b sica é simular um passeio aleatorio no espaço de  $x$  que converge para uma distribuiç o estacion ria, que é a distribuiç o de interesse no problema.

### 2.6.1 Cadeias de Markov

Uma cadeia de Markov é um processo estoc stico  $\{X_0, X_1, \dots\}$  tal que a distribuiç o de  $X_t$  dados todos os valores anteriores  $\{X_0, \dots, X_{t-1}\}$  depende apenas de  $X_{t-1}$ . Matematicamente,

$$P(X_t \in A | X_0, \dots, X_{t-1}) = P(X_t \in A | X_{t-1})$$

para qualquer subconjunto  $A$ . Os métodos MCMC requerem ainda que a cadeia seja,

- (a) homogênea, i.e. as probabilidades de transiç o de um estado para outro s o invariantes;
- (b) irredut vel, i.e. cada estado pode ser atingido a partir de qualquer outro em um número nito de iteraç oes;
- (c) aperiodica, i.e. n o haja estados absorventes.

Os algoritmos vistos aqui satisfazem estas condiç oes. Suponha que uma distribuiç o  $(x)$ ,  $x \in \mathfrak{R}^d$  seja conhecida a menos de uma constante multiplicativa porém complexa o bastante para n o ser poss vel obter uma amostra diretamente. Dadas as realizaç oes  $\{X^{(t)}, t = 0, 1, \dots\}$  de uma cadeia de Markov que tenha como distribuiç o de equilibrio, ent o, sob as condiç oes acima,

$$X^{(t)} \xrightarrow{t \rightarrow \infty} (x) \text{ e } \frac{1}{n} \sum_{t=1}^n g(X_i^{(t)}) \xrightarrow{n \rightarrow \infty} E_{\pi}(g(X_i)) \text{ quase certamente.}$$

Ou seja, embora a cadeia seja por de niç o dependente, a média aritmética dos valores da cadeia é um estimador consistente da média teorica. Uma quest o importante de

ordem prática é como os valores iniciais influenciam o comportamento da cadeia. A idéia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio. Assim, em aplicações práticas é comum que as iterações iniciais sejam descartadas, como se formassem uma amostra de aquecimento. O problema, então, consiste em construir algoritmos que gerem cadeias de Markov cuja distribuição converge para a distribuição de interesse.

### Precisão numérica

Na prática tem-se um número finito de iterações e tomando

$$\hat{g} = \frac{1}{n} \sum_{t=1}^n g(X_t^t)$$

como estimativa da  $E(g(X_i))$  deve-se calcular o seu erro padrão. Como a sequência de valores gerados é dependente pode-se mostrar que

$$Var(\hat{g}) = \frac{s^2}{n} \left[ 1 + 2 \sum_{k=1}^n (1 - k/n) \rho_k \right]$$

sendo  $s^2$  a variância amostral e  $\rho_k$  a autocorrelação amostral de ordem  $k$ . Se  $\rho_k \geq 0$  então  $Var(\hat{g}) \geq s^2/n$ .

### 2.6.2 Algoritmo de Metrópolis-Hastings

Nos algoritmos de Metropolis-Hastings um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade (processo de rejeição e aceitação). Este mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio. Suponha que a cadeia esteja no estado  $\theta$  e um valor  $\theta_0$  é gerado de uma distribuição proposta  $q(\cdot|\theta)$ . Note que a distribuição proposta pode depender do estado atual da cadeia, por exemplo,  $q(\cdot|\theta)$  poderia ser uma distribuição normal centrada em  $\theta$ . O novo valor  $\theta_0$  é aceito com probabilidade

$$\alpha(\theta, \theta') = \min \left\{ 1, \frac{(\theta')q(\theta|\theta')}{(\theta)q(\theta'|\theta)} \right\} \quad (2.6.2)$$

onde  $\theta$  é a distribuição de interesse.

Uma característica importante é que só necessita-se conhecer parcialmente, i.e., a menos de uma constante,  $j$  que neste caso a probabilidade (2.6.2) não se altera. Isto é

fundamental em aplicações Bayesianas onde não se conhece completamente a densidade a posteriori. Note também que a cadeia pode permanecer no mesmo estado por muitas iterações e na prática costuma-se monitorar isto calculando a porcentagem média de iterações para as quais novos valores são aceitos. Em termos práticos, o algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos,

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

Embora a distribuição proposta possa ser escolhida arbitrariamente na prática deve-se tomar alguns cuidados para garantir a eficiência do algoritmo. Em aplicações Bayesianas a distribuição de interesse é a própria posteriori, i.e.  $\pi(\theta|\mathbf{y}) = p(\theta|\mathbf{y})$  e a probabilidade de aceitação assume uma forma particular,

$$\alpha(\theta, \theta') = \min\left\{1, \frac{p(x|\theta')p(\theta')p(\theta|\theta')}{p(x|\theta)p(\theta)p(\theta|\theta)}\right\}. \quad (2.35)$$

O algoritmo é ilustrado no exemplo a seguir.

**Exemplo 2.4** (Ehlers, 2003) *Em uma certa população de animais sabe-se que cada animal pode pertencer a uma dentre 4 linhagens genéticas com probabilidades*

$$p_1 = \frac{1}{2} + \frac{\theta}{2}; p_2 = \frac{1-\theta}{4}; p_3 = \frac{1-\theta}{4}; p_4 = \theta/4$$

sendo  $0 \leq \theta \leq 1$  um parâmetro desconhecido. Para qualquer  $\theta \in [0, 1]$  é fácil verificar que  $p_i \geq 0$ ,  $i = 1, 2, 3, 4$  e  $p_1 + p_2 + p_3 + p_4 = 1$ . Observando-se  $n$  animais dentre os quais  $y_i$  pertencem à linhagem  $i$ , então, o vetor aleatório  $Y = (y_1, y_2, y_3, y_4)$  tem distribuição multinomial com parâmetros  $n, p_1, p_2, p_3, p_4$  e portanto,

$$p(y|\theta) = \frac{n!}{y_1!y_2!y_3!y_4!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} \propto (2 + \theta)^{y_1} (1 - \theta)^{y_2 + y_3} \theta^{y_4}.$$

Atribuindo uma priori para  $\theta \sim U(0, 1)$ , segue que a posteriori é proporcional à expressão acima. Tomando a distribuição  $U(0, 1)$  como proposta então  $q(\theta) = 1, \forall \theta$  e a probabilidade (2.35) se simplifica para

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{1, \frac{p(x|\boldsymbol{\theta}')}{p(x|\boldsymbol{\theta})}\right\} = \min\left\{1, \left(\frac{2+\boldsymbol{\theta}'}{2+\boldsymbol{\theta}}\right)^{y_1} \left(\frac{1-\boldsymbol{\theta}'}{1-\boldsymbol{\theta}}\right)^{y_2+y_3} \left(\frac{\boldsymbol{\theta}'}{\boldsymbol{\theta}}\right)^{y_4}\right\}.$$

Supondo uma observação de 197 animais com os números de animais nas categorias dados por  $y = (125, 18, 20, 34)$  e gerando uma cadeia de Markov com 1000 valores de  $\boldsymbol{\theta}$ , pode-se gerar uma amostra da posteriori, por exemplo.

### 2.6.3 Amostrador de Gibbs

O algoritmo de Gibbs é usado em problemas de estatística Bayesiana e pode ser utilizado na simulação da densidade a posteriori e na estimação dos parâmetros. O algoritmo é baseado no fato de que se a distribuição conjunta  $p(\boldsymbol{\theta}|x)$  for positiva em  $\Theta_1 \times \dots \times \Theta_k$ , com  $\Theta_i$  suporte da distribuição de  $\boldsymbol{\theta}_i$  para  $i = 1, \dots, k$ , então, esta é unicamente determinada pelas distribuições condicionais completas  $p(\boldsymbol{\theta}_i|x, \boldsymbol{\theta}(-i))$ ,  $i = 1, \dots, k$  onde  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)$  e  $\boldsymbol{\theta}(-i)$  representa o vetor  $\boldsymbol{\theta}$  sem a  $i$ -ésima componente, isto é,  $\boldsymbol{\theta}(-i) = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \dots, \boldsymbol{\theta}_k)$ . O algoritmo é, então, um esquema Markoviano dinâmico que requer a amostragem destas distribuições condicionais como se descreve em seguida.

#### Descrição do algoritmo de Gibbs

Seja  $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$  um valor arbitrário inicial para o vetor  $\boldsymbol{\theta}$ . Procede-se iterativamente da seguinte forma:

- obtém-se  $\boldsymbol{\theta}_1^{(1)}$  de  $p(\boldsymbol{\theta}_1|x, \boldsymbol{\theta}_2^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$
- obtém-se  $\boldsymbol{\theta}_2^{(1)}$  de  $p(\boldsymbol{\theta}_2|x, \boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_3^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$
- obtém-se  $\boldsymbol{\theta}_3^{(1)}$  de  $p(\boldsymbol{\theta}_3|x, \boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \boldsymbol{\theta}_4^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})$
- ....
- obtém-se  $\boldsymbol{\theta}_k^{(1)}$  de  $p(\boldsymbol{\theta}_k|x, \boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_{k-1}^{(0)})$ .

Completa-se assim uma iteração do esquema e uma transição de  $\boldsymbol{\theta}^0$  para  $\boldsymbol{\theta}^1 = (\boldsymbol{\theta}_1^{(1)}, \dots, \boldsymbol{\theta}_k^{(1)})$ .

- O esquema anterior é repetido com  $\boldsymbol{\theta}^1$  anteriormente obtido, como vetor inicial, para obter um novo vetor  $\boldsymbol{\theta}^2$  e assim haver uma transição de  $\boldsymbol{\theta}^1$  para  $\boldsymbol{\theta}^2 = (\boldsymbol{\theta}_1^{(2)}, \dots, \boldsymbol{\theta}_k^{(2)})$

- Itera-se  $t$  vezes este ciclo de geraç o de observaç oes aleatorias de cada uma das distribuic oes condicionais, produzindo assim a cadeia  $\theta^0, \dots, \theta^t$ .

A sucess o  $\theta^0, \dots, \theta^t$  é uma realizaç o de uma cadeia de Markov com espaço de estados  $\Theta$  e funç o de transiç o (definida anteriormente) dada por

$$p(\theta^t, \theta^{t+1}) = \prod_{i=1}^k p(\theta_i^{t+1} | x, \theta_j^t, j \geq i, \theta_j^{t+1}, j \leq i). \quad ( . 6)$$

Resultados teoricos (n o apresentados aqui. Ver Paulino et al. (2006)) permitem concluir que, quando  $t \rightarrow \infty$ ,  $(\theta_1^t, \dots, \theta_k^t)$  tende em distribuic o para um vetor aleatorio cuja funç o densidade de probabilidade conjunta é  $p(\theta | x)$ . Em particular,  $\theta_i^t$  tende em distribuic o para uma quantidade aleatoria cuja densidade é  $p(\theta_i | x)$  (densidade marginal a posteriori de  $\theta_i$ ) e

$$\frac{1}{t} \sum_{i=1}^t g(\theta^i) \rightarrow E[g(\theta | x)], q.c. \quad ( . )$$

para qualquer funç o  $g(\cdot)$ , onde  $E[g(\theta | x)]$  representa o valor esperado de  $g(\theta)$  em relaç o a distribuic o a posteriori  $p(\theta | x)$ . A partir da surgem quest oes relacionadas ao uso da metodologia de Gibbs. Para fazer inferencias sobre  $\theta$  usando uma amostra gerada por Gibbs, por exemplo, é necessario dar atenç o a algumas consideraç oes.

(a) Os vetores de  $\theta$  s o correlacionados. Como é que ent o se deve proceder para obter uma amostra aleatoria i.i.d da distribuic o a posteriori? Como utilizar a amostra obtida para fazer inferências?

(b) Como é que se pode controlar a convergência da cadeia de Markov para o estado de equilibrio?

(c) O que é que se pode dizer sobre a velocidade de convergência para a distribuic o de interesse e sobre a precis o das estimativas obtidas?

(d) E, nalmente, que fazer se uma (ou mais) das distribuic oes condicionais n o forem de f cil simulaç o, ou se forem apenas conhecidas a menos de uma constante normalizadora?

Estas s o algumas das quest oes que se põem e que necessitam ser devidamente resolvidas para que o método Gibbs seja a soluç o para os problemas encontrados na implementaç o da metodologia Bayesiana. É a volta destes que tem centrado grande parte da investigaç o nesse campo. Relativamente a quest o posta em (s), por exemplo,

existem essencialmente duas soluções possíveis, cada uma com seus pros e contras, nomeadamente, (i) a construção de apenas uma longa cadeia utilizando vetores espaçados de modo a anular o efeito da correlação, ou (ii) a construção de várias cadeias paralelas utilizando apenas o último vetor gerado de cada uma delas. No presente trabalho, foi utilizado o método do item (i) para gerar uma amostra no espaço paramétrico, necessária para o processo de inferência. A seguir este método é descrito com mais detalhe.

### Uma única cadeia

A ideia deste método é gerar uma longa cadeia a partir de um estado inicial, e usar para fazer inferências, apenas algumas das observações resultantes dessa realização. Assim, a partir do estado inicial de cadeia, designado por  $\theta^0 = (\theta_1^0, \dots, \theta_k^0)$ :

. Gera-se, usando o método Gibbs descrito, uma longa realização da cadeia com comprimento (número de iterações)  $t$ , com  $t = l + k^*m$  e onde  $l$  é o número inicial de iterações necessárias para que a cadeia atinja o estado de equilíbrio,  $m$  o número das iterações que vão ser usadas na aplicação do método de Monte Carlo e  $k^*$  o espaçamento entre iterações sucessivas delineado para eliminar a autocorrelação entre as consecutivas iterações. O período constituído pelas  $l$  primeiras iterações costuma ser designado por período de aquecimento (burn-in). Este período pode ser mais ou menos longo consoante o problema que se tem em causa. Há métodos de diagnóstico para verificar se a cadeia se encontra ou não no estado de equilíbrio.

. Extrai-se da realização original da cadeia um subconjunto de dimensão  $m$  contendo as observações  $\theta^{l+k^*}, \theta^{l+2k^*}, \dots, \theta^{l+k^*m}$ . Deste modo, obtém-se uma amostra  $\theta_1, \dots, \theta_m$  em que  $\theta_j = \theta^{l+jk^*}$  com base na qual são feitas as inferências.

A questão natural que se põe é de como escolher a dimensão  $m$  da amostra, o número inicial de iterações  $l$  e o espaçamento  $k$  entre iterações. Não há uma resposta simples e imediata para este problema. Primeiro, há a necessidade de abordar o problema de monitorização da convergência da cadeia de modo a escolher um valor apropriado para  $l$  que esteja claramente relacionado com o valor inicial atribuído a  $\theta$ . O espaçamento entre iterações depende muito da estrutura de correlação da cadeia gerada pelo método Gibbs. Correlações elevadas tendem a atrasar consideravelmente a convergência. Também, quanto mais elevada for a correlação, mais séria é a influência de maus valores iniciais. Geralmente um valor para  $k^* = 10$  ou  $20$  é considerado razoável na maioria das

aplicações, embora alguns autores tenham relatado a necessidade de valores de  $k^*$  próximos de 6. A escolha de uma dimensão apropriada  $m$  para a amostra depende muito da precisão que se pretende para a estimação. Pode-se dizer que valores de  $m = 500$  ou  $m = 1000$  são bastante típicos nas aplicações. Existem algumas fórmulas que relacionam  $l$ ,  $m$  e  $k^*$  quando o interesse reside em quantis da distribuição a posteriori. Contudo, não existe uma regra geral e cada problema individual tem de ser analisado cuidadosamente para uma escolha adequada de  $l$ ,  $m$  e  $k^*$ .

O modo de utilizar a amostra gerada pelo procedimento Gibbs para fazer inferência, passa com certeza por problemas semelhantes, embora talvez menos complicados, aos encontrados numa análise exploratória de dados multivariados. Os problemas mais complicados dizem respeito ao estudo da convergência da cadeia de Markov para o estado de equilíbrio. Embora se saiba que, em condições bastante gerais, a convergência seja geométrica, há contudo ainda muitos problemas por resolver, tais como a obtenção de estimativas gerais para taxas de convergência. Acoplado a esse problema existe outro, ainda mais importante, que é o da construção de métodos para detectar a convergência da cadeia de Markov para o estado de equilíbrio. Entre as várias soluções existentes na literatura será descrito o método proposto por Geweke ( ) que sugere métodos baseados em séries temporais.

### Método de Geweke

Seja  $g(\theta)$  a função do parâmetro cujo valor esperado se pretende estimar. Como se tem dito, o método Gibbs atinge este propósito de estimação, simulando valores  $\theta^t$  de uma determinada cadeia de Markov. A estimativa do valor esperado a posteriori de  $g(\theta)$  é então dada pela média ergódica dos  $g(\theta^{(j)})$ . Assim, considerando a função real  $g(\theta)$ , a sua trajetória  $g^1, g^2, \dots$  construída a partir de  $g^t = g(\theta^t)$ , define uma série temporal. O método de Geweke ( ) baseia-se precisamente na aplicação de técnicas usuais em séries temporais para averiguar a convergência do método Gibbs. Observa-se a série ao longo de um número  $N$  suficientemente longo de iterações e calcula-se a média  $g_a = \frac{1}{n_a} \sum g_{\theta^t}$  à custa de  $n_a$  das primeiras iterações, bem como a média  $g_b = \frac{1}{n_b} \sum g_{\theta^t}$  à custa de  $n_b$  das últimas iterações. Se a cadeia é estacionária, então, a média da primeira parte da cadeia deve ser semelhante à média da segunda parte da cadeia. Admitindo

que  $n_a/N$  e  $n_b/N$  são positivos e  $N \rightarrow \infty$  pode mostrar-se que

$$\frac{(g_a - g_b)}{\sqrt{(s_a^2/n_a) + (s_b^2/n_b)}} \rightarrow N(0, 1) \quad (2.6)$$

onde  $s_a^2$  e  $s_b^2$  são estimativas independentes das variâncias assintóticas de  $g_a$  e  $g_b$ .

Assim, de acordo com o resultado desta estatística pode-se averiguar se houve ou não indicação de convergência.

## 2.7 Estimador de máxima verossimilhança via MCMC

Cappe et al. (2006) propôs uma técnica para derivar um estimador de máxima verossimilhança, utilizando os métodos MCMC, aplicado a uma família de funções exponenciais do tipo,

$$f(x|\theta) = c(\theta)e^{\theta x - \nu(x)} = c(\theta)h(x|\theta). \quad (2.7)$$

A dificuldade encontrada nesses problemas é que, fora a maioria dos exemplos padrões, a constante normalizadora deve ser desconhecida ou difícil de computar.

### 2.7.1 Famílias de densidades normalizadas

Suponha a família de funções não-negativas  $h_\theta : \theta \in \Theta$

$P_\theta$  podem ser simuladas sem o conhecimento da normalizadora  $c(\theta)$  pelo algoritmo de Metropolis-Hastings. A estimação da máxima verossimilhança pode ser feita, de novo sem o conhecimento da normalizadora ou suas derivadas, através do uso desta simulação de Monte Carlo. Algo surpreendente, desde que há pouca estrutura matemática para se trabalhar, é que o Monte Carlo em máxima verossimilhança converge para qualquer família sob a continuidade do mapeamento  $\theta \mapsto h_\theta(x)$ .

A log-verossimilhança correspondente a uma observação  $x$  leva por conveniência a razão das log-verossimilhanças contra um ponto arbitrário eixo  $\psi$  (Geyer, 1996)

$$l(\theta) = \log\left(\frac{h_\theta(X_{obs.})}{h_\psi(X_{obs.})}\right) - \log\left(\frac{c(\theta)}{c(\psi)}\right) \quad (2.1)$$

que resulta em

$$= \log\left(\frac{h_\theta(X_{obs.})}{h_\psi(X_{obs.})}\right) - \log\left(E_\psi \frac{h_\theta(X)}{h_\psi(X)}\right) \quad (2.2)$$

desde que se tem

$$E_\psi \frac{h_\theta(X)}{h_\psi(X)} = \int \frac{h_\theta(x)}{h_\psi(x)} f_\psi(x) d\mu(x) = \frac{1}{c(\psi)} \int h_\theta(x) d\mu(x) = \frac{c(\theta)}{c(\psi)}. \quad (2.3)$$

Embora a notação sugere que  $\psi$  é um ponto no espaço paramétrico de interesse, isto não é necessário.  $h_\psi$  pode ser qualquer função integrável não-negativa tal que, para qualquer  $\theta \in \Theta$ , se  $h_\psi(x) = 0$  então  $h_\theta(x) = 0$  exceto, talvez, para  $x$  em um conjunto nulo que deve depender de  $\theta$ . Dada uma amostra  $X_1, \dots, X_m$  de  $P_\psi$  gerada pelo algoritmo de Metropolis-Hastings, a aproximação natural de Monte Carlo da log-verossimilhança é

$$l_m(\theta) = \log\left(\frac{h_\theta}{h_\psi}\right) - \log\left(E_{m,\psi} \frac{h_\theta(X)}{h_\psi(X)}\right) \quad (2.4)$$

onde  $E_{m,\psi}$  denota a esperança empírica com respeito a  $P_\psi$  definida por

$$E_{m,\psi} g(X) = \frac{1}{m} \sum_{i=1}^m g(X^i). \quad (2.5)$$

Se a cadeia de Markov  $X_1, X_2, \dots$  (amostra) gerada pelo algoritmo de Metropolis-Hastings é irredutível, então,  $E_{m,\psi} g(X)$  converge quase certamente para  $E_\psi g(X)$  para qualquer função  $g$  integrável. Em particular,  $l_m(\theta)$  converge quase certamente para  $l(\theta)$ , para qualquer  $\theta$  fixo. O ‘quase certamente’ aqui significa para quase todos os caminhos amostrais da simulação de Monte Carlo; a observação  $x$  é considerada fixa.

Fazendo-se  $\hat{\theta}$  ser o maximizador de  $l$  e fazendo  $\hat{\theta}_m$  ser um maximizador de  $l_m$ , Geyer e Thompson ( ) mostraram que, se a família normalizada é uma família exponencial, então  $\hat{\theta}_m$  converge para  $\hat{\theta}$  quase certamente.

## 2.8 Validação e seleção de modelos

Segundo Paulino et al. ( ) uma análise cuidadosa de dados deve encarar sempre o problema da determinação do modelo, isto é, o problema da avaliação e escolha do modelo que melhor represente a situação em estudo. Essa é uma análise complexa e nesse capítulo são dadas algumas abordagens, do ponto de vista Bayesiano, que procuram responder questões tais como,

- (a) O modelo em consideração é adequado?
- (b) Dada uma coleção de modelos, qual é o melhor modelo?

Várias metodologias para abordar esta questão têm sido sugeridas na já vasta literatura dedicada a esta temática. A abordagem inicial sugere a utilização de fatores de Bayes para a comparação de modelos. Assim, se se tiver, por exemplo, dois modelos paramétricos em competição,  $M_1$  e  $M_2$ , para os dados  $x$ , com parâmetros  $\theta_1$  e  $\theta_2$  e distribuições a priori  $h_1(\theta_1)$  e  $h_2(\theta_2)$ , respectivamente, as distribuições marginais de  $x$  são dadas por

$$p(x_i|M_i) = \int f(x|\theta_i, M_i)p_i(\theta_i)d\theta_i, i = 1, 2, \quad (2.6)$$

sendo o fator de Bayes a favor de  $M_1$  dado por  $B = \frac{p(x|M_1)}{p(x|M_2)}$ .

Um problema sério com o fator de Bayes é que, para muitos modelos de interesse, pelo menos parte da especificação a priori é vaga de modo que a distribuição a priori  $p(\theta)$  é impropria. Assim, mesmo que a distribuição a posteriori seja própria, a distribuição a priori (distribuição marginal dos dados) usada no cálculo do fator de Bayes é impropria. Para superar essa dificuldade várias modificações ao fator de Bayes foram propostas, sendo ainda a sua utilização e aceitação um ponto bastante controverso. Para além da questão de uma definição adequada de fator de Bayes, há a subsequente dificuldade da sua computação. Os métodos MCMC podem ser úteis no cálculo de  $p(x|M_i)$ , e consequentemente no cálculo do fator de Bayes, mas, em geral o método de Monte Carlo tende a ser instável quando usado no cálculo de  $p(x|M_i)$  e portanto os resultados

no caso contínuo. Alternativamente, pode-se fazer o estudo da adequabilidade e seleção de modelos usando distribuições preditivas a posteriori.

### 2.8.1 Medidas de diagnóstico - medida de adequabilidade de um modelo

Seja  $x = (x_1, \dots, x_n)$  a amostra observada, isto é, a amostra usada para construir, à custa do modelo em consideração, a distribuição a posteriori  $p(\theta|x)$ . Suponha que se tem  $y = (y_1, \dots, y_n)$  uma amostra de validação, isto é, uma amostra independente de  $x$  usada para validar o modelo em estudo. A correspondente distribuição preditiva de  $y$  é

$$p(y|x) = \int f(y|\theta)p(\theta|x)d\theta. \quad (2.8.1)$$

Esta distribuição preditiva permite uma avaliação do modelo no sentido em que se os dados  $y$  não estiverem de acordo com  $p(y|x)$ , a qual é calculada à custa do modelo usando  $x$ , então não é de se esperar que o modelo seja adequado. O valor médio e a variância preditiva de cada componente de  $y$ ,

$$E(Y_i|x) = \int y_i p(y|x) dy; \text{var}(Y_i|x) = \int (y_i - E(Y_i|x))^2 p(y|x) dy \quad (2.8.2)$$

são úteis para definir os resíduos Bayesianos padronizados,

$$d_i = \frac{y_i - E(Y_i|x)}{\sqrt{\text{var}(Y_i|x)}}. \quad (2.8.3)$$

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

Esta discussão assume a existência de uma amostra de observações independentes, o que não acontece com frequência na prática. Claro que se a amostra inicial for de dimensão elevada, há sempre a possibilidade de a particionar em duas de modo que uma parte sirva de amostra observada para construir a distribuição a posteriori e a outra de amostra de validação para obter a distribuição preditiva. Não sendo viável particionar a

amostra para fazer validação cruzada, pode-se optar por um tipo de *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 por todas as observações à excepção de  $x_i$  pode-se obter as distribuições preditivas condicionais  $p(x_i|x_{(-i)})$ ,

$$p(x_i|x_{(-i)}) = \frac{p(x)}{p(x_{(-i)})} = \int f(x_i|\theta, x_{(-i)})h(\theta|x_{(-i)})d\theta, \quad (2.6)$$

e os consequentes res duos Bayesianos de eliminação padronizados

$$d'_i = \frac{x_i - E(X_i|x_{(-i)})}{\sqrt{\text{var}(X_i|x_{(-i)})}}, i = 1, \dots, n, \quad (2.7)$$

onde os valores médios e variâncias são calculados, obviamente, à custa das correspondentes distribuições preditivas condicionais. Pode-se proceder novamente a uma validação informal do modelo à custa destes res duos. Por outro lado, os valores observados de  $p(x_i|x_{(-i)})$ , comumente designados por *ordenadas preditivas condicionais* (CPO), podem ainda ser usados num diagnóstico informal. Com efeito, estes valores são um indicador de verossimilhança de cada observação, dadas todas as outras observações e, portanto, valores baixos de CPO devem corresponder a observações mal ajustadas.

Outras medidas de diagnóstico podem ser consideradas para além dos res duos definidos. De um modo geral, pode definir-se uma medida de diagnóstico através do valor esperado preditivo de uma função de avaliação,  $g(x_i; x_{i.obs})$ , que meça a discrepância entre um valor  $x_i$  observado e o valor  $x_{i.obs}$  realmente observado. São exemplos dessas funções (Gelfand, 1996) e correspondentes medidas de diagnóstico:

- $g_1(x_i; x_{i.obs}) = x_{i.obs} - x_i$ ;  $d_{1i} = x_{i.obs} - E(X_i|x_{(-i)})$ , correspondendo ao res duo de eliminação Bayesiano não padronizado.
- $g_2(x_i; x_{i.obs}) = I_{(-\infty, x_{i.obs}]}(x_i)$ ;  $d_{2i} = P(X_i \leq x_{i.obs}|x_{(-i)})$ . Este res duo é um indicativo da localização do valor realmente observado de  $x_i$ , (a que aqui se chamou, para não dar lugar a confusões,  $x_{i.obs}$ ) em relação à distribuição de  $X_i$ , sob o modelo em consideração. Valores de  $d_{2i}$  próximo de 0 ou 1, são indicadores de observações mal ajustadas pelo modelo.

Uma vantagem em usar as distribuições preditivas  $p(x_i|x_{(-i)})$ , está no facto destas exis-

tirem mesmo que  $p(x)$  não seja finita (no caso de a distribuição a priori ser impropria). Além disso, é de notar que estas distribuições não são mais do que distribuições condicionais completas e, como tal, determinam univocamente  $p(x)$ , no caso em que esta distribuição esta bem definida.

### 2.8.2 Comparação de modelos

Como já se referiu, a comparação de modelos via fator de Bayes, sendo o método de eleição para um Bayesiano purista, não é contudo, do ponto de vista prático, um método em geral muito adequado dadas as dificuldades a este método associada. Há assim a necessidade de procurar soluções alternativas para a seleção de modelos. Tal como para a adequabilidade do modelo pode-se optar por métodos informais e gráficos. Vai-se mencionar aqui os métodos mais utilizados.

Existem alguns critérios baseados em aproximações ao fator de Bayes (BF) entre dois modelos  $M_1$  e  $M_2$ , nomeadamente o critério AIC (Akaike Information Criterion), e o critério BIC (Bayesian Information Criterion). Estes critérios têm como objetivo incluir a complexidade do modelo no critério de seleção. São critérios que ‘penalizam’ a verossimilhança e são introduzidos a seguir.

. Critério AIC

$$\Delta AIC = -2 \ln \left[ \frac{\sup_{M_1} f(x|\theta_1, M_1)}{\sup_{M_2} f(x|\theta_2, M_2)} \right] - 2(p_1 - p_2), \quad (2.8.1)$$

onde  $p_i$ ,  $i = 1, 2$  representa o número de parâmetros de cada modelo. Este critério é baseado em considerações frequentistas de estatística assintótica, mas se aproxima  $-2 \ln BF$  se a informação contida na distribuição a priori aumentar na mesma razão que a informação contida na verossimilhança.

. Critério BIC

$$\Delta BIC = -2 \ln \left[ \frac{\sup_{M_1} f(x|\theta_1, M_1)}{\sup_{M_2} f(x|\theta_2, M_2)} \right] - 2(p_1 - p_2) \ln(n), \quad (2.8.2)$$

onde  $n$  é a dimensão da amostra e  $p_i$ ,  $i = 1, 2$ , é o número de parâmetros de cada caso. Para amostras de dimensão elevada, pode-se mostrar que  $\Delta BIC$  é uma boa aproximação de  $-2 \ln(BF)$ . De modo a facilitar o cálculo de mínimos por simulação sugere-se uma

modificação neste critério calculando para cada caso  $M_i$  em competição

$$B\hat{I}C_i = 2E[\ln(L(\theta)|x, M_i)] - p_i \ln(n) \quad (. .)$$

e escolhendo o modelo que apresenta maior valor de  $B\hat{I}C_i$ . O problema na utilização deste critério reside em uma definição adequada de  $n$ , (dimensão da amostra) e de  $p_i$ , o número de parâmetros do modelo  $M_i$ .

### 2.8.3 Seleção de modelos via métodos MCMC

Apenas em casos simples é possível implementar as técnicas de diagnóstico e seleção de modelos sem o recurso a métodos computacionais. O objetivo desta seção é o de mostrar como é que os métodos de simulação descritos no presente e anteriores capítulos podem ser usados para o cálculo das quantidades envolvidas na avaliação de adequabilidade e na seleção de modelos. Suponha-se que, por métodos de simulação, obteve-se uma amostra  $\theta_j^*; j = 1, \dots, m$  da distribuição a posteriori  $p(\theta|x)$ . Um cálculo efetivo das medidas de diagnóstico definidas anteriormente passa pela resolução dos seguintes problemas:

- (a) Como estimar a densidade preditiva de interesse;
- (b) Como calcular o valor esperado da função de avaliação  $g(x_i; x_{i.obs})$  sob esta densidade;
- (c) Como amostrar da densidade preditiva.

#### Estimação da densidade preditiva a posteriori

Segundo Achcar et al. (2006), utilizando o método de particionar a amostra, *Jackknife*, descrito anteriormente a densidade preditiva para  $x_i$  dado  $\tilde{x}_{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  é

$$c_i := f(x_i|\tilde{x}_{(i)}) = \int f(x_i|\theta) p(\theta|\tilde{x}_{(i)})d\theta \quad (. .)$$

onde  $p(\theta|\tilde{x}_{(i)})$  é a densidade a posteriori para o vetor de parâmetros dado  $\tilde{x}_{(i)}$ . Usando o amostrador de Gibbs ou Metropolis-Hastings, (. .) pode ser aproximada por seu estimador de Monte Carlo,

$$\hat{f}(x_i|\tilde{x}_{(i)}) = \frac{1}{m} \sum_{j=1}^m f(x_i|\theta^j) \quad (. . 6)$$

onde os  $\theta_j$  s o os gerados pelo método Metropolis-Hastings,  $j = 1, 2, \dots, m$  e  $m$  corresponde ao tamanho da cadeia gerada. Pode-se utilizar  $c_i = \hat{f}(x_i|\tilde{x}_{(i)})$  na seleção de modelos, fazendo o produtorio  $P_l = \prod_{i=1}^n c_i(l)$  (onde  $l$  indica o modelo) e depois verificando qual é o maior valor entre os  $P_l$ . Pode-se ainda, como critério de escolha de modelo, considerar a verossimilhança marginal de um grupo de dados  $D$  para o modelo  $M_l$  dado por

$$P(D|M_l) = \int_{\theta_l} L(D|\theta_l, M_l) (\theta_l|M_l)d\theta_l \quad ( . . )$$

onde  $D$  s o os dados,  $M_l$  é o modelo especificado,  $\theta_l$  é o vetor de parâmetros em  $M_l$ ,  $L(D, \theta, M_l)$  é a função de verossimilhança, e  $(\theta_l|M_l)$  é a priori. O critério de fator de Bayes indica que o modelo  $M_1$  é melhor que  $M_2$  se  $P(D|M_2) < P(D|M_1)$ . Um estimador de Monte Carlo para a verossimilhança marginal  $P(D|M_l)$  é dado por

$$\hat{P}(D|M_l) = \frac{1}{m} \sum_{j=1}^m L(D|\theta_l^{(j)}, M_l) \quad ( . . )$$

onde  $\theta_l^{(j)}$ ,  $j = 1, 2, \dots, m$  s o gerados pelos métodos MCMC. Outro caminho para estimar a verossimilhança marginal  $P(D|M_l)$  é dado da seguinte forma. Considerando uma amostra da distribuição a posteriori, tem-se,

$$\hat{P}(D|M_l) = \left( \frac{1}{m} \sum_{j=1}^m \frac{1}{L(D|\theta_l^{(j)}, M_l)} \right)^{-1} \quad ( . . )$$

Neste caso, a função de importância é a distribuição a posteriori.

### Faixas de valores para o fator de Bayes

Fazendo  $B_{lk} = \frac{\hat{P}(D|M_l)}{\hat{P}(D|M_k)}$ , onde  $l$  e  $k$  representam os modelos, tem-se a seguinte sugestão de faixas de valores para a escolha do modelo, segundo Raftery ( ).

Quando  $B_{lk} > 1$  os dados favorecem  $M_l$  a  $M_k$ . E quando  $B_{lk} < 1$  ocorre o contrário. O uso do fator de Bayes para comparar teorias científicas, foi primeiro proposto por Jeffreys ( ) que em 6 propôs a seguinte regra de domínios para interpretar  $B_{lk}$ . Quando  $1 \leq B_{lk} \leq 3$ , há evidência para  $M_l$ , mas esta ‘não é mais do que uma mera referência’, quando  $3 \leq B_{lk} \leq 10$  a evidência é positiva, quando  $10 \leq B_{lk} \leq 100$  é forte e quando  $B_{lk} > 100$  é decisiva.

## Capítulo 3

# Modelos para populações de plantas daninhas

### 3.1 Modelo de regressão não-linear

Considere o modelo de regressão não-linear

$$x_i = f(y_i; \psi)$$

### 3.2 Densidade a priori para os parâmetros

Segundo Favoretti ( ) antes de analisar estatisticamente os dados de um particular modelo, deve-se incorporar o nosso conhecimento prévio a respeito de seus parâmetros. Uma maneira conveniente de se fazer isto é formular uma densidade de probabilidade para os parâmetros  $\theta$ . Essa densidade de probabilidade para o parâmetro  $\theta$  é identificada como densidade a priori e é denotada por  $p(\theta)$ . Quando não se tem nenhuma informação a respeito de  $\theta$ , deve-se formular uma priori, tal que toda informação acerca do parâmetro seja exclusivamente dos dados, ou seja, deve-se considerar uma priori não-informativa. Várias regras para a formulação de uma priori não-informativa estão disponíveis, entre as quais pode-se destacar a regra de Jeffreys (6). Neste trabalho, destaca-se o uso da densidade a priori não-informativa de Jeffreys.

Seja  $l(\theta)$  o logaritmo da função de verossimilhança para o parâmetro  $\theta$  de um modelo; a densidade a priori de Jeffreys para  $\theta$  é dada por

$$p(\theta) \propto \left\{ \text{Det}(I(\theta)) \right\}^{1/2} \tag{.5}$$

onde  $I(\theta)$  é a matriz de informação esperada (Fisher), dada por

$$E \left[ \frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s} \right] \tag{.6}$$

Quando  $\theta = (\alpha', \beta')$  e supondo independência a priori, isto é,  $p(\alpha, \beta) = p_1(\alpha)p_2(\beta)$ , para obter  $p_1(\alpha)$  e  $p_2(\beta)$  separadamente, a regra de Jeffreys trata os outros parâmetros como conhecidos. Por exemplo,

$$p_1(\alpha) \propto \left\{ \text{Det}(I_1(\alpha)) \right\}^{1/2} \tag{.6}$$

onde  $[I_1(\alpha)]_{r,s} := E \left[ - \frac{\partial^2 l(\alpha, \beta)}{\partial \alpha_r \partial \alpha_s} \right]$ .

Aplica-se, agora, essa idéia ao modelo ( . ) para obter a priori (Jeffreys) dos parâmetros. Para isso considera-se, primeiramente, a função de verossimilhança,

$$L(\theta, \sigma^2 | x, y) = (2\sigma^2)^{-n/2} e^{\left\{ -\frac{1}{2\sigma^2} S(\theta) \right\}} \tag{.7}$$

onde  $S(\theta) = \|x_i - f(y_i, \theta)\|^2$ , e o seu logaritmo é dado por

$$l(\theta, \sigma|x, y) = -n \ln \sigma - \frac{1}{2\sigma^2} S(\theta). \quad ( . . )$$

Expandindo  $S(\theta)$  diretamente por uma expansão quadrática em série de Taylor, pode-se obter

$$I_1(\theta) = \frac{1}{2\sigma^2} E \left[ \frac{\partial^2 S(\theta)}{\partial \theta \partial \theta'} \right] = \frac{1}{\sigma^2} F'(\theta) F(\theta) \propto F'(\theta) F(\theta) \quad ( . . )$$

onde  $F(\theta) = \left[ \left( \frac{\partial f_i}{\partial \theta_j} \right) \right]$ ,  $i = 1, \dots, n$  (numero de dados  $x_i$ ) e  $j = 1, \dots, p$  (número de parâmetros). Tem-se ainda que

$$I_2(\sigma) = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4} E[S(\theta)] = \frac{n}{2\sigma^2}. \quad ( . . )$$

Finalmente, tem-se que se  $\nu = (\theta', \nu)'$ , onde  $\nu = \sigma^2$ , então, de ( . . ) e ( . . ), a matriz de informação de Fisher é dada por

$$-E \left[ \frac{\partial^2 l}{\partial \nu \partial \nu'} \right] = \begin{pmatrix} \frac{1}{2\sigma^2} E \left[ \frac{\partial^2 S}{\partial \theta \partial \theta'} \right] & -E \left[ \frac{\partial^2 l}{\partial \theta \partial \nu} \right] \\ -E \left[ \frac{\partial^2 l}{\partial \nu \partial \theta'} \right] & -E \left[ \frac{\partial^2 l}{\partial \nu^2} \right] \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} F'(\theta) F(\theta) & 0 \\ 0' & \frac{n}{2\sigma^4} \end{pmatrix}. \quad ( . . )$$

Assim, uma priori não-informativa para  $(\theta, \sigma)$ , parâmetros do modelo ( . . ), onde  $\theta$  e  $\sigma$  são considerados independentes a priori é dada por

$$p(\theta, \sigma) \propto \text{Det} \left( F'(\theta) F(\theta) \right)^{1/2} \sigma^{-1}. \quad ( . . )$$

Para maiores detalhes ver Favoretti ( ).

### 3.3 Densidades a posteriori para os parâmetros

Uma vez definido o modelo para os dados e a distribuição a priori, deve-se combinar a informação prévia sobre o parâmetro  $\theta$  com a informação contida nos dados, obtendo assim uma distribuição a posteriori para  $\theta$ .

Se  $p(\theta, \sigma)$  representa a distribuição a priori dos parâmetros desconhecidos e  $p(y|\theta, \sigma)$  é a função de verossimilhança, frequentemente, escrita como  $L(\theta, \sigma|x, y)$ , pode-se obter

as distribuições a posteriori conjuntas e marginais para os parâmetros. Considerando o modelo não-linear dado na forma de regressão ( . ), onde cada  $y_i$  tem distribuição normal  $N(0, \sigma^2)$ , a função densidade a posteriori para  $\theta$  e  $\sigma$ , pela fórmula de Bayes ( . ), é dada por

$$p(\theta, \sigma | x, y) = \frac{L(\theta, \sigma | x, y)p(\theta, \sigma)}{\int \int_{\theta \in \Theta} L(\theta, \sigma | x, y)p(\theta, \sigma) d\theta d\sigma} \quad ( . )$$

em que  $L(\theta, \sigma^2 | x, y)$  foi definido e  $\Theta$  é o espaço paramétrico.

Combinando a informação dos dados ( . ) com a informação prévia ( . ), obtém-se a densidade a posteriori conjunta para  $(\theta, \sigma)$  dada por

$$p(\theta, \sigma | x, y) \propto \text{Det}\left(F'(\theta)F(\theta)\right)^{1/2} \sigma^{-(n+1)} \exp\left(-\frac{1}{2\sigma^2}S(\theta)\right). \quad ( . )$$

em que  $F(\theta)$  foi definido.

Integrando ( . ) com respeito a  $\sigma$  tem-se a densidade a posteriori para  $\theta$ , dada por

$$p(\theta, y) \propto \text{Det}\left(F'(\theta)F(\theta)\right)^{1/2} S(\theta)^{-n/2} \quad ( . )$$

ou  $p(\theta, y) = K^{-1} \text{Det}\left(F'(\theta)F(\theta)\right)^{1/2} S(\theta)^{-n/2}$ , onde

$$K = \int_{\Theta} S(\theta)^{-n/2} d\theta \quad ( . 6)$$

representa a constante normalizadora e pode ser calculada através de métodos numéricos. O resultado acima também é válido para o caso discreto substituindo-se a integral em ( . 6) por uma soma.

A distribuição a posteriori ( . ) pode ser usada para estimação pontual (por exemplo a média a posteriori  $E(\theta | y)$  pode ser usada como estimativa de  $\theta$ ) e para a construção de intervalos e regiões a posteriori.

Uma estimativa Bayesiana por intervalo pode ser obtida, através dos intervalos de credibilidade Bayesianos aproximados, onde utiliza-se métodos numéricos encontrando as áreas aproximadas abaixo da curva  $p(\theta | x, y)$ , ou seja, determina-se numericamente os pontos  $(\theta_*, \theta^*)$ , tais que  $P(\theta_* \leq \theta \leq \theta^*) = 1 - \alpha$ . Assim, poderia-se ter,

$$\int_{-\infty}^{\theta_*} p(\theta | x, y) d\theta = \alpha/2; \int_{\theta^*}^{\infty} p(\theta | x, y) d\theta = \alpha/2 \quad ( . )$$

onde  $p(\theta|y)$  é a densidade a posteriori unimodal de  $\theta$  e  $(\theta_*, \theta^*)$  representa o intervalo Bayesiano que contém o verdadeiro valor de  $\theta$  com probabilidade  $(1 - \alpha)$ .

Nas seções seguintes, ir o ser obtidas inferências sobre os parâmetros do modelo através de uma análise Bayesiana para o modelo (3.1), considerando uma densidade a priori não-informativa, explorando o método numérico de Simpson (Barroso et al., 2010).

### 3.4 Distribuição à posteriori do modelo de 2 parâmetros

#### 3.4.1 Uma densidade a priori não-informativa para $\theta = (s, a)$ e $\sigma$

A partir do momento em que se observa os dados, toda a informação sobre os parâmetros está contida na função de verossimilhança. A função de verossimilhança para os parâmetros  $s, a$  e  $\sigma^2$  do modelo (3.1), considerando uma amostra aleatória do tamanho  $n$  é dada por

$$L(s, a, \sigma^2|x, y) = (2\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(x_i - \frac{sy_i}{1+ay_i}\right)^2\right\}. \quad (3.2)$$

Uma densidade a priori para  $\theta' = (s, a)$  e  $\sigma$ , assumindo  $\theta'$  independente de  $\sigma$  é dada por

$$p(s, a, \sigma) = p(s, a)p(\sigma). \quad (3.3)$$

Considerando que o pesquisador não tenha informação prévia a respeito dos parâmetros, será utilizada a regra multiparamétrica de Jeffreys, para obtenção de uma priori não-informativa para  $s$  e  $a$ , e considerar uma densidade a priori não-informativa para  $\sigma$  proporcional a  $\sigma^{-1}$ . Assim, uma densidade a priori conjunta para os parâmetros (ver (3.3)) é dada por

$$p(s, a, \sigma) \propto \sigma^{-1} \left\{ \text{Det}(F'(\theta)F(\theta)) \right\}^{1/2}, 0 < a < 1, s < \infty; \sigma > 0 \quad (3.4)$$

onde  $\text{Det}(F'(\theta)F(\theta)) = s^2g(a)$ , e

$$g(a) = \left[ \sum_{i=1}^n \frac{y_i^4}{(1+ay_i)^4} \right] \left[ \sum_{i=1}^n \frac{y_i^2}{(1+ay_i)^2} \right] - \left( \left[ \sum_{i=1}^n \frac{y_i^3}{(1+ay_i)^3} \right] \right)^2. \quad (3.5)$$

Portanto, para o modelo ( . 6 ), a densidade a priori de Jeffreys para  $s, a$  e  $\sigma$  é dada por

$$p(s, a, \sigma) \propto s\sigma^{-1}[g(a)]^{1/2}, 0 < a < 1, s < \infty; \sigma > 0. \quad ( . 7 )$$

### 3.4.2 Densidade a posteriori conjunta para $\theta$ e $\sigma$

A densidade a posteriori conjunta para  $s, a$  e  $\sigma$ , considerando a priori ( . 7 ) e representando o vetor  $(x, y)$  por ‘dados’, é dada por

$$p(s, a, \sigma | \text{dados}) \propto p(s, a, \sigma)L(s, a, \sigma^2) \quad ( . 8 )$$

onde  $L(s, a, \sigma^2)$  é a função de verossimilhança ( . 5 ) e  $p(s, a, \sigma)$  representa a densidade a priori ( . 7 ). A densidade conjunta para os parâmetros do modelo ( . 5 ), substituindo ( . 8 ) e ( . 7 ) em ( . 5 ) é dada por

$$p(s, a, \sigma | \text{dados}) \propto s\sigma^{-(n+1)}[g(a)]^{1/2} \exp\left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n \left(x_i - \frac{sy_i}{1+ay^i}\right)^2 \right\}. \quad ( . 9 )$$

Integra-se ( . 9 ) em relação a  $\sigma$ , obtendo a densidade conjunta para  $s$  e  $a$ , dada por

$$p(s, a | \text{dados}) \propto s[g(a)]^{1/2}[s(s, a)]^{-n/2}. \quad ( . 10 )$$

onde  $0 < a < 1, s < \infty$  e  $g(a)$  é dada por ( . 5 ).

### 3.4.3 Densidades a posteriori marginais para $s$ e $a$

As densidades a posteriori marginais para  $s$  e  $a$  são obtidas integrando-se a densidade a posteriori conjunta ( . 10 ) com relação aos parâmetros  $a$  e  $s$ , respectivamente, isto é,

$$p(s | \text{dados}) \propto \int_0^{\infty} p(s, a | \text{dados}) da = s \int_0^{\infty} g(a)^{1/2} S(s, a)^{-n/2} da, \quad ( . 6 )$$

$$p(a | \text{dados}) \propto \int_0^{\infty} p(s, a | \text{dados}) ds = g(a)^{1/2} \int_0^{\infty} s S(s, a)^{-n/2} ds. \quad ( . 11 )$$

Como tem-se dificuldade na resolução analítica das integrais ( . 11 ) e ( . 6 ), será utilizado o método numérico de Simpson. Os detalhes do programa utilizado serão explicados no capítulo seguinte.

### 3.5 Distribuição à posteriori do modelo de 3 parâmetros

#### 3.5.1 Uma densidade a priori não-informativa para $\theta = (s, a, b)$ e $\sigma$

### 3.5.2 Densidade a posteriori conjunta para $\theta(s,a,b)$ e $\sigma$

A densidade a posteriori conjunta para  $s$ ,  $a$ ,  $b$  e  $\sigma$ , considerando a priori ( . . ) e representando o vetor  $(x, y)$  por ‘dados’, é dada por

$$p(s, a, b, \sigma | \text{dados}) \propto p(s, a, b, \sigma) L(s, a, \sigma^2) \quad ( . . )$$

onde  $L(s, a, b, \sigma^2)$  é a função de verossimilhança ( . . ) e  $p(s, a, b, \sigma)$  representa a densidade a priori ( . . ). A densidade conjunta para os parâmetros do modelo ( . . ), substituindo ( . . ) e ( . . ) em ( . . ) é dada por

$$p(s, a, b, \sigma | \text{dados}) \propto \sigma^{-(n+1)} g(s, a, b)^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( x_i - \frac{sy_i}{(1+ay_i)^b} \right)^2 \right\}. \quad ( . . )$$

Integra-se ( . . ) em relação a  $\sigma$ , obtendo a densidade conjunta para  $s, a$  e  $b$  dada por

$$p(s, a, b | \text{dados}) \propto g(s, a, b)^{1/2} [S(s, a, b)]^{-n/2} \quad ( . . )$$

onde  $0 < a < 1$ ,  $b < \infty$ ,  $s < \infty$  e  $g(s, a, b)$  é dada por ( . . ).

### 3.5.3 Densidades a posteriori marginais para $s, a$ e $b$

As densidades a posteriori marginais para  $s, a$  e  $b$  são obtidas integrando-se a densidade a posteriori conjunta ( . . ) com relação aos parâmetros ‘ $a$  e  $b$ ’, ‘ $s$  e  $b$ ’ e ‘ $s$  e  $a$ ’ respectivamente, isto é,

$$p(s | \text{dados}) \propto \int_0^\infty \int_0^\infty p(s, a, b | \text{dados}) da db, \quad ( . . 6)$$

$$p(a | \text{dados}) \propto \int_0^\infty \int_0^\infty p(s, a, b | \text{dados}) ds db, \quad ( . . )$$

$$p(b | \text{dados}) \propto \int_0^\infty \int_0^\infty p(s, a, b | \text{dados}) ds da. \quad ( . . )$$

E novamente utiliza-se o método numérico de Simpson para calcular essas integrais.

### 3.6 MCMC para os parâmetros e seleção dos modelos

De acordo com os algoritmos descritos na Seção 3.6 (Metropolis-Hastings e Amostrador de Gibbs) as cadeias de Markov do espaço paramétrico foram geradas a partir das densidades condicionais para os parâmetros como segue.

#### Condicionais para 2 parâmetros

$$p(s|a, \text{dados}) \propto s[s(s, a)]^{-n/2} \quad (3.6.1)$$

para gerar  $s$ , através do algoritmo de Metropolis-Hastings

$$p(a|s, \text{dados}) \propto [g(a)]^{1/2}[s(s, a)]^{-n/2} \quad (3.6.2)$$

para gerar  $a$ , também através do algoritmo de Metropolis-Hastings e

$$p(\sigma|s, a, \text{dados}) \propto \sigma^{-(n+1)} \exp\left\{\frac{1}{2\sigma^2} \sum_{i=1}^n \left(x_i - \frac{sy_i}{1+ay^i}\right)^2\right\} \quad (3.6.3)$$

para gerar  $\sigma$ , através do Amostrador de Gibbs, onde  $n$  é o tamanho da amostra  $(x_i, y_i)$  que referiu-se como 'dados' nas condicionais acima. Nas duas primeiras condicionais (3.6.1) e (3.6.2) não temos um núcleo de modo que venha a sugerir uma distribuição conhecida para gerar a cadeia, enquanto na terceira condicional (3.6.3), tem-se um núcleo de uma gama invertida para  $\sigma$ .

#### Condicionais para 3 parâmetros

$$p(s, a, b|\text{dados}) \propto g(s, a, b)^{1/2}[S(s, a, b)]^{-n/2} \quad (3.6.4)$$

para gerar  $s, a$  e  $b$  através do algoritmo de Metropolis-Hastings e

$$p(\sigma|s, a, b, \text{dados}) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(x_i - \frac{sy_i}{(1+ay^i)^b}\right)^2\right\} \quad (3.6.5)$$

para gerar  $\sigma$ , através do Amostrador de Gibbs. Nas condicionais para  $s, a$  e  $b$  (3.6.4) (que no caso são iguais) não tem um núcleo de modo que venha a sugerir uma distribuição

conhecida para gerar a cadeia, enquanto que na condicional ( . . ), tem-se um núcleo de uma gama invertida para  $\sigma$ .

### 3.6.1 Algoritmo MCMC

Para gerar a cadeia de parâmetros foi utilizado o método de Metropolis-Hastings, cujos passos de iteração são os dados pela seguinte sequência, considerando o modelo de parâmetros

- ) Inicializa-se o contador em  $t = 0$ , fazendo  $\theta^0 = (s^0, a^0)$ .
- ) Gera-se um valor  $s'$  da distribuição  $q(s)$  (chamada distribuição proposta).
- ) Com  $a = a^t$ , calcula-se a probabilidade de aceitação  $\alpha(s', s) = \min \left[ 1, \frac{p(s'|a, \text{Dados})}{p(s|a, \text{Dados})} \right]$  e gera-se um valor  $u$  da distribuição uniforme  $U(0, 1)$ .
- ) Caso  $u \leq \alpha$  então aceita-se  $s'$ , ou seja faz-se  $s^{t+1} = s'$ . Caso contrário faz-se  $s^{t+1} = s^t$ .
- ) Gera-se um valor  $a'$  da distribuição  $q(a)$ .
- 6) Com  $s = s^{t+1}$ , calcula-se a probabilidade de aceitação  $\alpha(a', s) = \min \left[ 1, \frac{p(a'|s, \text{Dados})}{p(a|s, \text{Dados})} \right]$  e gera-se um valor  $u$  da distribuição uniforme  $U(0, 1)$ .
- ) Caso  $u \leq \alpha$  então aceita-se  $a'$ , ou seja faz-se  $a^{t+1} = a'$ . Caso contrário faz-se  $a^{t+1} = a^t$ .
- ) Incrementa-se  $t$  para  $t + 1$  e retorna-se ao passo . . .

Obs . . : As funções propostas  $q(a)$  e  $q(s)$  são normais centradas em médias que foram obtidas pelo método numérico de Simpson.

Obs . . : Para o caso de parâmetros acrescenta-se um passo a mais para a geração do parâmetro  $b$  e o procedimento é análogo ao descrito anteriormente, com uma função proposta  $q(b)$  com as características citadas na Obs . . .

### 3.6.2 Densidades preditivas ordenadas (DPO)

Para fazer a seleção de modelos foram utilizadas as seguintes funções  $f(x_i|\tilde{\theta})$  presentes em ( . . ) e ( . . 6) para o cálculo das densidades preditivas como segue.

**Função para a DPO do modelo de 2 parâmetros**

$$f(x_i|\tilde{\theta}) = \frac{1}{\sqrt{2}\sigma} \exp\left[-\frac{(x_i - \frac{sy_i}{(1+ay_i)^b})^2}{2\sigma^2}\right]. \quad (. . .)$$

**Função para a DPO do modelo de 3 parâmetros**

$$f(x_i|\tilde{\theta}) = \frac{1}{\sqrt{2}\sigma} \exp\left[-\frac{(x_i - \frac{sy_i}{(1+ay_i)^b})^2}{2\sigma^2}\right]. \quad (. . .)$$

Assim, para obter-se a aproximação de Monte Carlo da equação  $c_i := f(x_i|\tilde{x}_{(i)}) = \int f(x_i|\tilde{\theta}) (\tilde{\theta}|\tilde{x}_{(i)})d\tilde{\theta}$  (ver . . .), considera-se a equação  $\hat{f}(x_i|\tilde{x}_{(i)}) = \frac{1}{M} \sum_{j=1}^M f(x_i|\tilde{\theta}^j)$  (ver . 6) onde os parâmetros são gerados pelas densidades condicionais a posteriori, descritos na seção anterior.

### 3.6.3 Métodos de seleção AIC e BIC

Foram usados ainda, para critérios de comparação, os métodos AIC e BIC (definidos na Subseção . . .) para selecionar modelos. E algumas aproximações foram consideradas segundo Barreto e Andrade (2006), para que isto fosse realizado. Desse modo tem-se:

**Para o caso AIC**

$$AIC = \ln(\sigma^2) + \frac{2p}{n} \quad (. . 6)$$

**Para o caso BIC**

$$BIC = \ln(\sigma^2) + \frac{p \ln(n)}{n} \quad (. . .)$$

onde  $p$  é o número de parâmetros mais o  $\sigma$  e  $n$  é o tamanho da amostra. E o valor para  $\sigma$  nas expressões acima é a média dos  $\sigma$ 's gerados na cadeia de Markov.

### 3.6.4 MCMC em estimador de máxima verossimilhança

Tendo as funções de verossimilhança dadas em (3.6.1) e (3.6.2), tem-se as seguintes funções de  $h_\theta$ ,  $h_\psi$  e  $P_\psi$ , descritas na Seção 3.6.4, obtidas com base nos artigos Geyer (1992) e Hurter e Wu (2004).

#### Para 2 parâmetros

$$h_\theta(dados) = (2\sigma)^{-n/2} \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(x_i - \frac{sy_i}{1+ay_i}\right)^2\right] \quad (3.6.3)$$

e para obter  $h_\psi$  foi tomado  $\psi_i = E(\theta_i | dados)$ ,  $(i^1, \dots, i^m)$ , para cada cadeia de  $\theta_i$  gerada, como um valor inicial para  $\psi$ . Desse modo, para a geração dos  $X_i^1, X_i^2, \dots, X_i^m$ , onde  $m$  é o tamanho da cadeia e  $i$  varia de 1 até  $n$  (amostra). Tem-se, então,  $m$  cadeias de  $n$  amostras e a função  $P_\psi$  fica sendo, para cada  $X_i$ ,

$$P_\psi(X_i) = \frac{1}{\sqrt{2\sigma^*}} \exp\left[-\frac{1}{2\sigma^{*2}} \left(X_i - \frac{y_{iobs} \cdot s^*}{1+a^* y_{iobs}}\right)^2\right] \quad (3.6.4)$$

onde  $y_{iobs}$  é o  $i$ -ésimo  $y$  observado e  $\psi = (s^*, a^*, \sigma^*) = (E(s|dados), E(a|dados), E(\sigma|dados))$

#### Para 3 parâmetros

$$h_\theta = (2\sigma)^{-n/2} \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(x_i - \frac{sy_i}{(1+ay_i)^b}\right)^2\right] \quad (3.6.5)$$

e para obter  $h_\psi$  foi tomado  $\psi_i = E(\theta_i | dados)$ ,  $(i^1, \dots, i^m)$ , para cada cadeia de  $\theta_i$  gerada, como um valor inicial para  $\psi$ . Desse modo, para a geração dos  $X_i^1, X_i^2, \dots, X_i^m$ , a função  $P_\psi$  fica sendo, para cada  $X_i$ ,

$$P_\psi(X_i) = \frac{1}{\sqrt{2\sigma^*}} \exp\left[-\frac{1}{2\sigma^{*2}} \left(X_i - \frac{y_{iobs} \cdot s^*}{(1+a^* y_{iobs})^{b^*}}\right)^2\right] \quad (3.6.6)$$

onde  $\psi = (s^*, a^*, b^*, \sigma^*) = (E(s|dados), E(a|dados), E(b|dados), E(\sigma|dados))$ .

## Capítulo 4

# Resultados

Foram escolhidas as seguintes espécies de plantas daninhas para inferência estatística: *Digitaria Ciliaries* (Capim Colchão) e *Panicum maximum* (Capim Guiné) de ‘folha estreita’ e *Euphorbia heterophylla* (Capim Leiteiro) de ‘folha larga’. Apresentam-se resultados para dois casos considerando-se os modelos de regressão (3.1) (Caso 1) e (3.2) (Caso 2) com base no modelo de produção de sementes (3.3) com taxa de crescimento (3.4). Os conjuntos de dados utilizados são apresentados na Tabela 4.1 e foram obtidos a partir de um experimento conduzido em campo<sup>1</sup>, em cultura de milho por plantio direto, na área experimental da Embrapa Milho e Sorgo, Sete Lagoas, MG. A Tabela 4.2 apresenta as estimativas Bayesianas (Lopes et al., 2016) para os parâmetros do Caso 1 e suas respectivas densidades estão representadas na Figura 4.1, com

$$f(Y_i, s, a) = \frac{sY_i}{(1 + aY_i)} \quad (4.1)$$

e a Tabela 4.3 apresenta as estimativas Bayesianas para os parâmetros do Caso 2 e suas respectivas densidades estão representadas na Figura 4.2, com

$$f(Y_i, s, a) = \frac{sY_i}{(1 + aY_i)^b}. \quad (4.2)$$

Para fins de comparação, é apresentado a Tabela 4.4 que faz referência a resultados clássicos obtidos a partir do aplicativo R<sup>2</sup>. Para este caso a base de dados foi a mesma

---

<sup>1</sup>EMBRAPA Projeto 55.2004.509.00: Rede de Conhecimento em Agricultura de Precisão para Condições do Serrado e dos Campos Gerais

<sup>2</sup><http://www.r-project.org/>

fornecida pela Tabela . . . Nesse caso foi considerado para o modelo de regressão, o teste de significância de seus parâmetros. Assim, o teste da hipótese

$$H_0 : \theta = 0 \quad ( . . )$$

$$H_1 : \theta \neq 0 \quad ( . . )$$

pode ser feito por meio da estatística t-Student, rejeitando-se  $H_0$  a um nível de significância desejado  $\alpha$  ( . . . ) e, concluindo-se que os parâmetros testados são significativos se  $t > t_0$ , onde  $t_0$  é o valor crítico da distribuição t-Student, com  $(n-p)$  graus de liberdade. Sendo assim, observa-se, da Tabela . . . que somente o valor do parâmetro  $a$  do Leiteiro não foi significativo. Para ver maiores detalhes dos resultados clássicos consulte Vismara ( . . . 6).

#### 4.1 Inferência Bayesiana via método de Simpson

Para obter os resultados das Tabelas . . e .6 foi feito um programa em linguagem FORTRAN (ver pseudo-código no Apêndice B), estando dividido da seguinte forma. Um programa principal que entra com os dados  $(x, y)$  da Tabela . . . , intervalos de integração e retorna os parâmetros estimados, mediana, desvio padrão e I.C. (Intervalo de Credibilidade). E programas secundários que são funções e subrotinas que fazem o cálculo de integrais (pelo método de Simpson).

#### 4.2 Inferência Bayesiana via método MCMC

Para obter os resultados das Tabelas . . e . foram feitos programas em Matlab (ver (Lopes e Oliveira, . . . )). As Tabelas . . e . apresentam as estimativas da esperança, mediana, desvio padrão e o intervalo de credibilidade dos parâmetros, calculados a partir da geração da cadeia de Markov. Esse programa também fornece resultados referentes às taxas de aceitação e rejeição para os parâmetros  $s$  e  $b$  calculados através do algoritmo de Metropolis-Hastings (para esse caso, até . . . de rejeição, tomou-se como nível (Besag, . . . )) e fornece informações a respeito da convergência da cadeia de Markov, que será explicado a seguir.

Na análise da convergência da cadeia foi utilizado o critério de Geweke, para uma única cadeia, comentado anteriormente. E para avaliar através desse critério foi utilizada uma função (já pronta) que contém a descrição do funcionamento do algoritmo e que sugere que para valores menores que a convergência da cadeia é aceita. A Tabela . fornece os resultados para as taxas de rejeição (que também podem ser visualizadas através das Figuras . e .6) e para a convergência das cadeias.

### 4.3 Inferência clássica via método de máxima verossimilhança usando MCMC

Foram elaborados programas (ver (Lopes e Oliveira, 2016)) que fornecem estimativas da função de máxima verossimilhança, segundo a aproximação de uma razão da log-verossimilhança que permite estimar os parâmetros utilizando os métodos MCMC. Ver as Tabelas . e .

### 4.4 Discussão dos resultados

Dos resultados clássicos mostrados na Tabela ( . ), obtidos pelo método dos mínimos quadrados (Vismara, 2016), observa-se que os intervalos de confiança contemplam o zero. O que quer dizer, em outras palavras, que abrangem valores negativos. Isso ocorre no parâmetro  $a$  que biologicamente tem o significado de relacionar a mortalidade das plantas com a dependência da densidade. Portanto, surge a , uma inconsistência, do ponto de vista biológico em termos dos resultados. Essa inconsistência é contornada através da inferência Bayesiana com o auxílio de técnicas para resolução numérica de integrais (método de Simpson) e o uso dos métodos MCMC, como pode ser visto nas Tabelas . e . que apresentam intervalos de credibilidade com faixas de valores maiores do que zero.

#### 4.4.1 Método numérico de Simpson

Através do método de integração numérica de Simpson foram obtidas funções para as posteriores, nos dois casos considerados. Observou-se através dos gráficos . e . que estas funções apresentaram comportamento de densidades vistas que decaíam para

zero. Pode-se notar também através destes gráficos, que algumas posteriores possuem comportamentos parecidos com as funções gama e normal.

#### 4.4.2 Taxas de rejeição e convergência das cadeias

##### Das taxas de rejeição

As taxas de rejeição dão informação de quanto a distribuição da cadeia gerada se aproxima da 'real' distribuição em questão, que no caso é a densidade a posteriori obtida através da combinação da verossimilhança com a priori Jeffreys. Sendo assim, fica evidenciado, através das Tabelas 4.4 e 4.5 que foram obtidos valores satisfatórios para os parâmetros gerados pelo algoritmo de Metropolis-Hastings, visto que o maior valor de taxa de rejeição foi de 0.04. Como a distribuição condicional de  $\sigma$  tem um núcleo de uma gama invertida, esse parâmetro foi gerado do algoritmo de Gibbs (que tem taxa de aceitação igual a 1). Além disso, os gráficos das Figuras 4.4 e 4.6 indicam que não houve

tem-se que o fator de Bayes dado pela relação entre a DPO do Caso 1 e a DPO do Caso 2 é 1.0 e 1.0 para o Capim Guiné e Leiteiro, respectivamente. Já para o caso do Capim Colchão pela proximidade de número característica a evidência de um caso ser preferido ao outro. Nesse caso, qualquer um pode servir para o ajuste dos dados.

### BIC e AIC

Valores baixos selecionados por estes critérios, significam melhor ajuste. Com isso verifica-se pela Tabela 4.4 que foram evidenciados valores menores para o Caso 1 para os conjuntos de dados do Capim Guiné e Leiteiro. Já para o ajuste do Capim Colchão os valores não indicam satisfatoriamente qual caso é melhor. Tem-se, então, uma concordância entre os critérios adotados

#### 4.4.4 Estimador de máxima verossimilhança usando o MCMC

Para obter os resultados do estimador de máxima verossimilhança foram geradas 'amostras' de uma cadeia de 10000 elementos, com um burn-in de 1000 e um espaçamento de 100 (para cada conjunto de amostras, ou seja, de cada espécie). O vetor de valores iniciais,  $\psi$  (média gerada pelo algoritmo de Metropolis-Hastings), para as funções  $h_\psi$  e  $P_\psi$  (Ver Subseção 4.6.) estão especificadas nas Tabelas 4.4 e 4.5.

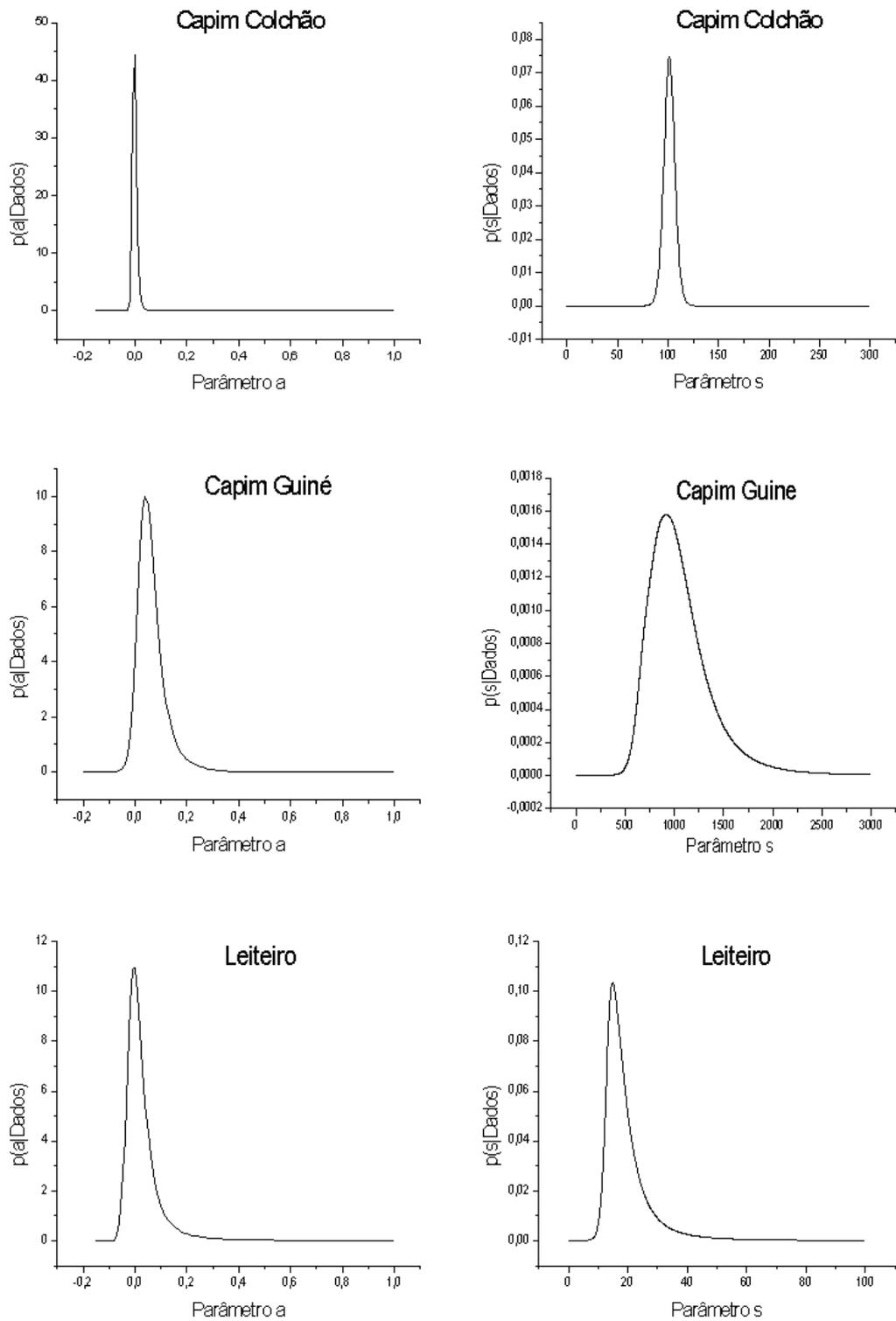


Figura 4.1: Gráficos das densidades à posteriori para o caso 1.

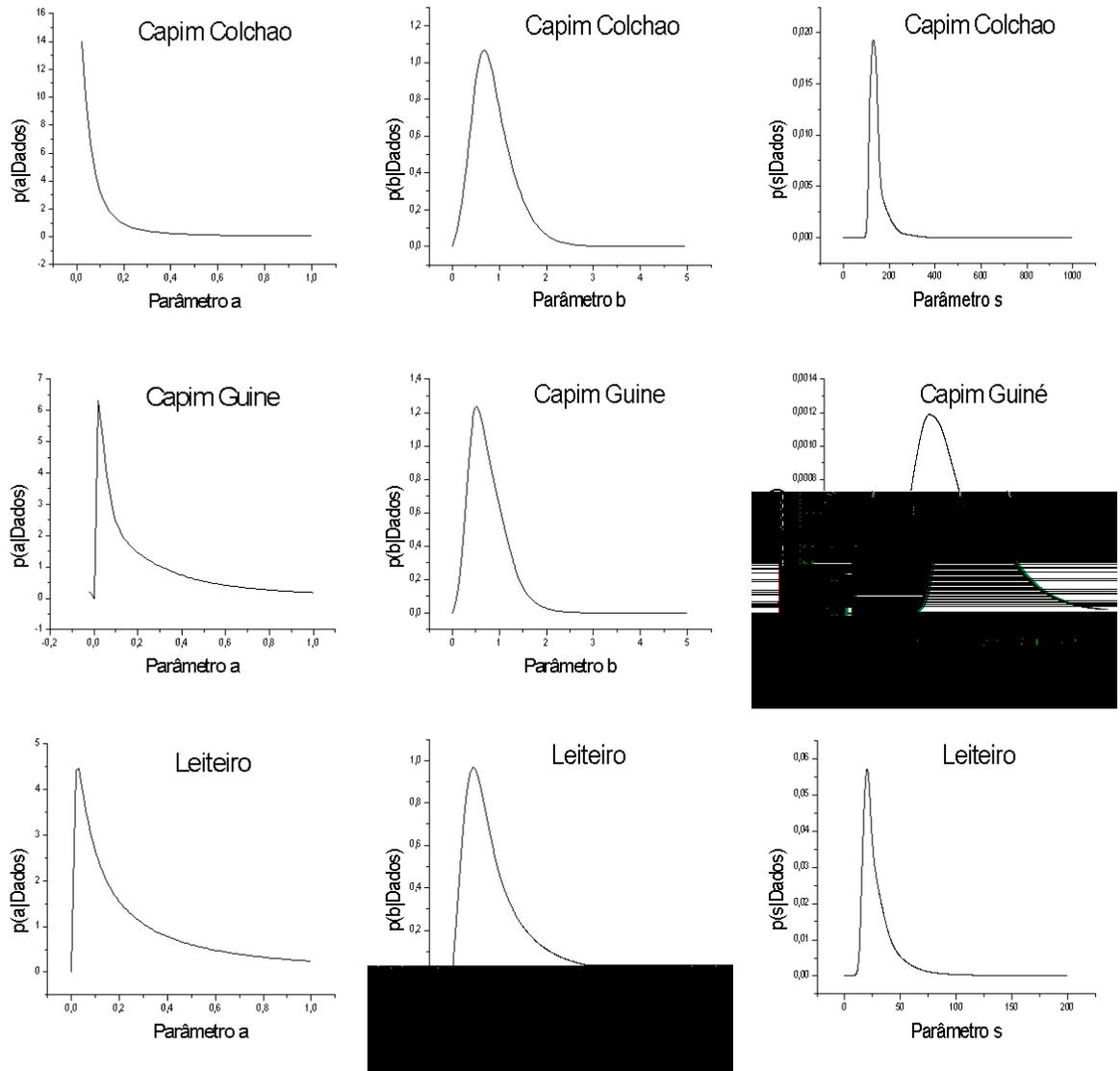


Figura 4.2: Gráficos das densidades à posteriori para o caso 2.

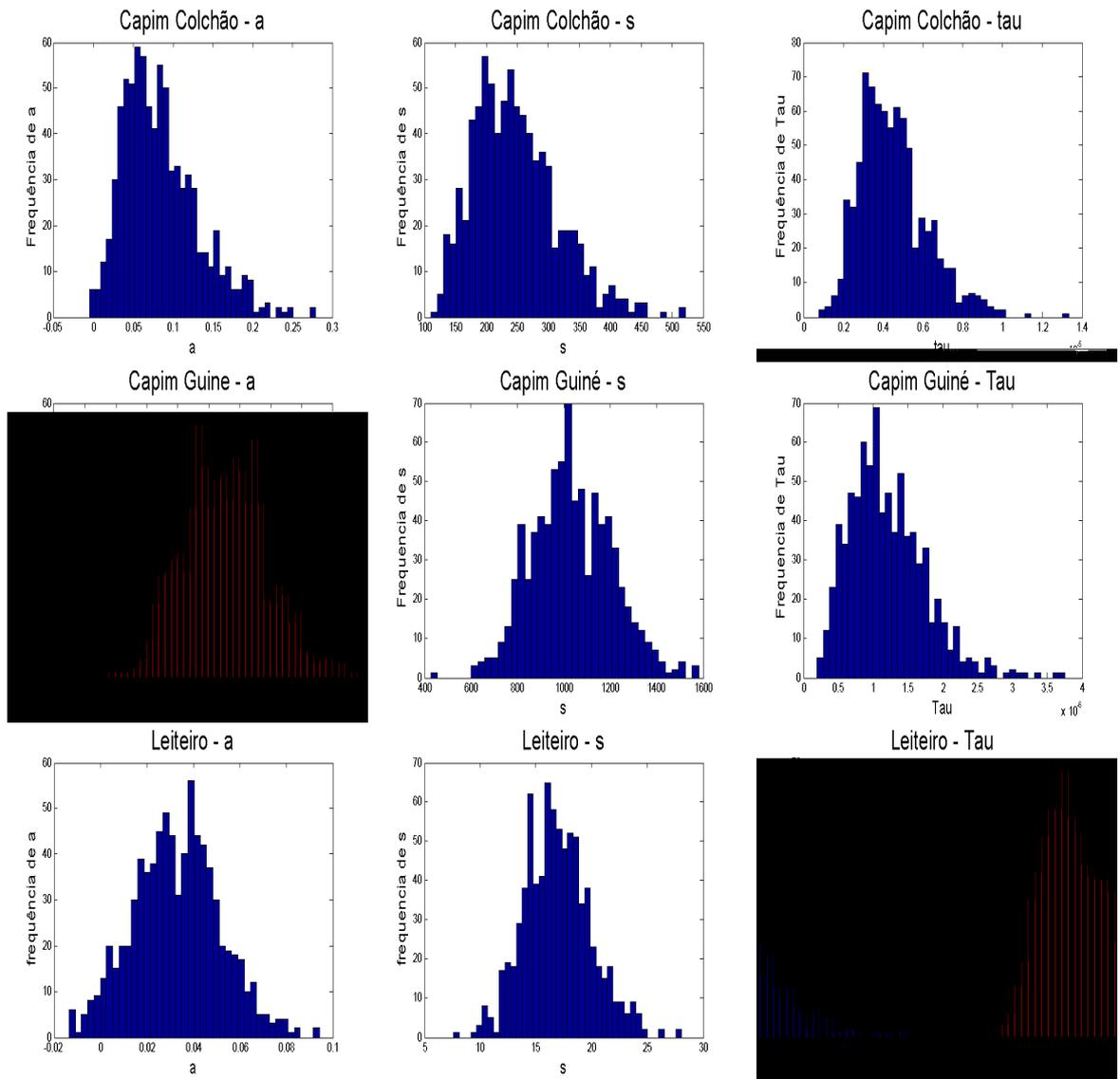


Figura 4.3: Gráficos dos histogramas dos parâmetros para o caso 1.

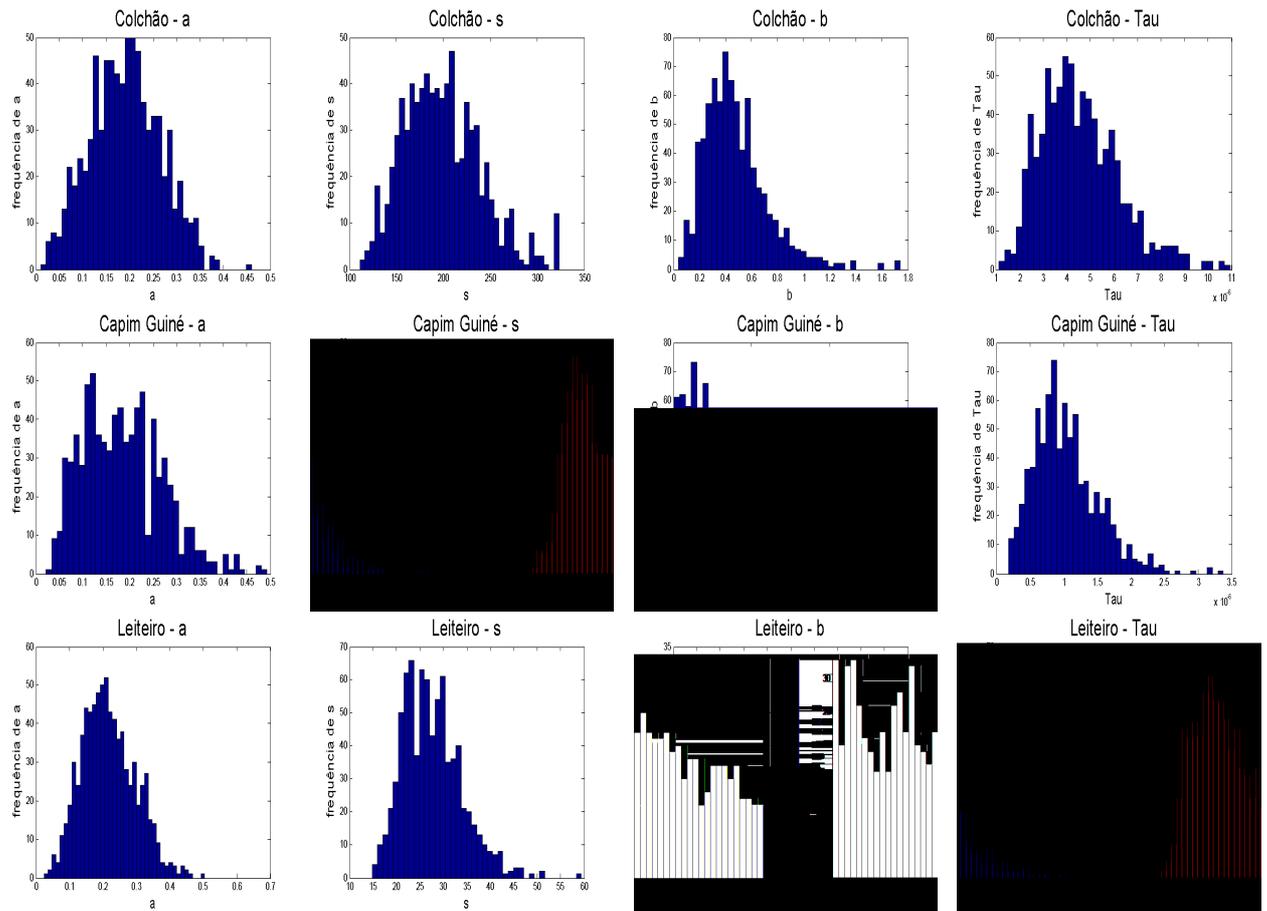


Figura 4.4: Gráficos dos histogramas dos parâmetros para o caso 2.

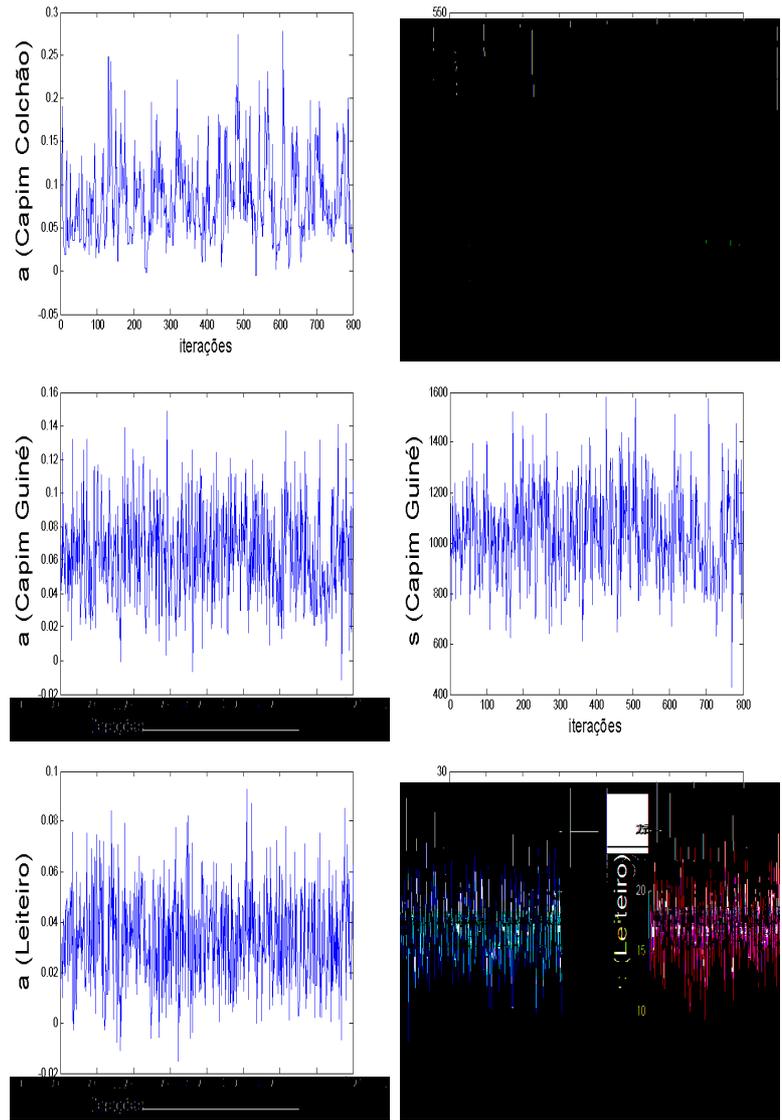


Figura 4.5: Gráficos das iterações dos parâmetros para o caso 1.

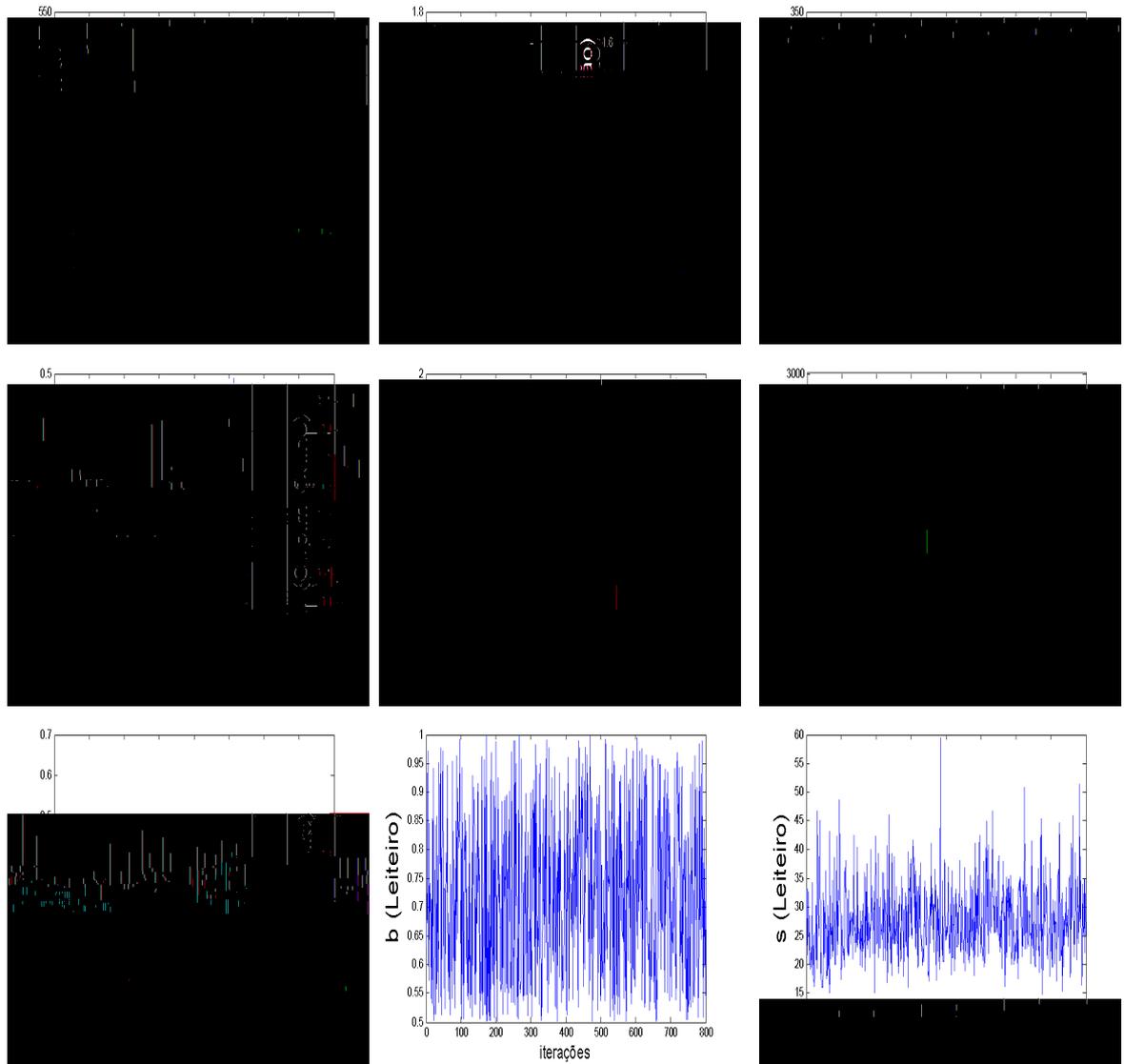


Figura 4.6:



**Tabela 4.4:** Resumo dos resultados de MCMC em EMV, para o caso 2.

Parâmetros	$\psi$ inicial	Estimativas
<b>Capim colchão</b>		
s	200.42	198.00
a	0.19	0.21
b	0.48	0.48
$\sigma$	473.50	475.00
<b>Capim guiné</b>		
s	1389.40	1391.00
a	0.18	0.20
b	0.86	0.86
$\sigma$	984.4	986.0
<b>Leiteiro</b>		
s	25.50	27.00
a	0.21	0.19
b	0.67	0.67
$\sigma$	25	27

**Tabela 4.5:** Resumo dos resultados via abordagem Bayesiana para o caso 1.

Priori Jeffreys	Parâmetros	
	s	a
<b>Capim colchão</b>		
Média	281.63	0.11
Mediana	218.00	0.063
Erro padrão	194.51	0.14
IC (95%)	[113.0 ; 880.5]	[0.0020 ; 0.56]
<b>Capim guiné</b>		
Média	1076.4	0.073
Mediana	1016.5	0.056
Erro padrão	345.8	0.064
IC (95%)	[620.5 ; 1908]	[0.0011 ; 0.23]
<b>Capim leiteiro</b>		
Média	19.61	0.061
Mediana	17.00	0.034
Erro padrão	9.24	0.091
IC (95%)	[11 ; 43.50]	[0.0002 ; 0.31]

onde IC é o intervalo de credibilidade Bayesiano.

Tabela 4.6: Resumo dos resultados via abordagem Bayesiana para o caso 2.

Priori Jeffreys	Parâmetros		
	$s$	$a$	$b$
<b>Capim colchão</b>			
Média	143.84	0.24	0.41
Mediana	136.43	0.15	0.29
Erro padrão	38.34	0.23	0.45
IC (95%)	[102.95 ; 285.86]	[0.0002 ; 0.9]	[0.049 ; 2.34]
<b>Capim guiné</b>			
Média	1300.18	0.22	1.21
Mediana	1244.29	0.14	0.87
Erro padrão	400.23	0.22	1.00
IC (95%)	[754.38 ; 2634.60]	[0.019 ; 0.92]	[0.22 ; 4.64]
<b>Capim leiteiro</b>			
Média	29.62	0.27	1.06
Mediana	24.62	0.17	0.78
Erro padrão	17.51	0.24	0.93
IC (95%)	[13.96 ; 104.24]	[0.019 ; 0.94]	[0.14 ; 4.49]

onde IC é o intervalo de credibilidade Bayesiano.

Tabela 4.7: Resumo dos resultados usando MCMC para o caso 1.

Priori Jeffreys	Parâmetros		
	$s$	$a$	$\tau$
<b>Capim colchão</b>			
Média	243.17	0.081	4.46e-6
Mediana	240.00	0.071	4.32e-6
Erro padrão	65.38	0.041	1.63e-6
IC (95%)	[193.33 ; 380.55]	[0.051 ; 0.168]	[3.28 ; 8.23]*e-6
<b>Capim guiné</b>			
Média	1015.8	0.061	1.22e-6
Mediana	1017.0	0.060	1.10e-6
Erro padrão	177.92	0.026	5.34e-6
IC (95%)	[896.24 ; 1390.1]	[0.040 ; 0.12]	[0.84 ; 2.45]*e-6
<b>Capim leiteiro</b>			
Média	17.25	0.034	0.0019
Mediana	16.95	0.033	0.0018
Erro padrão	3.02	0.018	9.07e-4
IC (95%)	[15.24 ; 23.92]	[0.021 ; 0.0691]	[0.0013 ; 0.0038]

onde IC é o intervalo de credibilidade Bayesiano e  $\tau = 1/\sigma^2$ .

**Tabela 4.8:** Resumo dos resultados usando MCMC para o caso 2.

Priori Jeffreys		Parâmetros		
<b>Capim colchão</b>	<i>s</i>	<i>a</i>	<i>b</i>	<i>tau</i>
Média	200.42	0.19	0.48	4.46e-6
Mediana	38.34	0.18	0.45	4.25e-6
Erro padrão	34.87	0.071	0.12	1.72e-6
IC (95%)	[172.68; 273.62]	[0.15 ;0.35]	[0.33 ;0.65]	[3.44 ;8.72]*e-6
<b>Capim guiné</b>	<i>s</i>	<i>a</i>	<i>b</i>	<i>tau</i>
Média	1389.40	0.18	0.86	1.03e-6
Mediana	1352.00	0.17	0.76	0.95e-6
Erro padrão	313.69	0.085	0.31	4.82e-7
IC (95%)	[1178.6 ;2106.0]	[0.11 ;0.36]	[0.63 ; 1.73]	[0.68 ;2.12]*e-6
<b>Capim leiteiro</b>	<i>s</i>	<i>a</i>	<i>b</i>	<i>tau</i>
Média	25.50	0.21	0.67	0.0016
Mediana	26.10	0.21	0.69	0.0014
Erro padrão	7.73	0.073	0.35	8.24e-4
IC (95%)	[19.81 ;44.72]	[0.16 ; 0.36]	[0.43 ; 1.48]	[0.00093 ; 0.0036]

onde IC é o intervalo de credibilidade Bayesiano e  $\tau = 1/\sigma^2$ .

**Tabela 4.9:** Convergência e taxas de rejeição para o caso 1.

Parâmetros	Taxa de rejeição	Convergência
<b>Capim colchão</b>		
<i>s</i>	0.71	0.062
<i>a</i>	0.68	0.041
$\sigma$	0	0,053
<b>Capim guiné</b>		
<i>s</i>	0.70	0.21
<i>a</i>	0.56	0.17
$\sigma$	0	0.020
<b>Leiteiro</b>		
<i>s</i>	0.42	0.066
<i>a</i>	0.71	0.061
$\sigma$	0	0.039

**Tabela 4.10:** Convergência e taxas de rejeição para o caso 2.

Parâmetros	Taxa de rejeição	Convergência
Capim colchão		
s	0.72	0.092
a	0.41	0.10
b	0.71	0.18
$\sigma$	0	0.018
Capim guiné		
s	0.71	0.051
a	0.62	0.091
b	0.67	0.14
$\sigma$	0	0.046
Leiteiro		
s	0.70	0.011
a	0.47	0.018
b	0.40	0.019
$\sigma$	0	0.00072

**Tabela 4.11:** Seleção de modelos

	Critérios		
	DPO	BIC	AIC
<b>Leiteiro</b>			
<b>2 Parâmetros</b>	7.22e-13	7.035	7.006
<b>3 Parâmetros</b>	3.009e-13	7.48	7.44
<b>Capim Guiné</b>			
<b>2 Parâmetros</b>	7.74e-29	14.35	14.28
<b>3 Parâmetros</b>	3.36e-29	14.76	14.67
<b>Capim Colchão</b>			
<b>2 Parâmetros</b>	3.20e-40	12.88	12.74
<b>3 Parâmetros</b>	3.30e-40	13.02	12.84

## Capítulo 5

# Conclusão

O presente trabalho aborda o problema de estimar parâmetros do modelo de produção de sementes de plantas daninhas. A partir da observação de amostras extra das destas populações em um experimento conduzido em campo obtiveram-se estimativas dos parâmetros do modelo via a inferência Bayesiana.

Destaca-se a técnica MCMC que constitui uma poderosa ferramenta na resolução de problemas formulados do ponto de vista Bayesiano. Esta técnica possibilitou resolver o problema de estimação de parâmetros de populações de plantas daninhas em um tempo computacional relativamente menor, além de ser de fácil implementação. Por outro lado o outro método considerado, o método de Simpson, não foi de fácil implementação para modelos de mais de parâmetros, tendo em vista a complexidade para a obtenção das posteriores.

Na seleção de modelos foram caracterizados, através dos critérios adotados, que o conjunto de dados do Capim Guiné e do Leiteiro foram melhores ajustados ao modelo de parâmetros. Já para o Capim Colchão ficou inconclusivo, ou seja, qualquer modelo poderia servir para ajustar os dados.

Finalmente, para a sequência do trabalho destacam-se alguns pontos como, considerar um modelo de produção de sementes mais complexo, que possa vir a conter mais informações, como por exemplo, incorporar a influência de outras espécies de plantas daninhas (multi-espécies de plantas daninhas competindo com a cultura), considerar também um modelo dinâmico para a produção de sementes e investigar o significado biológico da inclusão do parâmetro  $b$  considerando cada espécie.

# Apêndice A

## Noções de probabilidade

### A.1 Eventos

Em Meyer ( ) são apresentadas alguns tópicos importantes a serem considerados em um estudo introdutório à probabilidade. Um deles é o conceito de eventos. Um evento  $A$  (relativo a um particular espaço amostral  $S$ , associado a um experimento ) é simplesmente um conjunto de resultados possíveis. Na terminologia dos conjuntos, um evento é um subconjunto de um espaço amostral  $S$ . E isso significa que o próprio conjunto  $S$  constitui um evento, bem como o é o conjunto vazio  $\emptyset$ . Qualquer resultado individual pode também ser tomado como um evento. Alguns exemplos de eventos são dados a seguir, onde  $A_i$  representa o evento e  $E_i$  o experimento:

$E_1$ : Lançamento de um dado

$A_1$ : Um número par ocorre, isto é,  $A_1 = \{2, 4, 6\}$

$E_2$ : Lançamento de uma moeda

$A_2$ : Duas caras ocorrem  $A_2 = \{2\}$

$E_3$ : Teste de lâmpadas boas para o uso

$A_3$ : A lâmpada queima em menos de 3 horas  $A_3 = \{t | t \leq 3\}$

## A.2 Frequência relativa

Considere-se o seguinte procedimento: Suponha-se a repetição de  $n$  vezes o experimento  $\Omega$ , e sejam  $A$  e  $B$  dois eventos associados a  $\Omega$ . Admite-se que sejam, respectivamente,  $n_a$  e  $n_b$  o número de vezes que o evento  $A$  e o evento  $B$  ocorram nas  $n$  repetições.

**Definição A.2.1**  $f_a = n_a/n$  é denominada frequência relativa do evento  $A$  nas  $n$  repetições de  $\Omega$ . A frequência relativa  $f_a$  apresenta as seguintes propriedades, de fácil verificação:

(1)  $0 \leq f_a \leq 1$ .

(2)  $f_a = 1$  se, e somente se,  $A$  ocorrer em todas as  $n$  repetições.

(3)  $f_a = 0$  se, e somente se,  $A$  nunca ocorrer nas  $n$  repetições.

(4)  $f_a = 0$  se, e somente se,  $A$  nunca ocorrer nas  $n$  repetições.

(4) Se  $A_1, A_2, \dots, A_n, \dots$  forem, dois a dois, eventos mutuamente excludentes, então,  $P(\bigcup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$

Observe-se que da propriedade ( ), decorre imediatamente que, para qualquer n nito,  $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ .

A propriedade ( ) n o se seguir , no entanto, quando for considerado o espaço amostral idealizado, esta propriedade ser imposta e, por isso, foi inclu da aqui. A escolha das propriedades da probabilidade acima relacionadas é, obviamente, sugerida pelas correspondentes caracter sticas da frequência relativa. É poss vel mostrar que os números  $P(A)$  e  $f_A$  s o proximos um do outro (em determinado sentido), se  $f_A$  for baseado em um grande numero de repetições. É este fato a que se d a justi cativa de  $P(A)$  para se avaliar qu o veross mil é a ocorrência de  $A$ .

N o é mencionado como calcular  $P(A)$  e para maiores detalhes ver Meyer ( ). S o colocadas apenas algumas propriedades gerais que  $P(A)$  possui. A seguir ser enunciada e demonstrada v rias consequências relacionadas a  $P(A)$ , que decorrem das condições acima e que n o dependem da maneira pelo qual foi calculado  $P(A)$ .

**Teorema A.1** Se  $\emptyset$  for o espaço vazio, então  $P(\emptyset) = 0$ .

*Prova:* Para qualquer evento  $A$ , pode-se escrever  $A = A \cup \emptyset$  Uma

(ou)

### A.3 Probabilidade condicionada e eventos independentes

Segundo Costa Neto ( ) muitas vezes, o fato de car sabendo que certo evento ocorreu faz com que se modi que a probabilidade atribuida a outro evento. Denotar-se por  $P(E|F)$  a probabilidade do evento  $E$  sabendo-se que  $F$  ocorreu ou, simplesmente, probabilidade de  $E$  dado  $F$ .

Pode-se mostrar a coêrencia da relaço o, segundo a qual

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, P(F) \neq 0. \quad (\text{A. } )$$

Analogamente

$$P(F|E) = \frac{P(E \cap F)}{P(E)}, P(E) \neq 0. \quad (\text{A. } )$$

Das expressões acima resulta a regra do produto, que refere-se ao c lculo da probabilidade do evento de intersecço o,

$$P(E \cap F) = P(E)P(F|E) = P(E)P(E|F). \quad (\text{A. } )$$

Note-se que a ordem do condicionamento pode ser invertida. Para três eventos pode-se, por exemplo, escrever

$$P(E \cap F \cap G) = P(E)P(F|E)P(G|E \cap F). \quad (\text{A.6})$$

De forma semelhante a express o poderia ser generalizada para diversos eventos. Dois importantes teoremas s o os seguintes

(a) Teorema da probabilidade total. Seja  $E_1, E_2, \dots, E_n$  uma partiço o de  $F$  um evento qualquer de  $S$ . Ent o

$$P(F) = \sum_{i=1}^n P(E_i)P(F|E_i). \quad (\text{A. } )$$

Esse resultado pode ser demonstrado considerando-se o evento  $F$  sub-dividido em suas intersecções com os eventos  $E_i$  e aplicando-se as propriedades anteriores.

(b) Teorema de Bayes. Nas mesmas condições do teorema anterior:

$$P(E_j|F) = \frac{P(E_j)P(F|E_j)}{\sum_{i=1}^n P(E_i)P(F|E_i)}, j = 1, \dots, n. \quad (\text{A. } \bullet)$$

Esse resultado consegue-se facilmente do resultado anterior e demais propriedades.

Os dois teoremas acima são particularmente úteis no estudo de situações que se processam em duas etapas, a primeira das quais corresponde à ocorrência de um (e somente um, devido à sua natureza) dos eventos da família  $E_i$  dizendo o evento  $F$  respeito à segunda etapa. O primeiro teorema ensina como calcular a propriedade incondicional do evento  $F$ , isto é, não importando qual dos eventos da família  $E_i$  possa ter ocorrido. O teorema de Bayes por sua vez, mostra como calcular a probabilidade de que tenha sido o particular evento  $E_j$  da família  $E_i$  aquele que ocorreu, isto fornece informação de que o evento  $F$  ocorreu.

### A.3.1 Eventos independentes

Se  $P(E|F) = P(E|\bar{F}) = P(E)$ , o evento  $E$  é dito estatisticamente independente do evento  $F$ . Isso implica que o evento  $F$  será estatisticamente independente do evento  $E$ , o que é fácil provar.

Nas condições de independência os cálculos se simplificam, pois não é preciso se preocupar com probabilidades condicionadas.

Sendo independentes os eventos, a regra do produto fica

$$P(E \cap F) = P(E)P(F). \quad (\text{A. } \bullet)$$

sendo de imediata a generalização a vários eventos, ou seja:

$$P(E \cap F \cap \dots \cap K) = P(E)P(F)\dots P(K). \quad (\text{A. } \bullet\bullet)$$

## Apêndice B

# Algoritmo para o cálculo das posterioris pelo método de Simpson

A seguir, são apresentados alguns dos pseudo-códigos utilizados para resolver as integrais das posterioris pelo método de Simpson. Considerar-se a obtenção da posteriori  $p(a|Dados)$  (para o modelo de parâmetros). Inicialmente, é exibido as funções e sub-rotinas. Logo a seguir é exibida a rotina principal.

### B.1 Modelo de 2 parâmetros

#### B.1.1 Parâmetro $a$

\*\*\*\* Esta função retorna a função  $g(a)$  descrita em . \*\*\*\*\*

função FUNCAO(va, vs, z, y)

Declaração de variáveis

(va,vs): Parâmetros  $a$  e  $s$  respectivamente

S , G, S , S , S6, FUNCAO: Variáveis auxiliares

z,y: Dados (Ver a Tabela . )

Chama a subrotina SOMAS(va, vs, S , S , S , S6, z, y)

$$G \rightarrow S4 * S5 - S6 * S6$$

$$\text{FUNCAO} \rightarrow va * \sqrt{G} * \exp(-4.0 * \ln(S1))$$

m da funç o

funç o  $f(x, vs, z, y)$

Declaraç o de vari veis

$x, f$ : Vari veis auxiliares

$vs$ : Parâmetro a ser integrado

$z, y$ : Dados

$$f \rightarrow \text{FUNCAO}(x, vs, z, y)$$

m da funç o

\*\*\*\* A funç o *Simpson1* explicada a seguir, retorna valores para a posteriori  $p(a|Dados)$ .

funç o Simpson ( $a, b, n, teta, z, y$ )

Declaraç o de vari veis

$a, b, x, h, int1, d, e, i, n, Simpson$  : Vari veis auxiliares

$vs, z, y$ : Parâmetros e Dados

$$h \rightarrow \frac{(b-a)}{n}$$

$$int1 \rightarrow f(a, vs, z, y) + f(b, vs, z, y)$$

$$x \rightarrow a$$

Faça  $i = 1$  até  $(n-)$

$$x \rightarrow x + h$$

$$d \rightarrow i/2.0$$

$$e \rightarrow int(d)$$

Se  $(d = e)$  ent o

$$int1 \rightarrow int1 + 2 * f(x, teta2, z, y)$$

Sen o

$$int1 \rightarrow int1 + 4 * f(x, teta2, z, y)$$

Fim se

Fim faça

$Simpson1 \rightarrow \text{int} * h/3$

m da função

\*\*\*\* A função Simpson normaliza a  $p(a|Dados)$  \*\*\*\*

função Simpson (a,b,cont,pa)

Declaração de variáveis

$a, b, n, x, h, \text{int1}, d, e, Simpson2, i, \text{cont}$ : Variáveis auxiliares

$pa$ : Posteriori de  $a$

$h \rightarrow (b - a)/\text{cont}$

$\text{int1} \rightarrow pa(1) + pa(\text{cont})$

$x \rightarrow a$

Faça  $i = 1$  até (cont- )

$x \rightarrow x + h$

$d \rightarrow i/2.0$

$e \rightarrow \text{inteiro}(d)$

Se ( $e = d$ ) então

$\text{int1} \rightarrow \text{int1} + 2 * pa(i)$

Senão

$\text{int1} \rightarrow \text{int1} + 4 * pa(i)$

Fim se

Fim faça

$Simpson2 = \text{int1} * h/3$

m da função

\*\*\*\* A função Simpson retorna a estimativa do parâmetro a \*\*\*\*

função Simpson (a,b,cont,pa)

Declaraç o de vari veis

$a, b, n, x, h, int1, d, e, Simpson3, i, cont$ : Vari veis auxiliares

$pa$ : Posteriori de  $a$ .

$h \rightarrow (b - a)/cont$

$int1 \rightarrow a * pa(1) + b * pa(cont)$

$x \rightarrow a$

Faça  $i = 1$  até  $(cont - 1)$

$x \rightarrow x + h$

$d \rightarrow i/2.0$

$e \rightarrow inteiro(d)$

Se  $(d = e)$  ent o

$int1 \rightarrow int1 + 2 * x * pa(i)$

Sen o

$int1 \rightarrow int1 + 4 * x * pa(i)$

Fim se

Fim faça

$Simpson3 \rightarrow int1 * h/3$

m da funç o

\*\*\*\* A funç o Simpson retorna o erro padr o \*\*\*\*

funç o Simpson ( $a, b, cont, pa$ )

Declaraç o de vari veis

$a, b, n, x, h, int1, d, e, Simpson4, i, cont$ : Vari veis auxiliares

$pa$ : Posteriori de  $a$

$h \rightarrow (b - a)/cont$

$int1 \rightarrow (a^2) * pa(1) + (b^2) * pa(cont)$

$x \rightarrow a$

Faça  $i = 1$  até (cont- )

$x \rightarrow x + h$

$d \rightarrow i/2.0$

$e \rightarrow \text{inteiro}(d)$

Se ( $d = e$ ) ent o

$int1 \rightarrow int1 + 2 * (x^2) * pa(i)$

Sen o

$int1 \rightarrow int1 + 4 * (x^2) * pa(i)$

Fim se

Fim faça

$Simpson4 \rightarrow int1 * h/3$

m da funç o

\*\*\*\* A funç o Simpson retorna a variância e o intervalo de credibilidade (I.C) \*\*\*\*

funç o Simpson (a,b,cont,pa )

Declaraç o de vari veis

$a, b, n, x, h, int1, d, e, Simpson5, t1, t2, intmedia, i, cont$ : Vari veis auxiliares.

$med, m2, m1, teta11, teta12, teta21, teta22$ : Intervalos para a mediana e o I.C.

$pa1$ : Posteriori de a.

'Entrada dos valores  $m1$  e  $m2$ ' Para encontrar a mediana

'Entrada dos valores  $teta11$  e  $teta12$ ' (Para encontrar o valor de  $a$  que corresponde a 25% dos valores de  $p(a|Dados)$ )

'Entrada dos valores  $teta21$  e  $teta22$ ' (Para encontrar o valor de  $a$  que corresponde a 95% dos valores de  $p(a|Dados)$ )

$h \rightarrow (b - a)/cont$

$int1 \rightarrow pa1(1) + pa1(cont)$

$x \rightarrow a$

Faça  $i = 1$  até (cont- )

$x \rightarrow x + h$

$d \rightarrow i/2.0$

$e \rightarrow \text{int}(d)$

Se ( $e = d$ ) ent o

$\text{int1} \rightarrow \text{int1} + 2 * \text{pa1}(i)$

Sen o

$\text{int1} \rightarrow \text{int1} + 4 * \text{pa1}(i)$

Fim se

$\text{intmedia} \rightarrow \text{int1} * h/3$

Se ( $\text{intmedia} \leq m2$ ) ent o

Se ( $\text{intmedia} \geq m1$ ) ent o

$\text{med} \rightarrow x$

Fim se

Fim se

Se ( $\text{intmedia} \leq \text{teta12}$ ) ent o

Se ( $\text{intmedia} \geq \text{teta11}$ ) ent o

$t1 \rightarrow x$

Fim se

Fim se

Se ( $\text{intmedia} \leq \text{teta22}$ ) ent o

Se ( $\text{intmedia} \geq \text{teta21}$ ) ent o

$t2 \rightarrow x$

Fim se

Fim se

Fim faça

$\text{Simpson5} \rightarrow \text{int1} * h/3$

Escreve o intervalo de confiança  $t1, t2$  e a mediana  $med$ .

m da função

\*\*\*\* Esta subrotina auxilia no cálculo da  $g(a)$ \*\*\*\*\*

Subrotina SOMAS(teta, teta, S, S, S, S6, z, y)

Declaração de variáveis

$aux1, aux2, aux3, aux4, aux44, i, j, S1, S4, S5, S6$ : Variáveis auxiliares

$z, y$ : Dados

$va, vs$ : Parâmetros

$S1 \rightarrow 0; S4 \rightarrow 0; S5 \rightarrow 0; S6 \rightarrow 0$

Faça  $i = 1$  até

$aux1 \rightarrow (y(i) - va * z(i) / (1 + vs * z(i)))$

$aux2 \rightarrow z(i)^2 / ((1 + z(i) * vs)^2)$

$aux3 \rightarrow z(i) / (1 + vs * z(i))$

$aux44 \rightarrow (1 + z(i) * vs)^3$

$aux4 = z(i)^3$

$S1 = S1 + aux1^2$

$S4 = S4 + aux2^2$

$S5 = S5 + aux3^2$

$S6 = S6 + aux4 / aux44$

Fim faça

m da subrotina

O programa principal, a seguir, entra com os dados, chama as funções acima e re-

$inf1, sup1, inf2, sup2, H2, n1, cont, l, j, t, k$ : Variáveis auxiliares  
 $pa, pa1$ : Posteriori não-normalizada e normalizada  
 $vs$ : parâmetros  
 $z, y, Ea, Erro, retorno$ : Dados, estimativa, erro padrão, mediana e I.C.

'Leitura dos Dados'

Faça  $j = 1, 8$

Ler (e armazenar)  $z(j), y(j)$

Fim faça

\*\*\*\* Intervalos para a o método de integração de Simpson \*\*\*\*

'Ler limites inferior e superior de  $a$ ',  $inf2$  e  $sup2$

'Ler incrementos para  $a$ ',  $H2$

'Ler limites inferior e superior de  $s$ ',  $inf1$  e  $sup1$

'Ler o número de divisões para as integrais (em  $s$ )',  $n$

$H1 \rightarrow (sup1 - inf1)/n1$

$teta2 \rightarrow inf2; cont \rightarrow 0$

• Se ( $teta2 \leq sup2$ ) então

$t \rightarrow teta2$

$pa \rightarrow Simpson1(inf1, sup1, n1, teta2, z, y)$

$pa1(cont + 1) \rightarrow pa$

$teta2 \rightarrow teta2 + H2$

$cont \rightarrow cont + 1$

Volta para a linha •

Fim se

\*\*\*\*\* Normalização \*\*\*\*\*

$k \rightarrow Simpson2(inf2, sup2, cont, pa1)$

$t \rightarrow inf2$

Faça ( $l = 1$ ) até *cont*

$pa1(l) \rightarrow pa1(l)/k$

'Escrever (arquivo)'  $pa(1)$ ,  $t$

$t \rightarrow t + H2$

Fim faça

\*\*\*\*\* Estimativa, Erro Padr o, Variancia e I.C \*\*\*\*\*

$Ea \rightarrow Simpson3(inf2, sup2, cont, pa1)$

$Erro \rightarrow Simpson4(inf2, sup2, cont, pa1) - Ea * *2$

$retorno \rightarrow Simpson5(inf2, sup2, cont, pa1)$

'Escrever sa das',  $Ea$ ,  $Erro$

Fim programa principal

# Bibliografia

Achcar, J. A., Loibel, S. and Andrade, M. G. (2006). Interfailure data with constant hazard function in the presence of change-points, *REVSTAT, Statistical Journal* 4(1): 1-10.

Barreto, G. and Andrade, M. G. (2000). Bayesian inference and markov chain monte carlo methods applied to streamflow forecasting, *Proc. of the 6th Int. Conf. on Probabilistic Methods Applied to Power Systems (PMAPS'2000)*, Funchal, Ilha da Madeira, pp. FOR 1-6.

\*Disponível em <http://citeseer.ist.psu.edu/barreto00bayesian.html>

Barroso, L. C., Barroso, M. M. A. and Campos, F. F. (1998). *Cálculo Numérico*, Harper-Row do Brasil Ltda, São Paulo, SP.

Besag, J. (1994). Markov chain monte carlo for statistical inference.

\*Disponível em <http://citeseer.ist.psu.edu/besag94markov.html> (Pesquisado em 16/06/2006)

Cappe, O., Douc, R., Moulines, E. and Robert, C. (2006). On the convergence of the monte carlo maximum likelihood method for latent variable models, *Scandinavian Journal of Statistics* 29(1): 6-16.

Costa Neto, M. C. (1998). *Probabilidade - Resumos Teóricos - Exercícios Resolvidos*, Edgard Blucher Ltda, São Paulo, SP.

Cousens, R. and Mortimer, M. (1997). *Dynamics of Weed Populations*, Cambridge University Press, Cambridge, UK.

Doyle, C. J. (1992). A review of the use of models of weed control in integrated crop protection, *Agriculture, Ecosystems and Environment* 64(1): 6-15.

- Edelstein-Keshet, L. (1988). *Mathematical Models in Biology*, Random House, New York, NY.
- Ehlers, R. S. (2006). Métodos computacionalmente intensivos em estatística, Notas de aulas em disciplina ministrada no programa de bacharelado em estatística da universidade federal do paran, Disponvel: <http://www.est.ufpr.br/~ehlers/notas/mci.pdf> (pesquisado em 06/06/2006).
- Favoretti, A. C. (2004). *Modelos No-lineares: um Enfoque Bayesiano*, Dissertaço de Mestrado, Instituto de Cincias Matemticas e de Computaço, Universidade de So Paulo, So Carlos, SP.
- Geyer, C. J. (1992). On the convergence of monte carlo maximum likelihood calculations, *Journal of the Royal Statistical Society Series B-Methodological* 54(1): 6-11.
- Geyer, C. J. and Thompson, E. A. (1992). Constrained monte-carlo maximum-likelihood for dependent data, *Journal of the Royal Statistical Society Series B-Methodological* 54(1): 6-16.
- Hassell, M. P. (1975). Density dependence in single-species populations, *Journal of Animal Ecology* 44(1): 221-231.
- Huffer, F. W. and Wu, H. L. (1998). Markov chain monte carlo for autologistic regression models with application to the distribution of plant species, *Biometrics* 54(1): 16-23.
- Lopes, M. R. and Oliveira, V. A. (2006). Algoritmos para a implementaço de tcnicas MCMC.  
\*Disponvel em <http://www.sel.eesc.usp.br/rtsel>, Relatorio Tcnico
- Lopes, M. R., Vismara, L. S. and Oliveira, V. A. (2006). Estimaço de parmetros de modelos de populaçes de plantas daninhas usando o mtodo de monte carlo em cadeia de markov, *6 Congresso Temtico de Dinmica, Controle e Aplicaçes, 22-25 de maio (Resumos)*, So Jos do Rio Preto, SP.
- Meyer, P. L. (1990). *Probabilidade - Aplicaçes  Estatstica*, Ao Livro Tcnico S.A., Rio de Janeiro, RJ.

- Paulino, C. D., Turkman, A. A. and Murteira, B. (2006). *Estatística Bayesiana*, Fundação Calouste Gulbekian, Lisboa, PT.
- Raftery, A. E. (1987). Bayesian model selection in social research, *Sociological Methodology* 23: 6 .
- Sakai, K. (2006). *Nonlinear Dynamics and Chaos in Agricultural Systems*, Elsevier, Amsterdam, NL.
- Vismara, L. S. (2006). *Simulação da dinâmica de produção de sementes de plantas daninhas usando inferência Bayesiana*, Dissertação de Mestrado, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP.
- Wu, J. (2006). Optimal weed control under static and dynamic decision rules, *Agricultural Economics* 23( ): 6 .

# Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

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

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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